CONTENTS

ARTICLES

Recognizing the Societal Value in Information Privacy
James P. Nehf 1

Intergovernmental Cooperation, Metropolitan Equity, and The New Regionalism
Laurie Reynolds 93

NOTES & COMMENTS

So the Army Hired an Ax-Murderer: The Assault and Battery Exception to the Federal Tort Claims Act Does Not Bar Suits for Negligent Hiring, Retention and Supervision
Rebecca L. Andrews 161

Back to Prima Paint Corp. v. Flood & Conklin Manufacturing Co.: To Challenge an Arbitration Agreement You Must Challenge the Arbitration Agreement
Andre V. Egle 199

Researcher Liability for Negligence in Human Subject Research: Informed Consent and Researcher Malpractice Actions
Roger L. Jansson 229

Is Assent Still a Prerequisite for Contract Formation in Today’s E-conomy?
Melissa Robertson 265

Beggars Can’t be Voters: Why Washington’s Felon Re-enfranchisement Law Violates the Equal Protection Clause
Jill E. Simmons 297

Water, Property, and The Clean Water Act
Janis Snoey 335
University of Washington School of Law 2002–2003

OFFICERS OF ADMINISTRATION

Lee L. Huntsman, B.A., Ph.D., Interim President of the University
W. H. Knight, Jr., B.A., J.D., Dean, Professor of Law
Patricia C. Kuszler, B.A., M.D., J.D., Associate Dean for Research & Faculty Development, Professor of Law
Penny A. Hazellon, B.A., J.D., M.L.I., Associate Dean for Library & Computing Services, Professor of Law
Dexter Bailey, B.S., M.B.A., Assistant Dean for Development
Mary A. Hotchkiss, B.A., M.S.L.S., J.D., L.L.M., Assistant Dean for Academic Services
Sandra E. Madrid, B.S., Ph.D., Assistant Dean for Student Affairs and Administration

FACULTY EMERITI

William T. Burke, B.S. J.D., J.S.D.
Robert L. Fletcher, A.B., LL.B.
John Huston, B.A., J.D., LL.M.
Arval Morris, B.A., M.A., J.D., L.L.M., L.L.D.
Cornelius J. Peck, B.S., LL.B.
John R. Price, A.B., LL.B.
Luvern V. Rieke, B.S., LL.B., L.L.M., L.L.D.
Marjorie D. Rombauer, B.A., J.D.
Charles Z. Smith, B.S., J.D.
Frank W. Smith, Jr., B.A., J.D., L.L.M.

FACULTY OF LAW

Craig H. Allen, B.S., J.D., Professor of Law
William R. Andersen, B.S.L., LL.B., LL.M., Judson Falknor Professor of Law
Helen A. Anderson, B.A., J.D., Senior Lecturer
Robert T. Anderson, B.A., J.D., Senior Lecturer
Thomas R. Andrews, B.A., M.A., J.D., Professor of Law
Robert H. Aronson, B.A., J.D., Professor of Law
Melissa M. Berry, B.A., J.D., Lecturer
Karen Boxa, B.A., J.D., Associate Professor of Law
Steve P. Calandrillo, B.A., J.D., Assistant Professor of Law
Donald C. Clarke, A.B., M.Sc., J.D., Professor of Law, Wadu, Law School Foundation Scholar
Sam A. Donaldson, B.A., J.D., LL.M., Assistant Professor of Law
G. Meade Emory, A.B., LL.B., L.L.M., Professor of Law, Director of Graduate Program in Taxation
Joan M. Fitzpatrick, B.A., J.D., Jeffery and Susan Brotman Professor of Law
Julia Gold, B.A., J.D., Senior Lecturer
James H. Hardisty, A.B., L.L.B., Professor of Law
Gregory A. Hicks, B.A., J.D., Professor of Law
Roland L. Hjorth, A.B., L.L.B., Dean Emeritus and Garvey, Schubert, & Barer Professor of Law
Linda S. Hume, B.A., J.D., Professor of Law
Stewart M. Jay, A.B., J.D., William L. Dwyer Professor of Law
John M. Junker, B.A., J.D., Professor of Law
Lisa Kelly, B.A., J.D., Harry M. Cross Distinguished Visiting Professor of Law
Alan Kirtley, B.A., J.D., Associate Professor of Law, Director of Clinics
Richard O. Kimmert, B.S., M.B.A., L.L.B., Dwayne & Anne Gittinger Professor of Law
Deborah Maranville, B.A., J.D., Associate Professor of Law
Anna Mastroianni, B.S./B.A., J.D., M.P.H., Assistant Professor of Law
Kathleen M. McGinnis, B.A., J.D., Senior Lecturer
Jacqueline McMurtrie, B.G.S., J.D., Senior Lecturer
Peter Nicolas, B.A., M.P.P., J.D., Assistant Professor of Law
Kathleen O’Neill, B.A., J.D., Associate Professor of Law, Director of Legal Research and Writing
Roy L. Prosterman, A.B., L.L.B., Professor of Law
Amita Ramaswarya, B.A., M.A., J.D., Assistant Professor of Law
William H. Rodgers, Jr., B.A., L.L.B., Stimson Bullitt Professor of Environmental Law
Eric Schnapper, B.A., M.A., B.Phil., L.L.B., Pendleton Miller Professor of Law
Scott A. Schumacher, B.A., J.D., Lecturer
William B. Stoeckel, B.A., J.D., S.J.D., Judson Falknor Professor of Law Emeritus
Toshiko Takenaka, B.A., L.L.M., Ph.D., Associate Professor of Law, Director, CASRIP
Veronica L. Taylor, B.A., B.A., L.L.M., Professor of Law, Asian Law Program Director
Michael E. Townsend, B.A., M.A., Ph.D., J.D., Associate Professor of Law
Philip A. Trautman, B.A., J.D., Professor of Law
Lea B. Vaughn, A.B., J.D., Professor of Law
Walter J. Walsh, B.C.L., L.L.M., J.S.D., Assistant Professor of Law
Lis W. Wiehl, A.B., M.A., J.D., Associate Professor of Law
Jane K. Wynn, B.S., J.D., Professor of Law
Louis E. Wolcher, B.A., J.D., Professor of Law

ADJUNCT AND AFFILIATE FACULTY

G. Andrew H. Benjamin, B.A., M.A., J.D., Ph.D., Affiliate Professor of Law
Daniel M. Bodansky, A.B., M.P.H., J.D., Affiliate Professor of Law
Sharon E. Brown, B.A., M.A., J.D., E.D.D., Adjunct Research Associate Professor of Law
Daniel H. Foote, A.B., J.D., Affiliate Professor of Law
O’Connor, B.A., J.D., Visiting Assistant Professor of Law
Judy N. Osler, B.A., J.D., Associate Professor of Law
John O. Haley, B.A., L.L.B., L.L.M., Affiliate Professor of Law
Marc J. Hershman, A.B., J.D., Adjunct Professor of Law
Kenneth E. Himma, B.A., M.A., J.D., Ph.D., Adjunct Lecturer of Law
Nancy S. Jecker, B.A., M.A., Ph.D., Adjunct Professor of Law
CONTENTS

ARTICLES

Classes, Persons, Equal Protection, and *Village of Willowbrook v. Olech*  
Robert C. Farrell 367

Traditional Equity and Contemporary Procedure  
*Thomas O. Main* 429

What is the Rule of Law? Perspectives from Central Europe and the American Academy  
*Louis E. Wolcher* 515

NOTES & COMMENTS

The Foreseeability of Transference: Extending Employer Liability Under Washington Law for Therapist Sexual Exploitation of Patients  
*Timothy E. Allen* 525

A is Not A: Washington’s Unconstitutional Law of Single-Count, Single-Defendant Inconsistent Verdicts in *State v. Goins*  
*Natasha Shekdar Black* 557

Who Owns “The Law”? The Effect on Copyrights When Privately-Authored Works are Adopted or Enacted by Reference into Law  
*Katie M. Colendich* 589

Conundrums with Penumbras: Preembryos, the Right to Privacy, and Non-gamete Providers who Intend to be Parents  
*Lainie M. C. Dillon* 625

*Derek D. Green* 653

Copyright © 2003, Washington Law Review Association. Except as otherwise provided, the author of each article in this issue has granted permission for copies of that article to be made for classroom use, provided that (1) copies are distributed at or below cost, (2) the author and journal are identified, (3) proper notice of copyright is affixed to each copy, and (4) the Washington Law Review Association is notified of the use.
WASHINGTON LAW REVIEW

VOLUME 78
2003

2002–2003 EDITORIAL BOARD
Editor-in-Chief
KRISTIN ANN RELYEA

Managing Editors
KIMBERLY ALDERSON
REBECCA ANDREWS
SHANNON M. McMINEE
JUSTICE JOY RILLERA
CAMERON SMITH

Associate Editors-in-Chief
TIMOTHY M. MITROVIC
AIMEE SUTTON

Special Projects Editor
JANIS SNOEY

Executive Notes and Comments Editors
SARAH FARLEY KALTSOUNIS
TOBIAS KAMMER

Thesis Editors
ELIZABETH FARNAM
MEGAN WALSETH

Notes and Comments Editors
JANNA J. ANNEST
KRISTI DARNELL
CECILY FUHR
ROGER JANSSON
RYAN K. JENSEN
JILL SIMMONS

Executive Articles Editors
JAMES R. PUTNAM
MELISSA ROBERTSON

Articles Editors
AMY ALLSON
ANDRE EGLE
AARON MATTHEW LAING
JULIE NAPIER

EDITORIAL STAFF
TIMOTHY ALLEN
KIRSTEN AMBACH
NICK ANDERSON
SVETLANA ATTESTATOVA
NATASHA SHERKAR BLACK
RACHEL BYRNE
MICHAEL CAUGHEY
NICHOLE CHIAPPINI
KATIE COLENDICH
BILL DENTON
LAINIE DILLON
DEREK GREEN
REBECCA HARRISON
JENNIFER AMANDA KREBS
STEPHANIE MATHEY
MÖRK MURDOCK
ALICIA OZANICH
MICHAEL PEDHIRNEY
JOSEPH REEBERGER
ANDREW PEARSON RICHARDS
TERESA B. SUMEARLL
ZACHARY TOMLINSON
CRISTINA UNDERWOOD

BUSINESS STAFF
Faculty Business Manager
PENNY HAZELTON

Faculty Adviser
RICHARD O. KUMMERT

Business Office
JONATHAN FRANKLIN
FRAN JOHNSON
University of Washington School of Law 2002–2003

OFFICERS OF ADMINISTRATION
Lee L. Huntsman, B.A., Ph.D., Interim President of the University
W. H. Knight, Jr., B.A., J.D., Dean, Professor of Law
Patricia C. Kissel, B.A., M.D., J.D., Associate Dean for Research & Faculty Development, Professor of Law
Penny A. Hazleton, B.A., J.D., M.I.L., Associate Dean for Library & Computing Services, Professor of Law
Dexter Bailey, B.S., M.B.A., Assistant Dean for Development
Mary A. Hotchkiss, B.A., M.S.L.S., J.D., L.L.M., Assistant Dean for Academic Services
Sandra E. Madrid, B.S., Ph.D., Assistant Dean for Student Affairs and Administration

FACULTY EMERITI
William T. Burke, B.S.J.D., J.S.D.
Robert L. Fletcher, A.B., LL.B.
John Huston, B.A., J.D., LL.M.
Arval Morris, B.A., M.A., J.D., L.L.M., L.L.D.
Cornelius J. Peck, B.S., LL.B.
John R. Price, A.B., LL.B.
Steve J. Rieke, B.S., LL.B., L.L.M.
Marjorie D. Rombauer, B.A., J.D.
Charles Z. Smith, B.S., J.D.
Frank W. Smith, Jr., B.A., J.D., L.L.M.

FACULTY OF LAW
Craig H. Allen, B.S., J.D., Professor of Law
William R. Andersen, B.S.L., LL.B., LL.M., Judson Falknor Professor of Law
Helen A. Anderson, B.A., J.D., Senior Lecturer
Robert T. Anderson, B.A., J.D., Assistant Professor of Law, Native American Law Center Director
Thomas R. Andrews, B.A., M.A., J.D., Professor of Law
Robert H. Aronson, B.A., J.D., Professor of Law
Melissa M. Berry, B.A., J.D., Lecturer
Karen Boxa, B.A., J.D., Associate Professor of Law
Steve P. Calandrillo, B.A., J.D., Assistant Professor of Law
Donald C. Clarke, A.B., M.Sc., J.D., Professor of Law, Wash. Law School Foundation Scholar
Sam A. Donaldson, B.A., J.D., L.L.M., Assistant Professor of Law
G. Meade Emory, A.B., LL.B., L.L.M., Professor of Law, Director of Graduate Program in Taxation
Joan M. Fitzpatrick, B.A., J.D., Jeffery and Susan Brotman Professor of Law
Julia Gold, B.A., J.D., Senior Lecturer
James H. Hardisty, A.B., LL.B., Professor of Law
Gregory A. Hicks, B.A., J.D., Professor of Law
Roland L. Hjorth, A.B., LL.B., Dean Emeritus and Garvey, Schubert, & Barer Professor of Law
Linda S. Hum, B.A., J.D., Professor of Law
Stewart M. Jay, A.B., J.D., William L. Dwyer Professor of Law
John M. Junker, B.A., J.D., Professor of Law
Lisa Kelly, B.A., J.D., Harry M. Cross Distinguished Visiting Professor of Law
Alan Kirtley, B.A., J.D., Associate Professor of Law, Director of Clinics
Richard O. Kimmart, B.S., M.B.A., LL.B., D.Wayne & Anne Gittinger Professor of Law
Deborah Maranville, B.A., J.D., Associate Professor of Law
Anna Mastroianni, B.S./B.A., J.D., M.P.H., Assistant Professor of Law
Kathleen M. McGinnis, B.A., J.D., Senior Lecturer
Jacqueline McMurtrie, B.G.S., J.D., Senior Lecturer
Peter Nicolas, B.A., M.P.F., J.D., Assistant Professor of Law
Sean O’Connor, B.A., J.D., Visiting Assistant Professor of Law
Kathleen O’Neill, B.A., J.D., Associate Professor of Law, Director of Legal Research and Writing
Roy L. Prosterman, A.B., LL.B., Professor of Law
Anita Ramaswaty, B.A., M.A., J.D., Assistant Professor of Law
William H. Rodgers, Jr., B.A., LL.B., Stimson Bullitt Professor of Environmental Law
Eric Schnapper, B.A., M.A., J.D., Pendleton Miller Professor of Law
Scott A. Schumacher, B.A., J.D., Lecturer
William B. Stoebuck, B.A., J.D., S.J.D., Judson FalknerProfessor of Law Emeritus
Toshiko Takenaka, B.A., LL.M., Ph.D., Associate Professor of Law, Director, CASRIP
Veronica L. Taylor, B.B., B.A., L.L.M., Professor of Law, Asian Law Program Director
Michael E. Townsend, B.A., M.A., Ph.D., J.D., Associate Professor of Law
Philip A. Trautman, B.A., J.D., Professor of Law
Lee B. Vaughan, B.A., J.D., Professor of Law
Walter J. Walsh, B.C.L., L.L.M., S.J.D., Assistant Professor of Law
Lis W. Wehl, A.B., M.A., J.D., Associate Professor of Law
Jane K. Winn, B.S., J.D., Professor of Law
Louis E. Wolcher, B.A., J.D., Professor of Law

ADJUNCT AND AFFILIATE FACULTY
G. Andrew H. Benjamin, B.A., M.A., J.D., Ph.D., Affiliate Professor of Law
Daniel M. Bodansky, A.B., M.Phil., J.D., Affiliate Professor of Law
Sharon E. Brown, B.A., M.A., J.D., Ed.D., Adjunct Research Associate Professor of Law
Daniel H. Foote, A.B., J.D., Affiliate Professor of Law
John O. Haley, B.B., L.L.B., L.L.M., Affiliate Professor of Law
Marc J. Heuterman, A.B., J.D., Adjunct Professor of Law
Kenneth E. Himma, B.A., M.A., J.D., Ph.D., Adjunct Lecturer of Law
Nancy S. Jecker, B.A., M.A., Ph.D., Adjunct Professor of Law
CONTENTS

LUVERN V. RIEKE: IN MEMORIAM  William B. Stoebuck 689

ARTICLES

Fish as Pollutants: Limitations of and Crosscurrents in Law, Science, Management, and Policy  Jeremy Firestone
Robert Barber 693

NOTES & COMMENTS

Miranda’s Poisoned Fruit Tree: The Admissibility of Physical Evidence Derived from an Unwarned Statement  Kirsten Lela Ambach 757

Dr. Jekyll’s Waiver of Mr. Hyde’s Right to Refuse Medical Treatment: Washington’s New Law Authorizing Mental Health Care Advance Directives Needs Additional Protections  Nick Anderson 795

The Bonds of Joint Tax Liability Should Not Be Stronger than Marriage: Congressional Intent Behind § 6015(c) Separation of Liability Relief  Svetlana G. Attestatova 831

When Animals Invade and Occupy: Physical Takings and the Endangered Species Act  Rebecca E. Harrison 867


Aboriginal Title or the Paramountcy Doctrine? Johnson v. McIntosh Flounders in Federal Waters off Alaska in Native Village of Eyak v. Trawler Diane Marie, Inc.  Andrew P. Richards 939
University of Washington School of Law 2002–2003

OFFICERS OF ADMINISTRATION
Lee L. Huntsman, B.A., Ph.D., Interim President of the University
W. H. Knight, Jr., B.A., J.D., Dean, Professor of Law
Patricia C. Kuszler, B.A., M.D., J.D., Associate Dean for Research & Faculty Development, Professor of Law
Penny A. Hazleton, B.A., J.D., M.L.L., Associate Dean for Library & Computing Services, Professor of Law
Dexter Bailey, B.S., M.B.A., Assistant Dean for Development
Mary A. Hotchkiss, B.A., M.B.L., J.D., L.L.M., Assistant Dean for Academic Services
Sandra E. Madrid, B.S., Ph.D., Assistant Dean for Student Affairs and Administration

FACULTY EMERITI
William T. Burke, B.S. J.D., J.S.D.
Robert L. Fletcher, A.B., LL.B.
John Huston, B.A., J.D., LL.M.
Arval Morris, B.A., M.A., J.D., LL.M., LL.D.
Cornelius J. Peck, B.S., LL.B.
John R. Price, A.B., LL.B.

FACULTY OF LAW
Craig H. Allen, B.S., J.D., Professor of Law
William R. Andersen, B.S.L., LL.B., LL.M., Judson Falknor Professor of Law
Helen A. Anderson, B.A., J.D., Senior Lecturer
Robert T. Anderson, B.A., J.D., Assistant Professor of Law, Native American Law Center Director
Thomas R. Andrews, B.A., M.A., J.D., Professor of Law
Robert H. Aronson, B.A., J.D., Professor of Law
Melissa M. Berry, B.A., J.D., Lecturer
Karen Boxa, B.A., J.D., Associate Professor of Law
Steve P. Calandrillo, B.A., J.D., Assistant Professor of Law
Donald C. Clarke, A.B., M.Sc., J.D., Professor of Law, Wash. Law School Foundation Scholar
Sam A. Donaldson, B.A., J.D., LL.M., Assistant Professor of Law
G. Meade Emory, A.B., LL.B., LL.M., Professor of Law, Director of Graduate Program in Taxation
Joan M. Fitzpatrick, B.A., J.D., Jeffery and Susan Brotman Professor of Law
Julia Gold, B.A., J.D., Senior Lecturer
James H. Hardisty, A.B., LL.B., Professor of Law
Gregory A. Hicks, B.A., J.D., Professor of Law
Roland L. Hjorth, A.B., LL.B., Dean Emeritus and Garvey, Schubert, & Barer Professor of Law
Linda S. Humen, B.A., J.D., Professor of Law
Stewart M. Jay, A.B., J.D., William L. Dwyer Professor of Law
John M. Junker, B.A., J.D., Professor of Law
Lisa Kelly, B.A., J.D., Harry M. Cross Distinguished Visiting Professor of Law
Alan Kirtley, B.A., J.D., Associate Professor of Law, Director of Clinics
Richard O. Kimmert, B.S., M.B.A., LL.B., Dwayne & Anne Gittinger Professor of Law
Deborah Maranville, B.A., J.D., Associate Professor of Law
Anna Mastrianni, B.S./B.A., J.D., M.P.H., Assistant Professor of Law
Kathleen M. McGinnis, B.A., J.D., Senior Lecturer
Jacqueline McNamara, B.G.S., J.D., Senior Lecturer
Peter Nicolas, B.A., M.P.P., J.D., Assistant Professor of Law
Sean O’Connor, B.A., J.D., Visiting Assistant Professor of Law
Kathleen O’Neill, B.A., J.D., Associate Professor of Law, Director of Legal Research and Writing
Roy L. Prostman, A.B., LL.B., Professor of Law
Anita Ramesh, B.A., M.A., J.D., Assistant Professor of Law
William H. Rodgers, Jr., B.A., LL.B., Stimson Bullitt Professor of Environmental Law
Eric Schnapper, B.A., M.A., B.Phil., LL.B., Pendleton Miller Professor of Law
Scott A. Schumacher, B.A., J.D., Lecturer
William B. Stoebuck, B.A., M.A., J.D., S.J.D., Judson Falkner Professor of Law
Toshiko Takenaka, B.A., LL.M., Ph.D., Associate Professor of Law, Director, CASRIP
Veronica L. Taylor, LL.B., B.A., LL.M., Professor of Law, Asian Law Program Director
Michael E. Townsend, B.A., M.A., Ph.D., J.D., Associate Professor of Law
Philip A. Trautman, B.A., J.D., Professor of Law
Leo B. Vaughn, A.B., J.D., Professor of Law
Walter J. Walsh, B.C.L., LL.M. S.J.D., Assistant Professor of Law
Lisa W. Wiehl, A.B., M.A., J.D., Associate Professor of Law
Jane K. Winn, B.S., J.D., Professor of Law
Lea F. Wolcher, B.A., J.D., Professor of Law

ADJUNCT AND AFFILIATE FACULTY
G. Andrew H. Benjamin, B.A., M.A., J.D., Professor of Law
Daniel M. Bodansky, A.B., M.Phil., J.D., Affiliate Professor of Law
Sharon E. Brown, B.A., M.A., J.D., Ed.D., Adjunct Research Associate Professor of Law
Daniel H. Foote, A.B., J.D., Affiliate Professor of Law
John O. Haley, A.B., LL.B., LL.M., Affiliate Professor of Law
Marc J. Hershman, A.B., J.D., Adjunct Professor of Law
Kenneth E. Himma, B.A., M.A., Ph.D., Adjunct Lecturer of Law
Nancy S. Jecker, B.A., M.A., Ph.D., Adjunct Professor of Law
DEDICATION TO JOAN FITZPATRICK

Joan Fitzpatrick: In Memoriam  
Roland L. Hjorth  973

In Memory of Joan Fitzpatrick: A Former Student’s Perspective  
Elizabeth J. Kane  977

Obituary: Joan Fitzpatrick: Human Rights Lawyer and Activist, She Fought for Refugee Rights  
Irene Khan  981

ARTICLE

Unpacking New Policing: Confessions of a Former Neighborhood District Attorney  
Alafair S. Burke  985

NOTES & COMMENTS

Stephanie D. Matheny  1067

Abrogation or Regulation? How Anderson v. Evans Discards the Makah’s Treaty Whaling Right in the Name of Conservation Necessity  
Zachary Tomlinson  1101
2003–2004 EDITORIAL BOARD*

Editor-in-Chief
REBECCA E. HARRISON

Managing Editors
Svetlana Attestatova
Natasha Shekdar Black
Nichole Chiappini
Katie Co lendich
Alicia Ozanich

Executive Notes and Comments Editors
Derek D. Green
Andrew Pearson Richards

Thesis Editors
Lainie M.C. Dillon
Cristina Underwood

Notes and Comments Editors
Timothy Allen
Kirsten Ambach
Michael Caughey
Matt Kernutt
Michael Pedhirney

Executive Articles Editors
Jennifer Amanda Krebs
Zachary Tomlinson

Articles Editors
Rachel Byrne
William A. Denton
Michael Estok
Patrick McKenna
Mark Murdock
Kate Vaughan

Associate Editors-in-Chief
Nick Anderson
Joseph Rehberger

Special Projects Editor
Teresa B. Sumearll

EDITORIAL STAFF
Michael Beers
Matthew Berry
Sarah L. Bird
John Brust
Anne Chiu
Emily Cordo
Kimberly Cozzutto
Kelly Fennerety
Dorigen Fried
Josh Gaul
Megan Grembowski
Kyla Grogan
Lindsay Halm
Scott Holleman
Michelle Jensen
Melanie Mayer
Heather McKimmie
Lynette Meachum
Jennifer Murray
David Perkins
Tad O’Neill
Jill Val lely
Holly Vance
Jungmin Jennifer Yoo

*Washington Law Review issues are produced on a calendar-year schedule; the Editorial Board’s term of office, however, corresponds to the school year. Accordingly, editorial responsibilities for November issues are apportioned between consecutive Boards.
University of Washington School of Law 2003–2004

OFFICERS OF ADMINISTRATION

Lee L. Huntsman, Ph.D., Interim President of the University
W. H. Knight, Jr., B.A., J.D., Dean, Professor of Law
Patricia C. Kauzlar, B.A., M.D., J.D., Associate Dean for Research & Faculty Development, Professor of Law
Penny A. Hazelton, B.A., J.D., M.L.L., Associate Dean for Library & Computing Services, Professor of Law
Sandra E. Madrid, B.S., Ph.D., Assistant Dean for Student Affairs & Administration
Dexter Bailey, B.S., M.B.A., Assistant Dean for Development
Mary A. Hotchkiss, B.A., M.S.L.S., J.D., L.L.M., Associate Dean for Academic Services

FACULTY EMERITI

William T. Burke, B.S., J.D, J.S.D.
Harry M. Cross, B.A., J.D.
Robert L. Fletcher, A.B., LL.B.
John Huston, B.A., B.D., LL.M.
Arval Morris, B.A., M.A., J.D., L.L.M., L.L.D.
Cornelius J. Peck, B.S., LL.B.
John R. Price, A.B., LL.B.
Marjorie D. Rombauer, B.A., J.D.
Charles Z. Smith, B.S., J.D.
Frank W. Smith, Jr., B.A., J.D., L.L.M.

FACULTY OF LAW

Craig H. Allen, B.S., J.D., Professor of Law
William R. Andersen, B.S., LL.B., LL.M., Judson Falkner Professor of Law
Helen A. Anderson, B.A., J.D., Senior Lecturer
Robert T. Anderson, B.A., J.D., Assistant Professor of Law, Native American Law Center Director
Thomas R. Andrews, B.A., M.A., J.D., Professor of Law
Robert H. Aronson, B.A., J.D., Professor of Law
Kathryn Battuello, B.A., J.D., M.P.H., Research Assistant Professor of Law
Karen Boxx, B.A., J.D., Associate Professor of Law
Steve P. Calandrillo, B.A., J.D., Associate Professor of Law
Donald C. Clarke, A.B., M.S., J.D., Professor of Law
William Covington, B.A., J.D., Assistant Professor of Law
Sam A. Donaldson, B.A., J.D., M.L., Assistant Professor of Law
G. Meade Emory, B.A., J.D., LL.M., Professor of Law, Director of Gradate Program in Taxation
Julia Gold, B.A., J.D., Senior Lecturer
Robert W. Gormikiewicz, B.A., M.A., J.D., Research Associate Professor of Law, Director of Graduate Program in Intellectual Property Law & Policy
James H. Hardisty, A.B., LL.B., Professor of Law
Gregory A. Hicks, B.A., J.D., Professor of Law
Roland L. Hjorth, A.B., LL.B., Dean Emeritus, Garvey, Schubert & Barer Professor of Law
Linda W. Hume, B.A., J.D., Professor of Law
Stewart M. Jay, A.B., J.D., William L. Dwyer Professor of Law
John M. Junker, B.A., J.D., Professor of Law
Lisa Kelly, B.A., J.D., Professor of Law
Alan Kirtley, B.A., J.D., Associate Professor of Law, Director of Clinics
Richard O. Kummert, B.S., M.B.A., LL.B., D. Wayne & Anne Gittinger Professor of Law
Deborah Maranville, B.A., J.D., Professor of Law
Anna Mastroianni, B.S./B.A., J.D., Professor of Law
Kathleen M. McCrinius, B.A., J.D., Senior Lecturer
Jacqueline McMurtrie, B.G.S., J.D., Assistant Professor of Law
Peter Nicolas, B.A., M.P.P., J.D., Associate Professor of Law
Sean O'Connor, B.A., M.A., J.D., Assistant Professor of Law
Kathleen O'Neil, B.A., J.D., Associate Professor of Law, Director of Legal Research and Writing
Roy L. Prosterman, A.B., LL.B., Professor of Law
Amita Ramaswamay, B.A., M.A., J.D., Associate Professor of Law
William H. Rodgers, Jr., B.A., LL.B., Professor of Law
Eric Schnapper, B.A., M.A., B.Phil., LL.B., Professor of Law
Scott A. Schumacher, B.A., J.D., Lecturer
William B. Stoeckel, B.A., M.A., J.D., Professor of Law Emeritus
Toshiko Takenaka, B.A., LL.M., Ph.D., Professor of Law, Director, CASRIP
Veronica L. Taylor, B.A., LL.B., LL.M., Professor of Law, Asian Law Program Director
Michael E. Townsend, B.A., M.A., J.D., Associate Professor of Law
Philip A. Trautman, B.A., J.D., Professor of Law
Lea B. Vaughn, A.B., J.D., Professor of Law
Walter J. Walsh, B.C.L., LL.M, J.D., Associate Professor of Law
Lis W. Wiehl, A.B., M.A., J.D., Associate Professor of Law
June K. Winn, B.Sc., J.D., Professor of Law
Ron J. Whittenher, B.A., J.D., Assistant Professor of Law
Louis E. Wolcher, B.A., J.D., Professor of Law

ADJUNCT AND AFFILIATE FACULTY

G. Andrew H. Benjamin, B.A., M.A., J.D., Ph.D., Affiliate Professor of Law
Daniel M. Bodansky, A.B., M.Phil., J.D., Affiliate Professor of Law
Sharon E. Brown, B.A., M.A., J.D., Ed.D., Adjunct Research Associate Professor of Law
Daniel H. Foote, A.B., J.D., Affiliate Professor of Law
John O. Haley, A.B., LL.B., LL.M., Affiliate Professor of Law
Marc J. Hershman, A.B., J.D., Adjunct Professor of Law
Kenneth E. Himma, B.A., M.A., J.D., Ph.D., Adjunct Lecturer of Law
RECOGNIZING THE SOCIETAL VALUE IN INFORMATION PRIVACY

James P. Nehf*

Abstract: Much has been written about database privacy in the Internet Age, most of it critical of the way in which the American legal system addresses the issue. In this article, Professor Nehf maintains that one of the fundamental difficulties with the public policy debates is that information privacy is often discussed as a typical consumer problem rather than a problem of more general societal concern. As a result, arguments over appropriate resolutions reduce to a balancing of individual rights against more general societal interests, such as increased efficiency in law enforcement, government operations or commercial enterprise. Although privacy scholars discussed the “societal value” of information privacy in the 1960s and early 1970s, the concept was not fully developed. More recently, political theorists have revived the idea and argue the importance of recognizing privacy as a societal norm. Professor Nehf adopts a functional analysis that compares information privacy to other societal values, such as environmental protection, and concludes that privacy policy could take a different form if the issue were viewed in this way.

Information about us—countless bits and bytes1—exists in computer databases that are seemingly everywhere. Government records hold our salary histories and track our changes of address, and state motor vehicle departments maintain our driving histories. Much more information is now held in the private sector. It seems that everyone from the local grocery store to GOOGLE.com is collecting information about each of us at every point of contact.

1. For most computer users, the terms bit and byte are interchangeable. There is a technical distinction. A bit is simply a binary representation of 0 or 1. In current use, the word byte is a collection of eight bits. The word bit in this context appears to have been a creation of John Tukey, a computer technician who in 1949 was searching for a short term for a “binary digit.” Bit (contracted from binary digit) already carried the definition of a “small part.” The need for a term representing a collection of bits soon became apparent, since a set of bits was necessary to store, process or transfer any useful character. Although there is some dispute about the origin of “byte,” it seems that Werner Buchholz coined the term in the 1950s as a six-bit unit during the development of the IBM “stretch computer.” In 1956, the term was used in the development of the IBM System 360 to represent an eight-bit set, and has carried this meaning to the present. According to the Oxford English Dictionary, the word “byte” first appeared in the IBM “systems journal” of 1964. LINDA & ROGER FLAVELL, THE CHRONOLOGY OF WORDS AND PHRASES 246 (1999).
Americans have only the vaguest idea how much of their lives is recorded in databases, and how little control they have over the collection and sharing of that data. People understand that scattered data exist in government and business computers, but they are only beginning to understand the power of information technology, the widespread and fast-growing data aggregation industry, and the harm that can result from information collection and sharing.2

Although information about us has been recorded, filed, and manipulated by government authorities and businesses for decades, the database problem took a leap forward with the proliferation of Internet-linked computers3 in virtually every office and home. As new

2. See Jeff Sovern, Opting in, Opting Out, or No Options at All: The Fight for Control of Personal Information, 74 WASH. L. REV. 1033, 1057–63 (1999). Public concern about information privacy has remained consistently high over the last several years. See Am. Soc. Newspaper Editors, at http://www.epic.org/privacy/survey/ (April 2001) (51% of respondents “very concerned” and 30% “somewhat concerned” that a company might violate their personal privacy); First Amend. Center and Am. Soc. Newspaper Editors Freedom of Info. Comm., Freedom of Information in the Digital Age, at http://www.freedomforum.org/publications/first/foi/fointhedigitalage.pdf (April 2001) (61% of respondents were “very concerned” about “personal privacy,” 38% were “more concerned” about personal privacy since they gained access to the Internet); John Schwarz, Government Is Wary of Tackling Online Privacy, N.Y. TIMES, Sept. 6, 2001, at C1 (characterizing privacy as one of the most challenging issues for policy makers); Forrester Research, Inc., Forrester Technographics Finds Online Consumers Fearful Of Privacy Violations (Oct. 1999) (“Nearly 90% of online consumers want the right to control how their personal information is used after it is collected.”) (quoting Christopher M. Kelley, associate analyst in Technographics Data & Analysis), at http://www.forrester.com/ER/Press/Release/0,1769,177,FF.html (last visited Jan. 5, 2003); Susan Fox, Trust and Privacy Online: Why Americans Want To Rewrite The Rules, THE INTERNET LIFE REPORT, at http://www.pewinternet.org/reports/toc.asp?Report=19 (Aug. 20, 2000) (60% of all Americans are “very concerned about privacy”); Electronic Privacy Information Center (EPIC), Public Opinion on Privacy, at http://www.epic.org/privacy/survey/ (last updated July 16, 2002) (summarizing a collection of recent survey results).

A 1998 survey showed that 89% of the public was concerned about threats to personal privacy. ALAN F. WESTIN & DANIELLE MAURICI, E-COMMERCE & PRIVACY: WHAT NET USERS WANT 7 (1998). Another survey conducted in the same year found that 88% of consumers were concerned. See Executive Summary: 1998 Privacy Concerns & Consumer Choice Survey, at http://www.privacyexchange.org/iss/surveys/1298execsum.html (December 15, 1998). See also Alan F. Westin, Whatever Works: The American Public’s Attitudes Toward Regulation and Self-Regulation on Consumer Privacy Issues, Conference Report, DATA PROTECTION IN THE GLOBAL SOCIETY, available at http://www.privacyexchange.org/iss/confer/aicsberlin.html (November 15, 1996); Simson L. Garfinkel, How Computers Help Target Buyers, CHRISTIAN SCI. MONITOR, July 25, 1990, at 13 (quoting Bonnie Guiton, Special Advisor for Consumer Affairs to President George Bush: “A major concern of mine is that consumers are uninformed . . . . In most cases, they don’t even know that [information on them] is being collected.”).

3. Although the word “internet” dates from the late 19th Century (in an 1883 volume of NATURE, an author wrote of “[t]he marvelous maze of internetted motions”), the Internet as we know it began in 1969 when researchers at UCLA, Stanford, UC-Santa Barbara and Utah linked computers to form part of the Advanced Research Projects Agency. Until the early 1980s, however, the users of the
Information Privacy

technologies allow businesses and public authorities to collect and process information with increasing speed and sophistication, we can expect the data collection and dissemination problem to become more threatening in the years to come. There is an enormous amount of information about us in other people’s hands, and one thing is certain—some of us will be harmed by it. We just don’t know who, when, or how badly.

Our perception of the database problem is conflicting and ambiguous because many uses of our information are benign, or even beneficial. Yet, however, other uses (or misuses), such as identity theft, can have devastating consequences. The problem we currently face is thus not merely that a vast amount of information is resting in databases, but that we have very little control over that information—how it is used, shared, and manipulated—once it is “out there.” We are at the mercy of those who hold our data. We trust them to guard it and use it in ways that will help and not hurt us.

linked computer network were limited to a small number of scientists using large, mainframe computers. As smaller, less expensive computers became available to universities, businesses and individuals, commercial, consumer and academic exploitation of the Internet grew rapidly. Acceptance of the TCP/IP standard language in 1982 accelerated Internet usage, allowing previously incompatible systems to communicate with a common set of protocols. FLAVELL, supra note 1, at 255–56.

4. According to the Federal Trade Commission (FTC), identity theft represented 42% of all consumer fraud complaints in 2001, making it the fastest growing category. The FTC Theft Data Clearinghouse received more than 86,000 identity theft complaints in 2001, more than doubling the number from the previous year. See Kelly Lucas, Losing Identity, Saving Face, 13 Ind. LAWYER, July 3–16, 2002, at 9. See also the FTC website for identity theft at http://www.consumer.gov/idtheft/ (last updated Jan. 3, 2003). The identity theft problem hit the front pages in November 2002 when the federal government filed a criminal indictment against several defendants for allegedly stealing the personal, credit and banking information of approximately 30,000 individuals and selling it in a vast criminal conspiracy, resulting in millions of dollars of losses. United States v. Cummings, No. 02 MAG 2354 S.D.N.Y. (Nov. 22, 2002).

5. With modern data processing technology, the original reason for collecting the data becomes irrelevant. Once it has been stored, it can be used or shared in a variety of ways unconnected with the original purpose. See Spiros Simitis, Reviewing Privacy in an Information Society, 135 U. Pa. L. REV. 707, 711 (1987). According to a study of 1,017 Internet users, 86% of respondents were concerned that personal information about them or their families could end up in the hands of businesses or people they did not know. See Fox, supra note 2.

6. Disclosure of information can be intentional or accidental, but the potential for injury is the same in either situation. In July 2000, as part of its bankruptcy plan, online toy retailer ToySmart attempted to sell information in its customer list despite language in its privacy policy to the contrary. Greg Sandoval, ToySmart Creditor Targets Disney in Lawsuit, CNET News.com, at http://news.cnet.com/news/0-1007-200-2759944.html (Sept. 12, 2000). In April 2000, the DeBeers website, www.adiamondisforever.com, exposed the names, home addresses and e-mail addresses of 35,000 of its customers. See Stephanie Olsen, DeBeers Security Hole Reveals Customer Information,
Surveys dating from the 1970s show that public support for information privacy has been consistently strong. Americans will not tolerate abuse or misuse of information technology at the expense of their personal privacy, and they overwhelmingly support action to do something about it. Yet survey results also reveal that a great number of us understand that our interests in privacy must be balanced against other interests, i.e., the multitude of benefits resulting from more efficient government, business, and law enforcement functions when information in digital form is readily accessible. Government program administrators can process claims and detect fraud more easily if employment, personal history, and salary data are cross-referenced. Medical treatment can be more effective if physicians have access to our medical histories and prescription records online. Law enforcement can be strengthened if criminal records, employment, education, and immigration data can be accessed, matched, sorted, and correlated more easily. Direct marketing
Information Privacy

can be better tailored to individual preferences if our preferences are better known to the marketing organization.13

Since there are benefits and risks associated with information collection and data sharing, policy makers must attempt to strike a balance. In doing so, they must first define the problem. This is a critical step in the formulation of public policy because the way in which a problem is defined on the public agenda will affect its ultimate resolution.14 In the United States, information privacy has historically been defined as an individual concern rather than a general societal value or a public interest problem.15 This has influenced the resulting public policy solutions, yet it may not be the most effective way to approach modern privacy concerns.

Most consumer problems (e.g., defective goods, unfair trade practices, predatory lending, etc.) are viewed primarily as individual concerns. This means that public policy resolutions are characterized by laws that impose legal obligations on businesses, but injuries are seen as individual in nature (though sometimes aggregated for convenience in a class action). Enforcement largely depends on individuals recognizing an injury and seeking redress when the legal norms are breached. “Lemon Laws” or the consumer credit acts are examples of this regulatory approach. Although there may be important agency oversight (Federal Trade Commission, state attorneys general, or Federal Reserve Board), consumers assume a large responsibility for identifying their own injuries, policing the market by making informed decisions, and enforcing their rights, usually through litigation, when legal norms are breached.16

In contrast, when a problem is viewed as a general societal concern, and a resolution in the public interest is sought, enforcement of the legal norm is primarily through government agency oversight and regulation. Resolutions are characterized by the imposition of general standards, reporting requirements, periodic audits, government investigations, and

(2001), to aid law enforcement in identifying terrorist activities and other criminal enterprises. Much of the Act seeks to enhance law enforcement efficiencies by expanding government access to financial, student, medical, travel and other records for surveillance purposes. Id.

14. See REGAN, supra note 7, at xiii.
15. See discussion infra at Part II.C.
16. See infra note 342 and accompanying text.
remedies sought for the general welfare rather than for specific individuals. Moreover, to manage general societal concerns, our elected representatives create a regulatory regime in hopes of minimizing our collective injury before it occurs, not relying as heavily on individual enforcement and compensation for those who are injured by breach of the law. Environmental policy and food and drug laws are illustrative. We do not expect, as the primary control mechanism, individuals to seek redress for injuries resulting from contaminated water or dangerous pharmaceuticals. Although private remedies are an important supplement to the regulatory scheme, a foundation of government regulatory oversight is designed to minimize harm in the first instance.

During the 1960s and 1970s, there was some discussion in debates and commissioned studies about information privacy having a general societal value as well as being a matter of individual concern. Congress ultimately concluded, however, that privacy policy should begin with a voluntary, market-oriented approach, with reliance on individual self-policing as the dominant means of control. Proposals for a federal “Privacy Board,” for example, were rejected.

To this day, information privacy legislation in the United States has placed heavy reliance on individuals policing their own data records and protecting their own information from unintended use. For example, the

---


18. See Stephen Breyer, Analyzing Regulatory Failure: Mismatches, Less Restrictive Alternatives, and Reform, 92 Harvard L. Rev. 549, 556, 560 (1979) (recognizing that government action may be called for when individual use of available remedies is expensive or impractical as an effective means of addressing a problem).


20. See infra note 143 and accompanying text. Cf. Joseph I. Rosenbaum, Privacy On the Internet: Whose Information is it Anyway?, 38 Jurimetrics J. 565, 566 (1998) (noting that our notion of privacy has been and always will be a moving target, dependent on technological capability, societal values and cultural norms).

21. See infra notes 144–57 and accompanying text.

Fair Credit Reporting Act of 1970 (FCRA)\textsuperscript{23} imposes limits on the collection and sharing of credit histories by credit bureaus. Success of the FCRA depends largely on individuals monitoring compliance by keeping their credit reports complete and accurate. More recently, the Gramm-Leach-Bliley Act of 1999\textsuperscript{24} imposes limitations on the sharing of information in the financial services industry. The Gramm-Leach-Bliley Act allows affiliated companies to share data among each other and requires individuals to take affirmative action to prevent data sharing outside the affiliated group (so-called “opt out” legislation).\textsuperscript{25}

This article argues that, in the modern digital world, information privacy should be viewed as a societal value justifying a resolution in the public interest, much like environmental policy and other societal concerns, with less emphasis on individual self-policing and market-based mechanisms. In doing so, this article does not invoke abstract notions of natural rights, fundamental values, or the preservation of human dignity, as other defenders of privacy protection have argued in recent years.\textsuperscript{26} Its point is more functional and pragmatic. If we look at the way in which information is collected and used in today’s society, we see that the problems presented are not typical consumer issues that we can expect individuals to police for themselves with the aid of prohibitory laws. The policy issues have much more in common with societal problems that we have historically regulated in a fundamentally different way. Policy makers should recognize this relationship in the formulation of privacy legislation and create a regulatory environment that provides meaningful protection of our collective privacy interests.

Part I discusses the reasons why we have difficulty thinking about information privacy as a public policy issue, and concludes that we are only beginning to focus on the most troubling aspect of the problem—lack of control over our information once it has been revealed or

\begin{footnotes}
\footnotetext{23}{15 U.S.C. § 1681 (2000); see also generally United States Senate Committee on Government Operations, supra note 22.}
\footnotetext{24}{15 U.S.C. §§ 6801–6827.}
\footnotetext{25}{Id. § 6802.}
\footnotetext{26}{See, e.g., Richard S. Murphy, Property Rights in Personal Information: An Economic Defense of Privacy, 84 Geo. L.J. 2381, 2393 (1996) (discussing the economic implications of treating personal information as a property right); Glenn Negley, Philosophical Views on the Value of Privacy, 31 Law & Contemp. Probs. 319, 320 (1966) (modern thinkers have proposed and now reject the idea that privacy is a “natural right”); Moore v. City of East Cleveland, 431 U.S. 494, 503–05 (1977) (noting privacy is a basic value, part of the nation’s history and traditions); Edward J. Bloustein, Privacy as an Aspect of Human Dignity: An Answer to Dean Prosser, 39 N.Y.U. L. Rev. 962, 1000–07 (1964). See also infra at Part III.A.}
\end{footnotes}
gathered. Part II examines the origins of privacy law in the United States that resulted in our treating information privacy as a problem of individual, rather than societal, concern. The principle control mechanisms—a combination of self-regulation, individual self-policing, and market-driven controls—are a natural consequence of this conception of privacy, and they are unlikely to produce the degree of privacy protection that most Americans deserve and expect. Part III makes a case for viewing information privacy as a more general societal concern justifying a higher level of protection (or at least a different way of seeking public policy resolutions). This Article concludes that regulation similar to the European model of privacy protection, in which the issue is framed as a foundation of social protection, should prevail in the United States.

I. CONCEPTUALIZING THE DATABASE PRIVACY PROBLEM

A. Why Does Data Collection Bother Us?

Public concern about information collection and storage is an ancient problem. In the 11th century, William the Conqueror compiled a “Doomsday” survey, which collected information on each of his English subjects, ostensibly for taxation and other state purposes.27 For centuries thereafter, government authorities and commercial enterprises have collected, organized, and shared information about the general population or certain segments of it. Only recently, however, have new information technologies amplified the problem to an unprecedented degree.

Technology usually enables and empowers people to perform tasks that they could not previously perform, thus producing many benefits to individuals and society at large. Yet technological developments can also introduce previously unknown problems. For example, the printing press allowed information to spread rapidly among the citizenry and undermined the authority of those who previously controlled its

27. REGAN, supra note 7, at 69. The people of England referred to William’s survey as the Doomsday book, a Middle English spelling of doomsday. The root “dómi” meant “law or decree” during this period but it could also mean “judgment, sentence or condemnation.” See FLAVELL, supra note 1, at 30. Historical evidence suggests that William’s subjects probably were not concerned about the Doomsday book infringing on their privacy, since privacy as a societal concern developed centuries later. See FERNAND BRAUDEL, CAPITALISM AND MATERIAL LIFE, 1400–1800, 224 (1973) (describing the idea of privacy as an “an eighteenth century innovation”).
Information Privacy
dissemination. Yet the press could be censored, misinformation could be
spread, and publications could be used to expose damaging facts or
spread rumors about a person to a wider audience than was previously
possible. The telephone opened communication links to the world, but
phone lines could be tapped or recorded, calling records traced, and
personal information revealed. Radio and television have been powerful
instruments for education and information sharing, but also can be used
for manipulation, pacification, and political propaganda.28

Today’s information technologies present tradeoffs as well. While
some individuals object to the collection and distribution of personal
information for virtually all purposes,29 most recognize and appreciate
the many benefits of data storage and information sharing technologies.
We appreciate the efficiencies brought about by the revolution in
information processing, but we are concerned about how personal
information might be used. We may not care if a grocery store computer
“knows” that a woman buys three bottles of wine and a carton of
cigarettes each week if the store uses the information to send her some
useful discount coupons. However, most of us would object if the store

28. See generally RONALD J. DEIBERT, PARCHMENT, PRINTING, AND HYPERMEDIA:
COMMUNICATIONS IN WORLD ORDER TRANSFORMATION 47–110 (1997).

29. See Fox, supra note 2 (stating that 27% of Internet users are “hard-core privacy protectionists”
that would never provide personal information, 10% would be willing to provide it under the right
circumstances, and 54% of Internet users have provided personal information in order to use a Web
site). Alan Westin divided the American public into three groups regarding attitudes towards privacy:

“Fundamentalists,” about twenty-five percent of the American public, who rate privacy as an
extremely high value, are loathe to trade this for promised benefits to them or to society, and
generally favor legislative standards and government regulation.

At the opposite pole are the “privacy unconcerned,” about twenty percent of the American
public, who are generally ready to give personal information about themselves in order to get
consumer benefits and support government programs, and are not at all worried about
intrusiveness.

This leaves the “privacy pragmatists,” fifty-five percent of the American public and clearly the
“swing group” in setting public norms. The pragmatists are willing to listen to possible benefits
to them or to society from disclosing their personal information and weigh those values against
the important privacy interests involved. If they feel the benefits are meaningful, they next look
for meaningful safeguards—basically, the fair information practices elements—and decide
whether they trust these to be provided by private standards or whether they feel laws are
needed. Whether private standards are accepted generally depends on the trust the public has in
particular industries or government agencies to handle their information in a responsible way.

Alan F. Westin, Whatever Works: The American Public’s Attitudes Toward Regulation and Self-
Regulation on Consumer Privacy Issues, Conference Report, DATA PROTECTION IN THE GLOBAL
SOCIETY, available at http://www.privacyexchange.org/iss/confro/aisgsberlin.html (November 15,
1996).
gives or sells that information to a health insurance company for purposes of evaluating her insurable risk. We may not mind that a dinner guest can log onto Anywho.com and get a map to our home or office simply by inputting a last name and home town. Yet we likely would not feel the same about a stranger who overheard a young girl’s name in a casual conversation at the shopping mall.

Because of these concerns about how personal information is used, the most often quoted definition of database privacy is “the right to control information about ourselves.”30 The lack of control over information has long been a concern of privacy advocates. Many have invoked the “Big Brother” metaphor from George Orwell’s novel 1984 to describe the threat to privacy that databases present.31 Orwell’s fictional citizenry feared the totalitarian government in part because they lost control over vast amounts of personal information that Big Brother had collected, and

30. See, e.g., Jeffrey Rosen, The Purposes of Privacy: A Response, 89 GEO. L.J. 2117, 2120 (2001) (discussing privacy in terms of the “ability to exercise control over personal information”). One of the first to discuss information privacy in these terms was Charles Fried, in Privacy, 77 YALE L.J. 475, 482 (1968) (discussing information privacy as “the control we have over information about ourselves”) (emphasis in original). Privacy has been defined in various other ways. See JUDITH WAGNER DECEW, IN PURSUIT OF PRIVACY: LAW, ETHICS AND THE RISE OF TECHNOLOGY 58 (1997) (“whatever is not generally . . . a legitimate concern of others”); THOMAS M. COOLEY, THE LAW OF TORTS 29 (1888) (defining privacy as the “right to be let alone”); Ruth Gavison, Privacy and the Limits of Law, 89 YALE L.J. 421, 428 (1980) (“limitation of others’ access to any individual”).

31. The list of writers who have invoked the “Big Brother” metaphor in privacy literature is endless. See, e.g., REG WHITAKER, THE END OF PRIVACY: HOW TOTAL SURVEILLANCE IS BECOMING A REALITY 160 (1999) (chapter titled “Big Brother Outsourced: The Globalized Panopticon”); Daniel J. Solove, Privacy and Power: Computer Databases and Metaphors for Information Privacy, 53 STAN. L. REV. 1393, 1396 (2001) (“commentators have adapted the Big Brother metaphor to describe the threat to privacy”); A. Michael Froomkin, The Death of Privacy?, 52 STAN. L. REV. 1461, 1463 (2000) (noting that at least since George Orwell’s 1984, the image of the all-seeing eye has been synonymous with the power to exercise repression); Bryan S. Schultz, Electronic Money, Internet Commerce, and the Right to Financial Privacy: A Call for New Federal Guidelines, 67 U. CIN. L. REV. 779, 797 (1999) (“society inches closer to fulfilling George Orwell’s startling vision of a nation where ‘Big Brother’ monitors the who, what, where, when, and how of every individual’s life”); Katrin Schatz Byford, Privacy in Cyberspace: Constructing a Model of Privacy for the Electronic Communications Environment, 24 RUTGERS COMPUTER & TECH. L.J. 1, 50 (1998) (“Life in cyberspace, if left unregulated, thus promises to have distinct Orwellian overtones—with the notable difference that the primary threat to privacy comes not from government, but rather from the corporate world.”); Alan F. Westin, Privacy in the Workplace: How Well Does American Law Reflect American Values, 72 CHI.-KENT L. REV. 271, 273 (1996) (expressing skepticism that the creation of government data protection boards would be like “calling on ‘Big Brother’ to protect citizens from ‘Big Brother’”); See also McVeigh v. Cohen, 983 F. Supp. 215, 220 (D.D.C. 1998) (“In these days of ‘big brother,’ where through technology or otherwise the privacy of individuals from all walks of life are being ignored or marginalized”).
they feared how it might be used against them.\textsuperscript{32} Since Orwell’s Big Brother was a pervasive government surveillance mechanism, writers have sometimes referred to private sector databases as an amalgamation of “Little Brothers” or similar metaphors that invoke the same overpowering surveillance concern.\textsuperscript{33}

Along the same lines, writers in recent times have reached further back in literary history to recall Jeremy Bentham’s horrific “Panopticon” vision,\textsuperscript{34} which describes the ultimate utilitarian prison that consisted of a central watchtower surrounded by a multi-storied ring of prison cells. The inner wall of each cell is a clear window, floor to ceiling, facing the watchtower. Each prisoner is completely exposed through the window twenty-four hours a day, so a single guard in the watchtower can see every movement.\textsuperscript{35} Just knowing that one could be observed, a prisoner would behave in accordance with the expected norm.\textsuperscript{36} A modern variant of the Panopticon vision is the “nannycam,” a surveillance product marketed to anxious parents who want their child care provider to know that his or her every move is being monitored and recorded.\textsuperscript{37}


\textsuperscript{35}. The Panopticon metaphor in modern privacy literature is invoked almost as frequently as Big Brother. See, e.g., Alexander T. Nguyen, Here’s Looking at You, Kid: Has Face-Recognition Technology Outflanked the Fourth Amendment?, 7 VA. J.L. & TECH. 2, n.111 (2002); Jeffrey H. Reiman, Driving to the Panopticon: A Philosophical Exploration of the Risks to Privacy Posed by the Highway Technology of the Future, 11 SANTA CLARA COMPUTER & HIGH TECH. L.J. 27, 28 (1995).

\textsuperscript{36}. See WHITAKER, supra note 31, at 32; Michel Foucault, Surveiller et Punir: Naissance de la Prison, in DISCIPLINE AND PUNISH: THE BIRTH OF THE PRISON 200 (Alan Sheridan trans., 1979); Solove, Privacy and Power, supra note 31, at 1415 (discussing Foucault’s description and observing that the Panopticon is “so efficient that nobody needs to be in the tower at all”).

\textsuperscript{37}. See WHITAKER, supra note 31, at 80–81 (1999).
One difference between modern “dataveillance” and the pervasive observation of the Orwellian and Panopticon worlds is that we are “watched” not through a camera or guard tower, but by a computer collecting facts and data. The effect on human behavior may be similar, however. Data collection can restrain our free will. We experience this in everyday life. Parents of a teenage child may assume that if they learn a lot about their child’s behavior, living habits, and activities, and the child knows about the continual observation, the child is more likely to obey the rules of the house. Workers whose telephone calls are monitored may make fewer personal calls on the job. The database problem can thus be viewed as a shifting in the balance of power from the individual to the entities that collect and control information about us. Viewed in this way, the problem with databases is similar to the surveillance of Big Brother or the Panopticon. Databases are a form of observation that curtails individual freedom and enhances the power employers, governments, insurance companies, and others have (or hold) over our lives.

Yet as Daniel Solove observed, Big Brother, Panopticon, and other surveillance metaphors do not entirely capture the current concern with database technologies. Orwell’s Big Brother and Bentham’s Panopticon demanded obedience from their subjects and sought to control important aspects of their behavior. The goal was conformity and discipline. Through continual surveillance, the technologies policed individuals to the point where individualism would be suppressed. By constantly living under the possibility that one could be observed at any time, people would do what authorities wanted them to do.

The collection of information in cyberspace can only roughly be analogized to the Orwellian or Panopticon worlds. As we apply for jobs or government benefits or surf the Internet information about us is collected. We are being watched in a sense, and we do not know

39. See Paul M. Schwartz, supra note 17, at 560.
40. See Rosa Ehrenreich, Privacy and Power, 89 GEO. L. J. 2047, 2053 (2001) (“[W]hat certainly should bother us [about data collection] and confidentiality has a great deal to do with power. More specifically, what should bother us are balances of power that are damaging or inequitable.”).
41. Solove, supra note 31, at 1422.
42. See Rosenbaum, supra note 20, at 571 (“Individuals cruising on the information highways often are blind to the electronic footprints they leave. Every post to a bulletin board, every electronic message, every Web page accessed and item purchased can be monitored and tracked.”); see also infra notes 69–81.
precisely when or to what extent. But for the most part we are not being observed by anyone who has any interest in controlling our behavior, minimizing individualism, or keeping us from uprising against authoritarian power. With the exception of FBI-type of law enforcement surveillance, an issue outside the scope of this Article, the people collecting and sharing information today are usually just looking for better and more efficient ways to run their businesses or government departments or to market their goods or services.43 Motives are largely benign, or at worst greedy.

Solove argues that the database problem is better captured by Franz Kafka’s depiction of prosecutorial bureaucracy in *The Trial*.44 The protagonist, Joseph K, is arrested yet never told what crime he is accused of committing. The novel is an anxiety-ridden nightmare during which nameless authorities actually do little to Joseph K, although he is constantly fearful of what they might do. He feels powerless to find answers or participate meaningfully in the bureaucratic process. Joseph K becomes obsessed with his predicament. He wants the court to treat him like an individual and he wants the case to reach finality. However, he does not know who has information about him, what that information is, or how it might be used.45

*The Trial* captures the sense of helplessness and vulnerability we may experience when large bureaucratic organizations—or a multitude of smaller, private ones—collect information about us and possess the power to use it against our interests.46 Governments, employers, and businesses continually make decisions based on our data, and we may


46. The Kafka metaphor was invoked recently by a public university professor whose e-mails were read by university administrative officials pursuant to a state “open door” law that considered e-mail to be a public record. “I felt like a person in a Kafka novel,” the professor lamented after witnessing the perusal of his private messages by school administrators. See Andrea L. Foster, *Your E-Mail Message to a Colleague Could Be Tomorrow’s Headline*, CHRONICLE OF HIGHER EDUCATION, *available at* http://chronicle.com/free/v48/i41/41a03101.htm (June 21, 2002).
have no knowledge of the process or an ability to challenge outcomes. 47 Like Joseph K, we might not even know if and when important decisions are being made. We are at the mercy of an unknowable digitized process—a world where people feel powerless and vulnerable, without any meaningful form of participation in the collection and use of information that can cause them harm.

Yet *The Trial* metaphor also fails to describe the modern database problem as most of us perceive it because, unlike Joseph K, we voluntarily participate in the data collection system because we see benefits from participating. My son cannot get a job unless he fills out an application form. I do not receive discounts from the grocery store unless I use the store “convenience” card. My research assistant cannot get information from a web site without divulging some personal information. None of us can get money from the ATM without leaving footprints that reveal where we were and what we were doing at that point in time. We cannot use a credit card on line to avoid the shopping mall traffic without transmitting a name and card number. We are not coerced but *seduced* to reveal personal information by the pleasures we derive from living in the modern world and consuming the goods and services that others offer. A better literary metaphor would be from the story *Hansel and Gretel*. 48 We are happily eating all the cookies, candy, and gingerbread, enjoying what we think are the benefits of sharing personal bytes of data in the information society. As we do so, we may be fattening ourselves for someone else’s feast, unaware of the fate that may await us.

47. See Froomkin, *supra* note 31, at 1463 (employers continually seek new ways to monitor employees for efficiency and honesty; businesses search databases for information about new customers); Mary W.S. Wong, *Electronic Surveillance and Privacy in the United States After September 11 2001: The USA Patriot Act*, 2002 SING. J. LEGAL STUD. 214, n.142 and accompanying text (noting that one criticism of the FBI’s e-mail reviewing software program, the Carnivore system (referred to as “DCS 1000” by the FBI), is that the FBI does not disclose details about how the system works).

Information Privacy

There is another difference with database surveillance. Since marketers and other users of databases generally are interested in aggregating data and selling products or services, they do not usually care much about snooping into the private lives of particular individuals. We are being watched not by other humans, but by machines, which gather information, compute profiles, and generate lists for postal delivery, e-mailing, or phone solicitations. This impersonality makes the surveillance seem less personally invasive than the leering of a “Peeping Tom,” an FBI wiretap, or Aunt Edna looking at your credit card bill when she stops by for a visit. A large portion of our personal information involves facts that are hardly embarrassing at all: our financial information, race, marital status, hobbies, occupation, and the like.

Indeed, most information collected about us in cyberspace concerns relatively innocuous and boring facts and details. Even so, there is a real and justified concern about how even this seemingly innocent information might be used in ways we would not prefer. We hope that our data will be accurate, complete, relevant, and current in every database in which it resides. We want the data used only for the right purposes (those to which we consent or would consent if asked), and which are permitted by law. We want it used by the right people (those who need to use the data for permissible purposes), and by no others. If any of these conditions is missing, we feel that important rights and interests have been jeopardized. The surveillance seems less personally invasive than the leering of a “Peeping Tom,” an FBI wiretap, or Aunt Edna looking at your credit card bill when she stops by for a visit. A large portion of our personal information involves facts that are hardly embarrassing at all: our financial information, race, marital status, hobbies, occupation, and the like.

Indeed, most information collected about us in cyberspace concerns relatively innocuous and boring facts and details. Even so, there is a real and justified concern about how even this seemingly innocent information might be used in ways we would not prefer. We hope that our data will be accurate, complete, relevant, and current in every database in which it resides. We want the data used only for the right purposes (those to which we consent or would consent if asked), and which are permitted by law. We want it used by the right people (those who need to use the data for permissible purposes), and by no others. If any of these conditions is missing, we feel that important rights and interests have been jeopardized. Privacy literature is sprinkled with horror stories about inaccurate, incomplete, irrelevant, or derogatory information in files, and unauthorized access to files containing information that can be dangerous in the wrong person’s hands. The privacy in cyberspace is a concern that needs to be addressed.

49. See PAUL SIEGHART, PRIVACY AND COMPUTERS 76 (1976) (referring to this as the “instrumental” value in privacy protection); COLIN J. BENNETT, REGULATING PRIVACY 33–35 (1992).

50. See, e.g., William McGeever, Programmed Privacy Promises: P3P and Web Privacy Law, 76 N.Y.U. L. REV. 1812, n.24 (2001) (citing McVeigh v. Cohen, 983 F. Supp. 215 (D.D.C. 1998) (refrencing homosexual sailor who discussed his sexual orientation in a seemingly anonymous online profile, only to have it revealed to his military superiors, who commenced discharge proceedings)). For additional illustrations, see Margot Williams & Robert O’Harrow, Jr., Online Searches Fill in Many Holes, WASH. POST, Mar. 8, 1998, at A19, reporting on free web service that found consumer’s address, phone number, names, and addresses of 20 neighbors, and provided map and directions to consumer’s home; another service provided for $9.50 consumer’s previous addresses and for $12.00 consumer’s Social Security number and birthday; another service provided driving record for $15.50. In a 1999 article, Jeff Sovern warned about the types of lists that are for sale, including: lists of people who have bought skimpy swimwear; college students sorted by major, class year, and tuition payment; millionaires and their neighbors; people who have lost loved ones; men who have bought fashion underwear; women who have bought wigs; callers to a 900-number national dating service; rocket scientists; children who have subscribed to magazines or have sent in
reason data collection bothers us is not complicated: we live in fear that we may be the next storyline.

B. Evolution of the Database

Our fears are heightened by a vague awareness of the absolute enormity of information residing in databases. Federal agencies alone control hundreds of databases on immigration, bankruptcy, Social Security histories, military personnel, and countless other subjects of legitimate government activity. The federal government has a database containing the Social Security numbers, addresses, and wages of nearly everyone who obtains a job in the United States. State governments keep digitized records on prescription drug purchases, automobile ownership, car insurance, births, criminal records, marital status, real estate holdings, liens and easements, voter registration, worker compensation claims, and many other aspects of our lives that are recorded, stored, and sorted. Licensing offices keep records on a variety of occupations from bail bondsmen to beauticians. Federal and state governments are encouraged to seek new, more efficient ways of integrating and aggregating these databases to serve their public mandates, but we can never be sure who has access to all of that information, and whether adequate security procedures are in place to guard it from theft, sale, or unauthorized use.

rebate forms included with toys; people who have had their urine tested; medical malpractice plaintiffs; workers' compensation claimants; people who have been arrested; impotent middle-aged men; epileptics; people with bladder-control problems; buyers of hair removal products or tooth whiteners; people with bleeding gums; high-risk gamblers; people who have been rejected for bank cards; and tenants who have sued landlords. Sovern, supra note 2, at 1034.


Although the build up of government databases has been extraordinary, the most revolutionary developments have occurred in the direct marketing industry and the private sector trade in personal information. This is where the “information superhighway” merges into the Autobahn, and where the speed is limited only by the power of the data processing machines driving it. As computers, software and data manipulation methodologies grow more powerful and sophisticated, data collection in the private sector will be an increasingly dangerous threat to information privacy interests.

Direct marketing to individuals was an inefficient and comparatively costly business practice for most of the twentieth century. One of the reasons for its slow development was the low response rate compared to the cost of compiling contact lists, printing and mailing solicitation letters, and hiring workers to make individual contacts by phone. To increase positive response rates, marketers realized that they needed to target their customers more accurately. Firms began compiling research on consumer preferences in various geographic areas and devising ways to analyze, sort, and use the collected data more effectively.

Improvements in direct marketing were aided by the federal government. When the postal service began using the five-digit zip code in the 1960s, direct marketers began grouping consumers by zip code to

---


58. See Solove, supra note 31, at 1405–06; Sovern, supra note 2, at 1046–47; Michael A. Fisher, The Right to Spam? Regulating Electronic Junk Mail, 23 COLUM.-VLA J.L. & ARTS 363, 364 (2000) (noting that an important cost of marketing activity is the time and inconvenience suffered by consumers, particularly those who never requested the information); cf. David E. Sorkin, Technical and Legal Approaches to Unsolicited Electronic Mail, 35 U.S.F. L. REV. 325, 337–39 (2001) (stating e-mail “spammers,” unlike senders of traditional non-electronic communications, have little incentive to compare the expected benefits of the communication against the cost, as the cost of sending unsolicited bulk e-mail is negligible).

determine the best areas to target particular product lines.\textsuperscript{60} Sorting data by zip codes proved to be a rough but inexpensive way to reach certain demographic subgroups.\textsuperscript{61} A decade later, the government began selling census data in electronic form. In an effort to protect the privacy in individuals identified in the census, the Census Bureau sold the information in bundles of several hundred households, providing addresses but not names.\textsuperscript{62} Businesses reattached many of the names, however, by matching census addresses to addresses in telephone directories and other databases such as voter registration lists. As a result of these and other developments, by the early 1990s consumers on average were included in nearly one hundred mailing lists, and the number was growing at a rapid pace.\textsuperscript{63}

Cyberspace technologies and the widespread use of the Internet profoundly affected the data collection business by the late 1990s. Government agencies placed individual records on their websites, although this practice is now receding due to public complaints about the disclosure.\textsuperscript{64} These government-held records used to be maintained in filing cabinets physically scattered in offices across the country. As a legal matter, these records were available to the public, but as a practical matter they were accessible only to local authorities and the occasional news reporter or ardent researcher. Now many of these records can be searched by anyone with a personal computer and some basic instructions on the system’s search logic.\textsuperscript{65} To make searching for personal information even easier, several Internet websites collect and

\textsuperscript{60} Dick Shaver, The Next Step in Database Marketing: Consumer Guided Marketing: Privacy for Your Customers, Record Profits for You 27 (1996).


\textsuperscript{62} Shaver, supra note 60, at 29–32.


\textsuperscript{64} William Matthews, Access Denied, FEDERAL COMPUTER WEEK, at http://www.fcw.com/fcw/articles/2000/0529/cov-access-05-29-00.asp (May 29, 2000) (“Information once eagerly posted on government Web sites to promote environmental safety, assist military personnel or help retirees is now being viewed as dangerous if found by terrorists, hackers and other criminals. . . . [A]gencies and Congress are tightening controls over federal Internet sites. Federal Webmasters who once enthusiastically posted information now anxiously take some of it down.”).

\textsuperscript{65} See Solove, supra note 53, at 1139; Steven C. Carlson & Ernest D. Miller, Comment, Public Data and Personal Privacy, 16 SANTA CLARA COMPUTER & HIGH TECH. L.J. 83, 84 (1999) (warning that federal, state, and local governments are currently pursuing ambitious programs to upgrade and integrate their information technology systems into unified data networks, making vast holdings of data far more accessible to government officials and to the general public).
compile public records from across the country and sell it online from a single source.\footnote{KnowX.com and Locateme.com, for example, sell information on airplane ownership, court filings, death certificates, pilot licenses, judgments, liens, professional licenses, foreclosures, refinancings, driver and voter registrations, and credit “headers” (part of a credit report), at http://www.knowx.com (last visited Jan. 20, 2003); http://www.locateme.com (last visited Jan. 20, 2003). Focus USA claims to have information on 203 million people and offers demographic lists such as “Tech-Savvy Hispanics,” “Big-Spending Parents,” “Proven Patriots,” “Rural Riches,” and “Pet Lovers Online.” See http://www.focus-usa-1.com/lists_az.html (last visited Jan. 20, 2003). See generally Solove, supra note 31.}

Currently, most personal information in cyberspace is collected in one of two ways. An organization may directly solicit and collect information from individuals who contact the organization and provide information voluntarily.\footnote{See Solove, supra note 32, at 1094 (noting that Web sites collect data when people fill out online questionnaires pertaining to their hobbies, health, and interests).} Alternatively, and increasingly more common, the organization might surreptitiously track and record individual’s surfing activity on the Internet.\footnote{See Joel Reidenberg, Privacy in the Information Economy: A Fortress or Frontier for Individual Rights?, 44 FED. COMM. L.J. 195, 201–02 (1992).}

Direct solicitation of information has been with us for years in various forms. We have all completed job or credit applications or filled in surveys. Many consumers have completed and returned “warranty registration” cards to the manufacturer, which volunteer valuable data that can be used for marketing purposes.\footnote{Warranty registration cards for many products ask a host of lifestyle questions in addition to product-related queries. Other surveys are more direct. The author received a mailing called the “Consumer Product Survey of America” asking the “main grocery shopper in your household” to fill out the form and mail it. The questionnaire asked dozens of questions ranging from the kind of bladder-leakage products used to the name of the consumer’s auto insurance company. It also asked for name, home address, phone numbers and e-mail addresses. See Letter from Laura David, Shopper’s Voice, Consumer Product Survey of America (undated) (original on file with author).} In the modern age, more information is directly solicited online as an increasing number of websites require registration and the disclosure of personal information before a user can access the site’s content. Amazon.com, for example, uses registration information to help keep track of its customers’ purchases of books, CDs, electronics, toys, and other items.\footnote{Andrew Shen, Online Profiling Project, EPIC, at http://www.epic.org/privacy/internet/Online_Profiling_Workshop.PDF (last visited Dec. 17, 2002) (citing an article from The Economist in which Jeff Bezos, CEO of Amazon.com, describes Amazon as an “information broker”, acting as the connection between consumers looking for books and publishers looking for consumers; according to Bezos, Amazon’s vast database of customers’ preferences and buying patterns is tied to their e-mail and postal addresses); Alan Murray, Net Effect: Is Service Getting Too Personal?, WALL ST. J., July 19, 1999, at A1. “[T]he next wave of Internet innovation is in the area of personalized marketing and services. Companies such as
Surreptitious collection of information from web users is even more common. Many websites secretly track a customer’s surfing practices through the use of “cookies” and similar technologies.\textsuperscript{71} When a user explores a site, the user leaves electronic footprints behind. By following the footprints, the site can record information about the user, such as the Internet service provider used, and the type of hardware and software the user employed. The site can also record some behavioral information about the user’s Internet habits, such as the website previously visited, the amount of time spent on each web page, and the length of time spent visiting different parts of the site.\textsuperscript{72}

To make this information more useful, the web site might connect the “clickstream” data to particular Internet users.\textsuperscript{73} This can be done by either requiring users to register or branding them with cookies that will report identifying information back to the website the next time the user visits.\textsuperscript{74} Using either method, the site can compile a profile of individual Amazon.com are eagerly assembling and sorting massive amounts of information on customer preferences. Their aim is to know what book, record or other product you want before you know it, and then market it directly to you.”).

\textsuperscript{71} A “cookie” is a small file of codes that is dispatched to a user’s computer when a web page is viewed. The site puts an identification mark in the file, and the cookie is stored on the user’s hard drive. When the user visits the site again, the site locates the cookie and matches the file code with information previously collected about the user’s surfing activity. While privacy advocates object to the use of cookies, the problem with banning them is that they have practical uses other than secretly collecting information about surfing activity. They can store passwords, for example, which speeds access to frequently used websites. See generally Viktor Mayer-Schönberger, The Internet and Privacy Legislation: Cookies for a Treat?, 1 W. VA. J.L. & TECH. 1.1 (1997) (discussing cookie technology and its impacts on privacy); Lori Eichelberger, The Cookie Controversy, COOKIE CENTRAL, at http://www.cookiecentral.com/ccstory/cc2.htm (Apr. 8, 1998).

Indeed, the much maligned cookie can even enhance online privacy. The electronics retailer Future Shop switched to the use of cookies on its site in November 2000 after learning that unauthorized people could log on and view other customers’ names, addresses, phone numbers, and possibly credit card numbers. The company claimed that cookie technology would have prevented a breach of security and an invasion of privacy. See Future Shop Homepage, at http://www.futureshop.ca (last visited Jan. 20, 2003); T. Hamilton, Price Snafu Stings Web Retailer, THE TORONTO STAR Nov. 17, 2000, at C01.

\textsuperscript{72} See Solove, supra note 31, at 1411.

\textsuperscript{73} Jerry Berman & Deirdre Mulligan, Privacy in the Digital Age: Work in Progress, 23 NOVA L. REV. 551, 559 (1999) (explaining “clickstream” data are series of detailed transactional information that improve targeted online advertising; some firms, such as Adfinity, combine clickstream data, or “mouse-droppings,” with personal information collected from other sources to create profiles of a person’s Web browsing behavior).

\textsuperscript{74} See supra note 71. Internet users usually associate clickstream and cookie technology with private sector websites, but they have been frequently used in government websites as well. Despite a ban on federal government use of these information gathering tools, a survey in October 2000 found cookies in use in 11 of 65 government websites, 3 of them passing information to third parties without the permission of users. See Ronna Abramson, Federal Agencies Caught With Hand in the
interests, concerns, and general web surfing habits. Savvy online marketing firms can even draw inferences about how we respond to web page presentations. For example, an online travel service could keep track of every destination to which a person requested a fare or every city in which hotel information was sought. A medical information site could track the number of times a user linked to pages providing information on osteopathic remedies. Clickstream data can thus reveal lots of useful marketing information about all who use the Internet.

Another form of user-tracking technology is the “web bug,” also known as a web beacon or clear graphic image file (GIF) tag. Web bugs are image files secretly imbedded in a web page and are invisible to the person browsing the page. The bug sends information about the user’s browsing habits and interaction with the page back to the home server. Internet advertisers also can capture the search terms a person uses to find web sites on a subject of interest. The process, known as “banner ad leakage,” allows an advertiser to record search terms as the user submits them to the search engine. Banner ad leakage allows the advertiser to collect an enormous amount of potential marketing data and to tailor ads

---


76. See Solove, supra note 31, at 1412.

77. A web bug is invisible because it is only one pixel square in size and blends into the background on a web page or HTML e-mail message. The only way to detect a web bug is to locate the source code for the web page or e-mail message and discover that the web bug image comes from a different server than everything else. The server sending the bug might belong to an advertising network that uses it to obtain information, including the Internet Protocol (IP) address of the computer that accepted the web bug, the URL of the page on which the web bug appears, the time the web bug was viewed, the type of browser that accepted the web bug image, and any previously set cookie data. (The cookie can link the bug and the information it has obtained back to the online profile associated with that cookie.) Web bugs are common in HTML e-mail and are used to tell if an e-mail has been read or forwarded to another person. See R. Smith, FAQ: Web Bugs, at http://www.privacyfoundation.org/resources/webbug.asp (last visited Jan. 20, 2003); Robert O’Harrow Jr., Fearing a Plague of ‘Web Bugs’: Invisible Fact-Gathering Code Raises Privacy Concerns, WASH. POST, Nov. 13, 1999, at E01.

78. See Barber & Quinn-Barabanov, supra note 75.
to the user’s specific interests more quickly and accurately than cookie technology would permit. 79

The data aggregation industry also developed technology for sharing information between websites, thereby making the aggregated data considerably more useful to merchants, advertisers, direct marketers and other entities that see value in personal information. The best known provider of this type of service is probably DoubleClick, a service that distributes client advertisements to various web sites. 80 When a user clicks on a client’s advertisement banner on a web site, a message is automatically sent back to DoubleClick reporting that the banner had achieved some success with a particular user. This lets DoubleClick determine which ads are being seen and which user is seeing them. DoubleClick can then create a profile of a user and search its list of subscribing companies for advertisements that match the user’s interests. When the user browses the Internet later, the user will see advertisements tailored to his or her revealed preferences. Using this process, DoubleClick compiled eighty million customer profiles by the end of 1999. 81

Even with knowledge of the astounding amount of digital information that is collected, manipulated, and shared every day, we still have difficulty discussing precisely what harm is being done. We see the benefits of the data processing revolution, yet we still have an uneasy

79. See id.
80. See Berman & Mulligan, supra note 73.
81. Heather Green, Privacy Online: The FTC Must Act Now, BUSINESS WEEK, Nov. 29, 1999, at 48. DoubleClick’s notoriety is due in large part to an investigation of the company’s practices by the FTC in 2000. The FTC initiated a “routine” investigation after learning that the company planned to merge its database with the database of Abacus Direct Corp., a direct marketing company that had information on most U.S. households. Charles L. Kerr & Oliver Metzger, Online Privacy: New Developments and Issues in a Changing World, Third Annual Institute on Privacy Law: New Developments & Issues in a Security-Conscious World, 701 PLI/PAT 303, 330-31 (2002); Robert W. Hahn & Anne Layne-Farrar, The Benefits and Costs of Online Privacy Legislation, 54 ADMIN. L. REV. 85, 109 (2002). The FTC was concerned that DoubleClick might be disclosing sensitive information about consumers in violation of its privacy policy, which would be a violation of the FTC Act. See In re DoubleClick Inc. Privacy Litigation, 154 F. Supp. 2d 497, 505 (S.D.N.Y. 2001). The FTC closed its investigation in January 2001, finding no evidence of illegal conduct. Robert G. Bagnall, Privacy: Investment Company Regulation and Compliance, SG100 ALI-ABA 255, 265 (2002). DoubleClick ultimately declared that it would not pursue the plan to merge information, and agreed to require all new clients and web sites to disclose their use of DoubleClick’s services. Anthony Rollo, The New Litigation Thing: Consumer Privacy, 1301 PLI/CORP 9, 56-59 (2002); Shaun B. Spencer, Reasonable Expectations and the Erosion of Privacy, 39 SAN DIEGO L. REV. 843, 867 (2002). It also agreed to modify its privacy policy to explain its practices more clearly and to provide a better explanation of its opt-out procedures. See Barber & Quinn-Barabanov, supra note 75.
sense of foreboding. What exactly are we afraid of? This question is not easy to answer because we are not entirely sure how much of the collected data can be traced to us as individuals and used in a harmful way. Online firms, for example, maintain that the bits of information gleaned from cookies, web bugs, and the like cannot be associated with specific persons. At most, they can be used to identify computers at particular locations. That is hardly comforting news, however, because it will not be long before information about a computer location will be matched with individual owners or users of the computer.82 This uncertainty about how personal information might be used in the future is cause for concern,83 especially because once data is collected and stored, the shelf life is indefinite.

C. Conceptualizing the Harm

There are several ways to conceptualize the types of injury that can result from data collection and sharing.84 One of the most discussed harms is the mischaracterization of an individual.85 Other people,
businesses, or governmental institutions examine information that we generally regard as private, and they use that information to make judgments about us. Since the information used to form the judgment is not the complete set of relevant facts about us, we can be harmed (or helped) by the stereotyping or mischaracterization. Jeffrey Rosen observed in his influential work, *The Unwanted Gaze*, that “[p]rivacy protects us from being misdefined and judged out of context in a world of short attention spans, a world in which information can easily be confused with knowledge.”

Our data records cannot tell the whole story about us, yet we are frequently judged on the basis of small bits of information in important aspects of our lives.

We are aware of the mischaracterization problem in many areas of life. For example, in the legal education community, we sometimes measure the importance of law review articles by looking at the number and type of citations made to them in judicial opinions. Each spring, law school deans, faculties, and alumni lament (or quietly cheer) the release of a national “ranking” of law schools by *U.S. News and World Report*, which uses selected facts and figures to divide schools into four “tiers” of descending prestige. College athletic teams are ranked by a computerized data system that relies on a handful of variables, including margin of victory and the relative strength of each team’s opponents.

---

86. Kenneth Karst warned in the 1960s that one problem with a large database of personal information is that the facts in the database “will become the only significant facts about the subject of the inquiry.” Kenneth L. Karst, “The Files”: Legal Controls Over the Accuracy and Accessibility of Stored Personal Data, 31 L. & CONTEMP. PROBS. 342, 361 (1966). See Solove, supra note 31, at 1424.


88. See Solove, supra note 31, at 1424.


The reasons for stereotyping are easy to appreciate. The objective is to place people, groups, or other subjects in specific categories that can give us some measure of confidence that judgments based on the categorization will be well founded. For example, only if a consumer’s credit score rises above a certain threshold will she be considered a good credit risk and then receive a “pre-approved” credit card. The card issuer has confidence that individuals in the higher scoring groups will be more likely to pay their bills on time.91 While stereotyping has been with us for ages, computer databases encourage the practice by making it much easier to collect vast amounts of data quickly and sort it in countless ways. The resulting judgments might be proved correct in the aggregate (for example, the overall default rate on credit accounts may decrease), but they can be unfair in individual cases. A bankruptcy filing or criminal arrest record can be misleading without knowing the story behind it or the ultimate disposition of the case, but the record’s appearance on a credit report will adversely affect a credit decision regardless of the background details.

If mischaracterization were the main problem with information collection and storage, however, it is by no means clear that the appropriate public policy response would be to restrict data flows. A logical solution would be to encourage the collection of more, not less, information about each of us. We could reduce the likelihood of misjudgments if our data files were more complete, so we should encourage greater aggregation of data and more extensive sharing of information among data processors. The FCRA essentially embraces this approach. If a consumer believes that her credit report contains inaccurate or incomplete information, she can supply the missing facts and have them added to the report in hopes of providing a more accurate picture.92

91. See generally The Credit Scoring Site, at http://www.creditscoring.com (last visited Jan. 20, 2003); Daniel Mendel-Black & Evelyn Richards, Peering into Private Lives: Computer Lists Now Profile Consumers by Their Personal Habits, WASH. POST, Jan. 20, 1991, at H01 (“Details . . . are sorted, digested and compiled so that computers can plop you into neatly defined categories to help determine the likelihood that you’ll pay your Visa bill on time or buy a new brand of detergent or cigarettes within the next few months.”); David Rameden, When the Database Is Wrong . . . Do Consumers Have Any Effective Remedies Against Credit Reporting Agencies and Information Providers?, 100 COM. L.J. 390, n.10 (1995) (citing Consumer Problems With Credit Reporting Bureaus: Hearings Before the Senate Subcomm. on Consumers of the Comm. on Commerce, Science & Technology, 102 Cong., 2d Sess. 2 at 47 (1992) (testimony of Mary Santina, retail representative)); What Price Privacy?, 56 CONSUMER REP. 356, 357 (1991) (credit scoring example).

Other categories of injury from data collection and sharing have been described in anthropological or sociological terms. Loss of privacy can be seen as an affront to human dignity, a loss of personal autonomy, and other terms that reference the value of one’s “core self.” These concepts focus on the objectification of an individual resulting from pervasive surveillance and seemingly unlimited access to personal information in modern society. Protecting privacy, and contracting those aspects of our lives that are open to searching and monitoring, is an important way of showing that individuals are worthy of respect. The promotion of this ideal is central to the modern view of a liberal society. It bolsters John Stuart Mill’s vision that “over himself, over his own body and mind, the individual is sovereign.”

For most of us, though, the more troubling problem is not a mischaracterization of our complex persona or a reduction of our humanity to a set of electronic bits and bytes. The more cognizable and immediate problem with a loss of information privacy, and the problem that is most likely to produce a political resolution, is our inability to avoid circumstances in which others control information that can affect us in material ways—whether we get a job, become licensed to practice in a profession, obtain a critical loan, or fall victim to identity theft. We cannot avoid the collection and circulation of information that can profoundly affect our lives. We feel that we have little or no voice or choice in the data collection and sharing process. We do not know who has what information, how they got it, what purposes or motives those

---


94. See Jeffrey H. Reiman, Privacy, Intimacy, and Personhood, in PHILOSOPHICAL DIMENSIONS OF PRIVACY: AN ANTHOLOGY 300 (Shoeman ed., 1982).

95. WESTIN, supra note 85, at 32.

96. See Stanley I. Benn, Privacy, Freedom, and Respect for Persons, in PHILOSOPHICAL DIMENSIONS OF PRIVACY: AN ANTHOLOGY 223 (Shoeman ed., 1982); Rosen, supra note 30, at 2121.


98. In the area of privacy and technology, part of the explanation of why privacy did not draw more congressional advocates in the 1960s and early 1970s is that it is difficult to agree on a definition of the problems presented. See WESTIN, supra note 85, at 7 (“Few values so fundamental to society as privacy have been left so undefined in social theory or have been the subject of such vague and confused writing by social scientists.”); Judith Jarvis Thomson, The Right to Privacy, 4 PHILOS. & PUBLIC AFFAIRS 295, 295 (1975) (“[T]he most striking thing about the right to privacy is that nobody seems to have any very clear idea what it is.”); REGAN, supra note 7, at 3. The definition of privacy in the U.S. that has formed the basis of most of the policy discussion is the right to control information about and access to oneself. Id. at 4.
entities have, or what will be done with that information in the future. Even if the information in a database is accurate and complete in all relevant respects, it can still harm us if it falls into the wrong hands or if it is used for a purpose we did not envision when we disclosed it.

What compounds our discomfort is the likelihood that as technological developments improve, we can expect to lose more control over the collection and sharing of information about us. Advances in genomics are fueling the creation of DNA databases. The trend in

---

99. In 1969, Alexander Solzhenitsyn described the relationship between an individual and an authority that controls information about him:

As every man goes through life he fills in a number of forms for the record, each containing a number of questions. A man’s answer to one question on one form becomes a little thread, permanently connecting him to the local center of personnel records administration. There are thus hundreds of little threads radiating from every man, millions of threads in all . . . . Each man, permanently aware of his own invisible threads, naturally develops a respect for the people who manipulate the threads . . . and for these people’s authority.


100. See Robert S. Peck, The Right to Be Left Alone, 15 HUM. RTS. 26, 28 (1987) (“information collected for one purpose may be shared with other agencies and used for entirely different purposes . . . [and] assembled into a complete personality profile at the touch of a computer button”).

101. Many companies have begun commercializing genomic information for therapeutic and other purposes. As early as 1995, over 50 biotechnology companies were developing or providing tests to diagnose genetic disorders or to predict the risk of their future occurrence through identification of “susceptibility-conferring genotypes.” See Neil Holtzman, Are Genetic Tests Adequately Regulated?, SCIENCE 286, 409 (October 15, 1999).

The web site for Celera Genomics, Inc. describes its function as that of a “leading provider of information based on the human genome and related biological and medical information.” See http://www.celera.com (last visited Jan. 20, 2003); Scott Hensley, Celera’s Genome Anchors it Atop Biotech, WALL ST. J., Feb.12, 2001, at A3 (“Three-year-old Celera, it now is clear, has produced a [genome] map that drug and biotech companies, hungry for gene information that will help them find new treatments, are plunking down millions of dollars a year for the right to sift through.”). Other recent literature documents the worldwide scope of these activities. For example, an August 2000 article notes proposals for the creation of “phenotype” databases (databases containing information about the physical characteristics of patents for whom genetic information is known) in the United Kingdom, Italy, and Estonia; and a recent $200 million agreement between Reykjavik-based decode Genetics and Hoffman-LaRoche, Inc. based on a database based on the medical records of Iceland’s 275,000 citizens. See Daniel Machalaba, Burlington Northern Ceases its Genetic Testing, WALL ST. J., Feb. 13, 2001, at B10; Antonio Regalado, Medical Records, Inc., TECHNOLOGY REVIEW (July/August, 2000); Editorial, Gene Library, BOSTON GLOBE, Sept. 17, 2000.

Several states have taken action to guard genetic information more zealously than other medical records. See, e.g., Cal. CIV. CODE § 56.17 (West 2002); N.J. STAT. ANN. § 10:5-45 (West 2002); Cf. GA. CODE ANN. § 33-54-6 (2002) (research facilities may use the information derived from genetic testing for scientific purposes so long as the identity of any individual tested is not disclosed to any third party, except that the individual’s identity may be disclosed to the individual’s physician with the consent of the individual).
business is to use technology to create “human-centered computing” and new technologies designed to benefit consumers by making their lives more efficient and their work more productive. Yet while the potential efficiencies of new technologies are trumpeted, the societal cost of these technologies is hardly noticed. Examples include: biometric signatures that can make air travel, banking, and other activities more secure and efficient as they enhance customer convenience; “smart cards” that can be carried like a digital passport, which are useful because a stolen credit card is much less valuable when the person attempting to use it could be immediately identified as an imposter; palm or retinal scanners that businesses could install to verify every customer’s identity; and even “smart buildings” that can follow the whereabouts of all employees and visitors, all day long, and record the data for future use.

Video analysis and scanning systems can pick out a face in the crowd that matches a digital imprint of an escaped convict, a father behind on child support, a terrorist, or a missing person. Every time a person moves in front of a surveillance camera, a database could record the location of the person and at precisely what time she walked by.

105 See Deirdre K. Mulligan, Privacy Past, Present, and Future, THIRD ANNUAL INSTITUTE ON PRIVACY LAW: NEW DEVELOPMENTS & ISSUES IN A SECURITY-CONSCIOUS WORLD, 701 PLI/Pat 63 (2002).
106 See WHITAKER, supra note 31, at 140–42 (discussing new surveillance technologies that “render individuals ‘visible’ in ways that Bentham could not even conceive”); John D. Woodward, Biometric Scanning, Law & Policy: Identifying the Concerns-Drafting the Biometric Blueprint, 59 U. PITT. L. REV. 97, n.203 (1997) (hypothesizing that a state legislature could decide to require all children attending private day care to be biometrically scanned for identification purposes).
107 See WHITAKER, supra note 31, at 140–42.
Global Positioning System (GPS) technology\textsuperscript{108} can take the concept further. With a fairly low-cost receiver, a computer can pinpoint a person’s location anywhere on earth within a few feet. Lost automobile drivers and hikers in remote areas can benefit immensely from this technology. With some enhancements, a computer chip could be attached to a small child, an Alzheimer’s patient, or a parolee, and the person could be tracked anywhere he or she goes.\textsuperscript{109}

Of course, all of these developments have two sides. One is convenience, efficiency, and empowerment; the other is continual surveillance and loss of individual control once information is revealed. We like the information used for some purposes, but we dread its use for others. This dynamic is typical of most public policy problems that find a resolution in our legal system. How do our laws work to preserve the benefits of information collection and storage, but minimize the risks of its misuse?

II. THE LEGAL LANDSCAPE IN THE UNITED STATES

A. Inadequacy of Common Law Torts and the United States Constitution

Privacy law in the United States did not begin to develop until the middle of the twentieth century.\textsuperscript{110} As it developed, the legal doctrine addressed problems fundamentally different from those presented by digital databases. Courts and legislatures created a number of torts designed to redress injuries caused by unwelcome interlopers, such as an overly aggressive press or a political enemy, who invaded what Warren

\textsuperscript{108} A Global Positioning System (GPS) can transmit information about an object’s three-dimensional position, velocity and time to anyone equipped with a GPS receiver. GPS can also provide precise guidance and targeting information for missiles. See Shaun B. Spencer, Reasonable Expectations and the Erosions of Privacy, 39 SAN DIEGO L. REV. 843, 882–83 (2002) (explaining that the Department of Defense developed the GPS in the early 1970s as a satellite-based positioning and navigation system).

\textsuperscript{109} See WHITAKER, supra note 31, at 140–42; Froomkin, supra note 31, at 1496–98 (noting that satellite tracking is being used to monitor convicted criminals on probation, parole, home detention, and work release, at a daily cost of only about $12.50 per target); Mark G. Young, Note, What Big Eyes and Ears You Have!: A New Regime for Covert Governmental Surveillance, 70 FORDHAM L. REV. 1017, 1035–36 (2001).

\textsuperscript{110} Although the U.S. Supreme Court recognized a Fourth Amendment right to privacy as early as 1886 in Boyd v. United States, 116 U.S. 616 (1886), most of the case law development in tort and constitutional law emerged decades later. See generally Gerald G. Ashdown, The Fourth Amendment and the “Legitimate Expectation of Privacy,” 34 VAND. L. REV. 1289 (1981).
and Brandeis called “the sacred precincts of private and domestic life.” Depending on the jurisdiction, variously named “privacy torts” ordered the payment of compensation for the most egregious harms resulting from disclosure of facts that most would consider purely private matters. These torts included invasion of privacy, intrusion upon seclusion, public disclosure of private facts, false light or false publicity, and misappropriation.

Because these torts developed to address injuries resulting from the release of private and embarrassing facts, they were not intended, and are not well suited, to redress the harms caused by the collection and sharing of information in databases. Most of the injuries caused by the misuse of data in modern society are not particularly embarrassing or emotionally disturbing. Even when they are, because data can be aggregated, stolen, or transferred with the click of a keystroke, tracing the injury to the source of the information leak, and then establishing the requisite mens rea for the tort, will often be impossible. A remedy in


112. See REGAN, supra note 7, at 32-33; ARTHUR R. MILLER, THE ASSAULT ON PRIVACY: COMPUTERS, DATA BANKS, AND DOSSIERS 189 (1971) (“[M]ost significantly, the existing common law structure does nothing to give the data subject a right to participate in decisions relating to personal information about him, a right that is essential if he is to learn whether he has been victimized by a privacy invasion.”); Solove, supra note 31, at 1432.

Defamation is another tort that can be alleged in some cases of unauthorized information sharing. For example, privacy and defamation torts were being combined in cases where the use of a person’s name without consent was held to be offensive. See Harry Kalven Jr., Privacy in Tort Law—Were Warren and Brandeis Wrong?, 31 LAW & CONTEMP. PROBS. 326, 333 (1966). This comes close to judicial recognition of Charles Fried’s definition of privacy as “that aspect of social order by which persons control access to information about themselves.” Charles Fried, Privacy, 77 YALE L.J. 475, 493 (1968). Still, courts seldom adopted this reasoning. See REGAN, supra note 7, at 34.


114. Although many writers discount the effectiveness of privacy torts as a mechanism for redressing misuse of personal information, others believe that tort law has simply been underutilized in information privacy cases. See, e.g., The Privacy Torts: How U.S. State Law Quietly Leads the Way in Privacy Protection, at http://www.privacilla.org/releases/Torts_Report.html (July 2002) (maintaining that greater use of privacy torts would more effectively protect information privacy than would government regulation); Denise G. Callahan, Courts Make Better Privacy Law, 13 IND. LAWYER, Aug. 14-27, 2002, at 1 (noting that litigation may be the best way to discover harmful information practices, citing tobacco litigation as an example of successful litigation strategies).

115. See Solove, supra note 32, at 1085 (noting that harms resulting from the misuse of personal data often do not result from malicious intent or the desire for domination); Catherine Therese
tort, assuming precedent would support one, will often be unobtainable as a practical matter.

More recently, other torts have been alleged in litigation over the collection and misuse of information in databases. Although some show promise for database protection, to date they have proved unsuccessful. Fraud (or deceit), and unjust enrichment, for example, have been asserted in situations where information was collected surreptitiously or used in ways contrary to the data collector’s privacy policy. Fraud is difficult to prove because of the requirement that the tortfeasor act with intent to defraud or at least with reckless disregard for the truth. In the fast moving and often depersonalized world of digital data transfers, this can be difficult to establish.

The equitable doctrine of unjust enrichment has more promise as a vehicle for redressing the harms caused by unlawful data collection and sharing. Unjust enrichment requires a showing of economic benefit to one party at the expense of the other, plus a finding that the enrichment was somehow unfairly or unjustly obtained. In the context of digital


117. See e.g., Cao v. Nguyen, 607 N.W.2d 528, 532 (Neb. 2000) (stating that to maintain an action for fraudulent misrepresentation, a plaintiff must prove: (1) that a representation was made; (2) that the representation was false; (3) that when made, the representation was known to be false or made recklessly without knowledge of its truth and as a positive assertion; (4) that it was made with the intention that the plaintiff should rely upon it; (5) that the plaintiff reasonably did so rely; and (6) that the plaintiff suffered damage as a result); Suiter v. Mitchell Motor Coach Sales, Inc., 151 F.3d 1275, 1282 (10th Cir. 1998) (concluding that “intent to defraud” may be inferred if it is shown that the defendant lacked such knowledge only because he displayed reckless disregard for the truth or because he closed his eyes to the truth). See generally 37 AM. JUR. 2D, Fraud and Deceit § 107 (2002) (explaining that proof of a mere naked falsehood or representation is ordinarily not enough, but in addition to the false representation, the false statement must have been made intentionally to deceive).

118. See Lesser, supra note 116. A cause of action for unjust enrichment lies where someone has conferred a benefit and it would be inequitable or unjust for the recipient to retain that benefit. RESTATEMENT (THIRD) OF RESTITUTION AND UNJUST ENRICHMENT §§ 1, 2 (2000); United Coastal Indus., Inc. v. Clearheart Const. Co., 802 A.2d 901, 905 (Conn. App. Ct., Aug. 13, 2002)
databases, this doctrine has potential for success when, for instance, a web site violates its privacy policy by selling information to a marketing firm, or storing information that the site said it would not store.\textsuperscript{119} Aggrieved web site users would have to show that the site obtained an economic benefit from the information obtained, and that the collection or transfer of information was otherwise inequitable and unjust.\textsuperscript{120} Proving both elements is difficult in many circumstances, which may explain why unjust enrichment is generally considered a relatively ineffective doctrine in this and other contexts.

As a public policy matter a more important problem with common law principles is that they do not protect the individual prior to injury. When our privacy is threatened by public or private sector data collection and sharing, lawsuits only offer a means of redress after the invasion; they do not provide a sufficient deterrent that will prevent the invasion from occurring.\textsuperscript{121} Since most of us are not going to pursue damage remedies for privacy invasion except in the most extreme circumstances, these doctrines cannot do much to redress the dangers of information collection and exchange that most Americans fear.

Federal constitutional protections have proved equally unhelpful for most database privacy problems, at least at the federal level.\textsuperscript{122} Like tort

\textsuperscript{119} Lesser, supra note 116; Pamela Samuelson, \textit{Privacy as Intellectual Property?}, 52 STAN. L. REV. 1125, 1134 (2000) (arguing that “a property rights approach [will] . . . halt the unjust enrichment that compilers of personal information now enjoy”); Natalie L. Regoli, A Tort for Prying Eyes, 2001 J.L. TECH. & POL’Y 267, 287 (2001) (envisioning an Internet profiling tort that would provide restitution as a baseline recovery, and arguing that the unjust enrichment that was obtained at the user’s expense should be measured by the aggregate enrichment gained from all users’ information).

\textsuperscript{120} \textit{Restatement (Third) of Restitution, supra} note 118.


\textsuperscript{122} See Paul M. Schwartz & Joel R. Reidenberg, \textit{Data Privacy Law: A Study of United States Data Protection} 33–35 (1996) (discussing the limited availability of constitutional law in protecting privacy interests in the private sector). Indeed, to the extent the U.S. Constitution is invoked in privacy litigation or public policy debate, it is often to invoke the First Amendment in support of those who support less onerous restrictions on access to information and fewer impediments to exchange of information once they obtain it. See, e.g., Eugene Volokh, \textit{Freedom of Speech and Information Privacy: The Troubling Implications of a Right to Stop People
law, to the extent there is a constitutional remedy, it arrives only after the injury has occurred. Moreover, the U.S. Constitution protects only against unlawful government action, and many of the more powerful and potentially injurious databases are maintained in the private sector.123

The U.S. Supreme Court has recognized constitutionally protected privacy interests in various contexts, deriving the right from the First, Third, Fourth, Fifth and Ninth Amendments.124 However, most of the Court’s decisions have only marginal relevance to the problem of databases.125 The Fourth Amendment comes closest to encompassing a right of information privacy against government misuse of data, but thus far it has not been construed broadly enough to protect against most harms that result from the data collection, use, and distribution.126 The Supreme Court has assumed that privacy is about protecting highly personal information.127 Thus, we have no constitutionally protected expectation of privacy when we permit our information to be accessed by a third party (such as an online search engine) or when we voluntarily give the information to someone else (such as filling out a job


124. See REGAN, supra note 7, at 35.

125. The U.S. Supreme Court, for example, has recognized privacy interests such as associational privacy, NAACP v. Alabama, 357 U.S. 449, 462 (1958), political privacy, Watkins v. United States, 354 U.S. 178, 198–99 (1957), and the right to anonymity in public expression, Tolley v. California, 362 U.S. 60, 64 (1960) (distribution of handbills in public place). See also Couch v. United States, 409 U.S. 322, 327 (1973) (the privilege against self-incrimination "respects a private inner sanctum of individual feeling and thought").

126. See, e.g., Dow Chemical Co. v. United States, 476 U.S. 227, 231 (1986) (no violation of Fourth Amendment when EPA engaged in warrantless aerial photography of manufacturing facility); California v. Greenwood, 486 U.S. 35, 40–41 (1988) (no reasonable expectation of privacy in trash bags left at curb for pick-up because bags are "readily accessible to . . . scavengers, snoops, and other members of the public").

127. See infra note 132.
application). The Court has held, for example, that there is no expectation of privacy in the telephone numbers we dial because we have released the information to the phone company.\textsuperscript{128} There is no privacy right in the information contained in personal checks because they are not “confidential communications” once they are sent through the check collection process.\textsuperscript{129}

However, courts have upheld a constitutional right to information privacy in a narrow set of circumstances. In 1977, the Supreme Court recognized a right to information privacy, noting a constitutionally protected “individual interest in avoiding disclosure of personal matters.”\textsuperscript{130} Most of the decisions in this line involve a breach of confidentiality, or the risk of unwanted publication of very private facts. In particular, courts have occasionally found a constitutionally protected right to information privacy when the records involve highly personal issues such as sexual practices or medical conditions.\textsuperscript{131} Thus, while medical records might be protected from disclosure,\textsuperscript{132} arrest and conviction records are unprotected because they already appear in a public record.\textsuperscript{133} Outside the realm of health and sex information, courts have not found much protection within the United States Constitution against information collection and disclosure.

\textsuperscript{128} See Smith v. Maryland, 442 U.S. 735, 742 (1979). See also United States v. Miller, 425 U.S. 435, 442 (1976) (holding there is no expectation of privacy in a person’s financial records held by a third party).

\textsuperscript{129} Miller, 425 U.S. at 442. In response to the decision in Miller, Congress enacted the Financial Privacy Act in 1978, 12 U.S.C. § 3401 et seq. (2000), which provides some privacy protection for a customer’s banking records.

\textsuperscript{130} Whalen v. Roe, 429 U.S. 589, 599 (1977). Although the Court recognized the privacy right, it upheld a New York law that required the state to maintain computerized records for certain drugs. \textit{Id}.

\textsuperscript{131} See Solove, \textit{supra} note 53, at 1204–05.

\textsuperscript{132} See Doe v. Borough of Barrington, 729 F. Supp. 376, 382 (D.N.J. 1990) (police disclosure that a person had AIDS); Woods v. White, 689 F. Supp. 874, 875 (W.D. Wis. 1988) (prisoner has privacy right in medical records). The Supreme Court in \textit{Nixon v. Administrator of General Services}, 433 U.S. 425, 465 (1977) held that President Nixon had a privacy interest in his communications with family members, his physician and his minister, but not in communications within the scope of his official duties. Even with respect to Nixon's personal communications, however, the Court held that the privacy interest was outweighed by the public interest in obtaining full access to the documents. \textit{Id}.

\textsuperscript{133} See Russell v. Gregoire, 124 F.3d 1079, 1094 (9th Cir. 1997) (holding law requiring community notification of sex offenders did not violate privacy rights because the arrest and conviction records were “already fully available to the public”); Cline v. Rogers, 87 F.3d 176 (6th Cir. 1996) (holding that there is no protection for information already in public record).
Without a radical expansion of tort and constitutional law doctrine, these areas are not likely to provide effective privacy protection for most database problems. What people want when they demand privacy with regard to their personal data is confidence that information about them—even if it is not confidential or embarrassing per se—will be used only for the purposes they desire.134 There is a considerable loss of privacy when someone extracts even ordinary information buried in government or business records and uses it for purposes other than the purpose for which it was originally intended.

More importantly, as more and more information about us is maintained in databases, less and less information will be considered secret. The limited federal constitutional and common law protections that currently exist will become even less relevant. If privacy law is only concerned with protecting against the release of certain highly personal, non-public, confidential information, it protects very little information at all. And if we rely on individual enforcement of common law and constitutional norms through private litigation as the principal policing mechanism, we have hardly any legal safeguard whatsoever.

B. The Development of Legislative Solutions—the Public Sector and the Privacy Act of 1974

Concerns about information privacy as a political and social issue surfaced in the 1960s. Two factors combined to bring the issue to the public agenda. One was the rapid development in record-keeping systems in both government and the private sector. The other was the computerization of information storage, retrieval, and data processing.

During the period after World War II, government agencies expanded social welfare programs, consumers and businesses became more credit dependent, and the insurance industry grew rapidly. As a by-product of these and other developments, more records containing personal information were being collected and maintained by an increasing number of institutions, both public and private. At the same time, governments, financial institutions, insurance companies, and other entities that held personal information on large numbers of individuals began to see the benefits of converting their paper files and forms to

---

134. “Between data warehousing, profiling, and bankruptcy asset liquidations, American consumers perceive that they have lost control over their personal information. For e-commerce, this belief becomes an obstacle to the growth of online transactions.” Joel R. Reidenberg, E-Commerce and Transatlantic Privacy, 38 HOUSTON L. REV. 717, 722 (2001).
computer databases. Few legal norms constrained data collection practices during this era.

The first major catalyst for the privacy debate in Congress occurred in 1965. A report of the Social Science Research Council proposed the creation of a “Federal Data Center” that would coordinate the storage and use of government statistical information among several agencies that, until then, had been operating independently. The proposal was intended to increase the efficiency of government operations by aggregating data from various government sources and combining it into a single, more versatile system. However, the proposal raised concerns that too much information would be held and maintained by one centralized source, and more generally, about the impersonality of treating citizens as data entries rather than human beings. This led to a series of hearings that focused on invasion of privacy in the computerized world. Congress ultimately rejected the national data center idea, but government agencies continued to computerize information systems hoping to increase the efficiency of government operations.

Also influential in the development of early privacy legislation was a 1973 report by the Department of Health, Education and Welfare (HEW Report) entitled “Records, Computers, and the Rights of Citizens.” The report viewed information privacy as an important and growing societal

135. See Regan, supra note 7, at 69.
136. Id. at 70.
138. Rep. Cornelius Gallagher (D-NJ), chair of the House Special Subcommittee on Invasion of Privacy, expressed this concern about creating a Federal Data Center: “If safeguards are not built into such a facility, it could lead to the creation of what I call ‘The Computerized Man.’ The Computerized Man, as I see him, would be stripped of his individuality and privacy. Through the standardization ushered in by technological advance, his status in society would be measured by the computer, and he would lose his personal identity. His life, his talent, and his earning capacity would be reduced to a tape with very few alternatives available.” U.S. House Privacy Hearing, supra note 137, at 2.
139. Regan, supra note 7, at 73.
Information Privacy

problem, and concluded that the “natural evolution of existing law will not protect personal privacy from the risks of computerized personal data systems.” Its final recommendations included the adoption of a Code of Fair Information Practices to govern recordkeeping throughout the federal government. The report was influential in creating a framework for subsequent policy formulation in both the public and private sectors for years to come.

Early legislative initiatives followed the recommendations of the HEW Report and viewed information privacy as a societal value and called for a comprehensive regulatory response in the public interest. In 1974, Senators Sam Ervin (D-NC), Charles Percy (R-Ill.), and Edward Muskie (D-Maine), introduced a bill that was broad in scope, covering all information storage systems in federal, state, and local government, as well as the private sector. Its regulatory approach included the creation

141. HEW REPORT, supra note 140, at 37. The report also called for “[t]he development of legal principles comprehensive enough to accommodate a range of issues arising out of pervasive social operations, applications of complex technology, and conflicting interests of individuals, record-keeping organizations and society, will have to be the work of legislative and administrative rule-making bodies.” Id.

142. The proposed Code of Fair Information Practices set out certain fundamental principles of privacy protection:

There must be no personal record-keeping system whose very existence is secret.

There must be a way for an individual to find out what information about him or her is in a record and how it is used.

There must be a way for an individual to prevent information about him or her that was obtained for one purpose from being used or made available for other purposes without his or her consent.

There must be a way for an individual to correct or amend a record of identifiable information about him or her.

All organizations creating, maintaining, using, or disseminating records of identifiable personal data must assure the reliability of the data for their intended use and must take precautions to prevent misuse of data.

HEW REPORT, supra note 140, at 40–41. A similar set of guidelines was developed in England and Germany at about the same time, and the Organisation for Economic Cooperation and Development (OECD) issued eight similar principles in 1980. See OECD, Guidelines on the Protection of Privacy and Transborder Flows of Personal Data [hereinafter Guidelines], available at http://www.oecd.org/oecd/pages/home/displaygeneral/0,3380,EN-document-43-1-no-no-10255-0,00.html (1980); see also BENNETT, supra note 49, at 98–100. As the OECD was developing its Guidelines, a similar effort by the Council of Europe resulted in the Convention for the Protection of Individuals With Regard to Automatic Processing of Personal Data. The Convention had similar aims, but one difference is that the Guidelines are not legally binding on OECD member states, whereas the Convention is binding on all ratifying states. However, only 21 of the Council’s 41 member states ratified the Convention. See Council of Europe, Convention for the Protection of Individuals With Regard to Automatic Processing of Personal Data, E.T.S. 108, available at
of a national Privacy Board with authority to police breaches of the privacy rules and standards established by the law. It also proposed a set of fair information practices similar to those in the HEW Report, and granted individuals the rights to see and amend their personal files, and to be informed about the release or sharing of their information.144

These comprehensive privacy bills did not fare well in the legislative process. The private sector was eventually removed from the most seriously considered privacy bills. Influencing this decision was the lack of evidence that the private sector was misusing large amounts of personal information at the time.145 The burden was placed on privacy advocates to show that a serious problem existed in private sector databases, and the evidence was unconvincing.146 Although anecdotal accounts of improper information disclosure were offered, and some evidence of systematic problems with the integrity of private database was presented, the evidence was not persuasive enough to justify a heavy handed regulatory approach that included privately held data records.147

There were practical considerations as well. The private sector was considered too complex for a centralized regulatory system and it


144. REGAN, supra note 7, at 77. At the time, several important participants in the debate considered the establishment of a federal agency to oversee information privacy practices to be an integral part of an effective plan. Among them was Arthur Miller, who called for the appointment of an “information ombudsman.” See UNITED STATES. S ENATE COMMITTEE ON THE JUDICIARY, SUBCOMMITTEE ON CONSTITUTIONAL RIGHTS, Federal Data Banks, Computers, and the Bill of Rights, Testimony of Arthur Miller, 92d Cong., 1st sess., at 19 (1971). Others recommend a “regulatory commission with full powers over the collection, use and dissemination of personal information.” See Becker, supra note 143, at 713.

145. See Before the Ad Hoc Subcommittee on Privacy and Information Systemmes of the Senate Committee on Government Operations and the Subcommittee on Constitutional Rights of The Senate Committee on the Judiciary, Privacy—The Collection, Use and Computation of Personal Data: Joint Hearings, 93d Cong., 2d sess., 515, 450–51 (statements of the American Life Insurance Association and Department of Commerce).

146. See REGAN, supra note 7, at 78.

involved too many competing interests.\textsuperscript{148} Countless factors would have to be balanced if a regulatory regime was going to protect information privacy yet not impose high costs on businesses. Even Alan Westin, a strong privacy advocate during this period, concluded that this was not the right time for comprehensive regulation of the private sector.\textsuperscript{149}

With the private sector effectively removed from the legislation, support also weakened for the creation of a national Privacy Board, or a similar independent agency with supervisory authority over information privacy in the government sector.\textsuperscript{150} A new oversight agency would be costly and, agencies argued, unnecessary because they also regarded privacy as a high priority and could monitor their own compliance with legal mandates. A separate entity would only pit one agency against another, an inefficient course when all agencies shared the same privacy protection goals.\textsuperscript{151} At this pivotal point in the debates, the focus thus shifted to privacy as an individual concern, with emphasis on agency independence and legislative solutions that gave legal rights and remedies to individuals who would be left to police their rights under the law.\textsuperscript{152}

The ultimate legislative result came after more than ten years of debate,\textsuperscript{153} when Congress enacted the Privacy Act of 1974 (Privacy

\textsuperscript{148} See \textit{Regan}, supra note 7, at 79.

\textsuperscript{149} See \textit{HEW} Report, supra note 140, at 43; \textit{Regan}, supra note 7, at 78–79. This political dynamic in the development of privacy law has continued to this day. Industry lobbyists argue that information practices have not resulted in significant economic loss to individuals and that greater protection of privacy would cost the rest of society more than any harm done to the individuals affected. See Reidenberg, supra note 134, at 724; William S. Challis and Ann Cavoukian, \textit{The Case for a U.S. Privacy Commissioner: A Canadian Commissioner's Perspective}, 19 \textit{John Marshall J. Computer and Info. L.} 1, 4 (2000); Joel R. Reidenberg, \textit{Governing Networks and Rulemaking in Cyberspace}, 45 \textit{Emory L. J.} 911, 922 (1996).

\textsuperscript{150} \textit{Regan}, supra note 7, at 79.

\textsuperscript{151} \textit{Id.}

\textsuperscript{152} Influential in the debates was this statement by President Ford:

I do not favor the establishment of a separate Commission or Board bureaucracy empowered to define privacy in its own terms and to second guess citizens and agencies. I vastly prefer an approach which makes Federal agencies fully and publicly accountable for legally mandated privacy protections and which gives the individual adequate legal remedies to enforce what he deems to be his own best privacy interests.


\textsuperscript{153} Debates over privacy policy did not end in 1974. Congressional committees held more than 150 days of hearings dealing with privacy policy between 1965 and 1988, excluding those on the privacy aspects of the FCRA. During the same period, congressional staff released more than a dozen reports on the subject. See \textit{Regan}, supra note 7, at 7.
Act). In its final form, the law provided the lowest level of protection that policy makers were considering at the time. The statute covered only federal agencies and did not establish a separate entity to oversee government information processing practices. With regard to digital databases, the policy resolution was more procedural than substantive. From the beginning of the debates, database technology had been a primary concern. A stated objective of the Privacy Act was to restrict the government’s use of technology to invade privacy interests, and the Privacy Act even included a statement that digital technology “greatly magnified the harm to individual privacy that can occur from any collection, maintenance, use or dissemination of personal information.” The solution to this growing problem, however, was not to impose substantive limits on the use of technology in collecting and sharing information, but to create a handful of procedural safeguards to give individuals a right of access to their files, an opportunity to correct errors, and a right to demand disclosure about records under certain circumstances. The only notable substantive limit in the Privacy Act was the imposition of standards of fair information handling on federal agencies to reduce the likelihood of mistakes or inadvertent disclosure to unauthorized sources, and even here the standards were vague and riddled with loopholes.

The Privacy Act is generally considered weak and ineffectual by today’s standards. Although the privacy interest was properly defined throughout the years of legislative debates as the right to control information about oneself, the most seriously debated solutions merely quibbled over how individuals could effectively guard this right. On paper, the Privacy Act guaranteed access to one’s own records and the right to correct inaccurate or irrelevant information. But these rights were not easily exercised because the costs, in time and money, were high both for the individuals involved and the agencies covered by the law.

155. Regan, supra note 7, at 82.
157. REGAN, supra note 7, at 81–82.
158. See Gellman, supra note 156, at 195–96.
159. See id. at 196–98. Gellman deftly illustrates the shortcomings of the Privacy Act as a means of satisfying the list of six Fair Information Practices developed a year earlier in the HEW Report. See HEW REPORT, supra note 140.
160. See REGAN, supra note 7, at 100.
For example, individuals were given the right not to have their information shared among agencies without their consent, but obtaining consent from massive numbers of people was impracticable.\textsuperscript{161} It was suggested that individuals be notified prior to sharing their information, but privacy advocates questioned whether this would give individuals a bona fide choice, and agencies complained that the expense of mass paper notification would outweigh any efficiencies gained from database technologies.\textsuperscript{162}

The Privacy Act had other deficiencies. It had no specific enforcement or oversight structure for ensuring compliance with the statute’s limitations on information collection.\textsuperscript{163} It limited the internal use of personal information to those agency employees who had a need to know the information in the performance of their official duties, but this proved to be a vague standard incapable of systematic enforcement.\textsuperscript{164} There was no requirement that information be used only for purposes related to the original reason for gathering it.\textsuperscript{165} No administrative process provided a safeguard or review process for internal agency use. Individuals have occasionally objected to specific uses as unauthorized by law, but in most litigated cases the agency has prevailed.\textsuperscript{166}

Perhaps the most glaring failure of the Privacy Act is the manner in which it addresses the disclosure of information to external sources. Agencies have been permitted to disclose records for virtually any purpose if the agency can establish that disclosure is for “routine use,” or it can satisfy the statute’s procedural requirements for notice by advance

\begin{footnotesize}
\begin{enumerate}
\item[161.] See Regan, supra note 7, at 86–87.
\item[162.] See Regan, supra note 7, at 96–99. Ultimately “Data Integrity Boards” (the Boards) were created in the late 1980s as intermediaries between individuals and agencies that wanted to share information or run matching programs with databases maintained by other agencies. The Boards, which every federal agency doing computer matching was required to create, had authority to reject database matching programs before they were implemented. The idea was borrowed from the Defense Department, which had implemented a similar privacy board to comply with its obligations under the Privacy Act of 1974. See Regan, supra note 7, at 95–100. Gellman, supra note 156 at 200.
\item[163.] In contrast, the Paperwork Reduction Act, 44 U.S.C. § 3507, calls for a specific administrative approval process before a federal agency collects information covered by the Act. The Office of Management and Budget (OMB) oversees both the Paperwork Reduction Act and the Privacy Act, but it devotes far more resources to the Paperwork law. See Gellman, supra note 156, at 197.
\item[164.] See Solove, supra note 32, at 1166.
\item[165.] See Gellman, supra note 156, at 198–99; See Solove, supra note 32, at 1167.
\item[166.] See Department of Justice, Freedom of Information Act Guide and Privacy Act Overview 478–80 (1994); Gellman, supra note 156, at 198.
\end{enumerate}
\end{footnotesize}
publication in the Federal Register.\textsuperscript{167} Neither has been an effective limit on information collection and sharing.\textsuperscript{168} Consequently, the Privacy Act became more of a procedural impediment for federal agencies and a symbolic, but ineffectual law for citizens, far less protective of individuals’ privacy interests than other alternatives would have ensured.\textsuperscript{169}

Congress has not materially revised the Privacy Act since 1974 because efforts to strengthen the law have met strong resistance. Three years after the Privacy Act was enacted, a report of the Privacy Protection Study Commission (PPSC)\textsuperscript{170} resurrected the idea of creating a federal Privacy Board that would monitor and implement privacy legislation and advise on the privacy implications of proposed legislation.\textsuperscript{171} Advocates of the idea saw the Board serving as an “influential prodding structure,”\textsuperscript{172} which would have supplemented self-policing legislation such as the FCRA that calls upon individuals to monitor their own privacy interests. The recommendations of the PPSC were never enacted, however, due in part to strong opposition from government agencies, trade associations, and other organizations that were enjoying the benefits of more relaxed controls on information sharing.\textsuperscript{173} In addition, although President Carter supported efforts to protect privacy in principle, establishing a federal Privacy Board conflicted with his view of limited federal involvement in this and other policy issues at the time.\textsuperscript{174}

\textsuperscript{167} Limited oversight by the OMB and Congress also has some controlling effect on the external sharing of personal information. See House Committee on Government Operations, Who Cares About Privacy? Oversight of the Privacy Act of 1974 by the Office of Management and Budget and by the Congress, H.R. No. 98-455, 98th Cong., 1st Sess. (1983).

\textsuperscript{168} See Gellman, supra note 156, at 198.

\textsuperscript{169} See id.

\textsuperscript{170} The Privacy Protection Study Commission (PPSC) was established by the Privacy Act of 1974 to examine the need for privacy legislation governing the private sector and to review the need for a general oversight body to ensure compliance with privacy rules by the federal government. See Regan, supra note 7, at 81–82.


\textsuperscript{172} Id. See Regan, supra note 7, at 84–85.

\textsuperscript{173} See Regan, supra note 7, at 85.

\textsuperscript{174} See id. at 86. The decision to reject a federal Privacy Board was critical. Public policy analysts have long recognized that the institutions chosen to pursue a public policy goal will profoundly affect the ultimate public policy resolution. See J. Brooke Overby, An Institutional Analysis of Consumer Law, 34 Vand. J. Transnat’l L. 1219, 1232 (2001) (“Under an institutional
As it turned out, an important role for a national oversight body became apparent two years after the PPSC report was issued. In an effort to reduce welfare fraud, HEW created Project Match, which compared the digital records of federal employees to the records of individuals who received benefits under the entitlement program Aid to Families with Dependent Children. In March 1979, the Office of Management and Budget (OMB) issued guidelines for Project Match and allowed it to proceed over the objections of a few federal agencies and numerous privacy advocates. Critics observed that the Privacy Act prohibited the use of information for purposes other than those for which it was initially collected, unless the individual gave affirmative consent to the different use. The OMB guidelines permitted the matching so long as the agency complied with certain procedural requirements, including advance notice in the Federal Register, and established that the matching would have “demonstrable financial benefit” exceeding its costs. Agencies did not follow the procedural guidelines in many cases, however, and the public did not respond to the Federal Register notices. Because there was little congressional opposition to the matching program, the practice grew over time. The only group with a vested interest in protesting the program was the class of individuals under investigation, but they were not aware that their records were being matched until the results of the matching detected a conflict.
For the next several years, privacy did not disappear from the congressional agenda, but the importance of other governmental interests—particularly operational efficiency and improved law enforcement—took on greater importance. Catching “welfare cheats” through information sharing was more important than limiting governmental intrusions on privacy interests. Nevertheless, the increased use of computer matching during the late 1970s and early 1980s reawakened privacy advocates and reopened the debate. The number of computer matches performed by federal agencies had almost tripled by 1984, the Orwellian year of reckoning that once again brought Big Brother to the front pages and privacy concerns to the public policy stage. Also during this period, the capabilities of computerized databases were expanding rapidly, making it possible for governments and private organizations to monitor the activities of individuals to an unprecedented degree.

By 1986, increased concern about information privacy prompted the Office of Technology Assessment (OTA) to survey the practices of federal agencies on database sharing and privacy protection. The OTA’s analysis revealed that new applications of information technology were undermining the principal goal of the Privacy Act—to give individuals more control over their personal information. The study concluded that widespread information collection, sharing, and computer networking was leading to the creation of a “de facto national database” containing information about most every person residing in the United States. This was the same concern that had ignited the privacy debate more than twenty years earlier.

180. See Regan, supra note 7, at 92.
181. See id.
182. See id. at 94 and n.83.
185. See Regan, supra note 7, at 95.
186. See supra note 138 and accompanying text.
In response to the concern about data matching, Congress enacted the Computer Matching and Privacy Protection Act (CMPPA) in 1988.\textsuperscript{187} The legislation required agency review and cost-benefit analysis before computer matching could be performed, and it instituted other procedural requirements in an effort to curb abuses.\textsuperscript{188} But the ultimate effect of the CMPPA was not to limit, but to legitimize, computer database sharing in the federal government. Rather than address privacy and surveillance concerns, the law took the same approach as the Privacy Act by emphasizing procedural and administrative goals rather than imposing limits on what kind of data could be collected and how it could be used.\textsuperscript{189}

With respect to privacy protection in the federal government, not much has changed since the late 1980s. The PPSC report in 1977 observed that “neither law nor technology now gives an individual the tools he needs to protect his legitimate interests in the records organizations keep about him.”\textsuperscript{190} Twenty-five years later, this statement rings truer than ever.

C. Protecting Privacy in the Private Sector

On the subject of information privacy outside of government,\textsuperscript{191} the Privacy Act charged the PPSC with studying the issue and making recommendations to Congress.\textsuperscript{192} As the PPSC considered the relevant public policy interests in regulating information practices, it took the position that privacy was both a “societal value” and an “individual interest.”\textsuperscript{193} Because record-keeping relationships were “inherently

\begin{footnotesize}
\begin{itemize}
    \item \textsuperscript{187} 5 U.S.C. § 552(a) (2000). For a discussion of the legislative history of the CMPAA, see \textsc{Regan}, supra note 7, at 95–99.
    \item \textsuperscript{188} See \textsc{Regan}, supra note 7, at 96–97.
    \item \textsuperscript{189} See \textsc{David A. Flaherty}, Protecting Privacy in Surveillance Societies: The Federal Republic of Germany, Sweden, France, Canada, and the United States 357 (1989).
    \item \textsuperscript{190} PPSC Report, \textit{supra} note 171, at 8. See \textsc{Reidenberg}, \textit{supra} note 134, at 722.
    \item \textsuperscript{191} Although privacy protection laws in the United States usually apply to either the public or private sector, the line between the two in information sharing is often blurred. The FBI, for example, routinely purchases information from privately assembled databases in its crime investigations. See Glenn R. Simpson, Big Brother-in-Law: If the FBI Hopes to Get the Goods on You, It May Ask Choicepoint, \textsc{Wall St. J.}, Apr. 13, 2001, at A1; \textsc{Reidenberg}, \textit{supra} note 134, at 721.
    \item \textsuperscript{192} See PPSC Report, \textit{supra} note 171, at 621–38 (schedule of hearings from credit, banking, insurance and other economic sectors); \textsc{Regan}, \textit{supra} note 7, at 83.
    \item \textsuperscript{193} PPSC Report, \textit{supra} note 171, at 21.
\end{itemize}
\end{footnotesize}
social” the societal interest could not be ignored. The PPSC developed a list of “significant societal values and interests” against which privacy should be balanced, including First Amendment concerns, freedom of information policies, law enforcement priorities, cost implications, and federal-state governmental relations. But in deciding how these interests should be balanced, the PPSC ultimately lost sight of the “societal value” of maintaining privacy and treated privacy protection as largely a problem of individual concern.

As the PPSC viewed the policy problem, the goal was to strike the right balance between an individual’s interest in keeping his or her personal information private and the need for governments and businesses to gain access to that information for various societal purposes. Consistent with the philosophical literature on privacy that had been circulating during this period, the PPSC recognized in principle that privacy has a societal value apart from each individual’s own privacy interests, but the group did not develop that value further. Only the societal benefits to governments and businesses were taken into account.

Consistent with the view of privacy as a matter of individual concern, the PPSC Report ultimately concluded that privacy policy in the private sector should begin with a voluntary, market-oriented approach. As is the case with most market mechanisms, the conclusion was premised on the principle of individual choice and the idea that individuals could assess their own risks of harm. If individuals were given the right to assert their own privacy interests, they would take measures to protect themselves, and organizations that collect and maintain information would have incentives to honor privacy concerns voluntarily. Only if voluntary compliance proved ineffective would mandatory enforcement mechanisms be imposed. Moreover, existing federal agencies, such as the FTC and state insurance regulators, were seen as appropriate control mechanisms to the extent that some government oversight was needed. A

194. Id.
195. Id.
196. Id. at 29 (“A major interest that must be weighed in the balance of organizations’ needs for information against the individual’s interest in having his personal property protected is society’s interest in maintaining the integrity of the Federal system.”).
197. See REGAN, supra note 7, at 84.
198. PPSC REPORT, supra note 171, at 32.
This legislative evolution was not surprising given the political and economic conditions. Protecting information privacy threatened defined and influential stakeholders—government agencies, employers, marketing firms, law enforcement—all of whom were just beginning to see the advantages of information technologies. All had an interest in collecting and sharing as much information as possible. Each of these stakeholders thus had incentives to redefine the issue from the ideal of privacy as a foundational societal value to some lesser ideal that required the balancing of other societal concerns—efficiency, productivity, crime control, etc.—against the individual harms that might be caused by data collection and sharing.

The result was a legislative process that quickly involved the balancing of competing interests, and the focus of debates centered on whose particular interests would be jeopardized by limiting information collection and sharing, and whether jeopardizing those interests was worthwhile. Privacy advocates were again put on the defensive to carry the burden of showing how a particular data collection or sharing activity invaded privacy and, even if it did, showing that protecting the privacy interest was not outweighed by the interests of others in gaining access to the information in question.200 This was another burden privacy advocates could not carry. This political dynamic, which has since been repeated numerous times, has led to a set of privacy laws that are sectoral in their scope and largely consist of narrowly applicable privacy provisions that do not cover the vast array of today’s data collection and sharing activities.201 Even in the sectors covered by legislative enactments, there is a heavy reliance on market-based solutions and laws that require individuals to police their own data protection interests.

1. The Fair Credit Reporting Act and the Rise of Market-Based Solutions

As it turns out, the emphasis on market solutions to the privacy problem in the 1970s foreshadowed a general shift in the perception of

199. Id. See REGAN, supra note 7, at 84.
200. See REGAN, supra note 7, at 174–75.
201. See Bradley Slutskey & Allison Brantley, Privacy on the Internet: A Summary of Government and Legal Responses, 637 PLI/Pat 85 (2001); Reidenberg, supra note 134, at 726. For a list of federal privacy laws, see Gellman, supra note 156, at 202.
privacy invasion in the following decades. The anti-regulatory position was reinforced in 1997 when the Clinton White House released a report entitled “A Framework for Global Electronic Commerce.” The first principle stated in the Framework is that “the private sector should lead” the development of electronic commerce. The section on privacy calls on private industry to work with consumer groups to develop a self-regulatory environment.

With market-based solutions as the presumed control mechanism, stakeholders in the formulation of privacy policy, whether proponents of data collection, consumer groups, or privacy advocates, became partners with largely similar objectives, differing only over the details. Resulting regulatory schemes now involve a voluntary component that has the effect of neutralizing public concern but with few enforceable restrictions on the use and misuse of data. Privacy rights have been treated primarily as commercial policy problems, rather than fundamental civil rights. Debates are often framed as if privacy were a consumer issue (e.g., caller identification blocking). Resulting laws do little to prevent or limit the collection of information in the first instance, and their success depends on individuals to seek remedies when legal norms are broken, much as they do with other consumer laws.

This self-policing approach had an early precedent in the FCRA, the first major federal privacy legislation in the United States and arguably the most detailed to date. The FCRA limits the uses for which credit

---


204. Id.

205. Id. at II, Legal Issues, § 5, par. 17.

206. David Flaherty aptly observed in 1979 that the “public is being lulled into a false sense of security about the protection of their privacy by their official protectors, who often lack the will and energy to resist successfully the diverse initiatives of the ‘information athletes’ in our respective societies.” See Davies, supra note 202, at 156–57.


208. See Gellman, supra note 156, at 202. Privacy legislation at the state level has been sporadic, in part because some courts have held that state laws regulating information trafficking can violate the Commerce Clause. See ACLU v. Johnson, 194 F.3d 1149, 1160–63 (10th Cir. 1999); American Ass’n v. Pataki, 969 F. Supp. 160, 169 (S.D.N.Y. 1997). For information on state privacy laws generally, see ROBERT ELLIS SMITH, COMPILATION OF STATE AND FEDERAL PRIVACY LAWS (1997); Electronic Privacy Information Center, Privacy Laws by State, available at http://www.epic.org/privacy/consumer/states.html (last visited Jan. 20, 2003).
reporting companies can release the information in a consumer’s file. It provides that consumers may see their files and correct mistakes, and relies heavily on consumers to ensure the accuracy of their own records. Yet only in recent times, more than twenty years after the enactment of the FCRA, have a significant number of consumers learned how to locate their credit reports and take steps to ensure their accuracy and completeness. This improvement is due in part to the increased access to information about credit reports on the Internet. Even so, most consumers have no idea how to police their rights under the FCRA.

Even with its deficiencies, the self-policing scheme of the credit reporting system is at least feasible for many consumers because most


210. United States Public Interest Research Group, Mistakes Do Happen: Credit Report Errors Mean Consumers Lose [hereinafter PIRG], at http://www.pirg.org/reports/consumer/mistakes/index.htm (March 1998); (concluding that it is still difficult for many consumers to obtain their own credit report; participants in a survey had to make several calls, deal with busy signals, and remain on hold numerous times to obtain their credit reports). See Information Resources, A Summary of Your Rights Under the Fair Credit Reporting Act, available at http://www.informationresources.com/faircredit.htm (2002) (explaining a consumer’s right to view the contents of a credit report, and the conditions under which the report must be provided free of charge).

211. R. Ken Pippin, Consumer Privacy on the Internet: It’s Surfer Beware, 47 A.F. L. REV. 125, 146 (1999) (“Credit report information is becoming more accessible on the Internet as credit reporting agencies take advantage of this growing business medium. Although a credit report is only supposed to be available to authorized customers, over-disclosure and unauthorized disclosure are certainly possible, if not more likely, on the Internet.”).

212. PIRG, supra note 210, at Executive Summary (warning that ‘until policymakers and credit bureaus do what it takes to allow consumers to have free and easy access to their credit reports and set tougher standards to prevent and clean-up mistakes, too many credit reports will remain a ticking time bomb of dangerously inaccurate information,” recommending that each national credit bureau annually and automatically mail a copy of each consumer’s report, and urging that increased duties to ensure accuracy and avoid errors be imposed on banks, department stores and other firms that furnish information to credit bureaus); Consumers Union, Credit Bureau Nightmares: Victims Speak Out, available at http://www.consumersunion.org/finance/vict.htm (Sept. 29, 1997) (discussing and illustrating the difficulty of having mistakes removed from a credit report); Dagen McDowell, How to Fix Credit Errors: Be Ready to Dog the Process Every Step of the Way, ABCNEWS.COM, available at http://abcnews.go.com/sections/business/TheStreet/dagen000714.html (July 14, 2002).

213. Among its weaknesses, the FCRA permits credit reporting companies to sell the “credit header” portion of credit histories (which contains names, addresses, former addresses, telephone number, Social Security number, employment information, and birthdate) to various commercial entities. The FCRA does little to equalize the unbalanced power relationship between individuals and credit reporting companies, and the vast amount of information in credit reports can be obtained for many commercial purposes. See Susan E. Gindin, Lost and Found in Cyberspace: Informational Privacy in the Age of the Internet, 34 SAN DIEGO. L. REV. 1153, 1206 (1997).
injuries caused by inaccurate information in a credit report will be made known to a consumer shortly after the harm occurs. The statute requires a user of a credit report, such as a creditor, employer, or insurance company to whom the consumer has applied, to inform the consumer if adverse action was taken in reliance on information in that report.214 This disclosure must identify the source of the credit report and tell the consumer that she has a right to see her report and correct any inaccuracies.215 Thus, in the ordinary course of events, a consumer will learn that a wrong has occurred and will be able to identify the problem by tracing it to the reporting agency that issued the mistaken credit history. Outside of the scope of the FCRA, this kind of feedback information is much less accessible, if at all.

2. Other Privacy Statutes that Rely on Individual Self-Policing

Unfortunately, privacy statutes subsequently enacted follow the FCRA model in this respect and rely largely on individual self-policing as the primary control mechanism, but they do not create a similar framework for ensuring that an individual will learn about a problem when it occurs. In a few instances, these laws have been reasonably successful. For example, the Family Educational Rights and Privacy Act of 1974 (FERPA),216 regulates the disclosure of student records. The statute imposes controls on the release of student transcripts, disciplinary files, and other records without the student’s (or parent’s) consent.217 One of the principle reasons why the law protects student privacy interests is that there are few lawful reasons for disclosure. Accordingly, there are few opportunities for even honest mistakes to be made. In addition, information in student records is controlled by a well defined, relatively small set of information keepers, the educational institutions, who have few incentives to disclose information about their students anyway. Information in student records has long been considered personal,

---

214. 15 U.S.C. § 1681(m) (2000). For example, if a consumer applies for a store credit card and is denied, the store must inform the consumer that a credit report was relied upon in making the decision, and it must give contact information so the consumer can find out what information is in the report.
215. Id.
216. 20 U.S.C. § 1232(g) (2000). FERPA is also known as the “Buckley Amendment.”
Information Privacy

possibly embarrassing, and sometimes hurtful if disclosed. Consequently, schools usually prefer not to disclose student records, and they have resisted market incentives to do so. Indeed, the law gives schools legal cover to refuse information requests that they would prefer to decline.

Other privacy laws that rely on individual self-policing have proved less successful. The Cable Communications Policy Act of 1984 (CCPA) prohibits cable operators from disclosing information about the viewing habits of subscribers. It also requires cable companies to have privacy policies that inform subscribers about the nature of information collected and how it will be used. The law contains a broad exception, however, that permits the disclosure of personal information for a “legitimate business activity.” Moreover, the recently enacted USA Patriot Act reduces the privacy protections afforded to cable Internet users. Before the terrorist attacks of September 11, 2001, the CCPA required cable companies to notify and grant a hearing to cable subscribers when their confidential information was subject to disclosure to the government. The USA Patriot Act took those rights away from cable broadband subscribers.

In 1986, Congress updated wiretapping and clandestine surveillance limitations with the Electronic Communications Privacy Act of 1986 (ECPA). The ECPA extends the protections of the Federal Wiretap Act of 1968 to the government’s unauthorized interception of modern forms of communications, including cellular phones, e-mail, and computer transmissions. The focus of the law, which draws heavily on the Big Brother metaphor, is on eavesdropping and monitoring communications.

---

218 See, e.g., Porten v. Univ. of San Francisco, 64 Cal. App. 3d 825, 832, 134 Cal. Rptr. 839, 843 (1976) (recognizing cause of action for privacy invasion when university released student transcript to scholarship commission without authorization).
219 Market incentives in some instances might favor nondisclosure. To the extent an educational institution relies upon good alumni relations to finance its operations, disclosing student information to outside entities could damage the school’s ability to raise funds from its alumni base.
221 Id. § 551(a)(1).
222 Id. § 551(c)(2)(A).
between one person or computer and another.\textsuperscript{227} The ECPA does not otherwise limit the collection and use of personal data, and its usefulness as a limitation on cookie technology, web bugs, clickstream data recovery, and other surreptitious Internet data mining is uncertain.\textsuperscript{228} Moreover, the law governs only the disclosure of records to government entities, so in the absence of a state law, service agreement or privacy policy to the contrary, Internet service providers are free to share the e-mail of their subscribers with non-governmental entities.\textsuperscript{229} Even the limited protection of the ECPA is in some jeopardy in the aftermath of September 11, 2001 because law enforcement agencies perceive a need to monitor e-mail and other electronic communications more rigorously and share information more freely among each other.\textsuperscript{230}

A related federal statute is the Computer Fraud and Abuse Act of 1986 (CFAA),\textsuperscript{231} also known as the “anti-hacking” statute, prohibits persons from obtaining access to a computer without authorization. The applicability of the CFAA was at issue in \textit{In Re DoubleClick, Inc. Privacy Litigation}.\textsuperscript{232} In that case, the federal district court held that the CFAA did not bar the use of cookies and other data mining activities online, and that the statutory minimum of $5,000 in damages under the CFAA was not satisfied.\textsuperscript{233} Although litigation on the scope of this statute continues, the $5,000 minimum damage threshold significantly limits the

\textsuperscript{227} See Peter Murphy, \textit{An Examination of the United States Department of Justice’s Attempt to Conduct Warrantless Monitoring of Computer Networks Through the Consent Exception to the Wiretap Act}, 34 CONN. L. REV. 1317, 1321 (2002) (explaining purpose of the ECPA).

\textsuperscript{228} The ECPA contains two distinct causes of action relevant to data collection. The first prohibits the unauthorized interception of electronic communications in transit. 18 U.S.C. § 2511. The second prohibits unauthorized access to stored electronic communications. Id. § 2701. Plaintiffs have argued both sections in recent litigation. The majority of decisions have held that the ECPA is not violated by the use of cookies, web bugs and similar data mining tools. \textit{See In re Toys R Us, Inc., Privacy Litig.}, 2001 U.S. Dist. LEXIS 16947, at *28 (N.D. Cal. Oct. 9, 2001); \textit{In re DoubleClick, Inc. Privacy Litig.}, 154 F. Supp. 2d 497, 526 (S.D.N.Y. 2001); Chance v. Ave. A, Inc., 165 F. Supp. 2d 1153, 1163 (W.D. Wash. 2001). \textit{Cf. In re Intuit Privacy Litig.}, 138 F. Supp. 2d 1272, 1278 (C.D. Cal. 2001) (district court refused to dismiss a claim based on one aspect of the ECPA—unauthorized access to stored information—although the court agreed with the \textit{DoubleClick} decision on other claims).


\textsuperscript{230} \textit{See supra} note 12; Abraham McLaughlin, \textit{CIA Expands Its Watchful Eye to the U.S.}, CHRISTIAN SCIENCE MONITOR, Dec. 17, 2001 (describing how USA Patriot Act expands the authority of law enforcement agencies to share information).


\textsuperscript{232} 154 F. Supp. 2d 497, 519 (S.D.N.Y. 2001).

\textsuperscript{233} \textit{See EF Cultural Travel BV v. Explorica, Inc.}, 274 F.3d 577, 585 (1st Cir. 2001) (holding CFAA required proof of loss of at least $5,000).
usefulness of the CFAA. Some courts have held that damages under the CFAA may be aggregated in a class action, but they can only be aggregated with respect to a single act of wrongful conduct.\textsuperscript{234} Other courts have denied aggregation of claims altogether.\textsuperscript{235} Moreover, the CFAA has a mens rea requirement that is difficult to establish.\textsuperscript{236}

Congress enacted the Video Privacy Protection Act of 1988 (VPPA)\textsuperscript{237} in response to the disclosure of Robert Bork’s video rental information to reporters during his contested U.S. Supreme Court confirmation hearing in the Senate.\textsuperscript{238} The VPPA prohibits video stores from disclosing information about the titles of video cassettes rented or purchased unless the customer has given prior written consent.\textsuperscript{239} The VPPA relies primarily on customer self-policing for enforcement, and creates a private cause of action only against stores that knowingly make prohibited disclosures.\textsuperscript{240} The statute expressly permits disclosure of the subject matter of video sales and rentals to marketing firms, and to any person if the disclosure takes place in the “ordinary course of business.”\textsuperscript{241}

The Telephone Consumer Protection Act of 1991 (TCPA)\textsuperscript{242} addresses information privacy only at the margin. The TCPA permits individuals to sue a telemarketer for certain automated dialing calls and unauthorized faxes.\textsuperscript{243} The TCPA is concerned primarily with the aggravation of disruptive phone calls. It does not limit the collection, use, or transfer of

\begin{itemize}
\item \textsuperscript{235} In Thurmond v. Compaq Computer Corp., 171 F. Supp. 2d 667, 680 (E.D. Tex. 2001), the court held that the $5,000 aggregated loss must be to no more than one computer. See also In re DoubleClick, 154 F. Supp. 2d at 523. The Thurmond court stated that damages could not be aggregated among individual plaintiffs, relying in part on Attorney General Janet Reno’s statement that “we may need to strengthen the Computer Fraud and Abuse Act by closing a loophole . . . [which would] escape [from] punishment if no individual computer sustained over $5,000 worth of damage.” 171 F. Supp. 2d at 680–81.
\item \textsuperscript{236} See In re Pharmatrak, Inc. Privacy Litig., 2002 U.S. Dist. LEXIS 15293, at n.93.
\item \textsuperscript{237} 18 U.S.C. § 2710–11 (2000).
\item \textsuperscript{238} See Elbert Lin, Prioritizing Privacy: A Constitutional Response to the Internet, 17 BERKELEY TECH. L. J. 1085, n.430 (2002); Francoise Gilbert and Brad Laybourne, Privacy Issues for the Global Company, 724 PLI/PAT 291, n.4 (2002).
\item \textsuperscript{239} 18 U.S.C. § 2710(a)(4) (definition of “video service provider” covered by the law).
\item \textsuperscript{240} Id. §§ 2710(b)(1), 2710(c) (the aggrieved person may recover actual damages, statutory damages in the amount of $2,500, punitive damages and attorney’s fees).
\item \textsuperscript{241} Id. §§ 2710 (b)(2)(D)(ii), 2710(b)(2)(E).
\item \textsuperscript{242} 47 U.S.C. § 227 (2000).
\item \textsuperscript{243} Id. § 227(b) (actual damages or a statutory award of $500; treble damages for willful violations).
\end{itemize}
personal data. In the same vein, many states have recently passed “no call” legislation to supplement the federal law. These laws represent important consumer rights legislation because by registering on a state-administered list, an individual can keep her name and phone number off of most telemarketing calling lists, but they do little to protect the privacy of information in digital databases.

An important step in the protection of individual records was the Driver’s Privacy Protection Act of 1994 (DPPA), which forces states to obtain a driver’s consent before disclosing personal information. Although the DPPA is an important development in controlling government disclosures of “public records” to the private sector, it applies only to motor vehicle records, and it authorizes the sharing of information in many circumstances. The legislation exempts law enforcement authorities, the automotive industry, government agencies, the insurance industry, debt collectors, businesses that want to verify certain identifying information, researchers, private investigators, and several other categories of inquirers.

In 1996, Congress addressed the critically important issue of health information privacy in the Health Insurance Portability and Accountability Act (HIPAA). HIPAA did not contain detailed substantive restrictions on information sharing in the health sector, but it required the Department of Health and Human Services (HHS) to promulgate regulations dealing with the privacy of medical records. HHS published its regulations under HIPAA in December 2000. The privacy

244. See Solove, supra note 31, at 1442.
245. See, e.g., IND. CODE § 24-4.7-1 (2002). The law exempts several categories of callers, including a call made “primarily in connection with an existing debt or contract for which payment or performance has not been completed.” This would include calls from creditors with whom the consumer already carries outstanding balances. Since many telemarketing calls are solicitations from current creditors of the consumer (e.g., the creditor trying to sell credit insurance), this is a significant exception. It also exempts calls made by a charitable organization, a licensed real estate broker, a licensed insurance agent, and a newspaper of general circulation. Id. § 24-4.7-1(a).
247. Id. § 2721(b). See REGAN, supra note 7, at 103.
249. 45 C.F.R. § 164.500 (2002). For an account of the legislative history of the HIPAA statute, see REGAN, supra note 7, at 105–07. As stated in the Federal Register, “The use of these standards will improve the efficiency and effectiveness of public and private health programs and health care services by providing enhanced protections for individually identifiable health information. These protections will begin to address growing public concerns that advances in electronic technology and evolution in the health care industry are resulting, or may result, in a substantial erosion of the
rules, which take effect in April 2003, will govern the use and dissemination of health information and will apply to health plans, health care clearinghouses, and certain health care providers. HIPAA requires that a covered institution obtain consent or authorization prior to using or disclosing protected health information to carry out treatment, payment, or health care operations, and contains many other requirements to safeguard patient medical information. Even without written consent or authorization, there are provisions within HIPAA which permit dissemination of patient information when identifying information has been removed. Both civil and criminal penalties are provided for non-compliance with the mandates of the privacy regulations.


250. 45 C.F.R. § 164.104.

251. 45 C.F.R. §§ 164.506(a), 164.508(a)(1) (2001), and Subpart E of the regulation generally. Beyond HIPAA’s requirement for consent or authorization, and the necessary review of medical records for identifying information, the HHS rules require additional steps to safeguard patient medical information, including: the right to adequate notice to patients of the uses and disclosures of protected health information that may be made by the covered entity, and of the individual’s rights and the covered entity’s legal duties with respect to protected health information, Id. §164.520; the need to ensure that patients may receive an accounting of disclosures of protected health information made by a covered entity in the six years prior to the date on which the accounting is requested, with certain exceptions, Id. § 64.528; the need to maintain documentation (including policy and procedure documents, communications, and other specified documents) for a period of six years, Id. §164.530; and the need to provide training to all members of a covered entity’s workforce on policies and procedures with respect to protected health information required by HIPAA. Id. §164.530(b).

252. Regulations concerning these “de-identification” procedures specify numerous identification elements, including (a) names; (b) all geographic subdivisions smaller than a State, including street address, city, county, precinct, and zip code; (c) all elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death; and all ages over 89; (d) telephone numbers; (e) fax numbers; (f) electronic mail addresses; (g) social security numbers; (h) medical record numbers; (i) health plan beneficiary numbers; (j) account numbers; (k) certificate or license numbers; (l) vehicle identifiers and serial numbers, including license plate numbers; (m) device identifiers and serial numbers; (n) web Universal Resource Locators (URLs); (o) Internet Protocol (IP) address numbers; (p) biometric identifiers, including finger and voice prints; (q) full face photographic images and any comparable images; and (r) any other unique identifying number, characteristic, or code. See id. §164.514(b)(2)(i)-(ii).

253. Affected entities may be subject to civil monetary penalties of up to $25,000 per person, per year, per standard. 42 U.S.C. § 1320d-5 (2002). In addition, federal criminal penalties may be imposed for the knowing, improper disclosure of information or the obtaining of information under false pretenses, with higher penalties prescribed for offenses involving actions designed to generate monetary gain. Criminal penalties are up to $50,000 and one year in prison for obtaining or disclosing protected health information; up to $100,000 and up to five years in prison for obtaining protected health information under “false pretenses”; and up to $250,000 and ten years in prison for obtaining or disclosing protected health information with the intent to sell, transfer, or use it for commercial advantage, personal gain, or malicious harm. Id. § 1320d-6.
The Children’s Online Privacy Protection Act of 1998 (COPPA)\(^{254}\) was the first federal law to specifically address Internet privacy concerns. The Act, which was prompted by an FTC study that found widespread abuses of privacy interests on web sites directed at children,\(^{255}\) seeks to limit the collection of personal information about children under age thirteen. If a website or web page is aimed at an audience predominantly comprised of children, it must post a privacy policy and obtain parental consent for the collection, use, or disclosure of personal information.\(^{256}\) COPPA applies only to sites that are “directed to children” or sites where the operator has “actual knowledge” that it is collecting information from children under thirteen.\(^{257}\) Moreover, it is primarily a disclosure law that imposes a strict privacy policy on Internet sites within its reach, and it relies primarily on individual enforcement for its success. It also contains a number of important exceptions.\(^{258}\)

The Gramm-Leach-Bliley Act of 1999\(^{259}\) presented Congress with an opportunity to enact meaningful privacy legislation in the financial services industry, but the law serves primarily as an enabling statute that imposes few limits on the collection and sharing of information. The Gramm-Leach-Bliley Act expressly authorizes banks, insurers, investment companies, and other financial services organizations that are “affiliated” with each other, through common ownership or otherwise, to


\(^{255}\) In its 1998 report to Congress, the FTC concluded that self-regulation was not working to protect the privacy of children online. *FEDERAL TRADE COMMISSION, PRIVACY ONLINE: A REPORT TO CONGRESS*, at http://www.ftc.gov/reports/privacy3/priv-23a.pdf (June 1998). In a review of 212 American commercial web sites aimed primarily at children aged 15 and under, the FTC found that 186 of them (88%) collected personal identifying information and 188 (89%) collected personal information. Only 109 of the 188 contained a notice of even one of the commonly accepted fair information principles, and no site practiced the full range of those principles. *Id.* at 20, 31, 36.

\(^{256}\) The law requires “verifiable parental consent” before information is collected from children 12 and younger that would allow them to be contacted online or off-line. “Verifiable parental consent” is defined as:

[A]ny reasonable effort (taking into consideration available technology), including a request for authorization for future collection, use, and disclosure described in the notice, to ensure that a parent of a child receives notice of the operator’s personal information collection, use, and disclosure practices, and authorizes the collection, use, and disclosure, as applicable, of personal information and the subsequent use of that information before that information is collected from that child.


\(^{257}\) *Id.* § 6502(a).

\(^{258}\) *Id.* §§ 6502(b)(2)(A)–(E).

share “nonpublic personal information.”

Although affiliates have to tell customers that they are sharing this information, individuals cannot block the sharing of this “nonpublic” information collected by the affiliated institutions. Because the financial services industry is dominated by large conglomerates of affiliated entities, sharing of information is routine.

Under Gramm-Leach-Bliley, customers can opt-out of the disclosure of certain “non-experience” data among affiliates and certain data that the conglomerate wishes to share with non-affiliated third parties. To do so, however, individuals must read through the often lengthy privacy policies mailed by the financial institution and learn now to exercise their opt-out rights. Bank customers who have tried to navigate through the numerous “opt out” forms mailed in recent months know that this is not as simple as it sounds. Financial institutions have no incentive to make opting out easier.

---


262. 15 U.S.C. § 6802(b) (creating obligation to give consumers the opportunity to opt out, and providing exceptions to the general opt out rule).

263. Acting Comptroller of the Currency Julie Williams commented, “Most bank customers can’t ever recall seeing something like this . . . . [I]t has been known to happen that the affiliate-sharing ‘opt out’ disclosure is buried in the middle or near the end of a multi-page account agreement. For existing accounts, some institutions have gotten into the habit of reducing the required ‘opt out’ disclosures to the fine print along with a long list of other required disclosures. Few consumers are likely to have the fortitude to wade through this mass of legal verbiage, and fewer still will take the time to write the required ‘opt out’ letter. I have even heard of people getting two separate notifications covering different types of information, requiring two separate letters to opt out.” Office of the Comptroller of the Currency, U.S. Treasury Dep’t, Remarks by Julie L. Williams, Acting Comptroller of the Currency, Before the Banking Roundtable Lawyers Council, at http://www.occ.treas.gov/ftp/release/98%2D50a.txt (May 8, 1998); see Sovern, supra note 2, at 1087, 1088; David J. Klein, Keeping Business out of the Bedroom: Protecting Personal Privacy Interests from the Retail World, 15 J. MARSHALL J. COMPUTER & INFO. L. 391, 398 (1997) (“List creators generally place [opt-out provisions] in the fine print with other boilerplate terms of the contracts; thus the clause is not readily apparent to most consumers.”).

264. In the financial services industry, opting out usually requires the account holder to read through the privacy policy of the financial institution and call a toll-free number. In the privacy policy of Bank One and First USA, the opt-out information appears on the third and fourth pages. The policy warns the customer, “Choosing to opt out of this information sharing may limit opportunities for you to receive product and service information that may interest you.” It also states that if only one customer on a joint account opts out, the bank can continue to share information about the other joint account holders. See “Important Privacy Notice,” M51388 STI40485, Letter from Carter Warren, Chief Marketing Officer, Bank One, to James P. Nehf (on file with author).
Note the holes in this patchwork of sector-specific privacy laws. For adults, there is virtually no regulation of the collection and disclosure of information on the Internet. Many Internet sites have voluntarily published privacy policies, but even those policies can offer little privacy protection, and if a privacy policy is breached the individual has little recourse under current law. It is difficult to show economic injury from the breach, and while violation may be an unfair or deceptive practice under the Federal Trade Commission Act (FTCA), in most circumstances the FTC will only obtain injunctive remedies. In addition, no federal law and few state laws make it illegal for an employer to gather and compile personal information about employees, even if the information is unrelated to the job they do. Employers can monitor our family lives, check on organizations in which we belong, ask about our medical histories, listen to our phone calls, read e-mail, listen to voicemail, monitor our computer screens, install software that tracks and counts our keystrokes, require urine tests for drugs, and check our credit reports. Other classes of unprotected records include those maintained by online and offline merchants, records held by bookstores, department stores, restaurants and clubs, and personal information profiles assembled by database companies.

D. The Failure of Self-Policing and Market-Based Solutions

One of the problems with privacy laws and regulations is that they are usually written by policy makers who lack thorough knowledge about the operation of computers and information systems. Even when lawmakers have the requisite technical background, they must try to anticipate largely unknown technological developments. Resulting laws and regulations have therefore contained broad guidelines with sufficient


266. See Benjamin F. Sidbury, You’ve Got Mail . . . and Your Boss Knows It: Rethinking the Scope of the Employer E-mail Monitoring Exceptions to the Electronic Communications Privacy Act, 2001 U.C.L.A. J. L. & Tech. 5; Amanda Richman, Restoring the Balance: Employer Liability and Employee Privacy, 86 Iowa L. Rev. 1337, 1346-47 (2000) (“[R]estrictions on preemptive screening create incentives for employers to monitor employees post-hire in order to minimize the employer’s potential liability and protect others from personal harm. . . . More than 67% of employers monitor employees, a 3.9% increase since 1997, and e-mail monitoring nearly doubled between 1997 and 1999.”); S. Elizabeth Wilborn, Revisiting the Public/Private Distinction: Employee Monitoring in the Workplace, 32 Ga. L. Rev. 825, 826 (1998).

267. See Solove, supra note 32, at 1148.
latitude to embrace later developments, but few details or specific
directions to the information processing industry concerning permissible
or prohibited activity.268 Almost by default, the laws have placed an
ermous amount of trust in individuals and the marketplace to ensure
that privacy interests are protected.

This choice of policing mechanism was not irrational. Representatives
of free information flow have long said that the market will achieve a
socially acceptable proportion of information privacy and disclosure.269
In theory, both individuals and businesses will balance the value of
personal information, such as its commercial value in the marketplace,
against the value of keeping the information within the individual’s
control.270 There are market incentives for companies to keep their
collected data secret and to be honest about their data collection policies.
Conversely, there are incentives for individuals to limit the release of
their personal information to others and to monitor the use of information
that has already been released. In some instances, market mechanisms
have worked. In 1996, for example, the online news and legal search
engine, Lexis-Nexis, announced a new service called the P-TRAK
Personal Locator, which would have given subscribers access to the
addresses, maiden names, and Social Security numbers of millions of
individuals. After considerable adverse publicity, the company changed
its plans.271

A related argument in support of marketplace solutions is that the
technology industry itself will provide an acceptable degree of privacy
protection. For example, more widespread use of cryptography has been
suggested as a technique to protect some types of privacy invasion,
particularly in the telecommunications and Internet data transfer

268. See Colin J. Bennett, Convergence Revisited: Toward a Global Policy for the Protection of
Personal Data?, in TECHNOLOGY AND PRIVACY: THE NEW LANDSCAPE 99, 104 (Philip E. Agre &
Marc Rotenberg eds., 1998).

269. One of the strongest and most powerful proponents of market-based resolutions and self-
regulation is the Direct Marketing Association. See SCHWARTZ & REIDENBERG, supra note 122. The
Online Privacy Alliance, a group of large, global corporations and trade associations also promotes
industry self-regulation and issues privacy guidelines. See Online Privacy Alliance: Privacy Policy
Guidelines, available at http://www.privacyalliance.org/resources/ppguidelines.html (Nov. 15,
1999).

270. See Solove, supra note 31, at 1446–47.

271. Along similar lines, Lotus Development Corporation and Equifax Credit Corporation had
planned to market a CD-ROM in 1990 containing information on 120 million consumers. The
parties cancelled the plans after concerned individuals posted complaints on the Internet. Kim Bartel
Sheehan & Maria Grubbs Hoy, Flaming, Complaining, Abstaining: How Online Users Respond to
Privacy Concerns, 28 J. ADVERTISING 37 (Fall 1999). See Solove, supra note 31, at 1447.
industries. Savvy consumers can install and employ a wide range of products that allow for anonymous web surfing and defense against surreptitious data mining. Computer software can be crafted to act like an “electronic lawyer,” negotiating our privacy concerns with Internet sites. Products such as Anonymizer allow users to retain anonymity while surfing the Internet. P3P technology, which provides a standard language for web sites to encode privacy policies, allows web browsers to display privacy warnings to users and block cookies. To date, such systems have yet to be widely used. If they are to become the standard control mechanism against privacy invasion on line, a massive educational effort would be needed and a universal system would need to be developed that would be compatible with most Internet sites, relatively easy for consumers to use, and difficult for data seekers to evade.

272. See Bruce Schneier & David Banisar, The Electronic Privacy Papers 4 (1997) (arguing that the FBI and other law enforcement organizations are impeding the development of new technologies that would enhance privacy but might impede crime investigations and communication monitoring).

273. Kennedy & Meade, supra note 74, at 344.

274. Lawrence Lessig more quaintly calls this service an “electronic butler.” Lawrence Lessig, Code and Other Laws of Cyberspace 160 (1999). An individual sets his or her privacy preferences once—specifying how to negotiate privacy and what information the user is willing to give up—and from that moment on, when the user enters a site, the site and the user’s machine negotiate. Only if the machines can agree will the site be able to obtain the user’s personal data. Microsoft, AOL, and IBM worked to develop the “Platform for Privacy Preferences” software (P3P) along these lines several years ago. See Jeri Clausing, New Technology Is Aimed at Web Privacy, NYTMESS.COM, June 22, 2000, available at http://www.nytimes.com/library/tech/00/06/cyber/articles/22privacy.html.


278. See Exposure in Cyberspace, WALL ST. J., Mar. 21, 2001, at B1 (survey showing that almost 30% of computer users did not know about “cookies” and almost 40% had no idea how to deactivate them).

279. The P3P program, for instance, was criticized by privacy advocates for failing to comply with basic standards for privacy protection, and for employing a protocol that was too complex and confusing. See Electronic Privacy Information Center & Junkbusters, Pretty Poor Privacy: An Assessment of P3P and Internet Privacy, at http://www.epic.org/reports/prettypoornonprivacy.html (June 2000). See also Bennett, supra note 268, at 117; Reidenberg, supra note 134, at 729 (stating for technology to provide effective privacy protection, three conditions must be met: technology respecting fair information practices must exist, the technology must be deployed universally, and the technology must have a “privacy protecting default configuration” to ensure its widespread use).
While technology might hold the key to privacy protection in the future, it would require government intervention to require privacy technology as a standard installation or default preference in most computers.\textsuperscript{280} A year 2000 survey showed that less than half of Internet users—forty-three percent—even knew what a cookie was; only ten percent said that they had set their browser to block cookies.\textsuperscript{281} For now, economic incentives more often produce technologies that enhance data collection and sharing rather than restrict it. For example, a version of the Microsoft Internet Explorer came equipped with default settings that enabled hidden surveillance of users, and a version of Netscape Communicator reported back to Netscape every time a user read e-mail.\textsuperscript{282} Because personal information has become so valuable, technologies have developed that increase data collection and decrease our ability to monitor the data collection process. This makes privacy protection even more difficult for computer users who might be interested in curbing surreptitious information collection practices.\textsuperscript{283}

1. Reasons Why Market Solutions Fail to Protect Privacy Interests

The non-regulatory solutions to the privacy problem were promoted with good intentions, but the conditions of market failure are simply too

\textsuperscript{280} Time Magazine ran a feature story on information privacy in July 2001 that recommended to readers ten steps to protect privacy. See David Jackson, et al., \textit{Internet Security}, \textit{TIME MAGAZINE}, July 2, 2001, at 50. Except for the advice not to “download anything unless you trust the sender” and “be careful what you give out,” the recommendations were not likely to be known or deployed by many computer users for years to come unless required by law as a default preference in home and business computers. \textit{Id}. The list includes: installing a home firewall, changing browser preferences to delete a user’s e-mail address and replace it with a “false name and dummy e-mail account,” opting out of information sharing policies when given the choice (although the writer acknowledges that this “can be a chore”), resetting your browser to reject cookies or install software like “Cookie Crusher,” checking to make sure a website uses encrypted transfer software before giving sensitive information online, hiding your identity with an “anonymizer” program, and clearing your memory cache each time you surf the Internet. Even with all of this advice, Time gives its readers the “Bottom line: If it has to stay secret, don’t put it on a computer hooked up to the Internet.” \textit{Id}.

\textsuperscript{281} Fox, \textit{supra} note 2.


strong. Although market-based solutions can be effective in some areas of consumer law, they are not likely to limit the use and misuse of personal information. There are several reasons why.

First, the operations of the data collection and sharing industry are not transparent. The vast majority of data collection and sharing practices occur outside public view. How do we know what information about us is being collected and when our data is being used in a way we did not expect or authorize? If the collection and sharing of information is not transparent, and it is becoming less so as data mining technologies become more sophisticated, then unfair information practices—practices most of us would object to—will likely go unnoticed. If we are not aware of malfeasance, we cannot seek redress or stop it from recurring. Moreover, requiring a public protest each time a privacy invasion occurs is not an effective privacy policy. People should not have to start a public relations campaign whenever a dangerous privacy plan is exposed.

Second, individuals cannot effectively value their personal information. To be effective, market-based solutions presume that we can value our privacy rights in some meaningful way. Only then can we make intelligent choices about whether and how to share information. Since it is impossible to know where our information will end up and how it will be used, it is difficult to assess the risks associated with giving out the information or failing to monitor its use once we have released it. We might read a privacy policy on a web site, for example, and conclude that even though the site reserves the right to share our data, we consent to the policy because we are only providing “innocent” facts and details. We may perceive the risk as small compared to the benefits provided by the web site, not knowing how or when the seemingly innocent information might be shared or aggregated with other information in a combination that will cause harm. According to one commentator, it is possible to identify eighty-seven percent of the American population knowing only a person’s birth date, gender, and postal ZIP code.285

284. See generally Reidenberg, supra note 134, at 726.
285. Erik Sherman, It Doesn’t Take Much to Make You Stand Out, NEWSWEEK, Oct. 16, 2000, at 74N. (quoting Latanya Sweeney, an assistant professor of computer science and public policy at Carnegie Mellon University). According to the FTC, 99% of the “Most Popular” websites collect personal information, including e-mail address and other personal identification, from and about consumers. The FTC concluded that most of the sites “are capable of creating personal profiles of online consumers by tying any demographic, interest, purchasing behavior or surfing behavior information they collect to personal identifying information.” See UNITED STATES, FEDERAL TRADE COMMISSION, PRIVACY ONLINE: FAIR INFORMATION PRACTICES IN THE ELECTRONIC MARKETPLACE.
To create an appearance that individuals are knowingly revealing personal information, data collection businesses often post privacy policies voluntarily. These policies are hardly “arms length agreements” between users and the data collectors, so the user’s consent is illusory in most instances. Moreover, privacy policies tend to make vague promises that commit to very little. The privacy “commitment” of the GAP credit card, for example, promises virtually no privacy protection. The policy simply informs the card holder that GAP will “collect personally identifiable information about you . . . from a number of sources,” and “may use and share all of the information” for any reason not prohibited by law. In addition, businesses can and do change their privacy policies frequently. Unless a person rereads the policy at each interaction with the business, it will never know the current practice. And if a business violates its own policy, how does one discover this fact and the extent of any injury? These problems exacerbate the difficulty of valuing information privacy.

---


287. Most Web privacy policies are little more than confusing boilerplate for the typical Internet user. Reidenberg, supra note 134, at 727–28. See also MacDonnell, supra note 83 (TRUSTe, BBBOnline, WebTrust and other seal programs do not require compliance with OECD privacy guidelines, which are discussed supra note 142).

288. See Will Rodger, Privacy Isn’t Public Knowledge: Online Privacy Policies Spread Confusion With Legal Jargon, USA TODAY, May 1, 2000, at 3D.


290. The privacy policy of the GAP Credit Card, supra note 74, states, “We may amend this Privacy Policy at any time, and we will inform you of changes as required by law.” In December 2000, the online travel service Expedia changed its privacy policy to include the following: “Expedia.com reserves the right to modify or amend this Privacy Statement at any time and for any reason. If there are material changes to this statement or in how Expedia.com will use your [personal information], Expedia.com will prominently post such changes prior to implementing the change.” Expedia.com’s Privacy Pledge, available at http://www.expedia.com/daily/service/privacy.asp?fpr=480&CCheck=1 (last visited Jan. 27, 2003). Despite the assurance that changed privacy rules will only apply to data acquired after the change, practical difficulties may make it impossible for Expedia to fulfill this promise, and for users to monitor Expedia’s compliance. See Kennedy & Meade, supra note 71, at 321, 337.

Another aspect of the valuation problem is the difficulty in pricing specific pieces of personal information out of context. How do I value my monthly grocery store shopping list? One might say that its value is the amount of store discounts I receive each month because I voluntarily trade that information for these cost savings when I use my store “convenience” card. Yet to value this information with even remote accuracy, I need to know how the information will be used. The store might sell it to a marketing firm in the aggregate (along with data from all customers) without any name identification, in which case I might value its release nominally. If the store sells it to insurance companies and financial institutions with name identification, I would likely value it much higher. Most personal information may never come back to haunt us, but a few items of information could be used to wreak havoc if identity fraud occurs. Once the information is stored and capable of being accessed, we lose control over our fate. We often do not have enough information to evaluate this risk.

Voluntary “trustmarks” or “web seals” are not a suitable substitute for legal mandates. Seal programs can help web users gain confidence in the privacy practices of the sites they visit, but the most popular programs do not require that damage remedies be readily available to the victims of information misuse, and the scope of the web seal “guarantee” of privacy can be narrower than individuals might expect. For example, TRUSTe certifies sites that promise not to share information “used to identify, contact, or locate a person.” Yet reports show that most Internet users do not want Web sites tracking their movements even if


294. See TRUSTe Program Principles, available at http://www.truste.org/programs/pub_principles.html (last visited Jan. 20, 2003). TRUSTe is an independent, non-profit organization founded in 1997 by the CommerceNet Consortium and the Electronic Frontier Foundation. TRUSTe developed a license agreement that governs a licensee's collection and use of information and requires licensees to adhere to standards for notice, choice, access, and security. The program includes third party monitoring and periodic review of licensee information practices. See Pippin, supra note 211, at 132.
the site does not associate the data with a particular user’s identity. In addition, the most popular seal programs do not perform regular and rigorous audits on their client’s web sites to ensure that the web seal standards are being satisfied.

Third, there are accountability problems with data collection and sharing practices. Tracing an injury to a particular cause, source, or leak will often be impossible. In the context of identity theft, a typical complaint is, “How did the thief get this information about me?” With information about us residing in so many databases, if a problem does surface, there may be no way to locate the original source of the leak. Without accountability, market forces cannot effectively curb harmful behavior.

Fourth, if we want to participate in modern society, we have little choice but to reveal information about ourselves. If we want the job, the loan, or the medical care, we have to disclose information about ourselves and our lives. Market solutions presuppose choice, and where the choice to reveal information is limited, the market will fail to protect our interests.

Perhaps most fundamentally, however, one may ask why the burden should be on the individual to figure out how to keep others from getting, selling, and using information. Self-regulation assumes that information about us is a property right or commodity that can be bought and sold. Viewing privacy in this way produces an ineffective system of sporadic notice and illusory choice. This approach ignores other universally recognized principles of fair information practice such as minimizing the amount and type of data that can be stored, and restricting access to the

295. Reidenberg, supra note 134, at 727–28. See also MacDonnell, supra note 83, at 348–49 (explaining that TRUSTe, BBBOnline, WebTrust and other seal programs do not require compliance with OECD and Canadian privacy guidelines).

296. See MacDonnell, supra note 83, at 392. Auditing compliance with web seal mandates is a burdensome task and can be costly if the audit is rigorous. One solution to this problem was advanced by Colin Bennett in a slightly different context. Bennett’s scheme would involve three tiers of compliance with privacy standards: 1) a conformity of “policy”; 2) a conformity of “procedure”; and 3) a conformity of “practice.” Only an organization seeking certification in the third tier would undergo a complete privacy audit to ensure that it honor its representations in practice. See Colin Bennett, Prospects for an International Standard for the Protection of Personal Information, available at http://www.e-com.ic.gc.ca/english/privacy/632d29.html (August 1997).

297. Studies have shown that nearly 80% of identity theft victims do not know how or where the thief obtained their personal information. Lucas, supra note 4.

298. The FCRA, for example, mandates that individuals consent before an employer can obtain their credit report. 15 U.S.C. § 1681(b)(2) (2000). This consent is virtually meaningless if the person wants the job.
data that can legally be stored, organized, and sorted. Only by deemphasizing the commodity aspects of information privacy, and focusing on the societal value of keeping information private, can we create a different and more effective regulatory model. More “notice and consent” requirements are not likely to provide greater privacy protection.

Evidence showing the failure of market-based regulation has been gathered by the FTC. After years of promoting market based privacy measures and waiting for acceptable industry-regulated fair information practices, the FTC in its May 2000 report concluded that broad-based legislation is necessary to ensure fair information practices online. The agency noted private sector initiatives to develop self-regulatory regimes, but concluded that industry measures were far from adequate. The FTC report recommended that technology neutral legislation be enacted, and called for an implementing agency with regulatory and supervisory authority. New leadership at the FTC, however, has recently backed off of this recommendation, calling for increased enforcement of existing laws instead. The retreat is unfortunate because, for a short period at least, it looked as if we might be moving toward a privacy policy that is more compatible with the idea of privacy as a societal value. Without prompting from the FTC or another influential voice in the privacy debate, the status quo is likely to remain for some time.

299. Self-regulation also enables data collectors to change the rules after the data has been collected from individuals. Reidenberg, supra note 134, at 727.
300. A number of bills have been introduced that would impose various types of notice and consent requirements, including S. 1055, 107th Cong. (2001); H.R. 89, 107th Cong. (2000); H.R. 237, 107th Cong. (2000); and H.R. 2135, 107th Cong. (2001).
301. FTC REPORT, supra note 285.
302. In a random sample of 335 American commercial web sites that collected personal identifying information, only 20% applied fair information practices. The figure was higher, 42%, for 91 of the 100 busiest commercial sites. FTC REPORT, supra note 285, at 20.
303. Id. at iii, 36.
III. PRIVACY AS A SOCIETAL VALUE

A. Defining Information Privacy as a Societal Value

It has been four years since Sun Microsystems CEO Scott McNealy issued his now famous warning, “You have zero privacy anyway. Get over it.”305 So much information about us is already in government and private sector databases that it may be too late to rethink our approach to information privacy protection.306 Yet, if we begin to think about information privacy as an important societal value rather than a typical consumer law problem calling for a balance between business and consumer interests, we may be able to achieve several important goals. However, we must initially agree on a set of privacy objectives. I suggest three baseline goals for a national privacy policy.

1. Fundamentals of Effective Privacy Policy

First, while we have begun to recognize that certain uses of even non-confidential information can threaten privacy, it is not sufficient to impose limits on the use of information once it has been collected. Procedural safeguards alone cannot protect the confidentiality of information adequately. Thought must be given to the types of information that can be collected in the first instance—by public authorities and private sector enterprises—because once information is in a database, controlling its subsequent use is extremely difficult. For instance, why should we allow employers to collect virtually any information on applicants and current workers and store the information in a database? We should reevaluate our presumption that virtually any piece of personal information can be stored electronically.

Second, we should acknowledge that the current self-regulatory approach, which requires individuals to police the market to ensure that their data is not collected or disseminated, is ineffectual. A more complete range of enforcement schemes should be developed to control how information will be collected, used, and shared. To the extent market mechanisms are used, regulation mandating that consumers opt-in rather than opt-out, for example, can in some circumstances achieve a

306. See AMITAI ETZIONI, THE LIMITS OF PRIVACY 131 (1999) (observing that the genie is already out of the bottle).
more balanced approach to the disclosure of information. Opt-in systems place the incentive on entities that want to acquire personal information, in both the government and private sector, to make it as easy as possible for individuals to give meaningful consent to the collection and use of their information.

Third, a national Privacy Board, or an institutional network of privacy agencies, is necessary to ensure that government and private sector data collectors maintain fair information practices. Stiff penalties should be imposed upon those who breach privacy agreements with their customers (not just a finding under the FTCA that an unfair trade practice has been committed). Watchdog organizations and individual lawsuits cannot be relied upon to assume this responsibility. Organizations like the Electronic Privacy Information Center (EPIC) bring legal actions to accomplish privacy protection goals, but they work piecemeal, vary in their targets depending on the priorities of the organization bringing the suit, and succeed only when state or federal laws have been broken. Individual lawsuits are even rarer. More importantly, however, litigation should not be the primary enforcement mechanism for citizens who can rarely afford to sue the government or a large commercial enterprise.

These three changes would mark a fundamental shift from a presumption favoring the collection and sharing of personal information

307. A survey conducted in 2000 showed that 86% of respondents favored privacy policies requiring organizations to seek explicit permission before gathering any personal information. See Fox, supra note 5. Several states require affirmative consent from individuals before their personal information can be shared with others. See William M. Fay, Jr., Lost in Oz: There Is No Yellow Brick Road for State Lawmakers to Follow in Drafting Privacy Legislation for Insurers, 7 CONN. INS. L.J. 585, 604–18 (2000).

308. To the extent websites give users a choice, most use an “opt-out” system, i.e., unless a user affirmatively indicates that he or she does not want information sold or shared to another party, the site is free to circulate the information. An August 2000 survey indicated, however, that 86% of users favored “opt-in” policies that require affirmative consent from the user before information can be shared. See Dylan Tweney, The Rules for Writing a Privacy Policy, ECOMPANY NOW, available at http://www.ecompany.com/articles/web/0,1653,8297,00.html (Sept. 7, 2000).


Information Privacy

and toward a neutral position or a presumption against it. How can we get there from here? We should begin by convincing policy makers to look at information privacy in a different way. How an issue is defined on the public agenda is important to the politics of the law-making process and ultimately the policy resolution.\(^{310}\) When information privacy is defined as a matter of individual concern, it is difficult to see a broader purpose in controlling the problem. There is no polluted atmosphere or outbreak of disease that identifies the issue as one of general public concern demanding a societal, rather than individual, resolution.

2. Moving Toward a Societal View of Privacy

Priscilla Regan and others have argued that privacy serves not just individual interests, but also common, public, and collective purposes. Privacy is a common value because we all recognize its importance in our lives, a public value because it is necessary to the proper functioning of a democratic political system, and a collective value because technology and market forces make it increasingly difficult for any of us to have privacy unless we all have privacy at a similar level.\(^{311}\) If privacy is regarded as being of societal importance, different policy discourse and interest alignments are likely to follow, and this opens the way to serious consideration of different policy resolutions.

Theoretical underpinnings for a societal view of information privacy began in the 1960s and early 1970s. Privacy literature was abundant during this period, and several writers acknowledged that privacy was important to society at large, not just to individuals.\(^{312}\) Yet there was little development of the idea that privacy policy should be created with this perspective in mind. Alan Westin, quoting from Robert Merton’s *Social Theory and Social Structure*, wrote: “Privacy is not merely a personal predilection; it is an important functional requirement for the effective operation of social structure.”\(^{313}\) Abstract declarations of this sort failed

---

310. See Regan, supra note 7, at xiii, 220–31.
311. Id. at xv–xvi. See also Paul M. Schwartz, Privacy and Democracy in Cyberspace, 52 Vand. L. Rev. 1609, 1653 (1999) (stating that database privacy is necessary for democratic deliberation).
312. See, e.g., Westin, supra note 85, at 58; James Rachels, Why Privacy is Important, 4 Phil. & Pub. Affairs 323 (1975) (observing a “close connection between our ability to control who has access to us and to information about us, and our ability to create and maintain different sorts of social relationships”).
313. Westin, supra note 85, at 58 (quoting Robert Merton, Social Theory and Social Structure 375 (1957)). See Regan, supra note 7, at 32.
to influence the formulation of public policy. Even in the philosophical literature during this period, the “effective operation of social structure” quickly evolved into advocacy for privacy as a component of individual freedom and liberty. The result was an approach to privacy policy that focused on the individual and the legal protection of his or her rights versus the rights of others to use personal information for their own purposes. More recent writing has similarly emphasized privacy’s importance to the individual, and the policy debate has sought to balance the individual right to privacy against broader societal interests in information collection and sharing.

An effort to redefine the societal value of privacy was begun, however, by Regan and privacy proponents such as Colin Bennett. Bennett writes about the humanistic and political aspects of privacy. The humanistic value recognizes the loss of dignity, autonomy, or respect for the individual that results when we lose control over personal information. For some privacy advocates, the humanistic dimension is the only justification needed for protective public policy. The very collection of personal information, regardless of how it is used, contributes to the sense of alienation in post-industrial society. Information technology adds a new layer to the already impersonal

314. See Regan, supra note 7, at 32–33.
315. Political theorists identify two types of “rights”—civil liberties and civil rights. Privacy has often been characterized as a civil liberty, the right to be free from interference from other individuals, governments or organizations. See Isaiah Berlin, Two Concepts of Liberty, in Four Essays on Liberty 118 (1969) (privacy as a negative liberty); Vincent J. Samar, The Right to Privacy: Gays, Lesbians, and the Constitution 53 (1991) (privacy as a “negative freedom”). As such, it loses some of the legitimacy that civil rights have in American politics. Defining a problem as a civil right can be a successful political strategy (e.g., women’s rights, minority rights, rights of the disabled). Civil rights movements in the United States have usually assigned some benefit or status to a group rather than to an atomistic individual. As privacy began to be viewed as a policy matter of individual liberty, its claim to status as a civil right diminished. See Regan, supra note 7, at 4.
316. See, e.g., Julie Inness, Privacy, Intimacy, and Isolation 140 (1992) (privacy important because it gives individuals more control over the making of intimate decisions); Post, supra note 93, at 2091 (privacy as “important to individuals to resist misjudgments based upon private information,” and loss of privacy as “particularly hurtful to individuals”).
character of government bureaucracy and commercial relations with individuals. It contributes to an uneasy sense that “someone out there knows something about me,” a sentiment which alone should get the attention of policy makers.

A political aspect of privacy policy is at work as well. Privacy is rooted in the classical liberal belief in limited government and a general distrust of powerful institutions, whether they are public or privately owned. Information technology enhances the power of the government and commercial enterprises to obtain and manipulate information about us. As power shifts further away from the individual to large institutions that can affect the individual’s life and liberty, we have a collective cause for concern and a need for a political resolution.

This revival of a societal view of privacy coincided with the emergence of the “politics of ideas” movement, which became an important model for explaining policy making in the 1990s. The model received considerable attention and some popular appeal during the Clinton administration, though its application in practice has been uneven at best. On the theoretical side, the model has both normative and descriptive aspects. The notion that public policy should promote the general public good, rather than accommodate competing claims of influential stakeholders, has long been part of our political culture, if not our political reality. More recent scholarship, however, has emphasized the descriptive aspect. Ideas about what is good for society, rather than what emerges from a battle of individual self interests, can explain more policy making activity than interest group models would predict. Ideas associated with the public interest, commonly shared


322. BENNETT, supra note 49, at 30 (quoting DAVID BURNHAM, THE RISE OF THE COMPUTER STATE (1983)) (“The ‘rise of the computer state’ is regarded as a threat to the liberal values that have such a central place in the American Heritage.”). Viewed in this way, information privacy is “rooted in Lockean liberalism: inalienable human rights, limited government, the rule of law, and a separation between the realms of state and civil society.” Id. at 31.


324. On interest group politics in general, see THEODORE J. LOWI, THE END OF LIBERALISM: IDEOLOGY, POLICY, AND THE CRISIS OF PUBLIC AUTHORITY (1969); GRANT MCCONNELL, PRIVATE
principles of morality, justice, and the collective welfare occasionally do overcome narrower interests as lawmakers evaluate policy alternatives. Some of the work in this area has focused on deregulation, particularly in the trucking, airline, and telecommunications industries, where deregulation occurred despite initial opposition by powerful interest groups such as the targeted industry, labor unions, and regulatory oversight agencies.  

In the instances where fundamental policy change results despite the influence of well-organized special interests, three conditions seem to be present. First, an outside force or public event brings the problem to the front of the political agenda. Second, popular symbols or slogans emerge to give the idea political strength. Third, a forceful yet latent public interest pushes the idea through to ultimate enactment. Interestingly, all three seemed to be present in the information privacy debate of the 1960s and early 1970s. Rapid changes in information technology and the proliferation of large computer databases brought renewed interest to the problem and continued throughout the period of debates. The idea of privacy protection evoked powerful symbols in American culture. Big Brother, 1984, and A Clockwork Orange created fearful images that helped get the issue on the public agenda. Finally, as numerous public interest surveys revealed, there was and continues to be a strong public interest in protecting information and controlling the technologies that could compromise it. 

Despite this strong public interest in protecting privacy, strong, countervailing forces ensured that privacy as a political ideal and as an accepted social value would not have a powerful influence on policy making. By framing the debate in terms of individual rights—the right of an individual to control access to information about himself or herself—policy makers elicited two responses that narrowed the range of possible resolutions. First, privacy was viewed as a matter of “individual utility.” Each of us considers how valuable our own information is to

---


325. See generally MARTHA DERTHICK & PAUL J. QUIRK, THE POLITICS OF DEREGLATION (1985) (arguing that the idea of deregulation was powerful enough to prevail over stakeholders in business and government); TIMOTHY J. CONLAN, ET AL., TAXING CHOICES: THE POLITICS OF TAX REFORM (1990); Priscilla M. Regan, Ideas or Interests: Privacy in Electronic Communications, 21 POL’Y STUD. J. 450 (1993).

326. See REAGAN, supra note 7, at 18.

327. See id. at 176–77.

328. See id. at 178.
Information Privacy

us, and how we might be hurt if the information is shared. If individuals have not personally been affected in a negative way, then the problem is not as important and the person may disengage from the debate. Second, because rights are not absolute, privacy had to compete with other rights and social interests, such as the right of commercial speech and societal interests in efficient government or business activity. As a result, the influence of interest group politics could not be overcome. 329

Another circumstance that affected privacy policy resolutions during this period was the existence of information technology at the periphery of public awareness and consequently at the sidelines of political discourse. Despite the appeal of popular symbols like 1984, the costs, benefits, and broader implications of computer systems were largely indirect, if perceived at all. Without the consciousness-raising equivalent of a nuclear accident, most Americans did not view privacy as a political priority. 330

In addition, policy makers were not writing on a clean slate. Previous court rulings had discussed privacy and frequently balanced privacy “rights” against other societal interests. In 1976, the Supreme Court decided United States v. Miller, 331 holding that records in possession of a third party (a bank) are considered property of that party. Under these circumstances, the individual does not have a Fourth Amendment privacy interest in the records. 332 Thus, the idea of privacy as an individual right or individual liberty had already been curtailed.

One lesson from the earlier privacy debates is that privacy, when framed as a matter of an individual interest, does not easily tap into the idea that there is a broader public purpose being pursued. In fact, those seeking to weaken proposals to protect individual privacy have successfully framed their position as a commitment to the common good (i.e., better government service, stronger law enforcement capabilities, and a less encumbered business environment). To date, arguments for

329. See id.
330. BENNETT, supra note 49, at 22–23; JAMES N. DANZIGER, ET AL., COMPUTERS AND POLITICS: HIGH TECHNOLOGY IN AMERICAN LOCAL GOVERNMENTS 1 (1982) (“the costs and benefits and the broader impacts of computer systems are largely perceived as indirect and subtle, if they are indeed perceived at all”).
332. See REAGAN, supra note 7, at 179.
privacy protection have not successfully transcended this political dynamic.333

3. Defining Characteristics of Societal vs. Individual Concerns

In an effort to move the debate forward, the theoretical foundations for a societal view of privacy must be buttressed by more policy-oriented arguments. What are the basic characteristics of issues that we generally view as societal concerns, justifying a resolution in the broader public interest? Why do we not say, for example, that individuals should police their own environmental problems and seek redress if an industry breaks the law or violates its own clean-up program? Why is environmental protection viewed as a general societal problem calling for a regulatory response in the public interest?

There are sound reasons why a societal problem like environmental pollution calls for a societal resolution. Societal problems have six defining characteristics:

Involuntary and unavoidable risk. We are all more or less equally and involuntarily at risk simply by living in and sharing the same environment. We have no real choice in this regard, and we are all equally susceptible to injury if legal norms are breached. We do not know which of us will become ill if toxins are released, so we all have an equal interest in controlling the risk.334 While a person might take some steps to minimize individual exposure (e.g., eating healthy foods or moving to a city with cleaner air), we realize that such measures help, if at all, at the margin.335 Most risks are uncontrollable by individual action.

333. Legislation in the United States must survive many “veto points” in the legislative process. A bill can lose momentum as a result of committee procedures, bicameral approval difficulties, floor amendments in either chamber, failure of presidential signing, and possibly another set of hurdles in agency implementation. Even in the best of political climates, any attempt to pass a comprehensive data protection law would be challenged by an influential army of industry lobbyists. See Bennett, supra note 268, at 114.

334. See ROBERT V. PERCIVAL, ET AL., ENVIRONMENTAL REGULATION: LAW, SCIENCE, AND POLICY 4, 5 (3d ed. 2000) (listing the basic characteristics of environmental concerns as follows: collective risks, uncertainty of mechanism and effect, potentially harmful effects, irreversibility, and uncontrollability).

335. See, e.g., Daniel Machalaba, Local Ties: Decades of Mishandling Hazardous Cargo Leave Railroads a Toxic Legacy: Areas Near Rail Yards Face Possible Health Problems; Lawsuits Are on the Rise, WALL. ST. J., at A1, Feb. 3, 1999 (reporting that a resident who lived across from a rail yard for forty-six years used to eat watermelons and cantaloupes grown in her garden but now restricts her plantings to flowers, fearing contamination in home-grown produce; another resident says he will not allow his four children to play outside anymore for fear the ground is contaminated, explaining that he feel like they are “prisoners” in their own house).
**Difficulty in identifying individual harm.** When injuries occur, they are often not known or even knowable. A harm resulting from environmental contamination can be latent or, in a great number of cases, undiscoverable. Yet even if an injury is undetectable, we are nonetheless concerned about potential injuries or damage that is occurring without our knowledge. Indeed, unseen harms frighten us just as much as known ones.

**Obstacles to tracing injury to its cause.** If a person does discover an injury and suspects an environmental cause, tracing the harm to a particular source is often impossible. The source may be unknown, unknowable, or there may be many possible contributors so it is impossible to identify the perpetrator. We may identify toxins in ground water as the cause of a person’s cancer, but we will have difficulty proving causation against a particular farm, industry or other contaminating source. Waste secretly dumped often avoids detection.

---

336. See Thornton v. Roosevelt Hosp., 391 N.E.2d 1002, 1003 (N.Y. 1979) (plaintiff was exposed to a carcinogenic substance in 1954, but did not develop cancer until 1972); Le Vine v. Isoserve, Inc., 334 N.Y.S.2d 796 (N.Y. Sup. Ct. 1972) (plaintiff developed cancer nine years after exposure to radioactive isotope); Carl B. Meyer, *The Environmental Fate of Toxic Wastes, the Certainty of Harm, Toxic Torts, and Toxic Regulation*, 19 ENVTL. L. 321, 330 (1988) (reporting that the impact of toxic exposure is often intensified because many toxic substances are neither visible nor malodorous; frequently, victims neither suspect nor avoid exposure to toxic waste until its effects, enhanced by cumulative exposure, manifest themselves in acute or chronic discomfort or harm); Christopher W. Krueger, *Legislative Relief from Toxic Exposure: The Lifeguard Presumption Act*, 3 S. CAL. INTERDISC. L.J. 867, 872 (1994) (noting that the gestation period for cancer is indeterminable, appearing anywhere from six months to fifty years after the initial toxic exposure); CONGRESSIONAL RESEARCH SERVICE OF THE LIBRARY OF CONGRESS, COMM. ON ENVIRONMENTAL PUBLIC WORKS, U.S. SENATE, 96th Cong., 2d Sess., *Six Case Studies of Compensation for Toxic Substance Pollution* 43 (June 1980) (noting that harms at Love Canal (an abandoned hydroelectric canal in New York, where the Hooker Chemical Company placed harmful substances from 1942–53) arose over twenty-five years after the last dumping of hazardous waste).

337. See Bill Charles Wells, *The Grin Without the Cat: Claims for Damages from Toxic Exposure Without Present Injury*, 18 WM. & MARY J. ENVTL. L. 285, 310 (1994) (commenting that the development of an action or element of damages to compensate the plaintiff for fear of injuries that he has not yet suffered is a more recent development); Terry Morehead Dworkin, *Fear of Disease and Delayed Manifestation of Injuries: A Solution or a Pandora’s Box?*, 53 FORDHAM L. REV. 527, 567 (1984) (recognizing that although precedent exists for allowing toxic tort plaintiffs who sue for fear of disease to recover, many courts have opted to focus on requirements such as physical injury or other economic harm).

Inadequacy of money damages. Once someone is injured, and assuming we can identify the cause, money is a rough compensation for the harm, and it will not make the person whole, even if damages can be recovered. Money damages are well suited for economic injury, but they are at best a crude substitute if the harm is not easily translated into economic terms.\textsuperscript{339} In such cases, we are better off preventing the harm from occurring in the first place. Money cannot replace the loss.

Externalities. With societal problems such as environmental contamination, it is not possible to charge the full cost of the harm against the entity that caused it.\textsuperscript{340} Pollution imposes costs on others that are not easily recoverable. Unclean air can increase health care costs, raise expenses for cleaning buildings, and decrease work productivity for those who take ill. Without government intervention of some kind (such as a tax on toxic emissions), the costs are not borne by the entity causing them.

Non-economic value in preventing the harm. Many aspects of life are difficult to quantify in economic terms. Most of us would agree that there are important intangible benefits to having certain legal norms in place, whether or not we can identify an economic benefit from their existence. We gain pleasure from having a clean environment (fresh air, good fishing waters, etc.) even if we do not suffer any obvious or tangible adverse effects when they are gone. While some quality of life benefits might be discussed in economic terms (e.g., we might estimate the value of clean drinking water by looking at the revenues of the bottled water industry), such an analysis seems counterintuitive and unnecessary in

\textsuperscript{339} See Meyer, supra note 336, at 370 (stating that environmental and personal injury damages are complex and difficult to measure); Schroeder, supra note 338, at 589 (arguing that the goal of modern environmental regulation is to prevent harm to the environment before it occurs, with an implementation structure that includes prior approvals, permits that embody standards to be met, and the monitoring of compliance).

\textsuperscript{340} See EBAN S. GOODSTEIN, ECONOMICS AND THE ENVIRONMENT 33–39 (1995) (discussing why it is not possible for victims of negative externalities to simply band together on their own to prevent pollution); Kenneth S. Abraham, The Relation Between Civil Liability and Environmental Regulation: An Analytical Overview, 41 WASHBURN L.J. 379, 379 n.2 (2002) (externalities resulting from barriers to the imposition of liability on those who create environmental risk were the principal justifications for the development of environmental law, and especially for the enactment of the major federal regulatory statutes of the 1970s); Henry N. Butler & Jonathan R. Macey, Externalities and the Matching Principle: The Case for Reallocating Environmental Regulatory Authority, 14 YALE L. & POL’Y REV. 23, 29 (1996) (arguing that the economic goal of government regulation of pollution is to force polluters to bear the full costs of their activities).
many circumstances. We enjoy certain things because we are human, and our lives would be less fulfilled if they were gone.

In contrast to a problem of societal concern, a problem of individual concern such as a consumer warranty complaint or deceptive advertising program typically has a very different set of characteristics:

**Voluntarily assumed risk.** The risk of injury is usually created individually and assumed voluntarily. By purchasing a car, we create expectations of quality from a specific product or manufacturer; by signing a lease, we establish a landlord/tenant relationship with a management company we chose; by dropping off shirts at the cleaners, we entrust the local merchant with our property. We make choices and voluntarily assume certain risks when we do so, and in this way we distinguish ourselves from the rest of society.

**Discoverable injury.** The car does not work properly, or the landlord breaches the lease. Although we might not know exactly what went wrong until further investigation reveals the actual cause, we usually know that an injury of some kind has occurred. We can then try to trace its cause and seek appropriate redress.

**Fewer tracing obstacles.** With most consumer problems of individual concern, tracing the injury to a particular source is typically not the problem. We know who caused the injury because we know who the other party in the relationship was. Even if the source of the problem is not readily identifiable, the list of possibilities is usually small and finite. For example, in an unlawful debt collection practice the wrongdoer will

---


342. Many consumer relationships are based in contract, and the relationship is voluntarily undertaken. See Stephen J. Ware, Employment Arbitration and Voluntary Consent, 25 Hofstra L. REV. 83, n.134 (1996) (citing ALAN WERTHEIMER, COERCION 4 (1987) (noting that “the general assumption is that promises are binding . . . if, but only if, the relevant actions are voluntary”)); Robert E. Scott & William J. Stuntz, Plea Bargaining as Contract, 101 YALE L.J. 1909, 1919 (1992) (the central normative justification for contractual enforcement is facilitating the exercise of voluntary choice). But see Michael Philips, Are Coerced Agreements Involuntary?, 3 LAW & PHIL. 133, 133 (1984) (observing that the term “voluntary” could be used to describe all willful acts, i.e., as a synonym for “volitional”; Michael D. Bayles, A Concept of Coercion, in XIV NOMOS: COERCION 16, 18 (J. Roland Pennock & John W. Chapman eds., 1972) (“[A] man who is physically forced to squeeze the trigger of a gun does not do it voluntarily in any sense. But a man who fires a gun due to a threat does in one sense act voluntarily although he does not in another.”).
likely be either the creditor itself, a financial institution as the creditor’s assignee, or a collection agency to whom the creditor transferred the debt.

**Economic injury predominates.** Transactions usually involve payment for goods or services. If we are compensated for the economic losses, we have some confidence that justice was done. Although full compensation may be difficult to obtain given that we may have to pay attorney’s fees or we have difficulty proving consequential harm, injury is usually economic and money damages can compensate for it.

**Fewer external costs.** In most consumer transactions, externalities are usually nonexistent or a small part of the harm done. The injury caused by breaking the legal norm usually affects only a single individual or class of individuals. Ripple effects to society at large are usually not significant.

When viewed in this light, information privacy fits the “public interest” model better than the “individual concern” model:

- We are all equally at risk of injury from misuse of our data, and we cannot avoid the problem if we are to participate in modern society. Information about us is seemingly everywhere, and we can do little to minimize its collection and use.343

- Except in the most egregious situations, harms resulting from information misuse may never be known to us. So much of our data is being shared every day, yet we have no idea what the ramifications may be (good or bad) or what decisions are being made in reliance on it.344


344. See Adam S. Marlin, *Online Identity Theft a Growing Concern*, CNN.COM, at www.cnn.com/2000/TECH/computing/08/16/id.theft.offline.idg/index.html (Aug. 16, 2000) (describing how an identity thief obtained a doctor’s personal information from the Medical Board of California and another web site (medical license and Social Security number) and used them to buy medical supplies on his credit; by the time the doctor convinced the medical supply company that he had not made the purchases and learned that someone had stolen his identity, the identity thief had spent $185,000); PRIVACY RIGHTS CLEARINGHOUSE, IDENTITY THEFT VICTIMS’ STORIES, LEGISLATIVE TESTIMONY OF JOHN AND JANE DOE (testifying before the Maryland legislature that he was shocked to find out that he and his wife had no credit: “We were being accused of defaulting on loans, not making car payments, and overdue on credit card payments. We were suddenly being called by stores that we never heard of, banks demanding payment on cars or loans that we didn’t have, collection agencies demanding that we pay immediately on some account we never heard of, or face legal action against us.”) at http://www.privacyrights.org/victim5.htm (1999); Identity Theft
Information Privacy

- Even if we discover an injury from data sharing, tracing its cause to a particular information source or leak will likely be difficult, if not impossible.\(^{345}\) Obtaining effective redress will therefore be rare.\(^{346}\)

- Injury, while economic in some cases, can be very hard to undo. This is particularly true with identity theft, the loss of an employment opportunity, or harm to reputation caused by embarrassing information being revealed.

- Information misuse imposes significant external costs beyond the direct injury to the individuals involved. Financial institutions, for example, incur costs investigating claims of

---

\(^{345}\) See United States General Accounting Office, Report to Congressional Requesters, Identity Theft: Prevalence and Cost Appear to be Growing, GAO-02-363, (Mar. 1, 2002) [hereinafter GAO REPORT] (reporting that identity theft can cause potentially severe emotional or other nonmonetary harm in addition to economic harm; the leading types of nonmonetary harm cited by consumers were “denied credit or other financial services (mentioned in over 7,000 complaints), “time lost to resolve problems” (mentioned in about 3,500 complaints), and “subjected to criminal investigation, arrest, or conviction” (mentioned in almost 1,300 complaints)); See generally Scott Shorr, Personal Information Contracts: How to Protect Privacy Without Violating the First Amendment, 80 CORNELL L. REV. 1756 (1995) (assessing common-law and federal legislative remedies for commercial disclosures of information that violate personal privacy); Daniel J. Solove, Conceptualizing Privacy, 90 CAL. L. REV. 1087 (2002) (arguing that privacy law has fixed itself too firmly to certain conceptions of privacy, and as a result, has lost flexibility in dealing with emerging privacy problems). Cf. Peter Swire, Markets, Self-Regulation, and Government Enforcement in the Protection of Personal Information, available at http://www.ntia.doc.gov/reports/privacy/selfreg1.htm#1A (last visited Jan. 27, 2003) (explaining that the key market failures with respect to privacy are due to information and bargaining costs: “[A] company that acquires personal information gains the full benefit of using the information but does not suffer the full losses caused by disclosure. Because of imperfect monitoring, customers often will not learn of the disclosure and will not be able to discipline the company in the marketplace for its less-than-optimal privacy practices. Because the company internalizes the gains from using the information, but can externalize a significant share of the losses, it will have a systematic incentive to over-use private information.”).
credit card fraud and re-crediting customer accounts. Some of these costs are passed though to customers in the form of higher interest rates and incidental fees, so we all pick up a share of the expenses when privacy breaches cause harm.

- Most of us would say that we benefit in intangible ways just knowing that our data is reasonably secure, and is not being bought and sold without our permission. We would feel less vulnerable if we knew that our data was either not being collected or, at least, was protected from misuse. Such sentiments are worth respecting in their own right, but they can translate into economic benefits as well. If we felt more secure in our relationships with data collectors, we might use their services more. One of the impediments to the development of Internet commerce is the fear many people have that the information they transmit could be shared, misused, or stolen.

Even though information privacy has many of the defining characteristics of other societal values, this does not mean that a heavy-handed regulatory approach should be used to protect our privacy interests. One important difference between information privacy and

347. See Wood & Schecter, supra note 345, at 4 (“law enforcement consider banks and financial institutions to be the ‘victims’ in identity theft cases [because] they are frequently forced to absorb the costs of the thefts”).

348. See LoPucki, supra note 343, at 91 (stating that in dealing with the problem of identity theft, defrauded creditors are likely to employ legal and practical means that are cost effective, and pass the remaining costs on to their consumers); Kurt M. Saunders & Bruce Zucker, Counteracting Identity Fraud in the Information Age: The Identity Theft and Assumption Deterrence Act, 8 CORNELL J.L. & PUB. POL’Y 661, 663 (1999) (estimating that identity theft imposes a cost on consumers approaching $100 million annually); GAO REPORT, supra note 346 (stating that the American Bankers Association reported total check fraud-related losses reached $2.2 billion in 1999; Mastercard’s and Visa’s aggregated identity theft-related losses from U.S. operations rose to $114.3 million in 2000; total cost of one national consumer reporting agency’s Fraud Victim Assistance Department was $4.3 million for 2000).

Apart from cost-benefit analysis, economic arguments misconstrue the harm to society from the loss of confidence in information practices. Many view privacy as central to the democratic fabric of society. The misuse of personal information harms an individual and deserves protection regardless of how the misuse might benefit others. See Reidenberg, supra note 134, at 725.

349. FEDERAL TRADE COMMISSION, supra note 285, at 2 (reporting that 92% of consumers are “concerned,” including 67% who are “very concerned,” about the misuse of personal information online, and explaining that this apprehension likely translates into lost online sales due to lack of confidence in how online personal data will be handled; also cites a study that estimates potential losses in online retail sales due to privacy concerns will reach $18 billion by 2002).
societal problems like environmental pollution or unsafe food and drugs
is that privacy is seldom a matter of life and death. The effects of even a
widespread disclosure of personal information will not be as catastrophic
as global warming or a mass outbreak of contagious disease. Still, the
case for viewing information privacy as a societal value should not be
discounted. If the stakes do not seem quite as high, the appropriate policy
resolutions may differ but the regulatory function of government need
not be minimized and the policing of privacy interests need not be left to
market forces and individual enforcement initiatives. A less intrusive, but
nonetheless vital, governmental role may be in order.

Throughout our history, we have created administrative bodies to
implement national legal norms and regulate important societal values.
Not all have concerned basic health and safety issues. The Securities
Exchange Commission oversees the operation of our capital markets, the
Equal Employment Opportunity Commission ensures fair employment
practices, and the Comptroller of Currency watches over our national
banks. In all of these areas and many others, oversight became necessary
when policy makers realized that market forces could no longer
effectively protect important societal interests. As we witness the vast
expansion of digital databases and nearly complete loss of control over
the collection and dissemination of our personal information, we see that
the same conditions presently exist with our interest in information
privacy.

B. Looking to Europe for a Model

There is reason to believe that our approach to the database problem
would take a different form, and policy resolutions would be
recognizably different, if the issue were viewed in a different way. In
Europe, for example, privacy is treated as a political imperative,
anchored in fundamental human rights, and considered a matter of basic
“social protection.” National and regional governments are viewed as
key players in ensuring data protection, and the problem is considered an

---

important element of public law. Market forces and individual enforcement are important in the regulatory scheme, but they are not the primary policing mechanisms. Law and government are considered fundamental in ensuring shared norms of social and citizen protection.351

The most important act of legislation dealing with information privacy was the 1996 European Community Directive on Data Protection (EU Directive), which outlines the basic principles for European Union member countries.352 The EU Directive took effect in 1998, and although the Directive has its share of critics,353 it recognizes some key dimensions of the problem that are missing in United States privacy law.354 The Directive mandates that all fifteen EU Member States ensure that citizens have the right to access their data, the right to fix erroneous data, the right to a recourse for violations, and the right to keep the information from being used for any marketing purpose without their permission.355

More importantly, however, unlike most American privacy laws, the EU Directive applies essentially the same standards to private sector and

351. See Reidenberg, supra note 350, at 1347.
353. See, e.g., Cate, supra note 122 (arguing that the EU approach is not compatible with longstanding constitutional and political traditions in the U.S.); PETER SWIRE & ROBERT E. LITAN, NONE OF YOUR BUSINESS: WORLD DATA FLOWS, ELECTRONIC COMMERCE, AND THE EUROPEAN PRIVACY DIRECTIVE 14, 50–53, 191 (1998) (EU Directive may have limited usefulness outside the world of mainframe computers).
354. The progression from the 1975 Council Resolution to the Treaty of Amsterdam evidences an increasing centralization of consumer policymaking at the Community level, if only in theory. Overby, supra note 174, at 1241 (discussing the progression).
355. Article 2(a) of the EU Directive broadly defines protected information as "any information relating to an identified or identifiable natural person." Article 2(b) defines data "processing" as "any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as collection, recording, organization, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available . . . ." EU Directive, supra note 352, art. 2(a), (b).
government data bases. Cross-sectoral legislation in each Member State guarantees a set of fundamental privacy rights that ensure the fair treatment of personal information. Data protection laws define each citizen’s basic legal right to control personal information. Instead of beginning with a presumption of legitimacy for government and commercial enterprises that wish to collect and share information, the European approach seeks to strike a balance that provides for a high level of data protection for all EU citizens.

Information policies in Europe tend to have broad applicability and cut across economic sectors. There is an underlying presumption that the collection and sharing of personal information, particularly in the private sector, is not a standard practice that citizens must simply learn to accept. Information can be collected only for specified purposes, used in ways that are compatible with those purposes, and stored no longer than is necessary. Individuals must be told that information is being collected, the purposes of the data collection, and the person responsible for collecting and controlling the information after it has been stored. Affirmative consent is required in more situations when data is to be collected or shared, with less responsibility on the individual to opt-out of data sharing. Independent, national supervisory authorities oversee, investigate, and enforce legal norms. National “ombuds” serve as

356. Indeed, earlier drafts of the Directive placed stronger restrictions on the private sector than on governments. The final version treats them essentially the same. Moreover, the scope of protection in the Directive does not depend on the technique used to store information. Manual filing systems are covered as well as computerized systems. BENNETT, supra note 49, at 105–07.


359. See SCHWARTZ & REIDENBERG, supra note 122, at 13–14.

360. See id. at 15.

361. The EU Directive mandates that the national law of all Member States protect information about each identifiable individual even if the data is publicly available. Laws must also require an individual’s consent before processing personal information, except for the purposes contemplated by the original data collection. Member States can further restrict the processing of data deemed “sensitive” (e.g., medical information), and certain “black list” data is not collectable at all without the affirmative consent of the individual. This includes data revealing racial/ethnic origin, political views, religious beliefs, and membership in a trade union. EU Directive, Art. 8(1). See Reidenberg, supra note 134, at 732.

362. See Cate, supra note 122, at 186.
advocates for individuals who feel that a breach has occurred. Persons who process individual information in both the public and private sectors must comply with notice and reporting mandates so their activities can be monitored. Civil liability and “dissuasive penalties” must be available for noncompliance with legal norms.

C. Moving Toward a Global Information Policy in the United States

Circumstances already exist that may move the United States toward the European approach to information privacy. The EU Directive is far from perfect, particularly with regard to enforcement of its mandates, but it is clearly becoming the international model for data protection. We are seeing a convergence of information policy worldwide, and the United States is under increasing international pressure to conform.

Article 25 of the Directive states that transfers of information about EU citizens to a country outside the EU may take place only if the receiving country ensures an “adequate level of protection.” EU Member States are instructed “to take the measures necessary to prevent [the] transfer of data of the same type to the third country in question.”

363. In the 1970s, four Nordic countries (Denmark, Finland, Norway, and Sweden) established the office of consumer ombudsman, a supervisory body for overseeing the marketing of consumer goods and services. See generally Kjersti Graver, A Study of the Consumer Ombudsman Institution in Norway with Some References to the Other Nordic Countries I: Background and Description, 6 J. CONS. POL’Y 1 (1986). Similar offices have been created in other nations. See Ewa Letowska, The East Block's First Government Ombudsman, INT'L HERALD TRIB., Jan. 8, 1988, at 5.

364. See, e.g., Italian Data Protection Act (1996), in MARC ROTENBERG, THE PRIVACY LAW SOURCEBOOK 2000, available at http://www.privacy.it/legge675encoord.html (2000). The Forward to the Italian law proclaims that data should be processed “by respecting the rights, fundamental freedoms and dignity of natural persons, in particular with regard to privacy and personal identity.” Privacy is considered a “fundamental component of the “electronic citizenship.”” Id. Member States must also require any person processing personal information to notify the national supervisory authority, which is required to keep a public register of data processors. Reidenberg, supra note 98, at 733. States must delegate responsibility to one or more public authorities for monitoring the compliance with the law. EU Directive, Art. 28. These authorities must act with “complete independence” and must be given investigative authority and the power to bring legal proceedings. See Bennett, supra note 268, at 108.


368. EU Directive, Art. 25(1).

This means that national authorities in each Member State can prevent U.S. businesses from processing data of EU citizens if U.S. privacy protections are not up to EU standards. The consequences of this provision are severe for credit-granting and financial institutions, hotel and airline reservations systems, direct-marketing firms, insurance companies, and any commercial enterprise that relies on the flow of personal information from European sources. Europe has made it clear that it will not tolerate “data havens” that would compromise the personal information of its citizens.

The EU Directive now constitutes the “rules of the road” for the increasingly global nature of information processing. The Directive represents a multi-national consensus on the content of data protection rights, and has proved to be a valuable model for countries looking to enact their own data protection laws. Pressures for “policy convergence” worldwide have prompted other nations to adopt similar principles that give their citizens more control over personal information. Significant movement toward an EU-style data protection has already occurred in Canada, South America, and Eastern Europe, and the movement is spreading to other regions. The international


372. Id. at 111–12.

373. Convergence in this context means more than similarity at a given point in time. It points to a pattern of similar regulatory regimes developing over time, rather than a static condition. See BENNETT, supra note 49, at 111. Bennett identifies five principle causes for this policy convergence: (1) technological determinism, (2) the influence of pioneers in the field, (3) the interaction of a small group of international experts, (4) harmonization projects of international organizations, and (5) the accelerating pace of global commerce that forces states to make policy changes that conform to international norms. Id. at 116–17. Supporting evidence of policy convergence also comes from David Flaherty, who examined the workings of national and state data protection agencies in Germany, Sweden, France, Canada and the United States. FLAHERTY, supra note 84. For statements of privacy principles in an international context generally, see BENNETT, supra 49, at 96–115 (discussing convergent themes of openness (disclosure of the type of information collected and from what categories of individuals), individual access and correction, limits on what information can be collected (e.g., relevant and necessary to accomplish the limited purposes of the collecting entity), limits on how data can be used, limits on disclosing information to external sources, and security safeguards. See also OECD Guidelines, supra note 142.

374. By the end of the 1980s, most European countries applied the same data protection standards to both the public and private sector. The U.S., Canada, Australia and Japan rejected this approach, regulating the public sector with one set of laws (e.g., the Privacy Act in the United States) and the private sector with sector-specific laws and voluntary codes of practice. See Bennett, supra note 268, at 100. Not much had changed by the end of 1996. Of the 24 OECD countries, only six had failed to
harmonization of data protection laws can be attributed partly to pressure from Europe and the effects of Article 25, but it is fueled by the conceptual appeal of a comprehensive set of standards that were carefully crafted by the EU after years of study and debate. Those rules are becoming the standard for multi-national transactions in the increasingly global environment of offline and online data sharing. In this regard, United States information policy lags behind. However, we may not be behind for long. Many U.S. businesses are already affected by the European standards. Companies that handle information about European citizens must now certify compliance with European data protection principles. A Safe Harbor agreement between the Department of Commerce and the European Commission enact a comprehensive privacy law to all data processing entities: U.S., Canada, Australia, Japan, Greece and Turkey. As a Member State in the EU, Greece must now conform to the standards of the EU Directive. Id. at 113. In the U.S., the private sector continues to be regulated through an expanding but still incomplete patchwork of federal and state laws, with no general oversight agency for privacy compliance in the U.S., and few effective remedies. Id.


376 Mechanisms outside the law, such as “contractual” agreements between American businesses and data protection authorities in specific countries, can minimize privacy conflicts for e-commerce transactions, but an international treaty may ultimately be necessary to ensure the growth of trans-border information exchange. The U.S. could, for example, promote a “General Agreement on Information Privacy” within the WTO framework. See Reidenberg, supra note 134, at 747.

Of course, inferences about the actual level of data protection in any country cannot be drawn merely by reading the statutes. In this regard, data protection is different from other societal problems like environmental protection, where states might agree on a desirable level of a particular contaminant in the atmosphere and have a clear understanding of how to monitor and assess performance. See Bennett, supra note 268, at 119. Assessing the level of data protection in practice is a difficult problem of measurement and is not addressed in this article.

377 The EU Directive bars the dissemination of information about EU citizens to entities outside of Europe where looser protections are in place. Sweden, for instance, insisted that American Airlines delete all health and medical information (including dietary requests) they had gathered on Swedish passengers unless the airline obtained the consent of each passenger to allow them to keep the info in the database. See American Airlines v. Sabre Kammarratan i Stockholm (Admin. Ct of Appeal, Stockholm), Apr. 1997; Paul R. Prabhaker, Who Owns the Online Consumer?, 17 J. CONSUMER MARKETING 158, 161 (2000). European nations can thus use the Directive to exert significant pressure on U.S. companies.
assists businesses who want to comply, and nearly 200 American corporations have signed up. If companies find that the EU norms are not unduly burdensome, then resistance to a similar regime in the United States may weaken. Moreover, as American policy makers see that U.S. companies are giving EU citizens greater data protection than U.S. citizens, then pressure to change our laws may increase.

The Safe Harbor agreement was a response to the real possibility that Europe would prevent data flows to the U.S. and to pressure from online industries that did not wish to take that risk. The U.S. Department of Commerce negotiated with the European Commission for an agreement that would assure Europe that U.S. businesses could comply with Article 25 even if the U.S. did not change its privacy laws. In the agreement, the European Commission endorsed what amounts to a voluntary code of conduct that, the parties agreed, would meet the Article 25 standard. The Department of Commerce then established the Safe Harbor mechanism allowing American businesses publicly to commit to this code for the treatment of European data. If businesses make and adhere to the commitment, they can be assured of continuing data transmissions from Europe.

There are several reasons why a U.S. based business with European operations might want to certify compliance with the Safe Harbor agreement. It reduces the likelihood that European privacy authorities would target a company on the compliance list, thereby avoiding the interruption of data flows and any associated negative publicity. In addition, claims brought by EU citizens would be brought in U.S. courts. Drawbacks include the difficulty of complying with the Safe Harbor, especially for a large corporation that uses personal data in various ways that may not be allowed. Risk of prosecution under the Safe Harbor,


379. The Safe Harbor procedure and a list of companies that have agreed to adhere to the Safe Harbor Principles can be found at www.export.gov/safeharbor/ (last visited January 30, 2003).

380. Reidenberg, supra note 134, at 738.


382. Id.
either by consumers or the FTC, may be higher than the risk of prosecution by EU authorities.  

There are legal and practical problems with the Safe Harbor agreement, however, that may limit its effectiveness as a standard for data collection and widespread use in the United States. First, some European Member States have expressed concerns about the adequacy of the agreement. To the extent national privacy authorities find its data protection provisions inadequate, they can influence the way American businesses deal with information of European origin through threatened prosecution of American businesses. Second, the Safe Harbor agreement relies largely on the authority of the FTC to ensure compliance, but the jurisdiction of the FTC and it ability, as practical matter, to fill this role are questionable. Amendments to section 5 of the FTC Act in 1975 extended the jurisdiction of the FTC to unfair or deceptive acts and practices “in or affecting commerce,” but there is no evidence that Congress contemplated protecting foreign consumers or American businesses from foreign prosecution. Its purpose was to extend the jurisdiction of the FTC to protect American consumers from a broader range of unfair or deceptive practices by businesses. The claim that Safe Harbor comes within section 5 of the FTC Act is a departure from the purposes of the statute and could be subject to legal challenge.

---


384. Reidenberg, supra note 134, at 744. For many Member States, Safe Harbor weakens their data protection standards. It exempts public record information and any information processing called for by “conflicting obligations” or “explicit authorizations” in U.S. law. These vague authorizations could turn into large loopholes for U.S. businesses who claim they cannot comply with European standards because of some other agreement or U.S. law that imposes different demands. The Safe Harbor agreement also weakens European standards for redress. Under the EU Directive, victims must be afforded legal recourse and a remedy in damages. The Department of Commerce assured the European Commission that Safe Harbor and the U.S. legal system provided for remedies for individual European victims of Safe Harbor violations. In support of its claim that U.S. law provided adequate remedies for information privacy violations, the Department of Commerce made some misleading statements about the remedies available to aggrieved individuals. In fact, few effective remedies for privacy violations exist. See id. at 744–45.

385. The directive states that national “supervisory authorities” have investigative powers and the right to institute legal proceedings against violators of the privacy laws mandated in the directive. EU Directive, art. 28(3).


387. See Reidenberg, supra note 134, at 741.
In addition, the Safe Harbor agreement vastly overstates the extent of privacy protection offered by U.S. law, and in any event, it may not be broad enough in scope to have a significant effect on U.S. information policy. By its terms, it applies only to the activities of organizations that fall within the regulatory jurisdiction of the FTC and the Department of Transportation. As a result, many economic sectors will not be able to insulate themselves from EU challenges by committing to the voluntary code. Among these sectors are the financial services and telecommunications industries, which are excluded from FTC jurisdiction.

IV. CONCLUSION

If information privacy were viewed more as a societal problem than one of individual concern, privacy policy in the United States would not necessarily change. The stakes might simply be too small. Data privacy, while important to most people, is not generally regarded as an absolutely critical societal value. Even the most ardent privacy advocates would not put it in the same class as basic health and safety concerns. If the risks are viewed as real but not particularly important, then the need for a fundamental shift in policy or a strong regulatory response may be lacking even if we view the problem as one of general societal concern.

Moreover, even in the regulation of societal problems like environmental pollution, we often do not use a singular, comprehensive approach. We have sectoral laws, such as water, air, noise, and laws governing specific industries such as coal burning utilities, and policies that reflect compromises reached in part by looking at the costs and benefits of various alternatives. We do not expect absolutely clean air and water. We should not expect to keep information about us absolutely private. Tradeoffs are inevitable. Most of us do not want or expect to keep our information completely private. We benefit from the collection and sharing of information in many ways.

The question remains whether we will see a fundamental shift in the way information privacy is controlled in the United States, or whether our interests will continue to be bought, sold, and given away as freely as they have been in recent years. I have argued in this article that privacy

389. Reidenberg, supra note 134, at 743.
should be viewed as a societal value, and if it were understood in this way, a more comprehensive regulatory approach like the one in place in Europe and a growing number of other regions could emerge in the United States. No legal impediments to a European-style regulatory regime exist. Commercial speech and speech on matters of purely private concern enjoy less First Amendment protection than speech related to political discourse. Congress has enacted privacy legislation many times before with few constitutional conflicts, and it could rationally conclude that personal information about its citizens warrants more protective legislation than currently exists.

The obstacles to a more comprehensive approach to information privacy are rooted not in our laws but in our view of the appropriate function of government and the role of the private sector in ordering societal relationships. Legalities aside, our long history of general distrust in government solutions, coupled with our preference for open information flows and reliance upon market forces, make a comprehensive regulatory approach less likely as a practical and political reality.

Given the political history of the privacy debate in this country, no significant shift in U.S. policy seems likely to occur until some crisis or highly publicized event forces us to look at the issue from a new perspective. Indeed, in the current political climate, efforts to press a fundamental shift in policy appear to be losing momentum. With the Chairman of the FTC coming out strongly against new privacy legislation, the prospect for instrumental change seems even more

390. See Volokh, supra note 122, at 1055 (opposing information privacy rules on free speech grounds, but conceding that First Amendment limitations on nongovernmental gathering of information are unclear); Krotoszinski, supra note 56, at 242 (“[T]he states or Congress could enact privacy-protection laws that limit the legal means of obtaining information about non-public figures involving matters that are not of public concern.”). But see Cate, supra note 122 (the U.S. Constitution does impose restrictions on privacy legislation addressed to the private sector). Some courts have struck down state privacy laws under the Commerce Clause. See authorities cited supra note 208.


392. See Cate, supra note 122, at 219–25 (observing “four features of American society” that work against an EU style of privacy protection); BOARD OF GOVERNORS OF THE FEDERAL RESERVE SYSTEM, REPORT TO THE CONGRESS CONCERNING THE AVAILABILITY OF CONSUMER IDENTIFYING INFORMATION AND FINANCIAL FRAUD (March 1997) at 2 (“it is the freedom to speak, supported by the availability of information and the free-flow of data, that is the cornerstone of a democratic society . . . ”).
remote. Without a sense of urgency, special interest politics and a general anti-regulatory sentiment will likely dominate political discourse in the United States on this issue for the foreseeable future.

A change in perception can occur over time, however. If people start thinking about privacy as a general societal concern, the rhetoric of public debate can shift and the range of politically acceptable policy resolutions can expand. If no change occurs, we can expect to see more laws enacted periodically that purport to address privacy concerns in particular sectors, but individuals will still be expected to shoulder the burden of monitoring their own information, and market-based solutions will predominate. So long as information privacy is viewed largely as a matter of individual concern, individuals will be asked to carry the lion’s share of the burden. In time, we may get better at the task, especially as younger generations become more comfortable with the technologies that control the flow of our data. For now, we have little choice but to hope, wait and trust that the data collectors who are holding our personal information are guarding it securely and using it only for purposes we would prefer.

393 See supra notes 301–04 and accompanying text; Remarks of FTC Chairman Timothy J. Muris, supra note 304 (calling for increased enforcement of current laws rather than new legislation).

394 See Schriver, supra note 388, at n.189 (observing that most privacy legislation has been enacted in response to public scandals, thus explaining its patchwork quality).

395 In the past two years, dozens of privacy bills have been introduced in Congress. See supra note 300 and the website for the Electronic Privacy Information Center, available at http://www.epic.org/privacy/bill_track.htm (last visited, Jan. 20, 2003).
INTERGOVERNMENTAL COOPERATION, METROPOLITAN EQUITY, AND THE NEW REGIONALISM

Laurie Reynolds*.

Abstract: The economic gap between affluent suburbia and the urban core has recently received widespread attention among state and local government law scholars. Although the underlying normative arguments rest on very different rationales, scholars with a wide range of doctrinal approaches appear to have formed a consensus that the current concentration of wealth and resources in metropolitan areas is unacceptable. Their common goal of reducing regional disparities has made the scholarly dialogue a dispute over how, rather than whether, to achieve a better distribution. For many of what can be described as the “New Regionalist” scholars, voluntary intergovernmental cooperative efforts may appear to offer the potential to accomplish many of their stated goals. This Article examines the common types of intergovernmental cooperative efforts and concludes that they fail to correct, and often exacerbate, the socioeconomic gap. Thus, the regionalist agenda must be reworked to take account of the negative impacts that many of the highly touted regional governance efforts actually produce in metropolitan areas.

I. INTRODUCTION

Metropolitan America remains stubbornly resistant to attempts to limit local government proliferation and the political fragmentation and territorial overlapping it produces.¹ Frequent and repeated calls for local government consolidation,² regional legislatures,³ strengthened municipal

* Professor of Law, University of Illinois. Many thanks to Richard Briffault, John Lopatka, and Steve Ross for their very helpful comments on earlier drafts of this Article. I would also like to thank Jeanah Park and Jennifer Chavez for their outstanding, indeed invaluable, research assistance.

¹ In his evaluation of the repeated failures of consolidation movements, Anthony Downs ascribes the lack of political will for metropolitan government to the following factors: local government officials’ opposition to power sharing; residents’ fear of more remote, less responsive government; and suburban fears of wealth redistribution. His negative assessment of the potential for region-wide consolidation is categorical: “In short, almost no one favors metropolitan area government except a few political scientists and intellectuals. Proposals to replace suburban governments completely are therefore doomed.” ANTHONY DOWNS, NEW VISIONS FOR METROPOLITAN AMERICA 170 (1994).

annexation powers, and anti-sprawl growth policies have not altered the patterns of local government formation and growth. New municipalities continue to form at ever increasing distances from central cities. And demographic statistics confirm that as suburbanization trends proceed unabated, central cities in major metropolitan areas suffer the loss of high-taxpaying residents who prefer to live in smaller, less urbanized, more homogeneous communities. This increasing stratification between city and suburb may be the intended result of state laws pertaining to local government formation, which allow affluent, homogeneous enclaves to form their own government and thus prevent the redistribution of resources that occurs when wealthy and poor pay property taxes to the same general purpose municipality. In the alternative, it may merely reflect the nostalgic American dream of a single family home in a safe and small community. Whatever the underlying cause, the population trends document the overall out-migration of middle- and upper-class residents from central city to suburb. As a result, in terms of job creation, quality of schools, affordable housing, and income, to name a few.

3. See Briffault, supra note 2, at 1164–71.
6. The most recent census statistics suggest that, in terms of absolute numbers at least, the net out-migration from many major cities may have been reversed. In fact, between 1990 and 2000, only nine of the 50 largest U.S. cities experienced a population decline. U.S. CENSUS BUREAU, CENSUS 2000 REDISTRICTING DATA (2001). For the cities that have gained population during the most recent census period, the news is still mixed. For the most part, the newcomers tend to have lower incomes and education levels than the city residents who have left for the suburbs. See, e.g., Kenneth T. Jackson, Editorial, Once Again, the City Beckons, N.Y. TIMES, Mar. 30, 2001, at A23. See also Renee Elder, Population Swap Costs Metro; Nashville’s Average Income Keeps Falling as Higher-Income Families Migrate to Suburbs, THE TENNESSEAN, Aug. 19, 1997, at 1A. Between 1994 and 1995, for example, 7,600 people moved from central Nashville to outlying suburban areas, while 7,000 people moved in. The new residents had incomes that were 5% lower than the incomes of those moving out. The resulting decrease in the city’s average income will likely increase the demand for social services while decreasing the government’s tax revenues.
7. See Cashin, supra note 5, at 2011.
Intergovernmental Cooperation

few indicia of municipal well-being, the gap continues to widen between affluent suburbia (now often referred to as the “favored quarter”\(^{10}\)) and the rest of the metropolitan area.

Responding to this well-documented disparity in terms of wealth and service provision, and loosely connected under the broad doctrinal umbrella of the New Regionalism,\(^{11}\) a growing number of commentators

---


10. Myron Orfield, a Minnesota state legislator, was the first to document how the city-suburb schism is more accurately described as a gap between, affluent suburbia (the favored quarter) and the central city and adjacent ring of older suburbs. See generally MYRON ORFIELD, *METROPOLITICS: A REGIONAL AGENDA FOR COMMUNITY AND STABILITY* (1997). According to Orfield, all major metropolitan regions display a remarkably similar distribution of population and wealth: 20–40% live in central cities; 25–30% in older declining suburbs; 10–15% in low tax base suburbs; and the remainder, the favored quarter, in high tax base, wealthy suburbs. See Myron Orfield, *Conflict or Consensus? Forty Years of Minnesota Metropolitan Politics*, 16 BROOKINGS REV. NO. 4, at 31, 34 (1998). Orfield seeks to promote natural political alliances between central city and older suburbs, many of which face similar problems with aging infrastructure, high social service needs, increasing poverty and declining tax base. His recent book, *METROPOLITICS*, offers a strategy for those who seek stronger regional governance structures. His recipe for successful regionalization efforts can be summed up in one piece of advice for the central city: “It’s the Older Suburbs, Stupid.” Id. at 168. That phrase captures much of the current political reality of many major metropolitan areas. As the first ring of suburbs age, he argues, they have begun to show the same signs of decline and decay as the central city. Influx of the poor, aging infrastructure, and exodus of the mobile middle class to ever more distant suburbs, is a story told by central city and older suburbs alike. Id. at 47–51. To capitalize on what he sees as a natural alliance, Orfield recommends joint legislative efforts to seek the imposition of regional fair housing obligations, property tax sharing, and a redirection of government infrastructure spending from urban fringe to central city and inner suburban ring. Id. at 78–103. Though in practice, alliances between central cities and older suburbs have proved difficult to create and sustain, Orfield’s critique offers a perspective on regionalism that suggests the political feasibility of some regional efforts. See also MYRON ORFIELD, *CHICAGO METROPOLITICS: A REGIONAL AGENDA FOR MEMBERS OF THE U.S. CONGRESS* 27 (1998). For a recent article in the legal literature that explores Orfield’s theories in depth, see Cashin, *supra* note 4, at 1987 (describing favored quarter as “high-growth, developing suburbs that typically represent about a quarter of the entire regional population but that also tend to capture the largest share of the region’s public infrastructure investments and job growth”).

11. See generally, Kathryn A. Foster, *Regional Capital*, in *URBAN-SUBURBAN INTERDEPENDENCIES*, *supra* note 5, at 85 (noting that as to regional equity, “[a]n influential
now advocates structural reforms in local governmental law. While some continue to endorse full city-county consolidation or the creation of a new regional governmental unit, most recommend more flexible governance solutions. In the legal literature, the New Regionalism has yet to emerge as a clearly defined doctrinal movement, yet the term has begun to appear as a shorthand for the scholarship that examines and criticizes the allocation of regulatory power among and between state and local governments in metropolitan areas. Calls for full scale consolidations have largely disappeared, and the focus has shifted to more limited solutions, including the establishment of a regional legislature in metropolitan areas, a reformulation of the local franchise to allow cross-border voting, and the encouragement of voluntary intermunicipal "burden sharing."

(footnotes omitted)

empirical focus of the new regionalism tradition now links social equity to economic growth") (emphasis in original); Rosabeth Moss Kanter, Business Coalitions as a Force for Regionalism, in Reflections on Regionalism, supra note 9, at 157–59; Cashin, supra note 5, at 1988 n.11; Symposium, New Regionalism and Its Policy Agenda, 32 ST. & LOC. GOV’T REV. 158 (2000).

12. See Rusk, supra note 2, at 91–97; Savitch & Vogel, supra note 9, at 162 (describing the consolidationist approach).

13. See Margaret Weir, Coalition Building for Regionalism, in Reflections on Regionalism, supra note 9, at 127–53; Briffault, supra note 2, at 1165.

14. See Savitch & Vogel, supra note 9, at 161 (defining governance as "the notion that existing institutions can be harnessed in new ways, that cooperation can be carried out on a fluid and voluntary basis among localities, ... that people can best regulate themselves through horizontally linked organizations ... [and that] localities can provide public services without ... producing them"). In contrast to governance, the authors define government as "formal institutions and elections and established decision-making processes and administrative structures." Id. In fact, many of the New Regionalists’ "governance" proposals involve the creation of a regional special district, which is a formal institution that exercises many governmental powers and typically has its own bureaucratic administrative structure. See infra Part III.A.3.


16. See Briffault, supra note 2, at 1164–71.


18. See Clayton P. Gillette, Regionalization and Interlocal Bargains, 76 N.Y.U. L. REV. 190, 194 (2001). Professor Gillette coined the term for his consideration of interlocal efforts to redistribute wealth. He describes burden sharing as: “subsidies from some localities (typically suburbs) to others (typically central cities) that bear disproportionate redistributional burdens ... Such subsidies may entail accepting a fair share of undesirable land uses or a fair share of residents who need redistributional services ... or the dedication of tax revenues generated in suburbs to the central city.” Id. Although this Article may have some relevance for the first two types of subsidies
The current landscape of multiple, fragmented units of local government is the exclusive province of state law. At least since the United States Supreme Court’s 1907 decision in *Hunter v. City of Pittsburgh*, state governments are the undisputed masters of their political subdivisions. States determine the rules under which local governments are created, the powers they exercise, and the relationships they have to the other local governments in the state. In that capacity, the states have displayed a remarkably uniform unwillingness to force regional consolidation or regional redistribution of wealth on their political subdivisions. State statutory frameworks that determine the rules of local government formation and operation are decidedly anti-regional. In fact, a variety of legal rules shore up the insular and insulated status of American municipalities. The ease with which new governments are formed, the ways in which municipal incorporation allows a community to capture the wealth derived from its property tax levy, the general inability of existing municipal governments to annex development on their borders, and broad municipal powers to regulate land use development without consideration of its impact on the overall regional welfare, are a few of the more salient examples.

In apparent contrast to the localist predilection evidenced in those common state laws, many states have adopted statutes that announce strong state support for cooperative ventures among and between local government units. In the modern state and local government law

Professor Gillette describes, it focuses primarily on the redistribution of resources that may occur through intergovernmental cooperative efforts.

20. See, e.g., ARIZ. REV. STAT. § 11-952(A) (2001) (“[i]f authorized by their legislative or other governing bodies, two or more public agencies by direct contract or agreement may contract for services or jointly exercise any powers common to the contracting parties and may enter into agreements with one another for joint or cooperative action or may form a separate legal entity, including a nonprofit corporation, to contract for or perform some or all of the services specified in the contract or agreement or exercise those powers jointly held by the contracting parties”); FLA. STAT. ch. 163.01(2) (2001) (“[i]t is the purpose of this section to permit local governmental units to make the most efficient use of their powers by enabling them to cooperate with other localities on a basis of mutual advantage and thereby to provide services and facilities in a manner and pursuant to forms of governmental organization that will accord best with geographic, economic, population, and other factors influencing the needs and development of local communities”); IOWA CODE § 28E.4 (2002) (“[a]ny public agency of this state may enter into an agreement with one or more public or private agencies for joint or co-operative action pursuant to the provisions of this chapter, including the creation of a separate entity to carry out the purpose of the agreement. Appropriate action by ordinance, resolution or otherwise pursuant to law of the governing bodies involved shall be necessary before any such agreement may enter into force”); OR. REV. STAT. § 190.010 (1999) (“[a] unit of local government may enter into a written agreement with any other unit or units of local government for the performance of any or all functions and activities that a party to the
literature, the regionalist critique of intergovernmental cooperation is either generally positive or neutral. Some, for instance, find value in its voluntary nature, as well as in the fact that it frequently does not involve the creation of a new government entity. A less enthusiastic assessment asserts that autonomous and affluent local suburban governments are unlikely to resort to intergovernmental cooperation to solve intra-regional disparities. In general, though the regionalist commentary may not see intergovernmental cooperation as likely to enhance overall metropolitan equity, it typically does not reject intergovernmental efforts as anti-regional. This Article, in contrast, suggests that that assessment fails to probe the ways in which intergovernmental cooperation negatively affects the metropolitan landscape. Building on the suggestion made recently by Professor Gerald Frug, this Article’s evaluation of intergovernmental cooperative efforts leads to the somewhat counter-intuitive claim that intergovernmental cooperation may actually have a non-trivial anti-regional impact.

Part II begins with a brief summary of the debate in the legal literature between localists and regionalists, summarizing their normative bases and the empirical evidence on which their proposals rest. It then describes how the emerging New Regionalism builds on the longstanding disputes between, on the one hand, localists, who favor significant local government autonomy and object to many regionalization efforts, and, on the other hand, regionalists, who endorse varying degrees of curtailment of local autonomy in favor of regionwide regulatory, fiscal, or general governmental mechanisms. Though it is easy to identify clear divergences among the localist camps, and between the localists and those whose viewpoints are better described as regionalist, a review of the substantive proposals currently offered by each major doctrinal strain in local government law reveals some unmistakable convergence. That is, just as the regionalist scholars have moved away from calls for full scale governmental consolidation in

agreement, its officers or agencies, have authority to perform. The agreement may provide for the performance of a function or activity” ). See also ADVISORY COMMISSION ON INTERGOVERNMENTAL RELATIONS, INTERGOVERNMENTAL SERVICE ARRANGEMENTS FOR DELIVERING LOCAL PUBLIC SERVICES: UPDATE 1983 41 (1985) [hereinafter ACIR] (describing how states encourage intergovernmental cooperation through the use of incentive grants, direct financial assistance for planning, and technical assistance to local governments).

21. See Cashin, supra note 5, at 2027 (describing New Regionalism); see also Savitch & Vogel, supra note 9 at 161-65.

22. See Briffault, supra note 2, at 1144-51.

Intergovernmental Cooperation

metropolitan areas, so too the localists have tempered their support of local autonomy with a recognition of the need for some regionwide government action and/or apparatus. Thus, the initial parts of this Article will compare the analytical frameworks and specific proposals of the predominant strands of scholarship in this area, emphasizing how regionalism is very much on the minds of local government law scholars, localists and regionalists alike.

Against that general doctrinal backdrop of the New Regionalism’s tenets, normative bases, and goals, in Part III the Article focuses specifically on the role of intergovernmental cooperative efforts in regional governance. Proceeding to a critique of the most common types, it evaluates their strengths, weaknesses and policy implications. Following on that assessment, the Article suggests that the New Regionalists may have been too quick to conclude that intergovernmental cooperation is consistent with their stated policy goals. It describes how, in the existing legal framework, multi-purpose local governments are able to obtain the purported efficiency benefits of being part of a metropolitan region. At the same time, though, because the cooperative efforts are voluntarily undertaken by the participating entities, local governments are able to selectively pick and choose the parts of metropolitan governance they wish to join. That is, when it comes to services needing large capital expenditures and a supra-municipal service territory to reach acceptable levels of efficiency, local governments are quick to turn to their neighbors to establish regionwide solutions. When however, the regional agenda focuses on the existing urban crises in affordable housing, education, job creation, or other social services, the local government boundaries allow the more affluent municipal governments to refuse to participate in regional redistributive efforts. Thus, local governments in metropolitan areas are not required to take the bad with the good, to bear the costs while they enjoy the benefits of their position in the metropolitan region. Rather, under the framework existing in nearly all states, the applicable legal rules enable them to preserve and solidify their privileged status. Though this system may be

24. Because it focuses on intergovernmental cooperative efforts, this Article does not consider single purpose governments that are created directly by state action. See generally Hoogasian v. Reg’l Transp. Auth., 317 N.E.2d 534 (Ill. 1974). For a full description of the various types of special districts existing across the country, see Kathryn A. Foster, The Political Economy of Special-Purpose Government, in AMERICAN GOVERNANCE AND PUBLIC POLICY 7–15 (Barry Rabe & John Tierny eds., 1997).
perfectly consistent with their perceived self interest\textsuperscript{25} and with established state and federal\textsuperscript{26} legal doctrine, it vastly limits the opportunities and incentives for redistribution of wealth at the regional level, and contributes to the preservation of inequalities within the region.

This Article recognizes that wholesale abandonment of intergovernmental cooperation among general purpose local governments is even less likely to occur than the consolidation proposals the New Regionalism seeks to replace. Nor does the Article endorse such an unlikely reformulation of local government powers. Instead, it more modestly urges heightened awareness of the regional costs of intergovernmental cooperation and the ways in which it may actually hinder the achievement of regional equity. It joins the ranks of those who have argued that overall metropolitan equity cannot be achieved through voluntary cooperative efforts and calls for states to take action to correct the gap that their laws and urban policies have facilitated and preserved.

II. THE MANY FACES OF REGIONALISM

Though the term means different things to different people, regionalism is enjoying a revival on both the academic and political fronts. As it subjects metropolitan regions to reexamination with renewed vigor and enthusiasm, the New Regionalism may be a “mixed bag of old prescriptions and new remedies to address problems both new and longstanding.”\textsuperscript{27} Refining the regionalists’ traditional focus on the central city and its decline, New Regionalists expand their scrutiny to consider

\begin{itemize}
\item \textsuperscript{25} See Richard Briffault, \textit{Localism and Regionalism}, 48 BUFF. L. REV. 1, 27 (2000) (noting that resistance to regionalism in the political process is “largely a matter of the self-interest of those who benefit from the status quo, such as local elected officials, land developers, corporations that are the subjects of interlocal bidding, and the businesses and residents located in the high-tax base localities of the metropolitan area”). Briffault observes that in practice, localism is less about the normative values of efficiency, democracy, or community, and more about “preserving existing political control over local resources, protecting residents of high wealth localities from the needs of their lower-wealth neighbors, and providing opportunities for businesses to take advantage of the interlocal competition for tax base.” \textit{Id.}
\item \textsuperscript{26} See Richard Briffault, \textit{Our Localism: Part I—The Structure of Local Government Law}, 90 COLUM. L. REV. 1, 86–111 (1990) for a detailed analysis of how U.S. Supreme Court doctrine protects local autonomy and applies localist values to uphold local government actions with decidedly anti-regional impacts, specifically in the areas of land use regulation and education finance.
\end{itemize}
Intergovernmental Cooperation

how the phenomenon of suburban growth forms a crucial part of urban and metropolitan development. The resulting endeavor, which searches for solutions to metropolitan problems in the absence of metropolitan governments, has become, in the words of noted commentators, “so proactive, so compelling, so urgent.”

A. In Defense of Local Government Fragmentation

Local government law has been the focus of a long and rich debate between localists and regionalists, a debate that seeks to determine the proper allocation of governmental power by the state to its political subdivisions. In the modern era, the post-World War II baby boom and

28. Savitch & Vogel, supra note 9, at 161.

29. The debate over whether decentralized, independent local government units are preferable to centralized, higher level government units was left unresolved at the founding of the nation. As a result, it is no surprise that the debate has continued unabated to the present. Though the Constitution makes no mention of local governments, their creation, or their status, foundational writings articulate the differing viewpoints on the debate between local autonomy and more centralized levels of government power. James Madison, writing in the Federalist Papers, defended centralization of governmental power to protect against the tyranny of the majority. In his view, it is more difficult for a faction to control a centralized unit of government than a smaller, more decentralized one. See THE FEDERALIST NO. 10 at 46-47 (James Madison) (Gary Willis ed., 1982). Moreover, he asserted, because a higher level of government has a larger territory and higher population within its territorial jurisdiction, it is more likely that government offices will be filled with qualified individuals. See id. at 47. Though Madison’s arguments were directed at the debate between allocation of power between the states and the national government, his insights on the dispute over the vertical allocation of power have broader relevance for the debate over the allocation of power at the local level, particularly in large metropolitan areas. In contrast to Madison’s endorsement of centralization, Alexis de Tocqueville’s commentaries on early 19th century America made the classic case for localism. ALEXIS DE TOCQUEVILLE, DEMOCRACY IN AMERICA (G. Lawrence trans., J.P. Mayer ed., 1969) (1848). Tocqueville’s preference for decentralization of power to the levels of government closest to the people were based on his belief that local government provides greater, rather than lesser, protection against the tyranny of powerful interests. See id. at 89. Fundamentally, the strength of local government in Tocqueville’s analysis lies in the ease with which the citizenry can participate in government when power is centered at the local level. See id. at 68–70, 189–95. In Tocqueville’s view, active local governments, with their ability to entice citizen participation because of their direct and immediate contact with the citizenry, provided the crucial protection against tyranny. See id. at 192. Though views of both Tocqueville and Madison were formulated at a time when population centers were separated by large distances and the impacts of their decisions could largely be limited to the territory and the people under their jurisdiction, the insights and the arguments articulated in this early localism debate have continuing endurance for current urban policy. For a review of the philosophical and historical underpinnings of the decentralization debate in local government law, see generally WILLIAM D. VALENTE, DAVID J. MCCARTHY, JR., RICHARD BRIFFAULT AND LAURIE REYNOLDS, STATE AND LOCAL GOVERNMENT LAW 1–24 (5th ed. 2001); Richard Briffault, Our Localism: Part II—Localism and Legal Theory, 90 Colum. L. Rev. 346, 403–04, 416 (1990); Gerald E. Frug, The City as a Legal Concept, 93 Harv. L. Rev. 1059, 1071–73 (1980); Cashin, supra note 4, at 1986–87.
the accompanying explosive growth of suburbia produced a new demographic reality for the localism debate. Central cities found themselves surrounded by increasing numbers of municipal governments, autonomous in the sense that they were able to control the territory and tax the resources within their borders, but integrated in the sense that they were a functional part of the metropolitan region. The well-documented loss of central city wealth and families to the suburbs and the inability of cities to share in the suburban prosperity that was occurring all around them, set the stage for the most recent iteration of the battle between localists and regionalists.

Localism is most prominently defended by two distinct doctrinal camps that come together in their conclusion that strong local governments are highly preferable to centralization in metropolitan regions. However, the two localist doctrines are based on divergent political ideologies. First, the so-called “public choice” approach, as articulated in the work of Professor Clayton Gillette, uses legal economic analysis to defend localism as the most efficient way of providing services. It heralds the competition in the marketplace of local governments as providing effective checks against government inefficiencies and abuses of power. In marked contrast, Professor Gerald Frug’s “participation theory” of localism rejects the notion that a city’s most fundamental role is that of service provider33 and urges enhanced local powers as a mechanism for empowering communities and their

30. For a description of the ways in which the housing market responded to a shortage of approximately 6,000,000 housing units in 1947, see JACKSON, supra note 5, at 231–45. Professor Jackson identified five characteristics of the postwar suburbs: increasingly distant from central cities, low density development, architectural sameness in construction, low prices associated with mass production, and racial and economic homogeneity. See id. The growth of suburbia continues to outpace central city growth. See id. at 283.

31. I use the term “public choice” as a shorthand convenience to describe generally those whose work finds support in Charles Tiebout’s theory, which is described infra at notes 35–38 and text accompanying. Professor Gillette’s work constitutes the most prominent articulation of the Tiebout theory as applied to questions about distribution of powers to local government in the legal literature, yet he seems to take issue with the use of the term. In a recent article, Professor Gillette referred to another commentator’s “critique of what he labels the ‘public choice approach’ to interlocal relationships.” Gillette, supra note 18, at 246. I use the term here with the purely descriptive intent described above.

32. In Our Localism: Part II—Localism and Legal Theory, supra note 29, at 346, 393–94, Professor Briffault used that term to describe the theory of localism propounded by Professor Frug. In the urban development literature, one study uses “metropolitan ecology” as a term to describe Frug-like belief “in the importance of local autonomy and small-scale governance,” see Foster, supra note 24, at 44–47. Professor Frug does not label his own theory.

citizens through meaningful involvement in grassroots levels of government.\textsuperscript{34} Both models continue to animate the localism debate as it turns its attention to the metropolitan regions of the 21st century.

Published in 1956, Charles Tiebout’s \textit{A Pure Theory of Local Expenditures}\textsuperscript{35} remains the seminal academic foundation of the economic defense of localism in urban government. According to Tiebout’s model, local governments will strive to provide a desired mix of services in order to retain their constituents, the individual citizens whom he described as “consumer-voters.”\textsuperscript{36} Competition between and among government units should produce greater efficiency in the provision of public services as well as more variety in the range and level of services offered by different government units. Crucial to Tiebout’s theory was the ability of citizen consumers to translate their preferences for a particular mix of public services into a choice of local government by exercising their power of “exit,” thereby ensuring ongoing competition among municipalities to attract and retain taxpayer citizens. Although Tiebout did not take a position in the political debate between localism and greater centralization in metropolitan America, his theories led to the conclusion that a greater number of competing local government units would produce higher citizen satisfaction and greater government efficiency. Successive generations of scholars have modified and refined Tiebout’s theory,\textsuperscript{37} yet the basic premises continue as the foundation for

---

\textsuperscript{34} Compare the following two assertions about the proper scope of local governmental power; they illustrate well the ideological chasm between the two approaches. On the one hand, the public choice theory rests on the normative claim that: “The primary function of a municipality is to provide local public goods.” Robert C. Ellickson, \textit{New Institutions for Old Neighborhoods}, 48 DUKE L. J. 75, 88 (1998). In contrast, Professor Frug’s defense of localism on the basis of participation insists that: “[T]he role that cities ought to play in American society . . . is community building.” Jerry Frug, \textit{The Geography of Community}, 48 STAN. L. REV. 1047, 1048 (1996). By community building, Frug means “the cultivation and reproduction of the city’s traditional form of human association.” Id. at 1077.

\textsuperscript{35} Charles M. Tiebout, \textit{A Pure Theory of Local Expenditures}, 64 J. POL. ECON. 416 (1956).

\textsuperscript{36} Id. at 419.

\textsuperscript{37} Professor Clayton Gillette is the leading exponent of Tiebout’s theory in the legal literature. See, e.g., Clayton P. Gillette, \textit{Opting Out of Public Provision}, 73 DENV. U. L. REV. 1185 (1996) (applying Tiebout’s theory to the privatization of municipal services.). At the same time, Gillette has frequently recognized that Tiebout’s assumptions do not reflect current realities. See, e.g., Gillette, supra note 18, at 197–210; Clayton P. Gillette, \textit{Reconstructing Local Control of School Finance: A Cautionary Note}, 25 CAP. U. L. REV. 37, 40 (1996) (recognizing that “the Tiebout world, however, is obviously not the world in which we live”); Clayton P. Gillette, \textit{Courts, Covenants, and Communities} U. Chi. L. REV. 1375, 1389 n.59 (1994) (noting that Tiebout’s model predicts the correct allocation of public services only so long as a series of externalities, and choice among substantial numbers of localities); Clayton P. Gillette, \textit{Equality and Variety in the Delivery of Municipal Services}, 100 HARV. L. REV. 946, 956 (1986) (stating that the documented inequality of municipal services provision itself disproves the assumptions underlying the Tiebout hypothesis).
those who herald the efficiency advantages of decentralization of power to local governments and reject arguments for regionalization or centralization of governments in metropolitan areas. 38

Scholars have criticized Tiebout’s theory from two main vantage points. Some question the accuracy of Tiebout’s assumptions about human behavior, while others disagree with the theory’s normative bases and practical results. Various commentators, for instance, have challenged the premise that local government units are formed to provide a specific mix and level of services. They point to evidence that local governments are often created for the sole, less benign, purpose of excluding poor and minority residents. 39 Others have rejected Tiebout’s assumption that individuals choose a municipal home that satisfies their desires for a particular mix and level of services, noting that choice is an inaccurate term to describe the plight of the poor who live with inadequate service levels. 40 Critics also fault Tiebout for ignoring how the costs of exercising that choice fall disproportionately on those with the fewest resources. 41 They question too whether the much noted lack of

More recently, Professor Lee Anne Fennell has documented the limits of both exit and voice in the economic analysis of local government services. See Lee Anne Fennell, Beyond Exit and Voice: User Participation in the Production of Local Public Goods, 80 TEX. L. REV. 1 (2001). Professor Fennell asserts that the exit-voice phenomena cannot capture the fact that local citizens have a dual role, as both consumers and producers, with regard to important services such as schools and public safety. See id. at 11–12.

38. See Ellickson, supra note 34, at 82-85; Gillette, supra note 18, at 208–10.

39. See, e.g., NANCY BURNS, THE FORMATION OF AMERICAN LOCAL GOVERNMENTS: PRIVATE VALUES IN PUBLIC INSTITUTIONS 117 (1994) (noting how the formation of local government units creates boundaries that exclude others, she claims “Americans have discovered in local institutions effective barriers to racial and economic integration”); GARY J. MILLER, CITIES BY CONTRACT 63–84 (1981) (describing how wealthy enclaves use municipal incorporation to avoid redistribution of wealth to poorer citizens; concluding that city formation in these cases “had little to do with the gratification of distinct collective tastes for public goods.” For additional economically based criticism of Tiebout, see generally Foster, supra note 24, at 39–41.

40. Moving is an expensive proposition, both in terms of the financial costs and the personal costs it imposes on citizens. As Professor Frug has noted, “[p]eople who live in unsafe neighborhoods or send their children to inadequate schools don’t do so because they have taste for them . . . . If they had a choice . . . , they would prefer better schools and less crime.” See Frug, supra note 33, at 31. Along those same lines, some have rejected the Tiebout theory because of its unrealistic assumption of full knowledge on the part of citizens. See GREGORY R. WEHNER, THE FRACTURED METROPOLIS: POLITICAL FRAGMENTATION AND METROPOLITAN SEGREGATION 17 (A. Gary Dworkin ed., 1991).

41. See, e.g., Briffault, supra note 29, at 420 (stressing that the exercise of the choice of municipality comes at different prices to different citizens). For the poor, the costs of moving, the constraints imposed by the need to be close to jobs, and the limited availability of affordable housing in many municipalities means that “poorer, less educated potential movers will have fewer options and will be forced to bear more costs if they attempt to move.” Id. See also id. at 422 (noting that wealthy communities, with their high levels of property wealth, can tax at low rates and generate
Intergovernmental Cooperation

affordable housing in suburban areas makes moving an illusion even for many moderate income individuals who are willing and able to assume the costs of the move.\footnote{See Miller, supra note 39, at 204–06.} And finally, even some who share Tiebout’s enthusiasm for the role of market mechanisms in achieving efficiency in the provision of public services suggest that the theory improperly emphasizes the consumer-voter’s likely exercise of choice in the selection of local government. Instead of focusing exclusively on the threat of “exit,” they argue, the theory should recognize the importance of citizen “voice” in motivating local government officials.\footnote{See Briffault, supra note 29, at 420.}

The second line of critical attack on Tiebout’s model questions the normative acceptability of the metropolitan landscape the theory produces. Though the criticism is wide-ranging, it coalesces around four main points. First, the critique asserts the highly fragmented local government world envisioned by Tiebout inevitably results in a self-destructive competitive “race to the bottom,” as municipalities try to out-bid each other in the incentives they are willing to offer to entice business and the property wealth it brings into their jurisdictions.\footnote{See Frug, supra note 33, at 33–34.}

Second, it allows self-contained local government units in a metropolitan region to take actions with negative spillover impacts on their neighbors.\footnote{See Briffault, supra note 2, at 1132–33; Briffault, supra note 29, at 429–30. This critique applies to both Frug’s and Tiebout’s theories of localism. As a matter of political theory, it argues, localism is flawed. Because government decisions should be made at the level that is coterminous with the impacts of its decision, for many issues, the region or metropolitan level is the proper level for decisionmaking. See Briffault, supra note 25, at 20–23.}

Third, in the view of some critics, Tiebout’s depiction of local communities as aggregations of like-minded individuals primarily concerned with the consumption of public services denigrates and trivializes humanity and ignores other important local government functions.\footnote{See Frug, supra note 33, at 28–31. Professor Frug bemoans Tiebout’s reliance on the dues-paying mentality of citizenship and its abandonment of norms of equality, “replacing the one-person, one-vote principle associated with democracy with the one-dollar, one-vote rule of the marketplace.” Id. at 31.}

Finally, some commentators argue that Tiebout’s theory unacceptably assumes, indeed encourages, socioeconomic segregation and the preservation of the widening gap between wealthy suburb and...
declining city.\(^47\)

In contrast to Tiebout's efficiency based analysis, the second prominent localist doctrine rests on the conception of local government as the place where democratic ideals can most easily be realized. This participation theory of localism originated in the modern legal literature with Professor Gerald Frug's 1980 publication of *The City as a Legal Concept*.\(^48\) That article has become a classic in local government law scholarship as a preeminent justification of localism. Though Frug favors the same redistribution of wealth as many of those scholars who support centralization of power as a means towards that end,\(^49\) he views highly centralized levels of government as unresponsive, ineffective, and impersonal. In this view, centralized governments are incapable of building the sense of community so necessary to defeat the alienation and loss of collective well-being that plague modern metropolitan areas. Municipal governments could offer the opportunity for community building and meaningful participation, Frug asserts, but the rules of local government law make them powerless.\(^50\) In turn, city powerlessness produces citizen apathy and destroys their incentive to participate in local government. Without grassroots participation in local government, powerful elites are able to exercise real power at the expense of the community's broader collective interest. In Frug's view, then, enhanced local power will both increase the individual's connectedness with and involvement in the business of government, and achieve higher levels of social justice and equality.\(^51\)

---

\(^47\) See PEIRCE, *supra* note 9, at 17 (arguing that the huge socioeconomic gap between poor cities and wealthy suburbs constitutes a major barrier to economic prosperity); WEIHER, *supra* note 40, at 87–115 (1991) (political fragmentation produces segregation along municipal borders); Briffault, *supra* note 29, at 420, 425 (criticizing public choice model for its acceptance and reinforcement of inequality among local government units); Frug, *supra* note 33, at 33–35 (criticizing ability of wealthy suburbs to enhance wealth by excluding poor and spending tax revenues only on themselves).


\(^49\) Gerald E. Frug, *Against Centralization*, 48 BUFF. L. REV. 31, 32 (2000) (critiquing the view that centralization is the only alternative to the current fragmentation of America's metropolitan areas).

\(^50\) See Frug, *supra* note 29, at 1062–67. Frug describes how cities are subject to the states' absolute discretion to delegate power to municipal governments and how federal constitutional limits have similarly weakened municipal authority. *Id*.

\(^51\) Professor Richard Thompson Ford has also championed the preservation of local autonomy while recognizing the negative consequences of local insularity in metropolitan regions. He opposes the creation of regional governments, because he believes that "we will lose the opportunities for participatory, or at least responsive, democratic government, effective place based political initiatives, and civic interaction and identification with the public sphere. Meanwhile government
Intergovernmental Cooperation

As Frug’s theory has acquired prominence, his premises have been questioned on several fronts. In his important articles on localism, Professor Richard Briffault argued that Frug’s description of city powerlessness does not reflect the political reality of major metropolitan regions. In Briffault’s view, affluent suburban governments, far from displaying the powerlessness Frug bemoaned, actually exercise substantial autonomy and independence. Second, Briffault rejected Frug’s assertions that localism will enhance citizen participation or local autonomy, arguing (1) that the presence of a large number of small local governments in metropolitan regions narrows, rather than increases, the range of local powers; and (2) that citizens in metropolitan regions have bonds to numerous jurisdictions, thus weakening the ties with a home city that the localist theory envisions. Third, Briffault challenged the will become more distant, more bureaucratic and less responsive.”


52. See Briffault, supra note 26; Briffault, supra note 29.

53. See Briffault, supra note 26, at 1–3, 56; Briffault, supra note 29, at 389–406.

54. See Briffault, supra note 26, at 18–58. Using school financing schemes and land use regulation as examples, Briffault asserted that local powerlessness exists more as a matter of legal doctrine than as a description of the ways in which the favored quarter currently exists in most states. The allocation of local government powers in major metropolitan areas, he argued, has created powerful, affluent suburban municipalities with significant autonomy and control over the wealth within their territorial borders and over the course of development that territory will follow. In contrast to the black letter law of local powerlessness at the whim of the state creator, Briffault showed how common state statutory schemes for local government incorporation, see id. at 73–77, annexation, consolidation, see id. at 77–81, and home rule, see id. at 9–11, display a commitment to local autonomy.

Similarly, Briffault’s analysis of the Supreme Court’s deference to local autonomy in a variety of cases revealed similar staunch support of local insularity. Using a variety of justifications, the Court has upheld local power against attack from different quarters. For instance, in Village of Belle Terre v. Boraas, 416 U.S. 1, 9 (1974), the Court appeared to rely on a notion of local government as an autonomous community with the power to maintain the community character it desires. Briffault discussed a number of cases displaying similar localist tendencies. See Briffault, supra note 26, at 86–96. In Briffault’s view, the Court elevated the autonomous status of local governments above the interests of those who were adversely affected by their actions. Based on his review of the Court’s opinions, Briffault concluded that not only is the localist claim of local powerlessness inaccurate as a descriptive matter, it has also ignored the many ways in which local governments have been able to exercise power so as to impose negative extraterritorial impacts on other areas within the metropolitan region. See id. at 109–11.

55. See Briffault, supra note 29, at 407. In a similar vein, others have pointed to the loss of sense of place, and the alienation and dulling sameness of suburban communities. See, e.g., RICHARD MOE & CARTER WILKIE, CHANGING PLACES: REBUILDING COMMUNITY IN THE AGE OF SPRAWL (1997); Bruce Katz & Jennifer Bradley, Divided We Sprowl, THE ATLANTIC MONTHLY, available at www.theatlantic.com/atlantic/issues/99dec/9912katz.htm (Dec. 1999); James Howard Kunstler,
claim that enhanced local powers will result in redistribution of wealth, pointing out that local governments now have and always have had the power to redistribute and arguing that their failure to engage in greater redistribution is explained, not by a lack of power, but rather by their fear of losing high tax-paying citizens. Finally, some of the normative critique leveled at the Tiebout theory applies to Frug’s participation theory as well. That is, the critics question Frug’s underlying judgment that municipalities should receive enhanced powers to act in metropolitan regions. They note the spillover effects of many local decisions, suggesting that enhanced local power in the metropolitan setting improperly allows an individual unit of government to impose negative costs on neighboring local governments. According to that critique, localism becomes a perpetrator of inequalities rather than a mechanism for maximizing the community’s overall welfare.

B. Regionalist Solutions for Metropolitan Localism

Professor Richard Briffault’s path-breaking articles brought regionalism squarely front and center in the local government law literature. In Our Localism—Part I and Part II, both published in 1990, Briffault offered an extensive critique of localism and its consequences for overall metropolitan well-being. First, Briffault turned his attention to a rebuttal of the localists’ claim of local powerlessness, documenting the very real and important local government powers routinely exercised by

---

56. See Briffault, supra note 29, at 407–10. Another commentator has taken this critique one step further, rejecting the basic underlying assumption of the participation theory of localism that local governments are “natural benevolent institutions that are the perfect training ground for democratic citizens.” See Burns, supra note 40, at 116. In contrast, Burns argues that local governments are often created for anti-democratic, exclusionary reasons. See id.

57. See Briffault, supra note 29, at 426–28; Cashin, supra note 5, at 2012–15; Briffault, supra note 2, at 116–17; Briffault, supra note 26, at 6–7.

58. Under this view, one possible solution short of total metropolitan area consolidation would be the creation of a regional government unit with the power to address only those policy matters with substantial region wide effects. See Briffault, supra note 25, at 21–22 (“In metropolitan areas, democracy requires giving the regional electorate a voice in local decisions that have regional consequences. Only by widening the scale of participation to include all those affected by local actions can local decision-making in metropolitan regions be made truly democratic.”).

59. Briffault, supra note 26, at 1.

60. Briffault, supra note 29, at 407. Since publication of the Localism articles, Professor Briffault’s defense of regionalism has been a constant theme of his scholarship. In particular, two later articles elucidate the normative bases and describe the solutions he suggests for implementation of a regionalist agenda. See generally Briffault, supra note 25; Briffault, supra note 2.
Intergovernmental Cooperation

municipal governments across the country.61 In the second part of the work, Briffault’s *Localism* rejected the normative assumptions of localist doctrine.62 Finally, Briffault offered some tentative suggestions for reform of the localism dilemma in metropolitan America. As described in *Localism* as well as in later articles, he concluded that the formation of a regionwide, politically accountable local government unit is the only realistic means of ensuring that the impacts of government decisions in metropolitan areas are coterminous with their political boundaries.63 Noting that neither the small size of existing local governments nor their sheer numbers in metropolitan areas was itself inconsistent with regional well-being, Briffault stressed that his critique was based on the ease and frequency with which local governments in metropolitan areas are able to take actions whose impacts extend far beyond their borders.64 In his view, then, regional government units are necessary, not as a substitute for nor as a consolidation of existing local governments, but rather as an additional, politically accountable legislative body, whose functions would be limited to issues with demonstrable regionwide impacts.65 As he recently noted, “[R]egionalism is . . . localism for metropolitan areas.”66

For a large number of commentators, Briffault’s work offered tantalizing possibilities for ways in which social policies could be improved with regionalist solutions. The 1990s saw a resurgence of

---


62. That critique has been described in this Article’s earlier review of localism. *See supra* Part II.A. Briffault’s critique of the public choice defense of a marketplace of municipal governments in competition for citizens asserts that it is based on erroneous assumptions about human behavior and inaccurate assessments about the freedom of many individuals to act as an enlightened consumer in the choice of municipal service provider. *See supra* notes 38–47, and accompanying text. Though Frug’s participation defense of local government was sensitive to and critical of the ways in which affluent autonomous localities can preserve their privilege and imposed costs on other local government territories in the region, Briffault concluded that the theory is, like Tiebout’s approach, seriously flawed. In Briffault’s view, Frug’s localism is based on a misguided perception of the local government as community and home for the vast majority of suburban Americans, rather than one that perpetuates inequalities within a region. *See supra* notes 52–58 and accompanying text. Fundamentally, he claimed, both localist theories ignore the harm caused when the impacts of the local governments’ exercise of power extend far beyond their borders, exacerbating regional inequalities. *See id.*


64. *See Briffault, supra* note 29, at 426–30.

65. *See Briffault, supra* note 2, at 1165.

66. Briffault, *supra* note 25, at 1. Crucial powers to be exercised at the regional level, in Briffault’s view, include land use, local taxation, and service provisions that could be enhanced with a supra-municipal territorial base. *See Briffault, supra* note 2, at 1166.
regionalist proposals. One author, for instance, argued that regional government is the *sine qua non* of an effective policy to combat poverty.67 Another asserted that low and moderate income housing policy is doomed to failure without regional control of land use and taxation powers.68 Yet another commentator urged the establishment of a regional tax base. She argued that regionwide taxation would more evenly distribute resources and spread the costs of the negative impacts associated with high levels of urbanization and concentrated poverty.69 For these scholars, regional efforts are preferred because of their potential to promote equity and social justice. In their view, regional action is the only way to combat metropolitan America’s pronounced racial segregation and sharp disparities in local wealth, local tax bases, and the quality of local services.

Though many scholars, policy analysts, and urban planners agree that regional approaches are preferable to the status quo, regionalization has its critics as well. Especially when regionalism is defined as advocating the creation of a new general purpose regionwide government, localists argue that regionalism would unwisely deprive citizens of their choice of service providers. In their view, it also would eliminate the advantage of having a large number of small government units that compete to attract citizens by differentiating themselves with the type and quality of services they offer.70 Others assert that bigger government brings a reduced sense of community empowerment and decreases government responsiveness.71 Moreover, the political hurdles are enormous. Entrenched local government bureaucracies and the apparent public preference for continued suburbanization are but two of the major obstacles to metropolitan governance.72 In the view of one commentator,


70. For a recent articulation of the case against regionalism, see Gillette, supra note 18, at 197–210.


72. Briffault identified the enormous practical political barriers to the implementation of his regional agenda. Fundamentally, he observed, “Localists do not become regionalists simply because they live in metropolitan regions.” Briffault, supra note 25, at 2. Existing local bureaucracies are not likely to support relinquishment of significant regulatory and taxing powers. Nor are many segments of the private sector, such as “land developers, corporations, . . . and the businesses and residents located in the high-tax base localities,” *id.* at 27, likely to agree to reorganization of the government
Intergovernmental Cooperation

It is the minds of the American public, and not the governmental framework in which it operates, that imposes the major barrier to regionalism: “[I]t is unlikely that any structural innovation, like the establishment of metropolitan government” can change the political and social attitudes that have led to segregated suburbs; the need is not “to recast the structure of municipal government, but . . . to revise the preferences of the American people; in this area, organizational reform cannot substitute for the alteration of popular predilections.”

C. Framing the New Regionalism Inquiry

In various branches of urban development studies, academic commentators have heralded the arrival of the New Regionalism. Like many of its doctrinal predecessors, the New Regionalism is an outgrowth of earlier metropolitan strategies; it is a term that catches a wide range of ideologies within its sweep. In general, however, those who congregate under the broad umbrella of the New Regionalism have focused their attention on the disparities between central city and favored quarter: in terms of wealth, job creation, earning power, and racial and socio-economic integration, the central city and the inner suburban ring are falling dramatically behind their wealthy suburban counterparts. From a historical perspective, the New Regionalism turns its attention to the radical transformation of major metropolitan areas that began in post-World War II America. It confronts an urban landscape that displays a continually growing number of local government units located at ever

73. See Edward Zelinsky, Metropolitanism, Progressivism, and Race, 98 COLUM. L. REV. 665, 667 (1998); see also Anita Summers, Regionalization Efforts Between Big Cities and Their Suburbs, in URBAN-SUBURBAN INTERDEPENDENCIES, supra note 5, at 181 (noting failure of most consolidation votes); Robert D. Yaro, Growing and Governing Smart: A Case Study of the New York Region, in REFLECTIONS ON REGIONALISM, supra note 9, at 44, 70–71.

74. See Cashin, supra note 5, at 1988; Foster, supra note 11, at 83; Savitch & Vogel, supra note 9, at 158–61. See also supra note 11 and accompanying text.

75. Commentators have defined New Regionalism as “a set of policies designed to reduce inequality arising from the way the metropolis developed and to improve the overall quality of life.” See H.V. Savitch & Ronald K. Vogel, Metropolitan Consolidation Versus Metropolitan Governance in Louisville, 32 ST. & LOC. GOV’T REV. 198 (2000). Other authors have noted New Regionalism’s focus on reducing the gap between urban core and favored quarter. See, e.g., Foster, supra note 11, at 85 (noting that New Regionalism “links social equity to economic growth”); Bennett Harrison, It Takes a Region (or Does It?): The Material Basis for Metropolitanism and Metropolitics, in URBAN-SUBURBAN INTERDEPENDENCIES, supra note 5, at 143–45.
increasing distances from central cities. It documents the undisputed gap between suburban prosperity and central city decline. At the same time, however, it notes the repeated failure of local government consolidation efforts in major metropolitan areas and stresses the permanence of existing multi-purpose local governments. Thus, it looks for regional answers to metropolitan area inequalities through a lens that is tempered with the pragmatic realization that proposals to eliminate existing local government units are unlikely to succeed.

Though loosely united in their articulation of policy goals, the New Regionalists embrace a wide range of divergent strategies. In fact, rather than a cohesive ideology with a well-accepted policy agenda, New Regionalism refers more precisely to shared concerns and goals for metropolitan area equity. At the level of specific proposals, the variety is significant. Though some continue to advocate consolidation between central city and the county it inhabits, most New Regionalists reject the so-called “big box” government approach in favor of more flexible, informal, governance arrangements. Examining possible solutions such as the “multitiered” approach, “linked functions,” or “complex

76. See, e.g., Mark S. Rosentraub, City-County Consolidation and the Rebuilding of Image: The Fiscal Lessons from Indianapolis’s UniGov Program, 32 ST. & LOC. GOV’T REV. 180, 189–90 (2000) (concluding that Indianapolis’s revival could have happened without consolidation, noting the many other factors that produced economic growth and pointing out that the fastest growing areas in the Indianapolis region are beyond the borders of UniGov); Savitch & Vogel, supra note 75, at 210 (concluding that consolidation in the Louisville area would not achieve metropolitan equity). One study of several city-county consolidations concluded that supporters of consolidation do not seek to equalize suburban-city disparities. Successful consolidation agendas focused on “local services, governmental ‘turf,’ taxes, and race”). See Arnold Fleischmann, Regionalism and City-County Consolidation in Small Metro Areas, 32 ST. & LOC. GOV’T REV. 213–24 (2000). An analysis of the Jacksonville, Florida consolidation concluded that the benefits of consolidation are overstated. See Bert E. Swanson, Quandaries of Pragmatic Reform: A Reassessment of the Jacksonville Experience, 32 ST. & LOC. GOV’T REV. 227, 236 (2000). In contrast to the pervasive anti-consolidation approach to metropolitan regions, David Rusk continues to advocate large scale local government consolidation in metropolitan areas. Rusk, the former mayor of Albuquerque, has become a vocal proponent of metropolitan governments. He urges state governments to consider four different approaches: 1) unifying local governments, either by empowering urban counties, by consolidating county and municipal governments, or by creating a regional government that extends throughout the multiple cities and counties that comprise most major metropolitan areas; 2) authorizing municipal annexation without approval of property owners; 3) limiting the creation of new municipalities; and 4) promoting public partnerships for joint action among local governments. Rusk, supra note 2, at 90–102.

77. A recent symposium issue on the New Regionalism explored these and other proposals for regional redistribution of wealth, service provision, and regulatory powers. See Symposium, New Regionalism, supra note 11.

78. See Savitch & Vogel, supra note 9, at 162–63. The multitiered approach to regionalism allocates power among existing and/or new layers of government according to the scope of the impact of the power being exercised. Metropolitan area government “is supposed to better deal with
Intergovernmental Cooperation

networks, they tend to endorse the establishment of limited purpose governments or the execution of voluntary intergovernmental agreements over proposals for wholesale consolidation of existing units of government. As a result, New Regionalism offers a middle ground for the dispute over the allocation of state and local government power. It attracts localists by recognizing that large scale consolidations are unlikely to be a successful political strategy, and that large numbers of local governments are an enduring feature of the metropolitan region. At the same time, New Regionalism appeals to the critics of localism, by documenting the ways in which fragmented local government authority and the phenomenon of suburbanization have contributed to central city decline and a widening gap between urban core and affluent suburbia.

1. The Normative Bases of the New Regionalism

Though the New Regionalists may agree about the ultimate goals for metropolitan America, the normative claims on which their arguments rest are quite diverse. At least four criteria underlie their proposals: efficiency, economic interdependence, participatory democracy, and equity. With regard to economic efficiency, the New Regionalist questions the traditional public choice localist claim that a multiplicity of autonomous local governments maximizes local wealth and efficiency by forcing local government units to compete with each other to attract citizen taxpayers. They point to recent empirical studies that suggest that fragmentation actually creates inefficiencies by allowing local governments to impose costs on others, and by allowing suburbs to

issues that cut across a number of local jurisdictions or involve redistributive policies.” Id. at 163. Lower level, municipal governments retain control over “labor-intensive services, which call for close relationships between service deliverers and citizen-consumers.” Id. at 162. The authors note that metropolitan level governments “often find themselves crushed between the grindstones of local and higher levels of government” and conclude that their record is mixed. Id. at 163.

79. See id. This approach results in a functional consolidation for service provision across jurisdictions, usually through the formation of interlocal service agreements. No new levels of government are required. In practice, the authors claim that linked functions are frequently perceived as an incomplete, temporary solution to a regional problem. For a review of the experiences of several cities that have used the linked functions approach, see Timothy D. Mead, Governing Charlotte-Mecklenburg, 32 ST. & LOC. GOV’T REV. 192, 194–96 (2000); Savitch & Vogel, supra note 75, at 198–212.

80. See Savitch & Vogel, supra note 9, at 164 (defining complex networks as “large numbers of independent governments (voluntarily) cooperating through multiple, overlapping webs of interlocal agreements”). Complex networks produce “numerous jurisdictions with overlapping services,” id., thus maximizing the citizens’ range of choices.

81. See Valente et al., supra note 29, at 350–52; Briffault, supra note 25, at 15–26; Cashin, supra note 5, at 2042–47. See also Summers, supra note 73, at 182.
maximize their own wealth at the expense of the central city they
surround.82 The New Regionalists also claim that excessive
fragmentation of local government units establishes barriers to the
efficient provision of some services.83 In metropolitan areas especially, a
large territorial base would better distribute the costs of infrastructure
and produce a higher quality, if not less expensive, service.84 Currently,

nearly all participants in the New Regionalist debate appear to concur
that some type of regionwide action would enhance regional efficiency
and provide a necessary antidote to extensive local autonomy.

The second regionalist norm, the asserted economic interdependence
of the metropolitan region, rests on evidence that suburban economic
well-being suffers as the gap between suburb and city widens. According
to the suggested “interdependence imperative,”85 central city health
becomes an item of suburban self-interest; suburbia ignores the fate of
the city at its peril and to the detriment of its own prosperity. One well
known commentator has identified numerous ways in which suburban
welfare depends on a strong central city, including: (1) central city
image is crucial to regional welfare; (2) new regional employers will
need to tap city markets to fill their work force; (3) failure to address
inner-city social problems will come back to haunt all taxpayers in
the form of higher costs for prisons and welfare; (4) inner-city crime
affects the image of the entire region; (5) environmental issues can
only be addressed regionwide; and (6) regional cooperation will bring

82. See, e.g., Briffault, supra note 25, at 18–20; Cashin, supra note 5, at 2000–01. Several
empirical studies support those claims. See, e.g., Joseph Persky & Wim Wiewel, The Distribution
of Costs and Benefits Due to Employment Deconcentration, in URBAN-SUBURBAN
INTERDEPENDENCIES, supra note 5, at 67 (concluding that expansion of suburban manufacturing
employment opportunities imposes substantial costs on inner city residents and represents a “subsidy
from low- and moderate-income black and Hispanic residents of the city and inner suburbs to
stockholders elsewhere in the nation”); Richard Voith, The Determinants of Metropolitan
Development Patterns: What are the Roles of Preferences, Prices and Public Choices, in URBAN-
SUBURBAN INTERDEPENDENCIES, supra note 5, at 50, 78 (study of recent transportation investments
in the Philadelphia area shows that cities subsidize suburban transit development, leading to
inefficient suburbanization, diminished opportunities for central city, and a clear causal link between
city and suburban dependency).

83. See, e.g., Briffault, supra note 2, at 1166.

84. A recent book by Kathryn A. Foster on special districts concludes, somewhat surprisingly
perhaps, that services provided by regional single purpose government units are actually more
expensive than having the general purpose government provide the same service. Foster, supra note
24, at 148–88. Foster speculates that the higher cost is due to “inflationary influence of political
isolation, functional specialization, and administrative and financial flexibility,” id. at 174–76. She
recognizes, however, that the higher cost may merely reflect the fact that regional special districts
provide a higher quality of service. See id. at 184.

85. See PEIRCE, supra note 9, at 131.
enhanced political clout. In addition, the economic well-being of the suburbs appears to correlate significantly with the prosperity of their central cities. Stated simply, the claim is that healthier metropolitan regions contain more prosperous central cities. In fact, in the twenty-five metropolitan areas with the most rapid income growth, central city incomes also increased. Conversely, as the gap between central city and suburban prosperity widens, the overall absolute level of suburban wealth tends to be lower; in the eighteen metropolitan areas that recently experienced declines in income, central city income also decreased in all but four instances. Thus, the reduction of socio-economic disparities in metropolitan regions can be justified as a matter of favored quarter self-interest, rather than exclusively dependent on municipal selflessness and charitable intent. If in fact the fate of the region depends on the health of the central city, redistribution from favored quarter to urban core is imperative.

Third, the New Regionalism rejects the localist claim that ideals of democratic participation militate against metropolitan area government reform. Recognizing that localists have long used the preservation of responsive, accountable units of government as an argument against regionalization, some New Regionalists argue that, in fact, a regionally bounded multi-purpose government would better promote those
longstanding democratic ideals.  

Equity, the fourth criterion advanced by many New Regionalists, is rooted in the assertion that the current economic schism between city and affluent suburb is fundamentally unfair. In this view, local government boundaries enhance the preservation of privilege and prevent the redistribution of wealth in metropolitan areas. Though the norms of efficiency, economic interdependence, and democratic participation can be debated with economic and statistical analyses, equity stands alone with its roots in moral convictions about fairness and justice. It starts with a recognition of the fact that suburbanization and the central city decline that accompanied it were the results of government policies that transferred large amounts of public resources from city to suburb. Metropolitan highway systems, government funding of commuter rail systems, education spending, government mortgage financing, and the privileged tax status of real property tax payments \(^{92}\) represent huge public subsidies of the suburbs at the expense of the cities. \(^{93}\) But the strength of the equity norm does not depend on empirical corroboration. More simply, it is rooted in the conviction that regional redistribution of wealth is “the right thing to do.” \(^{94}\)

91. See, e.g., Briffault, supra note 2, at 1164–70; Cashin, supra note 5, at 2035–47.

92. Because local property tax payments are deductible for federal income tax purposes, and because local school funding relies heavily on the property tax, the deductibility provides a huge indirect federal subsidy to schools. As Richard Rothstein explained: “A family in the 28% bracket that pays $1000 in local property taxes for public schools can deduct that payment on its income tax returns, reducing its income tax bill by $280. Of the $1000 going to schools, the family pays only $720 out of its earnings. The federal government contributes the $280 balance.” See Richard Rothstein, How the U.S. Tax Code Worsens the Education Gap, N.Y. TIMES, Apr. 25, 2001, at A17. A Stanford University economist theorizes that this federal subsidy makes families more willing to pay higher taxes for better schools. Id.

93. See, e.g., Cashin, supra note 5, at 2003–14 (citing various examples where the “favored quarter” receives a disproportionate share of public infrastructure funds for roads, highways, expensive sanitary sewage treatment systems); DUANY ET AL., supra note 5, at 8–15; Richard Voith, The Determinants of Metropolitan Development Patterns: What are the roles of Preferences, Prices, and Public Choices, in URBAN-SUBURBAN INTERDEPENDENCIES, supra note 5, at 78–80 (claiming that in the case of Philadelphia and possibly other cities, the central city subsidized public transit for suburbanites); ORFIELD, METROPOLITICS, supra note 10, at 71 (noting that in the fee structure of regional sewage disposal, Minneapolis and St. Paul pay $6 million more in fees than they incur in costs).

94. In the words of one urban economist who recently concluded that the regionalists’ economic interdependency argument is at best subject to empirical dispute:
2. **New Regionalism in the Legal Literature**

Although New Regionalism is fairly new to the legal literature and the incidents of self-identification are rare, some of the academic commentary has begun to use the term. Professor Sheryll Cashin’s recent article described the movement in the following terms: “[The New Regionalism] attempts to bridge metropolitan social and fiscal inequities with regional governance structures, or fora for robust regional cooperation, that do not completely supplant local governments.”

Irrespective of the label used, the current work of both localists and regionalists can fairly be said to come under the wide doctrinal umbrella of the New Regionalism. Public choice theorists, for instance, temper their enthusiasm for the asserted efficiency benefits of a fragmented metropolitan region, proposing voluntary cooperation among and between units, and suggesting that wealthier local governments will recognize that their own welfare is maximized by a regionwide sharing of burdens.

In addition, those who evaluate local government structures from the public choice perspective applaud the use of regional governance efforts for the provision of regionwide services, especially those that require large investments in capital infrastructure. Similarly, New Regionalism’s tenets appear to have appealed to such staunch localists such as Professor Frug, who recognizes that local autonomy produces unacceptable negative impacts in a metropolitan region and proposes ways in which a regionwide entity could produce a more equitable distribution of burdens. Though still fundamentally opposed...
to wide-scale consolidation of metropolitan area governments, both Frug and the public choice scholars admit the need for some regional entities or agreements. And finally, Richard Briffault’s critique of localism anticipated the New Regionalism’s acceptance, indeed support, of local autonomy. Briffault’s pragmatic assessment of the likely failure of general purpose regional governments, coupled with his endorsement of regional legislatures with well defined powers, would leave untouched the large number of pre-existing autonomous local government units in metropolitan areas. 99 Localist and regionalist doctrine alike, thus, enter the 21st century tempered with a recognition that local government law and theory must take into account both the reality of metropolitan regions and the interconnectedness among units of government, as well as the stubborn permanence of those units of local government.

Notwithstanding the emerging consensus on the importance of the region in metropolitan areas, it would be foolish to claim that localists and regionalists have put aside their differences and found salvation in the New Regionalism. 100 Rather, the allure of the New Regionalism may be simply that it holds something for everybody, that its wide ranging normative concerns and diversity of proposals allow all who study urban development to find something in the New Regionalist agenda to support. Those who emphasize economic efficiency arguments are likely to champion voluntary cooperation to produce supra-municipal service providers. In contrast, those who stress the economic interdependence norm are likely to support regionalist proposals that rely on enlightened suburban self-interest to cure metropolitan area inequality. The participation localists, who place primary importance on the role of small

100. In fact, a comparison of the localist approaches to the New Regionalism reveals major differences in the ways in which scholars support regionalization. While public choice advocates such as Professor Gillette seem to endorse only those regionalist efforts that preserve local autonomy, see Gillette, supra note 18, at 197–210, Professor Frug and the participation theorists advocate a real infusion of regional values into local decisionmaking, see Gerald E. Frug, Against Centralization, 48 BUFF. L. REV. 31, 36–37 (2000) and Jerry Frug, Decentering Decentralization, 60 U. CHI. L. REV. 253, 271–312 (1993).
Intergovernmental Cooperation

governmental units in the preservation of democracy, will undoubtedly favor those regionalist proposals that tinker the least with local autonomy. And finally, those who come to regionalism with an eye toward ensuring a closer tie between government’s territory and the scope of its impact are likely to defend regionalism reforms that entail the establishment of new, regional government bodies.101

III. INTERGOVERNMENTAL COOPERATION IN METROPOLITAN AREAS

The preceding description of New Regionalists’ emerging shared articulation of policy goals for metropolitan areas lays the background for this Part, which begins with a brief overview of the legal structural framework within which intergovernmental cooperation is established and operates. It then describes and evaluates the major types of intergovernmental cooperation techniques available in metropolitan areas. Finally, this Part reviews the New Regionalism’s analysis of intergovernmental cooperation and notes the incomplete and/or inaccurate assumptions on which it is based. It concludes by arguing that these voluntary efforts are actually unlikely to promote, and may in fact work counter to, many of the New Regionalism’s goals.

A. Types of Intergovernmental Cooperation Efforts

All intergovernmental cooperative efforts operate against a backdrop of state enabling authority. The possible sources of authority are wide-ranging, and may include a state constitutional provision102 municipal home rule powers,103 a general state statute enabling intergovernmental

101. As the preeminent proponent of regionalism in local government law, Richard Briffault recently summarized the normative defense of a redistribution of government power in metropolitan areas: “in contemporary metropolitan areas, the economically, socially, and ecologically relevant local area is often the region. Consequently, in metropolitan areas, concerns about efficiency, democracy, and community ought to lead to support for some shift in power away from existing localities to new processes, structures, or organizations that can promote decision-making on behalf of the interests of a region considered as a whole. Regionalism is, thus, localism for metropolitan areas.” Briffault, supra note 25, at 2.

102. See, e.g., ILL. CONST. art. 7, § 10; FLA. CONST. art. 8, § 4.

103. See Cape Motor Lodge, Inc. v. City of Cape Girardeau, 706 S.W.2d 208, 212–13 (Mo. 1986) (en banc) (upholding intergovernmental agreement as a valid exercise of home rule power without independent statutory authorization).
cooperation, or a specific enabling act for a particular type of intergovernmental action. In some cases, independent statutory restrictions, the primacy of other provisions of state law, or the interaction between different intergovernmental cooperation statutes


106. See, e.g., City of Hamilton v. Public Water Supply Dist., 849 S.W.2d 96 (Mo. Ct. App. 1993). In this case, the city’s contract to supply water to a public water district stated that no capital costs would figure into the rates charged by the city. When the city undertook to raise rates to cover the cost of repaying bonds for capital improvements, the water district filed suit to block the rate increase. The appellate court upheld the city’s power to ignore the contract limitation, noting that a state statute required the city to set rates at a level high enough to recoup the costs of financing the bonds. See id. at 101. See also Jefferson v. Mo. Dep’t of Natural Res., 863 S.W.2d 844, 849-50 (Mo. 1993) (local government must follow statute authorizing creation of intergovernmental solid waste management districts and cannot rely on state constitutional provision for intergovernmental cooperation); Info. Techs., Inc. v. St. Louis County, 14 S.W.3d 60, 62 (Mo. Ct. App. 2000) (application of state statutory requirements on bidding for public contracts depended on whether intergovernmental contract involved governmental function).

107. A contract for services, though adhering to statutory requirements on intergovernmental cooperation, may be invalidated for noncompliance with other statutory directives. See e.g., W. Wash. Univ. v. Wash. Fed’n of State Employees, 58 Wash. App. 433, 440, 793 P.2d 989, 992 (1990) (invalidating university contract with city for the provision of police services because of overriding statutory mandate that disbanding of university police department must adhere to labor rules regarding layoffs and interpretation of university police collective bargaining agreement). See also Elk Grove Township Rural Fire Prot. Dist. v. Vill. of Mount Prospect, 592 N.E.2d 549, 552 (Ill. App. Ct. 1992) (intergovernmental agreement involving promise to execute future tax levies invalidated by general principles of Illinois law; agreement improperly “denies prospective administrations and taxpayers any input into future levies . . . [and] is therefore void ab initio”); City of New Smyrna Beach v. County of Volusia, 518 So.2d 1379, 1382 (Fla. Dist. Ct. App. 1988) (Florida constitutional requirement for a dual referendum to uphold intergovernmental transfer of “function and power relating to services” not applicable to intergovernmental transfer of regulatory powers). Compare, e.g., Farlouise v. LaRock, 315 So.2d 50, 56 (La. Ct. App. 1975) (state bidding requirements not applicable to industrial district created by municipality), with Smith v. Intergovernmental Solid Waste Disposal Ass’n, 605 N.E.2d 654, 663 (Ill App. Ct. 1992) (statutory requirements for competitive bidding applies to regional special district).

108. In State v. Plaggemeier, 93 Wash. App. 472, 474–75, 969 P.2d 519, 521 (1999), five local governments entered a mutual aid agreement for the purpose of reducing drunk driving in their area. Two state enabling statutes were relevant to the scope of the agreement. First, under Washington’s Mutual Aid Act, a local sheriff or chief of police can give consent to mutual aid agreements with neighboring municipalities, see Wash. Rev. Code § 10.93.070 (2002). Under Washington’s Interlocal Cooperation Act, Wash. Rev. Code § 39.34 (2000), however, all intergovernmental cooperative efforts require ratification by the governing legislative bodies. In its decision to uphold an arrest made pursuant to a Mutual Aid Agreement that had not been ratified by the participants’ legislative bodies, the court invalidated all other aspects of the agreement, which included the establishment of an interjurisdictional task force. Plaggemeier, 93 Wash. App. at 476–81, 969 P.2d
Intergovernmental Cooperation

may limit local governmental flexibility. Moreover, general state constitutional limits on government power, such as the anti-delegation doctrine, the prohibition of special commissions, the prohibition of special legislation, restrictions on government subscription of stock, public purpose requirements, debt limitations, and the general state requirement that local governments not contract away their police


110. See, e.g., Specht v. City of Sioux Falls, 526 N.W.2d 727, 731 (S.D. 1995) (invalidating regional authority as unconstitutional special commission); but see Local 22, Phila. Fire Fighters’ Union, 613 A.2d at 526 (rejecting special commission challenge).


113. See, e.g., Mulkey v. Quillian, 100 S.E.2d 268, 270–71 (Ga. 1957) (intergovernmental contract involving state loan to political subdivision of state violates state constitutional requirement that tax revenues be spent for a public purpose).

114. See, e.g., Nations v. Downtown Dev. Auth., 338 S.E.2d 240, 243 (Ga. 1986) (holding intergovernmental contracts clause of state constitution limited by separate constitutional provision requiring voter consent to city decision to incur debt pursuant to intergovernmental agreement). But compare with Ambac Indem. Corp. v. Akridge, 425 S.E.2d 637, 640 (Ga. 1993) (county’s contract for services with regional special district, which involved issuance of revenue bonds for financing of waste disposal system, did not constitute debt in violation of state constitution).
power,\textsuperscript{115} may reduce the breadth of state enabling authority. In addition, states disagree on whether both entities in an intergovernamental agreement must independently have the power to engage in the activity that is subject to the agreement, or whether it is sufficient for one unit to be enabled. Known as the debate between the “mutuality of powers”\textsuperscript{116} approach and the “power of one unit”\textsuperscript{117} approach, the principle also has a substantial impact on the breadth of potential intergovernamental cooperative efforts. Within those general parameters, however, local discretion is quite broad.

According to the categorization adopted by the Advisory Commission on Intergovernmental Relations\textsuperscript{118} intergovernmental cooperative efforts

\begin{footnotesize}
\begin{enumerate}
\item Under the “mutuality of powers” approach to intergovernmental cooperation, governments can only contract with each other for services if each governmental unit has independent authority to engage in the subject of the contract on its own. See, e.g., United Water Res., Inc. v. N. Jersey Dist. Water Supply Comm’n, 685 A.2d 24, 31 (N.J. Super. Ct. App. Div. 1996) (concluding that intergovernmental cooperation act “was not intended as a vehicle to enhance the enumerated powers granted to local units”); Durango Transp., Inc. v. City of Durango, 824 P.2d 48, 49 (Colo. Ct. App. 1991) (applying COLO. CONST. art. XIV, § 18(2)(a), which authorizes intergovernmental cooperation and contracts to provide any function, service, or facility lawfully authorized to each of the cooperating or contracting units). The issue is not always as straightforward as it might appear. The Washington Interlocal Cooperation Act, WASH. REV. CODE § 39.34 (2000), for instance, clearly articulates a mutuality of powers standard. When the city of Spokane created a public facilities district for the purpose of financing and building a new public arena jointly with the county, it proceeded to condemn private property for the district to use as the arena site. The landowners challenged the condemnation, arguing that, because the district has no powers of condemnation, the interlocal cooperation act prohibited the city’s actions. The court upheld the city’s actions, saying that the its role as a “partner” and not as an “agent” in the agreement meant that the mutuality provision did not prohibit the condemnation. See Schreiner v. City of Spokane, 74 Wash. App. 617, 624, 874 P.2d 883, 888 (1994).
\item Under the “power of one unit” approach, so long as one of the contracting governments has the power to undertake the activity, an intergovernmental contract will be upheld. The result is an increase in governmental power. See, e.g., County of Wabash v. Partee, 608 N.E.2d 674, 679 (Ill. App. Ct. 1993) (“The very purpose of [the Intergovernmental Cooperation provision of the Illinois Constitution] is to allow a local government to do indirectly that which it cannot do directly, as long as it is otherwise lawful.”). For a comparison of the two approaches to intergovernmental cooperation, see ACIR, supra note 22; VALENTE ET AL., supra note 29, at 407–14.
\item A survey of these state statutes on intergovernmental cooperation, conducted by the Advisory Commission on Intergovernmental Relations, concluded that intergovernmental cooperation efforts are widely used and suggested a three part categorization that describes the most prevalent types of intergovernmental cooperative efforts: contracts for services; joint provision of services; and the creation of a new governmental entity. See ADVISORY COMMISSION ON
\end{enumerate}
\end{footnotesize}
Intergovernmental Cooperation

can be characterized as contracts for services; joint provisions of services; and the creation of a new unit of government.\textsuperscript{119} In addition, Professor Clayton Gillette has recently described a fourth type of intergovernmental cooperation, which he refers to as “voluntary burden sharing.”\textsuperscript{120} The following analysis will consider whether any of these efforts constitute a realistic mechanism for reducing regional disparities. It will suggest that the New Regionalism’s generally positive assessments of intergovernmental cooperation are flawed. The availability of voluntary regionwide cooperative efforts may improperly remove an incentive to meaningful regional burden sharing and may facilitate the ongoing off-loading of metropolitan burdens onto the less affluent segments of metropolitan areas. In fact, by allowing independent local governments to participate in metropolitan governance only when it benefits their own short-term interests, intergovernmental cooperation may exacerbate the metropolitan regional inequality that the New Regionalism seeks to eliminate.

1. \textit{Contracts for Services}

Pursuant to statutes or state constitutional provisions, local
governments in most states are empowered to contract with another local government to receive services provided by one of the contracting units. Presumably, two general purpose local government entities, such as municipalities or counties, will not enter into a contract for services unless the receiving entity can buy the service at a price that is lower than what it would have to spend to produce and provide the service itself. At the same time, the providing entity is unlikely to contract to provide the service unless it derives a benefit as well, for instance, by being able to exploit excess capacity that would otherwise go unused, by being able to price the service at a level that allows it to derive a profit, or by enhancing the welfare of its own citizens by providing the service on a larger scale.

Consider, for instance, the facts of Durango Transportation, Inc. v. City of Durango. In that case, a city-county agreement required the city to extend its municipally owned and operated bus service outside of the city but within the county’s territorial limits. Questioning the legality of the intergovernmental contract, a private company sued to stop the municipal service provider from encroaching on its own service territory within the county. Under the terms of that intergovernmental contract, a Transit Advisory Board was established to provide advice and recommendations to the city, which had the ultimate decisionmaking power. The Board, in turn, was comprised of individuals appointed by both the city and the county. The county had no “financial or

121. According to an ACIR study undertaken several decades ago, virtually all states authorize intergovernmental cooperation, either by constitutional provision, statutory enabling acts, or both. See ACIR, supra note 20, at 91.

122. This Section will focus on services provided pursuant to intergovernmental contracts between general purpose local governments. It is worth noting that these intergovernmental contracts may deal with either ongoing services (as is the more common case discussed in the text) or for the provision of a particular capital improvement, see, e.g., Skybort Props. of Or. v. Multnomah County Drainage Dist., 844 P.2d 909, 912 (Or. Ct. App. 1997) (culverts) and Beckwith v. County of Stanislaus, 345 P.2d 363, 367 (Cal. Ct. App. 1959) (bridges). Although local governments frequently contract with private service providers, those contracts for services are beyond the scope of this Article. Contracts for services involving a regional special district are considered at text accompanying notes 174–205, infra. Closely related to intergovernmental contracts for services, and with potential impact on regional well-being are intergovernmental boundary agreements. Under those intergovernmental contracts, two or more governmental units pledge to respect the others’ annexation authority in territory as allocated under the agreement. In Village of Long Grove v. Village of Kildeer, 497 N.E.2d 319, 321 (Ill. App. Ct. 1986), the court held that boundary agreements are not included within the scope of Illinois’ Intergovernmental Cooperation Act. Subsequently, the Illinois legislature specifically authorized the execution of intergovernmental boundary agreements. See Groenings v. City of St. Charles, 574 N.E.2d 1316, 1320 (Ill. App. Ct. 1991). This Article will not analyze this type of intergovernmental contract.

Intergovernmental Cooperation

management participation”¹²⁴ nor did it assume any “potential liability” from the operation of the bus system.¹²⁵ Although the Durango court upheld the legality of the intergovernmental contract, it did not discuss the underlying motivation for its formation. A wide range of explanations is plausible. It is possible, for instance, that the expanded service would not require the purchase of additional buses, or that the city could price the county service at a level that would produce a profit for city revenues, or that the enhanced service territory met the demand of city citizens for county-wide public transportation. On the other side of the bargain, the county may not have had the financial capability nor the political will to establish its own transportation service. In any event, its transfer of power to the city and its own limited involvement suggests that the county was confident in the city’s ability to protect the welfare of county residents.

A survey by the Advisory Commission on Intergovernmental Relations identified several local disincentives to intergovernmental contracts for services. For one thing, the receiving unit is likely to be wary of the potential loss of local autonomy and control that results from the relinquishment of its authority over the service production and provision.¹²⁶ The other major disincentive, according to the survey, is that local governments fear that their contracting partners will not apportion costs equitably.¹²⁷ Notwithstanding these perceived drawbacks, intergovernmental contracts for services are widely used across the country. The question remains, however, whether they enhance or decrease the general regional welfare.

Intergovernmental contracts for services would seem to provide an opportunity for maximized efficiency and overall enhancement of service delivery. In his important study of municipal incorporation in Los Angeles County, however, Gary Miller analyzed a series of intergovernmental contracts for services that had the opposite effect.¹²⁸ Known as the Lakewood Plan in honor of the first community to use an intergovernmental contract to avoid annexation by larger cities in the county, Los Angeles County became the provider of many services to newly incorporated municipalities. Citizens living in the unincorporated areas of Los Angeles County had three options—to remain in unincorporated county territory, which typically meant no local control

¹²⁴ Id. at 53.
¹²⁵ Id.
¹²⁶ See ACIR, supra note 20, at 39–40.
¹²⁷ Id. at 40.
¹²⁸ See MILLER, supra note 39.
over services or land use; to be annexed by the existing cities; or to incorporate as separate municipalities. 129 Before the Lakewood Plan, the costs of incorporation were substantial, requiring the new municipality to establish departments to provide police, fire, and other basic urban services. However, Los Angeles County, was willing to contract with newly incorporated municipalities to continue providing the services these communities had received from the county prior to their incorporation. 130 With the problem of providing municipal services solved, and with the attractive opportunity to form communities whose property tax revenues would now be kept largely within the community rather than redistributed throughout the county, numerous incorporations occurred, typically of wealthy, homogeneous residential areas. In fact, not only did the Lakewood Plan stop the redistribution of revenues that had occurred prior to incorporation, Miller’s analysis showed that in fact the county priced the services at such a low level that non-Lakewood Plan parts of the county were providing a substantial subsidy to the Lakewood Plan communities. 131

129. See id. at 21.

130. See id. at 10–22.

131. Miller claims that the county charged Lakewood Plan cities an artificially low rate and then spread the total cost across the entire county. See id. at 22–26. To some extent, this phenomenon occurs naturally in most U.S. counties, because of the ways in which counties service two distinct populations. First, they provide county-wide services, such as the judicial system, to all county residents. In addition, they serve as the general service provider to the unincorporated areas of the county. If the county tax rate is equal for all county residents, city residents, who do not receive many county services, subsidize county residents. In the Lakewood Plan, this subsidization extended to contract cities, for which the price of services was typically set at pre-incorporation levels. In fact, a private study suggested that the subsidization of contract cities exceeded $5 million a year for police services alone. Id. at 24. Miller suggests that the county was in fact eager to set its prices at this low rate to cement its position as the dominant service provider and to guarantee the preservation of a large bureaucracy. See id. at 22–23. The real losers, of course, were the cities in Lakewood County that provide their own services while paying taxes to the county. Because county officials are elected by all county voters, it may well be that the non-Lakewood Plan cities did not have a strong enough political voice to end the subsidy. Of course, the county may have had other, more benign, reasons to set the contract price so low. For one thing, the county may have doubted its continued ability to serve its own residents if its territorial base were to lose the population of the newly incorporated areas. That loss might have brought the county’s service base beneath the critical mass necessary to provide services on an efficient level. As a result, the county may have concluded that it was ultimately to the benefit of its citizens in the unincorporated area to provide service to nonresidents, even when the contract price imposed a loss on the provider.

The Supreme Court of Iowa appears to have prohibited a “Lakewood-like” arrangement proposed by a group of citizens seeking to incorporate in Citizens of Rising Sun v. Rising Sun City Development Commission, 528 N.W.2d 597, 600 (Iowa 1995). In that case, the proponents of incorporation asserted that the newly incorporated entity would contract with all pre-existing service providers to obtain all required municipal services. In its affirmation of the denial of the incorporation petition, the court concluded that petitioners had not met their statutory burden of
Intergovernmental Cooperation

Rather than the win-win situation hypothesized earlier for intergovernmental contracts for services such as the one described in Durango, the Lakewood Plan suggests that not all contracts may be so benign. The Lakewood Plan allowed wealthy enclaves to incorporate without having to face the costs associated with producing services for their constituents. At the same time, the act of municipal incorporation allowed them to capture their high property value for their own municipal taxation purposes, while artificially reducing the costs they would incur to provide the services their residents demanded. Though the Lakewood Plan communities were undoubtedly better off because of the high tax revenues and decreased service needs that resulted from their incorporation, not all segments of the area were so fortunate. In fact, the county suffered the loss of substantial tax revenues and lost money in the provision of the contract for services, while cities in the county were deprived of the opportunity to annex the highly valued properties.

A second potential weakness of intergovernmental contracts for services is the reduction of the legal oversight and potential for legal redress that might otherwise be available to regulate government provision of services. Normally, when a local government provides services to individuals living outside its territorial borders, a range of judicial doctrines and statutory limits apply to regulate the provider’s relationship with the service recipient. Rates are frequently required to be “reasonable and uniform,” and courts may invalidate what they perceive as municipal attempts to gouge nonresident recipients of municipal services. When the nonresidents receive service pursuant to intergovernmental contract, however, these limitations do not apply, perhaps because of the assumed equal bargaining position of the two contracting units of local government. Similarly, although a municipality has no general duty to provide services to nonresidents, it may have such showing that an incorporating entity has the ability “to provide customary municipal services.” Id. at 602.

132. See id. at 20–22.

133. See id. at 100.

134. See, e.g., Township of Aston v. S.W. Del. County Mun. Auth., 535 A.2d 725, 728 (Pa. Commw. Ct. 1985) (quoting state requirements that a municipality’s water rates for individual nonresident customers must be “reasonable and uniform”); Platt v. Town of Torrey, 949 P.2d 325, 311 (Utah 1997) (ruling town may charge nonresidents a higher rate, but there must be a reasonable basis for difference—court also notes that reasonableness standard applies to city’s provision of services to its residents as well).

135. See City of Texarkana v. Wiggins, 246 S.W.2d 612, 628 (Tex. 1952).

a duty in particular instances. Under the “holding out” doctrine, courts have determined that when a municipality agrees to provide a particular service to nonresidents, it will have impliedly assumed a duty to provide service to all similarly situated nonresidents.\footnote{137} The holding out exception also appears inapplicable to intergovernmental contracts that establish exclusive service territories or condition nonresident receipt of municipal service on annexation to the providing municipality.\footnote{138}

In addition, under the law of many states, municipal service providers are exempt from the regulatory regime that would apply to private providers of utilities and transportation services. This exemption often extends to a municipality’s provision of services to nonresidents pursuant to an intergovernmental contract for services.\footnote{139} It is true that lawsuits brought by frustrated private sector competitors who seek to provide the same services may provide some check against governmental abuse in this context.\footnote{140} If the rationale for the lack of regulatory oversight over municipal utilities is that the political process operates to ensure the fair treatment of citizens, however, it is not immediately obvious how the intergovernmental nature of the contract removes the need for regulatory oversight of a government’s agreement to provide services to nonresidents.

And finally, to the extent that Professor Frug’s critique of localism offers a persuasive rearticulation of the role of local governments as community builders,\footnote{141} any decision by a local government to divest

\footnotesize{\begin{flushleft}
137. For brief descriptions of the holding out exception to the rule that municipal governments have no duty to extend services to non-residents, see, e.g., Town of Rocky Mount v. Wenco of Danville, 506 S.E.2d 17, 20–21 (Va. 1999) and Barbaccia v. County of Santa Clara, 451 F. Supp. 260, 264 n.2 (N.D. Cal. 1976).


139. See City of Durango v. Durango Transp. Inc., 807 P.2d 1152, 1160 (Colo. 1991) (Public Utilities Commission has no jurisdiction over municipally owned bus service providing extraterritorial service pursuant to intergovernmental contract). Contra Texas Water Comm’n v. City of Fort Worth, 875 S.W.2d 332, 355 (Tex. Ct. App. 1994) (holding that state commission has jurisdiction to review rates charged by city to another city pursuant to intergovernmental contract for sewer services, notwithstanding a provision in state Interlocal Cooperation Act that the power to enter an intergovernmental contract “prevails over a limitation in any other law.” TEX. GOV’T CODE ANN. § 791.026 (Vernon 1994)).

140. See, e.g., Durango Transp. Inc., 807 P.2d at 1152; Lockheed Info. Mgmt. Servs. Co. v. City of Inglewood, 948 P.2d 943, 949–51 (Cal. 1998) (rejecting the challenge of a company that bid on city contract for the processing of parking tickets to the city’s decision to award the contract to another municipality).

141. For a full development of the argument that cities should focus on the development of strong personal ties between citizens and government and among citizens generally, see Frug, supra note 34, at 1081–1107 (1996).
\end{flushleft}}
Intergovernmental Cooperation

itself of control over the provision of services only contributes to the government’s own further disempowerment and increasing irrelevancy in the eyes of its citizens.\textsuperscript{142} Though Frug was referring specifically to the way in which special districts exercise power over city services and residents, syphoning off what would otherwise be important municipal functions, his critique applies equally to the self-inflicted divestment that occurs when local governments voluntarily transfer power to other co-equal units of government: “Power and participation are inextricably linked: a sense of powerlessness tends to produce apathy rather than participation, while the existence of power encourages those able to participate in its exercise to do so.”\textsuperscript{143} By removing responsibility from the general purpose government, intergovernmental contracts for services may reinforce citizen opinions of local government irrelevance and contribute to the widespread lack of interest in municipal affairs as expressed in the low (and generally decreasing) rates of voter participation at the local level. Moreover, following along the lines of a related Frug critique, the emphasis on municipal services and their severance from other important local government functions operates to the detriment of the role of government as community builder and enhances the “consumer-oriented vision”\textsuperscript{144} of local government. And with that vision comes a fundamental shift in citizen mentality, in which citizens choose their municipalities in the same way they choose their television sets, by evaluating where their dollars will go the farthest and where they will maximize the “get what you pay for” assessment of local government services.\textsuperscript{145}

Though the above critique may temper the enthusiasm with which intergovernmental contracts are embraced by state legislatures and judiciaries, the New Regionalist might be willing to tolerate the drawbacks in exchange for the enhanced regional equity the contracts are said to promote. On closer inspection, however, several potential anti-regional impacts can be identified. For one thing, as others have noted, because intergovernmental contracts for services typically encompass only one subpart of a metropolitan region, the well known prisoners’ dilemma makes it unlikely that an intergovernmental contract for services will be used to promote regionalization.\textsuperscript{146}

\begin{itemize}
  \item \textsuperscript{142} See Frug, supra note 29, at 1065.
  \item \textsuperscript{143} See id. at 1070.
  \item \textsuperscript{144} See Frug, supra note 33, at 31.
  \item \textsuperscript{145} See id. at 29.
  \item \textsuperscript{146} See Cashin, supra note 5, at 1988. As she explains: “the classic collective action problem occurs when the interest of each individual is too small relative to the costs of participation to justify
Second, the New Regionalist critique would do well to analyze the likelihood that the providing entity in an intergovernmental contract will promote the welfare of its newly enhanced territory, much less the welfare of the entire region of which it is a part. In that vein, consider the Durango intergovernmental contract.\footnote{\textsuperscript{147}} In an earlier Colorado Supreme Court opinion in that litigation, which had denied the applicability of state regulatory oversight of the intergovernmental contract at issue in Durango, the court noted that the city’s proposed county bus service would provide transportation to nearby ski slopes and to the area’s airport.\footnote{\textsuperscript{148}} Although the opinion did not mention whether the frustrated competitor had provided more extensive county-wide bus service, it is at least possible that these two presumably more profitable lines subsidized its provision of transportation services to other parts of the county. The city’s incentive to use the profitable routes for subsidization, of course, is quite different than that of the private provider. For the city, the profit derived from the ski slopes and airport routes are more likely to be used to subsidize intra-city bus service; the city would seem to have no incentive to provide county service more widely unless its own residents would profit from that enhanced service.\footnote{\textsuperscript{149}} Though this calculation may be perfectly sensible from the city’s own self-interested vantage point and the welfare of its citizens, the broader welfare of county residents appears to go unprotected.

Similarly, consider the intergovernmental contract for water and sewer services described by the Tenth Circuit in an unpublished opinion, his or her participation, so there is no incentive to take individual action. The collective interest of all the individuals combined may be very great, but because they are separated from each other, they either do not see the benefits of organizing or are unable to organize.” \textit{Id.} at n.9. \textit{See also} Briffault, \textit{supra} note 2, at 1147; Harrison, \textit{supra} note 75, at 147–49. \textit{But see} Gillette, \textit{supra} note 18, at 246–50 (arguing that prisoners’ dilemma may not apply to some intergovernmental burden sharing).

\textsuperscript{147}. For a description of the intergovernmental contract in that case, see \textit{supra} notes 123–125 and accompanying text.


\textsuperscript{149}. If the city residents and the county residents have the same interests, the political process, by protecting the residents’ interests, will also protect the interests of nonresidents. Although in some intergovernmental contracts for service that congruence may exist, the circumstances of \textit{Durango} are unlikely to be such a case. With regard to items such as route designation and scheduling, city and county interests are probably markedly different. In a different context, Professor Clayton Gillette has noted that residents may sometimes be satisfactory defenders of nonresident interests. Gillette gave the following example in a discussion of municipal taxation of nonresidents: “shopowners who fear that nonresident customers will gravitate to suburban shopping malls, or employers of nonresidents who fear that they will have to increase wages or lose nonresident employees to suburban employers, should serve as pretty good proxies for those who initially would bear the incidence of the tax.” \textit{See} Gillette, \textit{supra} note 18, at 210.
Haik v. Town of Alta. \(^{150}\) In that case, the Town of Alta refused to issue a building permit for the development of several lots near the local ski resort. As a prerequisite to the permit, the property owners had to obtain sewer and water services from Salt Lake, the service provider to Alta. Pursuant to an intergovernmental contract between Alta and the City of Salt Lake, however, decisions about the extension of those services were within Salt Lake’s exclusive control. \(^{151}\) When the landowner sued Alta to compel the extension of services, the court upheld the town’s right to rely on the terms of the intergovernmental contract. \(^{152}\) Because the court was not concerned with the potential regional impact of this contract, its opinion did not evaluate whether Salt Lake City could be expected to protect the welfare of the territory beyond its borders that formed the basis of the intergovernmental contract. \(^{153}\) In any dispute, however, the city is likely, and indeed should be expected, to protect the interests of its residents over the interests of other parties. \(^{154}\)

It seems unrealistic to expect that regionwide regulatory power can be exercised fairly by a governmental entity that makes up only a part of the region being regulated. Political reality and the check of the ballot box appear to guarantee that elected officials will exercise their duties to the region through the lens of their constituents’ self-interest. Intergovernmental contracts cannot transform a municipal entity into a regional entity; they merely enlarge the service area of the provider, leaving intact the preexisting representational allegiance to the subunit the provider represents. Contracts for services between general purpose local governments, then, do not create a mechanism that will bring a regional approach to the evaluation of regional problems. In the best of situations, the regional welfare will not suffer. In many others, however, the potential for elevation of the provider’s welfare at the expense of the broader region is likely, and perhaps inevitable.

Finally, intergovernmental contracts for services can be faulted for

---


151. _Id._ at *1.

152. _Id._ at *4.

153. Environmentalists have long argued for a more ecologically based distribution of regulatory power over water resources, thus making the watershed an attractive territorial border. See generally R.W. Adler, _Addressing Barriers to Watershed Protection_, 25 _Envtl. Law._ 973 (1995). It is doubtful, however, that Salt Lake City would be the proper entity to control that larger territory. In any event, the agreement between Salt Lake and Alta did not encompass the entire watershed, but merely extended Salt Lake’s power beyond its territorial limits to cover Alta and its environs.

154. See Briffault, _supra_ note 25, at 28; Briffault, _supra_ note 29, at 447–54. So long as state law authorizes local governments to undertake actions that impose costs on or ignore the regional welfare, those actions are likely to continue.
allowing the realization of several anti-regional incentives. By offering alternatives to service production, municipalities are better equipped to avoid annexation and consolidation into broader regional units. In reality, the potential for intergovernmental contracts for services is just one card in a deck that is already solidly stacked against regionalization in metropolitan areas. So long as the rules of annexation, incorporation, and property tax distribution allow the formation of self-contained, autonomous units within a metropolitan region, intergovernmental contracts for services will remain an attractive option for local government entities.

2. Joint Provision of Services

As its name suggests, and in contrast to the contracts for services described above, a joint services agreement requires that all government participants be involved in both the production and the provision of the service that is the subject of the agreement. Just as with contracts for services, however, the categorization is not rigid. In some instances, joint services agreements share characteristics with contracts for services. Similarly, the distinction between a joint services agreement and the creation of a new entity may blur. In a Nebraska case, for instance, the court was called upon to resolve an intergovernmental dispute arising out of a joint agreement that established a civil defense agency for a county and a municipality within its borders. Though the agency was in some ways an independent entity, both government participants retained considerable power and ongoing involvement in the agency. In fact, it was precisely this ambiguity that gave rise to the lawsuit, when the county purported to exercise its power as employer to fire the director without the city’s approval. The court noted that the intergovernmental

155. See PPC Enters., Inc. v. Texas City, 76 F. Supp. 2d 750 (S.D. Tex. 1999). In this case, two Texas cities passed identical fireworks bans followed by a mutual agreement that one of the cities would enforce the ordinance in the other’s extraterritorial jurisdiction. The court upheld the interlocal agreement against challenges that it violated the due process and equal protection clauses of the federal constitution, that it was unconstitutionally vague, and that it exceeded the cities’ power under the state’s Interlocal Cooperation Act, TEX. LOC. GOV’T CODE, § 217.042 (Vernon 1998). See 76 F. Supp. 2d at 755–59. For an example of an agreement that actually shows features of all three categories of intergovernmental cooperation, see Murphy v. City of Topeka-Shawnee County Department of Labor Services, 630 P.2d 186 (Kan. Ct. App. 1981), in which a city and county entered a “Cooperative Agreement for the Administration of Employment and Training Services” pursuant to which the county transferred significant authority to the city, yet at the same time retained important powers over program evaluations and funding. Id. at 189. The agreement also created a joint City-County joint department, which the court described as “solely an agency of the city.” Id.

156. Creation of a new entity is discussed infra Part A.3.
agreement provided that employees of the agency served “during the pleasure” of the county and of the city. As a result, it upheld the county’s prerogative to act individually to terminate the agency’s director without the city’s concurrence.\textsuperscript{157} The active governmental participation and retention of government authority displayed in this case is a salient feature of joint services agreements; though precise categorization may be impossible, some generalizations do appear.\textsuperscript{158}

At the risk of oversimplification, and in light of the above caveat, joint services agreements most typically involve one of two fact patterns. Most common of all, perhaps, are the many mutual aid provisions executed by neighboring municipal governments. Pursuant to these agreements, local governments cross-cede extra-jurisdictional enforcement powers to police, fire, and emergency service providers. Mutual aid agreements are typically designed to allow for emergency back-ups, or to authorize a municipal officer to complete a traffic or criminal chase begun in his or her own territory.\textsuperscript{159} Courts are usually

\textsuperscript{157} Heinzman v. Hall County, 328 N.W.2d 764, 768–69 (Neb. 1983). For an example of other intergovernmental agreements that display characteristics of a joint services agreement as well as involving intergovernmental creation of a new entity, see, e.g., Borough of Lewistown v. Pennsylvania Labor Relations Board, 735 A.2d 1240 (Pa. 1999). In this unusual case, three municipal governments created a regional police department. Pursuant to the intergovernmental agreement establishing it, the police department was under the control of an appointed board of directors. The municipalities delegated to the board “all the functions, powers and responsibilities which the municipalities respectively have with respect to the operation, management, and administration of a municipal police department,” \textit{id.} at 1241 n.3. In resolving the labor dispute before it, the court held that the constituent members of the regional department were also deemed to be the employers of the complaining police officers. \textit{See id.} at 1244–45. \textit{See also} City of Oakland v. Williams, 103 P.2d 168, 172 (Cal. 1940) (upholding joint action taken by seven municipalities to plan for joint sewage disposal services; agreement created joint agency and then transferred all decisionmaking authority to one of signatory municipalities); Magnon v. Acadian Metro. Code Auth., 413 So.2d 972, 973 (La. Ct. App. 1982) (joint services agreement for licensing contractors resulted in creation of intergovernmental public corporate body, the Metropolitan Code Authority); Ky-Ind. Mun. Power Ass’n v. Pub. Serv. Co. of Ind., Inc., 393 N.E.2d 776, 779 (Ind. Ct. App. 1979) (joint agreement created an agency that was an “instrumentality of the cities,” indicating ongoing municipal involvement and control over the agency, notwithstanding its corporate existence as a separate legal entity).

\textsuperscript{158} In terms of frequency, the ACIR study on intergovernmental cooperative efforts noted that local governments tend to use joint service agreements for the same types of services, for the same reasons, and with approximately the same frequency as they use contracts for services. \textit{See} ACIR, \textit{supra} note 20, at 30–35. The survey did note, however, that local officials evaluated contracts for services more positively than joint service arrangements. \textit{Id.} at 43. Unfortunately, it did not indicate the basis of that more favorable assessment.

quite deferential to the goal of intergovernmental cooperation in this context and routinely reject challenges to official actions taken pursuant to these agreements.\footnote{In fact, one New Jersey court upheld the validity of a search and seizure conducted by one police officer within the jurisdiction of another city, even though the two municipalities operated under an informal, unwritten policy of mutual aid. See State v. Montalvo, 655 A.2d 476, 480 (N.J. Super. Ct. App. Div. 1995). See also Commonwealth v. Mays, 431 A.2d 322 (Pa. Sup. Ct. 1981) (rejecting challenge to police officer’s extraterritorial arrest that had not followed mutual aid agreement provision that police assistance would be provided to neighboring units “upon proper request” by one of the units). But see State v. Allen, 790 So.2d 1122, 1124–25 (Fla. Dist. Ct. App. 2001) (invalidating seizure of marijuana by city police officer at county residence; quoting mutual aid agreement provision that city police “shall not routinely commence investigations of crimes or actively seek out criminal activity outside of their primary jurisdiction,” and stressing that court would not allow mutual aid agreements to expand police power jurisdiction).}

The second common type of joint services agreement, broader in its scope than the typical mutual aid agreement, facilitates the implementation of a jointly articulated program or regulatory goal. Consider, for instance, the facts of \textit{State v. Plaggemeier},\footnote{93 Wash. App. 472, 969 P.2d 519 (1999).} which involved a five-party intergovernmental contract to implement a county-wide drunk driving effort. In addition to the agreement’s mutual aid provision,\footnote{Id. at 474–75, 969 P.2d at 521. The court actually invalidated that part of the agreement, not because it was beyond the scope of intergovernmental cooperation, but rather because the legislative bodies of the participating entities had not ratified the mutual endeavor. \textit{Id.} at 481–82, 969 P.2d at 524. Under the state’s Mutual Aid Act, the individual sheriff or chief of police can enter binding mutual aid agreements with neighboring municipalities, see WASH. REV. CODE § 10.93.070 (2002). The more general Interlocal Cooperation Act, WASH. REV. CODE § 39.34 (2000), however, requires that intergovernmental cooperative efforts receive ratification by the governing legislative bodies.} it established a joint task force and other coordinating mechanisms.\footnote{In Teamsters Union Local 117 v. Port of Seattle, No. 36366-2-I, 1996 WL 523973 (Wash. Ct. App. 1996) (unpublished opinion), the Port of Seattle (owner of the Seattle-Tacoma Airport) leased property to the city of SeaTac for municipal operation of a park. \textit{Id.} at *1. The joint agreement established “concurrent law enforcement jurisdiction” in the park, but specified that emergency 911 services would be provided by the city, whereas the Port would “retain primary responsibility for all law enforcement related to airport operations.” \textit{Id.}} Other common types of joint services agreements involve some division of labor between the participating governments based on the expertise or specialized interests of the different units. An agreement between the Port of Seattle and the City of SeaTac, for instance, authorized joint provision of some services while transferring authority over other services to one of the two participating units of government.\footnote{In \textit{State v. Plaggemeier}, 93 Wash. App. 472, 969 P.2d 519 (1999), the court used the mutual aid agreement to uphold the arrest of the defendant by a police officer acting outside his city limits. See \textit{id.} at 483, 969 P.2d at 525.}
In both of these cases, units of local government were able to advance their own constituency’s health, safety, and welfare by acting jointly to articulate and implement common policies.

Because joint services agreements may blur or eliminate the lines between provider and recipient of services, difficult interpretative questions may arise about the scope of the powers being shared by the participating entities. A comparison of two judicial opinions nicely illustrates the analytical problem. In *In re Condemnation of 30.60 Acres of Land*, the court upheld the joint condemnation of land for construction of a school and public park facility pursuant to an agreement between a township and a school district. Under applicable state law, the township had condemnation power exclusively for park purposes, while the school district was authorized to condemn land only for use as a public school. Focusing on the fact that each local governmental unit possessed condemnation power, the court found that the agreement was lawful. It rejected the landowner’s argument that, because the school district has no power to engage in park purposes and the township cannot build schools, the agreement had improperly enhanced each entity’s government powers through the subterfuge of an intergovernmental contract.

In contrast, the Supreme Court of Nebraska reached the opposite result in *Gallagher v. City of Omaha*, when it invalidated a similar joint agreement. In that case, the city and a state university negotiated an intergovernmental use agreement whereby the city authorized the university to construct and use parking facilities on city park land. During the week, from 7 a.m. until 10:00 p.m., the facilities were to be used exclusively by the university; at all other times, the usage was joint. Although the court noted that the city was authorized to provide parking for its park patrons, it concluded that the interlocal agreement constituted an improper diversion of park lands. Thus, for the *30.60 Acres* court, the joint exercise of powers could extend to include actions that only one of the units could have lawfully undertaken while acting individually. In contrast, the *Gallagher* court implicitly rejected that argument and

---

165. This discussion patterns the “mutuality of powers” vs. “power of one unit” debate in intergovernmental contracts for services, see supra notes 116–117, but the courts do not appear to use those labels to delineate the disputes in the context of joint services agreements.
167. See id. at 245.
168. See id. at 244.
169. 204 N.W.2d 157 (Neb. 1973).
170. Id. at 160–62.
invalidated the joint agreement because, though all actions were within the scope of power of one of the contracting entities, neither unit had the power to engage in all of the activities undertaken pursuant to the contract’s terms.

The decision to enter a joint services agreement requires a commitment to a formal, ongoing relationship with another unit of local government. To a lesser extent than in the typical contract for services, governments entering joint services agreement assume a certain risk of incompatibility in the execution of the agreement. Presumably, each unit has satisfied itself that its partner’s policies are consistent with its own; otherwise, the mutual provision of a service or enforcement of extraterritorial powers runs the risk that the “foreign” entity will violate the other’s policies. Once the agreement is operational, moreover, all government participants are actively involved in the ongoing implementation of its terms.171

Although they may be more cumbersome to implement and execute, joint services agreements may offer several attractive qualities to local governments. First, and in contrast to contracts for services, they are less likely to involve a relinquishment of local police power or other governmental autonomy, because each entity is an active participant in the provision of the service or in the implementation of the policy under contract. In addition, the ongoing involvement, though it may be more costly, has the added benefit of making it more likely that each unit will be able to ensure compliance with its own goals and interests. Again in

171. This continuous involvement of multiple entities of government may create difficult questions regarding legal liability. For example, the nature of each entity’s involvement in the joint agreement may have important implications for tort immunity or workers compensation issues. See, e.g., Drain v. Galveston County, 999 F. Supp. 929, 937 (S.D. Tex. 1998) (city may be liable for actions of county employee who responded pursuant to mutual aid agreement); Hauber v. County of Yakima, 107 Wash. App. 437, 440, 27 P.3d 257, 260 (2001) (extent of recovery by widow of firefighter depends on whether deceased was acting pursuant to normal firefighter duties, a mutual aid agreement, or as a member of a special search and rescue team); Ewing v. State, 757 So.2d 843, 847 (La. Ct. App. 2000) (holding joint agreement to pay for and cover drainage ditch insufficient to impose tort liability on contributing entities); N.H. Ins. Co. v. City of Madera, 192 Cal. Rpt. 548, 550 (Cal. Ct. App. 1983) (holding city immune while engaged in fighting fires extends to county firefighters who responded to alarm pursuant to mutual aid agreement); Garcia v. City of South Tucson, 640 P.2d 1117, 1120 (Ariz. Ct. App. 1981) (city police officer injured while responding to mutual aid agreement request not limited to workers compensation relief because mutual aid agreement does not transform requesting city into statutory employer of injured officer); Murphy v. City of Topeka-Shawnee County Dep’t of Labor Servs., 630 P.2d 186, 189, 192–93 (Kan. App. 1981) (determining program administered solely by city, though jointly sponsored by city and county may create county liability under Workmen’s Compensation Act for alleged retaliatory discharge); Lauria v. Borough of Ridgefield, 291 A.2d 155, 158 (N.J. Super. Ct. App. Div. 1972) (only holding responding municipality to be liable to estate of firefighter killed during fire in summoning municipality pursuant to mutual aid agreement).
Intergovernmental Cooperation

contrast to contracts for services, and because of the cooperative implementation of the agreement, the danger of subversion of the receiving entity’s interests is minimized.

From a regional standpoint, a review of the case law and the literature suggests that joint services agreements have not been used to provide a truly regional solution to metropolitan area problems. Rather, and perhaps because of the need for ongoing monitoring and involvement, the agreements tend to be narrowly defined and limited. Generally, joint services agreements involve, not an attempt to implement a grand joint vision for improving the overall welfare of the constituents of all participating governments, but rather a carefully delineated description of specific ways in which joint action will complement pre-existing local government policy. In terms of the New Regionalists’ hopes for regionwide articulation of a common agenda, however, joint services agreements offer one potential advantage over contracts for services. As Professor Cashin has noted, “locational sorting” on the basis of race and socioeconomic class is extremely strong in metropolitan areas. To the extent that joint services agreements require meaningful cross-boundary collaboration and joint exercise of powers, they may create an opportunity for highly segregated local governments to become acquainted with “the other.”

3. Creation of a New Government Entity

Despite the anti-regionalist ethic that pervades local government law, the United States has no shortage of regionwide governmental units. With a few notable exceptions, however, these are not general purpose,

172. See Cashin, supra note 5, at 2016.

173. Professor Cashin describes how the desire for lower taxes, avoidance of redistribution of tax dollars, and racial exclusionary intent are important catalysts for “sectoral segregation” in metropolitan areas. Id. at 2016. In his analysis of that phenomenon, Professor Frug used the metaphor of local governments as exclusive country clubs, where residents equate taxes with dues. See Frug, supra note 33, at 29. These critiques are based on the reality that territorial boundaries in metropolitan America, notwithstanding the frequency with which they are crossed, are nevertheless quite effective in establishing the “us” vs. “them” mentality at the level of local government.

174. The Minneapolis area’s Met Council, Portland’s Metropolitan Service District, and Seattle’s Metro are examples of single purpose regional districts that have evolved into government units with a broader mandate. For a discussion of the history and scope of operation of these entities, see
or even multi-purpose, elected entities. Rather, the well documented government growth industry across the country consists of single purpose governments, generically known as special districts or public authorities. While the numbers of other local government units remains more or less constant, special districts continue to increase.

Special districts vary greatly in terms of the territory they service: while some are sub-local in scope, others are coterminous with existing

VALENTE ET AL., supra note 29, at 425–66. See also Briffault, supra note 2, at 1118–19; Cashin, supra note 5, at 2028–29.

175. Although the special district is but one type of single purpose local government unit, the Article uses the terms interchangeably.

176. Regional special districts are sometimes referred to as “quasi-municipal corporations;” though in some ways they act like municipalities, their powers may be quite narrowly focused. Special districts may have a range of revenue raising powers, see, e.g., ILL. COMP. STAT. 19/160.1 (1996); KAN. STAT. ANN. §§ 19-2769-70 (1995), the ability to condemn property, see, e.g., WASH. REV. CODE § 14.07.020 (1993); Cherokee Village Homeowners Protective Ass’n v. Cherokee Village Road & Street Improvement District No. 1, 455 S.W.2d 93, 95 (Ark. 1970), and the power to enact and enforce ordinances, Village of Glencoe v. Metro. Sanitary District of Greater Chicago, 320 N.E. 2d 524, 528–29 (Ill. App. Ct. 1974).

The legal status of regional special districts has been litigated in a variety of contexts. If, for instance, a regional special district is deemed to be a municipal corporation, it will not be entitled to assert the state’s sovereign immunity; if it constitutes an instrumentality of the state, in contrast, sovereign immunity will attach. See, e.g., Calvert Invs. v. Louisville & Jefferson County Metro. Sewer Dist., 805 S.W.2d 133, 138–39 (Ky. 1991) (regional special district, as a municipal corporation, is liable under state tort law). The scope of potential liability may also depend on its characterization. For instance, a Texas court concluded that a regional transit authority performed essential public functions, and thus was protected by a statutory damages cap available to governmental, but not proprietary, activities undertaken by local governments. Salvatierra v. Via Metro. Transit Auth., 974 S.W.2d 179, 183 (Tex. Ct. App. 1998). See also Scott v. Shapiro, 339 A.2d 597, 599 (Pa. 1975) (Southern Pennsylvania Transportation Authority not a state instrumentality and thus immune from state sunshine laws requiring disclosure and transparency in state proceedings); Fisher v. S.E. Pa. Transp. Auth., 431 A.2d 394, 397 (Pa. Commw. Ct. 1981) (Authority not subject to state law that provides full pay to state employees while they are on active duty with any reserve unit of the armed services).

177. Public authorities, which are single purpose governments designed to be led by professional, apolitical managers, became popular in the 1920s as part of reform movements intending to take power away from corrupt political bosses. These newly created entities, which typically have no taxing power but are able to borrow and issue bonds, took control over a number of public sector functions. They were seen as embodying the “values of efficiency, neutral competence, and professionalism in municipal affairs . . . .” Foster, supra note 24, at 18. The success of the Port Authority of New York (and later of New Jersey) spurred widespread adoption of this model of government. Id.

178. Between 1952 and 1997, the number of special purpose governments nearly tripled, rising from 12,340 to 34,683. See U.S. CENSUS BUREAU, 1997 CENSUS OF GOVERNMENTS, Volume I, Government Organization 4, 6 (2002). Over that same time period, subcounty general purpose local governments increased from 34,009 to 36,001. Id. at 4. In comparison to other units of local government, the proportion of spending by special districts has increased two to three times more quickly. See Foster, supra note 24, at 1–3.
municipal boundaries, and still others extend beyond the borders of existing municipal governments. Approximately eighty percent of all single purpose governments are created at the supra-municipal level, thus encompassing the territory of more than one general purpose local government. These regional special districts either serve some section of a county, an entire county, or multiple counties. In addition to the differences in their geographical scope, single purpose governments vary with regard to their revenue raising abilities: while some have general taxation powers, others are restricted to user fees or other more specific revenue raising devices. A similar range of variation occurs in their manner of creation. Some are formed directly by state legislation, some by petition of citizens, and others by voluntary local government action. Because this Article focuses on the regional impact of intergovernmental cooperation, it does not specifically evaluate regional special districts created by the state. Much of the following critique,

179. The remaining 20% are either co-terminous with municipal borders or cover a section of a municipal government. See Foster, supra note 24, at 123.

180. Special districts exercise a wide range of revenue raising powers, and the scope of those powers appear unaffected by whether the district has appointed or elected commissioners. See note 186, infra. Many special districts whose commissioners are appointed exercise general powers of taxation. Courts have generally approved of this combination. See, e.g., Solomon v. N. Shore Sanitary Dist., 269 N.E.2d 457, 464 (Ill. 1971); Madison Metro. Sewerage Dist. v. Stein, 177 N.W.2d 131, 138–39 (Wis. 1970); Buffalo, Dawson, Mechanicsburg Sewer Comm’n v. Boggs, 470 N.E.2d 649, 651–53 (Ill. App. Ct. 1984). Some special districts are limited to recouping their operating costs through charging user fees. See, e.g., Beatty v. Metro. St. Louis Sewer Dist., 867 S.W.2d 217, 221 (Mo. 1993) (describing district’s power to collect user charges and per unit charges for capital improvements, but holding that district must get voter approval to raise the rates). Some are allowed to levy property taxes. See, e.g., City Brookfield v. Milwaukee Metro. Sewerage Dist., 491 N.W.2d 484, 492, 498 (Wis. 1992) (authorizing district to charge for capital costs on the basis of property value); WASH. REV. CODE § 70.44.060(6) (2002) (authorizing hospital district to levy a property tax “not to exceed fifty cents per thousand dollars of assessed value, and an additional annual tax on all taxable property within such public hospital district not to exceed twenty-five cents per thousand dollars of assessed value”). Other special districts impose general charges on property and activity within the district. See, e.g., Anema v. Transit Constr. Auth., 788 P.2d 1261, 1263 (Colo. 1990) (affirming district’s ability to finance planning of a transit system by imposing a charge on all commercial property and by levying an “employment assessment” on each employer). Still others can impose sales taxes. See, e.g., Camp v. Metro. Atlanta Rapid Transit Auth., 189 S.E.2d 56, 61 (Ga. 1972) (upholding legislative delegation of sales tax power to regional special district to fund rapid transit system).


183. See, e.g., Goreham v. Des Moines Metro. Area Solid Waste Agency, 179 N.W.2d 449, 452 (Iowa 1970). For a general survey of the ways in which regional special districts are created, see Foster, supra note 24, at 7–15.
However, may be relevant to an assessment of those state-created districts as well.\textsuperscript{184}

Many commentators have criticized single purpose governments as anti-democratic and anti-localist.\textsuperscript{185} These critics have noted, for instance, that special districts are unaccountable\textsuperscript{186} and invisible.\textsuperscript{187} They

\textsuperscript{184} In fact, some of the following critique of regional special districts may have less to do with their voluntariness and more to do with shortcomings of regional special districts in general. Nevertheless, because of the enthusiasm with which New Regionalists have turned to regional special districts in their search for regional redistribution, this Section will suggest a number of criticisms that may apply primarily to state-created regional special districts. In addition, it is important to keep in mind that every regional special district will not display all of the weaknesses identified in the following section.

\textsuperscript{185} See, e.g., Burns, supra note 39; Frug, supra note 15, at 1781–88; Gillette, supra note 18, at 204–06.

\textsuperscript{186} Commissioners or trustees of regional special districts are often appointed, either directly by the state or by elected officials of the government units that are located within the district’s territory. Many single purpose local governments are immune from the judicial doctrine that requires proportionally elected, rather than appointed, officials. As first established in Avery v. Midland County, 390 U.S. 474, 485–86 (1968), the one person, one vote principle applies to general purpose local government units. In a series of cases, the Court has articulated an exception to that principle to allow restrictions on or elimination of the franchise for “special limited purpose districts” whose actions disproportionately affect those who are assessed financial charges by the government unit. See Salyer Land Co. v. Tulare Lake Basin Water Storage Dist., 410 U.S. 719, 728 (1973); Ball v. James, 451 U.S. 355, 370 (1981). For a full discussion of the difficult legal questions surrounding the right to vote and limited purpose governments, see Richard Briffault, Who Rules at Home: One Person/One Vote and Local Governments, 60 U. Chi. L. Rev. 339 (1993); Richard Briffault, A Government For Our Time? Business Improvement Districts and Urban Governance, 99 Colum. L. Rev. 365, 434–46 (1999); see also Valente et al., supra note 29, at 81–147. So long as the government unit does not exercise general governance powers, states retain broad discretion over the manner in which special district commissioners are selected. Courts routinely reject challenges to appointed special district commissioners. See, e.g., City of Atlanta v. Metro, Atlanta Rapid Transit Auth., 636 F.2d 1084, 1089 (5th Cir. 1981) (holding one person, one vote principle not applicable to transit authority); Eastern v. Canty, 389 N.E.2d 1160, 1165–69 (Ill. 1979) (upholding method of appointing sanitary district commissioners; extent of county boards’ appointment powers determined by percentage of total assessed property each county had in district’s territory); Van Zanen v. Keydel, 280 N.W.2d 535, 539 (Mich. Ct. App. 1979) (upholding appointment of commissioners of metropolitan district with powers to establish and operate parks; one county-one commissioner allocation valid notwithstanding wide divergence of population among member counties). If a special purpose district acquires additional powers and expands its functions, however, at some point the norm of one person, one vote will apply. Cunningham v. Municipality of Metropolitan Seattle, 751 F. Supp. 885, 889–90 (W.D. Wash. 1990) traces the history of Seattle’s “metropolitan municipal corporation,” a local government entity (Metro) created in 1958 to provide regional sewer services and focus on the area’s urgent water quality problems. Pursuant to state law, members on Metro’s governing council were appointed by the legislative bodies of the participating government units. By the late 1980s, Metro’s functions had grown to include the provision of mass transit services and the development of regional land use planning policies. See id. at 890. In Cunningham the court found that one person, one vote applied to Metro and invalidated the manner of selection of its council members. As the unelected commissioners of many special districts are granted substantial control of regional services and revenues, and as many regional special districts are granted substantial taxing
are frequently immune from land use regulations and free to disregard local concerns. They may allow local governments to pass the buck on difficult policy questions, to avoid tort liability, and to avoid state authority, metropolitan area voters lose the ability to define and shape the important government services that affect their daily lives. See VALENTE ET AL., supra note 29, at 439–47.

187. Even when special district commissioners are elected, the elections can rarely attract more than a 5% voter turnout. See BURNS, supra note 39, 12–13 (1994). Burns also explains her claim that the invisibility of special districts has both positive and negative effects: “the benefits of special districts are that they can fund and provide services and infrastructure; they are able to get things done in a fragmented American polity. The difficulties are two: They do this while no one watches except interested developers, and they are gradually becoming the realm where much of the substance of local politics happens. Thus local politics becomes quiet, not necessarily through . . . consensus . . . , but rather through the invisibility of special district politics.” Id. at 117. Moreover, because their purpose is so narrowly defined, regional special districts frequently remain below the metropolitan area political radar screen. Even for the diligent citizen, identification of all local government units in a metropolitan area may prove daunting. See WEHNER, supra note 40, at 17–19 (summarizing several studies that describe the complexity of mapping local government units in metropolitan areas and noting the frequency with which citizens are unaware of local government structures that apply to them).

188. Coupling immunity from zoning laws with the narrow purpose of the special district’s mission creates potential for enormous conflict at the local level. See, e.g., Austin Indep. Sch. Dist. v. City of Sunset Valley, 502 S.W.2d 670, 675 (Tex. 1973) (determining regional school district able to build large football stadium and field house in violation of small town’s zoning ordinance); Evanston v. Reg’l Transp. Auth., 559 N.E.2d 899, 905 (Ill. App. Ct. 1990) (holding RTA’s proposal to construct a bus storage and maintenance facility immune from city zoning ordinances); City of Heath v. Licking County Reg’l Airport, 237 N.E.2d 173, 179 (Ohio Misc. 1967) (ruling airport authority’s proposal to widen runways immune from local zoning).

189. Transferring the responsibility for a contentious local issue to an appointed body may ease the political heat on elected officials. For example, in Barnes v. Department of Housing and Urban Development, 341 N.W.2d 766, 768 (Iowa 1984), the court essentially refused to let participating municipalities transfer decisionmaking responsibility over the siting of public housing to a regional housing authority they had created pursuant to the state’s intergovernmental cooperation statute. Because the cities themselves would have been unable to approve a low-income housing project without city council approval, the housing authority was similarly limited. Id. at 768. Thus, the court concluded that the housing authority’s decisions required approval by the city councils of all participating entities. Id. This opinion seems inconsistent with two other Iowa Supreme Court interpretations of the status of government entities created pursuant to intergovernmental cooperation. In Goreham v. Des Moines Metropolitan Area Solid Waste Agency, 179 N.W.2d 449, 455 (Iowa 1970), the court upheld the creation of a regional solid waste district against challenges that the participating local governments had improperly delegated power to that agency. Similarly, in Allis-Chalmers Corp. v. Emmet County Council of Governments, 355 N.W.2d 586, 589 (Iowa 1984), the court recognized that the county council of governments had broad powers, including [T]he right to acquire and dispose of property, the right to enter contracts, the right to operate a solid waste disposal and collection service, . . . the right to fix and charge fees for its services, the right to establish a budgeting system . . . , the right to borrow money and issue bonds, the right to provide for remedies in the event of default . . . .

Id. at 589. Presumably, if the city itself were to exercise those powers, many, if not all of the actions would require city council action. It is perhaps no coincidence that the court was less deferential to the regional housing authority’s claimed power to site low income housing; the Barnes case, ironically, represents a rare example of a group of municipalities joining together to provide a social
limitations on borrowing and spending. Though touted as a way to reduce the cost of service provision, services provided by special districts may actually be more expensive than those produced and provided directly by the general purpose government itself. And for the

service on a regionwide basis. It appears that some state courts may share the anti-regional bias of their citizens when it comes to municipal attempts to provide a regionwide solution to difficult social problems. Or it may be that the court was reluctant to let elected municipal officials avoid the political accountability that comes with being forced to take a position on contentious issues such as public housing. In general, though, the unaccountability phenomenon may be doubly perverse. General purpose governments become less accountable to their own constituency by transferring power to local government units that themselves are frequently unaccountable.

190. Parties injured by negligent actions taken by regional special districts or who seek to challenge other aspects of the district’s operations are likely to be limited to the intergovernmental entity as a source of recovery. As a separately organized governmental body, it is a distinct unit of government. Courts have been reluctant to impose liability on the constituent members of the special district. See, e.g., Allis-Chalmers Corp., 355 N.W.2d at 591 (government creators of joint county council not liable on suit for breach of contract filed against county council).

191. In essence, regional special districts may offer an attractive way for a local government unit to circumvent taxing, borrowing, or spending limits. Creation of a new unit of government brings with it an untapped line of taxing, borrowing, and spending opportunities. See Burns, supra note 39, at 16, in which she described the common wisdom that formation of special districts is “largely a technical financing maneuver.” Though Burns recognizes the strength of that dynamic, she believes that it is an incomplete explanation, because it ignores the real, political struggle that frequently underlies the formation of special districts. Id.; see also Foster, supra note 24, at 16–17. Foster also describes how, in the state of New York, public authorities are the only units of local government able to issue revenue bonds that are not backed by the full faith and credit of the lending unit. Since all other units of government have strict limits on the amount of full faith and credit debt they can incur, public authorities become an even more attractive mechanism for financing the construction of capital infrastructure. See id. at 111. Courts typically reject arguments that a special district was created for the sole purpose of evading other provisions of state laws. See, e.g., Rider v. City of San Diego, 959 P.2d 347, 350–51 (Cal. 1998) (rejecting plaintiffs’ arguments that special district created to finance a new convention center was no more than a “hollow shell that exists only on paper,” noting that the newly created district was run, financed, and managed by the city that created it solely for the purpose of avoiding restrictions on municipal debt); Johnson v. Piedmont Mun. Power Agency, 287 S.E.2d 476, 480–81, 484 (S.C. 1982) (holding state constitutional requirement for voter approval of local government debt not applicable to intergovernmental agency; while dissenting Judge Littlejohn characterizes the intergovernmental entity as an “alter ego created for the purpose of doing indirectly that which the Constitution forbids municipalities to do directly”); Goreham v. Des Moines Metro. Area Solid Waste Agency, 179 N.W.2d 449, 463 (Iowa 1970) (Becker, J., dissenting) (criticizing the creation of the new regional special district as a “barely-disguised technique for debt ceiling avoidance”). For those who view these state limits on local revenue powers as archaic and unresponsive to modern economic realities, evasion of the limits may seem desirable. By providing an easy end run around a state imposed limitation, however, courts have allowed legislatures to avoid grappling with difficult policy issues about the proper parameters of local government borrowing. See generally Valente et al., supra note 29, at 671–716 for analysis of the growing use of “non-debt debt” as a way for municipal governments to raise revenue without running afoul of state law limits.

192. In her recently published book, Professor Kathryn Foster documented that services provided by special district cost more than those provided by multi-purpose municipal governments. See Foster, supra note 24, at 148–85. As Professor Foster noted, however, higher cost does not
participation theory of localism, which stresses the importance of the local government as an accountable, responsive unit of government, special districts can be criticized for the ways in which they disempower the general purpose government that created them. 193

necessarily indicate less efficiency; that is, regional special districts may provide a higher level and quality of service than the general purpose government service provider. Id. at 184. Foster’s findings, however, should give pause to the many supporters of regional special districts who tout their ability to achieve cost savings in the provision of urban services. See, e.g., Connelly v. Clark County, 307 N.E.2d 128, 131 (Ill. App. Ct. 1974) (noting that cost savings was an important rationale when reviewing the history of the Illinois constitutional amendment authorizing intergovernmental cooperation).

193. By assuming control of a function that would otherwise be left to the multi-purpose local government, the creation of a regional special district strips the participating municipal governments of some of their powers and excludes them from participation in the regional enterprise of the district. It is true that municipalities may voluntarily relinquish power to other government units by transferring power to a regional special district, for some of the reasons noted earlier in this Section. The fact of voluntariness, however, does not disprove the underlying claim that the transfer of power to an obscure single purpose government may create a municipal government that is less important and less relevant in the eyes of its citizens. In some major metropolitan areas, where regional special districts number in the hundreds, it is not unreasonable to speculate that the loss of local control over and input into numerous important regional issues contributes to the well-documented loss of citizen interest and involvement in their municipal governments. In the legal literature, Professor Gerald Frug is a passionate supporter of increased local power. See generally GERALD E. FRUG, CITY MAKING: BUILDING COMMUNITIES WITHOUT BUILDING WALLS (1999). He has long pointed to local government powerlessness as one of the main reasons for citizen cynicism and apathy. See Frug, supra note 29.

In some situations, states or citizens may have the statutory authority to create a regional special district that will take over service provision from a general purpose municipal government. See, e.g., Reg’l Serv. Auth. v. Bd. of County Comm’rs, 618 P.2d 1105 (Colo. 1980) (describing regional service authority created by petition of citizens over objection of county board). Of course, state-created special districts are frequently done with no regard for concerns about municipal disempowerment.

Courts have generally rejected challenges that the creation of regional special districts impermissibly interferes with home rule authority or local autonomy generally. But compare Seto v. Tri-County Metro. Transp. Dist. of Or., 814 P.2d 1060, 1065 (Or. 1991) (rejecting municipality’s argument of impermissible interference with home rule powers), with Four-County Metro. Capital Improvement Dist. v. Bd. of County Comm’rs of County of Adams, 369 P.2d 67, 73 (Colo. 1962) (opposite conclusion).

In addition to the somewhat ephemeral disempowerment argument, the proliferation of regional special districts may deprive the municipalities in a region of an important bargaining chip as they seek to bring urbanized development within their borders. If many basic urban services are provided by special districts, rather than by municipalities, property owners in the unincorporated territory on the fringes of metropolitan areas have less incentive to seek annexation to the general purpose local governments they abut. See Fond du Lac Metro. Sewerage Dist. v. Miller, 166 N.W.2d 225, 229–30, (Wis. 1969) (acknowledging the problem, but noting judiciary’s obligation to uphold the formation of the special district under the terms of state law). Many state statutes express a clear desire that urbanization occur within municipal government borders, rather than in unincorporated areas of the county. See, e.g., Colo. Rev. Stat. § 31-12-102 (West 2002); Minn. Stat. § 414.01 (2001); N.C. Gen. Stat. §§ 160A-33,160A-45 (2001); N.D. Cent. Code § 40-51.2-02 (1983) Nev. Rev. Stat.
Notwithstanding these many criticisms, much of the New Regionalist literature touts regional special districts as a means of achieving regionalism while preserving localism. And at first glance, regional special districts appear to promote many of the New Regionalists’ goals. Regional special districts are an attractive way to free local government services from the artificial barriers created by political boundaries, which usually do not reflect the natural geographic or economic territory that would achieve greater efficiencies of scale. Moreover, regional special districts appear to corroborate the New Regionalist claim of the economic interdependence of the metropolitan area. After all, by providing a service to all members of a region that would otherwise not be available on a smaller scale, regional special districts promote the welfare of favored quarter and urban core alike. And finally, regional special districts, though they may add another layer of government, do not disturb the pre-existing landscape of independent local government units. For many New Regionalists, then, regional special districts present a cost-free win-win situation for all constituent governments in a metropolitan region. 194 This enthusiastic endorsement of voluntary regional governance efforts, however, merits rethinking.

First, regional special districts are myopic. By definition, they are concerned with one issue only, be it transportation, sewage, water, flood control, or a similar regional infrastructure problem. They ignore broader questions about the general welfare, focusing instead on their own limited mandate. 195 Consider the following description of regional special districts in California:

The South Coast Air Quality Management District exercises such immense powers over vehicle use, traffic congestion, land use, and job growth that many people call it a de facto regional government—albeit unelected and essentially unaccountable. The L.A. region also has multihundred-million or billion-dollar-a-year special districts in charge of transportation, water supply, waste disposal. Each agency’s professionals do what seems logical from their own narrow point of view—building roads or transit lines, cleaning up L.A.’s putrid air, dealing with toxic wastes, for example. But not one of them is entrusted with the whole—seeing whether and how the pieces fit together. Cumulatively, for example, they spend $71 million a year on planning activities, virtually none of it coordinated.

PEIRCE, supra note 9, at 318.

Similarly, Professor Foster concludes that “The Achilles’ heel of specialized service delivery is its inability to coordinate the planning, budgeting, and delivery of services throughout a metropolitan area . . . When coordination problems occur in a specialized world with separate water, sewer, utility, and highway districts, . . . these problems are predictable outcomes of institutional autonomy combined with functional specialization.” See Foster, supra note 24, at 230.

268.572 (1997). At the same time, though, the laws allowing for the creation of special districts to provide regionwide services severely limit the municipality’s ability to achieve that state goal.

194. See supra notes 76–78 and accompanying text.

195. Consider the following description of regional special districts in California:

The South Coast Air Quality Management District exercises such immense powers over vehicle use, traffic congestion, land use, and job growth that many people call it a de facto regional government—albeit unelected and essentially unaccountable. The L.A. region also has multihundred-million or billion-dollar-a-year special districts in charge of transportation, water supply, waste disposal. Each agency’s professionals do what seems logical from their own narrow point of view—building roads or transit lines, cleaning up L.A.’s putrid air, dealing with toxic wastes, for example. But not one of them is entrusted with the whole—seeing whether and how the pieces fit together. Cumulatively, for example, they spend $71 million a year on planning activities, virtually none of it coordinated.

PEIRCE, supra note 9, at 318.

Similarly, Professor Foster concludes that “The Achilles’ heel of specialized service delivery is its inability to coordinate the planning, budgeting, and delivery of services throughout a metropolitan area . . . When coordination problems occur in a specialized world with separate water, sewer, utility, and highway districts, . . . these problems are predictable outcomes of institutional autonomy combined with functional specialization.” See Foster, supra note 24, at 230.
say that regional special districts are greedy. As Professor Kathryn Foster has established, regional special districts capture a bigger slice of the budget than occurs when those same services compete for funds from a general purpose local government. 196 Isolated, single purpose districts are able to avoid what is known as “full-line forcing,” 197 that is, having to compete for their budget dollars with demands for a variety of other services being funded by the same local government. Overall, services provided by special districts get a higher percentage of the available revenues than would be allocated to that service if it were provided by the general purpose government. As a necessary corollary, then, because the total amount of available revenues does not necessarily change when service provision is shifted to a special district, the funds available for services not provided by the special district may experience an overall decrease.

Third, regional special districts may allow a small segment of the community to act in disregard of the broader general welfare. When citizens are able to establish the borders of a regional special district, they can draw the lines in ways that will enhance the value of the property within the district, which in turn will frequently determine the revenue-raising capabilities of the district. Conferring district-drawing ability on self-interested property owners, though it may appear to facilitate regionalization, does not necessarily guarantee fair allocation of regional burdens. Rather, it merely substitutes one narrow self-interested
calculation for another. 198

Fourth, regional special districts sustain a myth of self-sufficiency. Though they generally require massive infusions of government funds to pay for their initial capital construction, 199 they typically impose user fees to cover at least a share of their operating costs. As citizens have to reach for their wallet each time they use the service, they come to believe that these services fall into the “get what you pay for” model of local government. Thus, the prevalence of fees as a revenue raising mechanism for regional special districts may have a significant impact on the mindset of the individual service user. If the regional services that I use are funded by fees, the reasoning goes, why should I support redistribution of my tax dollars to fund regional social services, all of which will be funded by tax revenues rather than by user fees? 200 Resistance to increased taxation for social services grows because of the belief that taxes are used only to redistribute wealth to the poor, while users pay their own way for other regional services. This anti-tax mentality, however, ignores the infusion of government money that subsidizes regional infrastructure.

Fifth, and perhaps somewhat perversely, regional special districts may impose greater costs on urban residents than on suburban residents. In part, this is due to the higher cost of providing services to low density sprawling suburbs than to dense urban areas. 201 In part, too, it may be due

198. In State ex. rel. Angel Fire Home & Land Owners Ass’n v. South Central Colfax County Special Hospital District, 797 P.2d 285, 287 (N.M. App. Ct. 1990), for example, objecting property owners argued that their property had been included in a proposed hospital district solely because of its high assessed value. They complained that though residents of their area contained only one fourth of the electorate for the hospital district, they owned one half of the district’s tax base. Moreover, they argued that they would not benefit from the construction of the hospital, noting that their homes were substantially closer to an existing hospital than the site of the proposed hospital. Id. The court upheld the formation of the district under state law. Id. at 289. So long as citizens have the power to form regional special districts, however, state law will give them an incentive to draw district boundaries so as to maximize the district’s property wealth without regard for the underlying fairness of their decision.

199. See Foster, supra note 24, at 14.

200. Foster noted that most analysts agree that, in contrast to financing services by user fees and charges, “tax financing promotes communal responsibility for important social services and likely narrows service disparities.” Id. at 107. However, one important study identified three advantages to the use of user fees over taxes: 1) they prevent waste; 2) they ensure that private benefits will not be subsidized by the public as a whole; and 3) they allow governments to recoup expenses from individuals and groups that are not within the government’s taxing territory, such as nonresidents and tax exempt organizations. These benefits can only be attained, however, if the fee is set at the level that reflects individual, rather than community benefit. See Colman, supra note 197, at 9–10. Moreover, the study recognized that user fees conflict with equity norms when poor citizens are assessed fees for essential services. Id. at 13.

201. See Duany et al., supra note 5, at 108.
Intergovernmental Cooperation

to the difficulty of calculating the costs and benefits of any large regional infrastructure, which may in turn lead to unintended overallocation of costs. Or it may reflect a pro-suburban bias against central city areas. Whatever the underlying reason, recent empirical studies make the claim that the urban core has subsidized some regionally provided services.

202. For example, in a lawsuit against a metropolitan area mass transit district, challenging the allocation of costs as between city riders and suburban commuters, the Second Circuit noted the many subtle costs and benefits that may have been reflected in the computation of the fares. See N.Y. Urban League, Inc. v. State of New York, 71 F.3d 1031, 1038–40 (2d Cir. 1995).

203. Many other longstanding patterns of revenue distribution reflect similar anti-urban biases. For instance, with regard to the federal tax treatment of local property tax revenues, Richard Rothstein’s short op-ed piece describes a study by a Stanford economist that shows how the federal tax treatment of local property taxes produces a huge indirect subsidy for wealthy districts. Princeton, N.J., for instance, because of its high local property taxes (deductible for federal income tax purposes), got $2399 in per-student federal aid. Camden, with high direct federal grants through the Title I program, got only $1140 per student in federal aid. See Richard Rothstein, How the U.S. Tax Code Worsens the Education Gap, N.Y. Times, Apr. 25, 2001, at B8. Similarly, the federal commitment to single family housing and its popular mortgage guarantee programs have provided a tremendous transfer of wealth to suburban Iowa. See DUANY ET AL., supra note 5, at 7–8; JACKSON, supra note 5, at 190–215, (documenting how these longstanding and extremely expensive federal programs contributed to urban decay and implemented a strong bias in favor of single family suburban development). Between 1933 and the late 1960s, almost half of all suburban housing was financed by federal loans. JACKSON, supra note 5, at 215. Government spending on thousands of miles of interstate highways constitutes another subsidy of suburbia to which the urban core contributed. See DUANY ET AL., supra note 5, at 8 (noting that the Interstate Highway Act of 1956 provided for 41,000 miles of roadway, 90% of which was paid for by the federal government, at an initial cost of $26 billion. The authors also note how common parlance embodies this anti-urban bent, noting the distinction between the use of the terms “highway investment” and “transit subsidy.” Id. at 96. They quote one study that concludes that government subsidies for highways and parking total somewhere between 8% and 10% of the GNP, totaling approximately $5,000 per car per year. Id. at 94; see also JACKSON, supra note 5, at 250 (stating that in post World War II America, 75% of government transportation expenditures went to building highways, in contrast to 1% dedicated to urban mass transit). These programs constitute a significant subsidization of suburban amenities by central city residents. And finally, one author has claimed that cities bear an additional redistributive burden because of the amount of money they must devote to serving their poor. Though most direct funding for poverty relief comes from state and federal sources, cities devote more than 12% of their own source revenues to public welfare, health and hospitals which are utilized by the poor. See SUMMERS, supra note 73, at 183; see also supra note 93.

204. Commentators have identified several instances in which central city residents subsidize the provision of suburban infrastructure. One study of the Minneapolis area, for instance, concluded that the central urban areas paid more than $6 million more in sewer fees than the costs they create; households in the growing suburban areas received subsidies from those central city users ranging from between $10 and $136 per household per year. See ORFIELD, METROPOLITICS, supra note 10, at 71. Another concluded that corporate relocation from city to suburb creates a subsidy by low and moderate income minority city residents of the suburban redevelopment. See Persky & Wiewel, supra note 82, at 50–69. See generally Cashin, supra note 5, at 2004–15. City subsidization of suburbia is not a new phenomenon. One study of school funding in the Atlanta area concluded that in 1937, more than 50% of the taxes collected for schools in unincorporated areas around Atlanta came from Atlanta taxpayers. See JACKSON, supra note 5, at 132.
Sixth, more generally, and perhaps most crucially for the New Regionalism’s goal of reducing metropolitan area disparities between favored quarter and the rest of the region, regional special districts remove much of the incentive for true regional burden sharing. If the regional services that are desired and used by the wealthier segments of the metropolitan area can be obtained without loss of local government autonomy, without loss of control of local property tax revenues, and without the imposition of any cost beyond the cost of the service itself, there is no incentive for regionwide action to achieve metropolitan equity. By creating regional special districts for the provision of desired services, the favored quarter governments are able to limit their participation in regionwide endeavors according to their own perceived self-interest. They are able to capitalize on the benefits of their presence in a large metropolitan area, yet they are not required to act as a part of

Similar claims can be made with regard to the funding of regional transportation services. See, e.g., N.Y. Urban League Inc. v. State of New York, 71 F.3d 1031, 1036 (2d Cir. 1995). In that case, plaintiffs alleged that city users paid a disproportionate share of the operating costs of the Metropolitan Transit Authority, an umbrella agency that includes the New York City Transit Authority and the operating authorities of several commuter transit lines. Although the U.S. Supreme Court appears to have removed the legal basis of the plaintiffs’ Title VI claim, when it held in Alexander v. Sandoval, 532 U.S. 275, 286–87 (2001), that Title VI does not provide a private right of action, the facts of the NYUL case support the assertion made in the text of this Article, that city users may in fact subsidize suburban users of regional services. Under state law, the MTA must fully fund its operating costs; in 1995 it proposed several system-wide fare increases to cover projected deficits. 71 F.3d at 1035. The fare proposals established a 20% increase in city fares, with a proposed 8.5% increase in commuter service fares. Consider the following additional facts. City transit serves 1.5 billion passengers per year; commuter lines have 1.35 million riders. City operating expenses were $3.1 billion annually, whereas the commuter lines had operating expenses of $1.4 billion. The two services projected different operating deficits: $316 million for the city compared to $72 million for the commuter lines. The proposed fare increases would generate approximately $277.3 million; the 20% city fare increase would generate $274 million, whereas the 8.5% commuter increase would generate $3.3 million. These figures suggest several salient comparisons. City transit has ten times as many passengers as commuter lines, yet its operating expenses are only 2 ½ times the size of the commuter lines’ NYUL, 71 F.3d at 1033-35. The proposed fare increase would generate almost 88% of the city’s projected deficit, but less than half of the commuter lines’. Nevertheless, the Second Circuit rejected the claim of discrimination, stating that the proof offered did “not reveal the extent to which one system might have higher costs associated with its operations—costs stemming from different maintenance requirements, schedules of operation, labor contracts, and so on.” Id. at 1038. Thus, the cost differential appeared to be tied to either a difference in quality of service or inefficiencies in the provision of the services; the fact of a higher subsidy of commuter lines remains unchallenged. For additional discussion of the ways in which city service users may provide a disproportionately high subsidy of metropolitan area services, see generally Kevin L. Siegel, Discrimination in the Funding of Mass Transit Systems, 4 HASTINGS W.-N.W. J. ENVTL. L & POL’Y 107 (1997). Similar conclusions with regard to the development of mass transit in Philadelphia were discussed in Richard Voith, The Determinants of Metropolitan Development Patterns: What are the Roles of Preferences, Prices, and Public Choices?, in URBAN-SUBURBAN INTERDEPENDENCIES, supra note 5, at 71–82.
the region when it comes to finding solutions for the pressing educational, housing, and other social needs that inevitably exist in large metropolitan areas. Able to pick and choose the areas of regional governance in which they want to participate, they maintain their privileged position without having to contribute to the overall regional welfare. Thus, regional special districts offer an attractive way for the favored quarter to preserve its privilege and avoid participation in solutions to regional problems that would require redistribution from favored quarter to urban core.205

4. Voluntary Interlocal Burden Sharing

In a recent article, Professor Clayton Gillette explored the New Regionalist emphasis on achieving metropolitan equity and pointed to the growing evidence behind the claim that suburban prosperity is closely linked to central city health.206 In light of that clear suburban self-interest in solving urban problems, Gillette went on to speculate why examples of formal intermunicipal “burden sharing,” which he defines as “interlocal agreements to alleviate socio-economic disparities within a region,”207 are so rare. Dismissing the hypothesis that suburban indifference to urban ills or antipathy to urban residents could be the explanation, since, as he had already noted, the suburbs had a selfish reason to be concerned about the health of the city, Gillette theorized that the costs of contracting must be the real barrier to interlocal burden sharing. Specifically, Professor Gillette suggested that legal principles (such as non-delegation, lending of credit, and public purpose doctrines), organizational structures (such as fixed geographical municipal boundaries), and the costs of verifying contract compliance, were the

205. In a recent article, Kathryn Foster attempted to identify characteristics that correlate with high levels of “effective regional governance.” See Foster, supra note 11, at 83–87 (explaining desired outcomes for regional governance consist of high levels of economic performance, social equity and regional articulation). Foster then hypothesized that those metropolitan regions with high levels of effective regionalism would display high concentrations of “regional capital.” Id. at 90–96. The ability to execute intergovernmental agreements, and the presence of many specialized regional government structures were identified as positive measures of regional capital. See id. Somewhat surprising to Foster, her analysis did not reveal high correlations between regional capital and effective regional governance; that is, regions identified as accomplished in regional governance did not display significantly more indicia of regional capital than the unaccomplished areas. Id. at 113. If this Article is correct in its suggestion that intergovernmental cooperation and specialized regional governance mechanisms are anti-regional in impact, the lack of correlation Foster observed may be due to a mis-identification of the proper measures of regional capital.

206. See Gillette, supra note 18, at 241–47.

207. See id. at 194.
likely culprits. Rather than revamp the rules to eliminate these substantial disincentives to redistributional regional efforts, however, he suggested that informal voluntary agreements may be at least a partial antidote. Gillette’s proposal puts him in direct opposition to local government scholars who claim that voluntary efforts will never be an effective means for achieving metropolitan equity and call instead for directive state or federal action.

Professor Gillette’s suggestion that voluntary intermunicipal burden sharing could promote regional equity is attractive on several fronts. Most importantly, and consistent with many of the New Regionalist goals for metropolitan areas, it preserves existing local autonomy, it is based on voluntary action rather than coercive direction from a higher level of government, and it requires the creation of no new governmental entities in the regions that are already the most saturated with units of government. For several reasons, however, Gillette’s proposal seems unlikely to have a substantial impact on the regional problems he identifies.

First, consider Gillette’s suggestion that informal agreements may be preferable to formal intermunicipal burden sharing agreements. That endorsement is in turn based on the claim that existing legal rules and government structures create obstacles to formal contracts for intergovernmental cooperation for regional burden sharing. We are left to wonder, however, why those same rules and structures have not been an impediment to the formation of the numerous formal regional agreements and entities that currently exist in metropolitan areas for regionwide provision of services. Regional special districts are routinely created for provision of capital- or infrastructure-intensive services, and the applicable enabling legislation is frequently broad enough to allow the use of that structure for regionwide provision of the kind of burden-sharing Gillette seeks to foster. When intergovernmental benefits the affluent segment of suburbia, formal legal requirements do not appear to

---

208. See id. at 212–31.
209. See id. at 263–69. Professor Gillette does not suggest that voluntary arrangements can completely solve metropolitan area problems, but rather that “in some situations, interlocal bargains are likely to lead to a better allocation of local resources than we would expect from centralization.” Id. at 196.
210. See, e.g., Briffault, supra note 29, at 431–35 (arguing that voluntary cooperation among local government units is unlikely to remedy metropolitan area disparities in wealth and quality of services); Briffault, supra note 2, at 1149 (“As long as cooperation is voluntary, no locality will cooperate with another unless it sees that it will benefit from such cooperation.”); Cashin, supra note 5, at 2030–33 (describing proposals for voluntary burden sharing as “fanciful”).
211. Gillette, supra note 18, at 263–69.
impose insurmountable barriers to regionalism.

Second, Professor Gillette claims that voluntary redistributive efforts are superior to solutions mandated by higher levels of governments. Underlying this preference is his assertion that wealthier local governments should be willing to impose a cost on themselves (by executing an agreement to burden share with less affluent localities) in order to receive the vague, indeed indefinite, benefit of enhanced regional well-being.\textsuperscript{212} It is unclear that local governments can be expected to identify their self-interest over such a long term time line.\textsuperscript{213} Even if the municipality were to make the long term calculation, however, the possibility is high that many more immediate demands for local funds would trump the call for voluntary burden sharing, especially in this era of tax caps and increased state and federal mandates.\textsuperscript{214}

Third, Professor Gillette’s preference for preservation, as opposed to rearticulation, of the current background rules of local government law appears inconsistent with the examples he offers in support of voluntary burden sharing. His article describes several voluntary interlocal burden sharing agreements to illustrate local government recognition of the importance of regional well-being and the assumption of responsibility for achieving it.\textsuperscript{215} In reality, though, the examples he gives are predictable reactions based on the immediate, short term self-interest of the “giving government.” That is, voluntary burden sharing appears to occur when the background legal rules have been modified to give the “receiving government” the right to extract something from the giving government that the giving government values more highly than the price of informal burden sharing. If, for instance, a municipality has the power to annex land in the surrounding unincorporated county, and if the county’s residents are opposed to annexation, the county’s agreement to

\begin{footnotesize}
212. See id. at 232–40.
213. Policy making in the U.S. appears inextricably tied to short-term, rather than long term, interests. See Gottlieb, supra note 9, at 38 (noting that “a short-term focus is endemic to American politics and society”). In fact, one of Gillette’s examples of voluntary burden sharing suggests that municipal governments have considerable difficulty in properly identifying their own long term interests. In 1982, the city of Charlottesville, in an agreement with Albemarle County, relinquished its power to annex additional territory in exchange for the county’s agreement to share tax revenues with the city. Twenty years later, Charlottesville is landlocked, and its income has declined by 31% relative to county income. Its poverty level is now twice as high as the county’s, whereas in 1980 the difference was slight. Recent attempts to have the city revert to town status, thus making it a part of the county, have failed. See “Does the City of Charlottesville Have a Future as a Town,” reprinted at www.virginia.edu/insideuva/textonlyarchive/93-09-17/1.txt (last visited Jan. 13, 2003); Lucy v. County of Albemarle, 516 S.E.2d 480, 487 (Va.1999).
214. See COLMAN, supra note 197, at 1, 10–12.
215. Id. at 234–36.
\end{footnotesize}
share revenue with the city in exchange for the city’s agreement to forego annexation is easy to understand. Undoubtedly, it reflects the county’s calculation that it is better to share revenues than to engage in lengthy, and ultimately futile, battles to resist annexation. \(^{216}\) Similarly, several municipalities within the jurisdiction of a regional district with regulatory land use powers might be expected to establish a common fund to offset the negative impacts of the district’s decisions. This joint burden sharing, rather than being motivated by a desire to achieve regional equity, is better understood as a rational method to allocate the future costs of unknown future regulation, much the way that joint municipal insurance pools do. \(^{217}\) Yet another example of voluntary burden sharing Professor Gillette provides is in fact mandated by state law. \(^{218}\)

While Professor Gillette rejects a reworking of the background rules of local government formation as too costly, the examples he uses to illustrate implementation of voluntary burden sharing are for the most part negotiated in jurisdictions that have in fact changed those background rules. Thus, it appears that voluntary burden sharing typically occurs only when the state redefines the background legal rules to create an incentive for burden sharing that generally does not exist. If the current rules allow the imposition of the costs Gillette seeks to remediate, and if his examples of voluntary burden sharing occur in jurisdictions where the usual background rules have been changed, it is difficult to understand the strength of the case for preservation of the status quo.

Finally, and more broadly, Professor Gillette’s approach to redistribution of wealth and equalization of services in metropolitan areas appears to downplay the evidence that the current admitted need for “burden-sharing” was in no small part created by the efficiency

\(^{216}\) Under the laws of most states, municipalities do not have involuntary annexation powers. See generally Reynolds, supra note 4. In Virginia, however, annexation can be ordered by a court irrespective of the wishes of the residents. See VA. CODE ANN § 15.2-3211 (1997). The revenue sharing agreement between Albemarle County and Charlottesville, then, was a bargain struck against an unusual set of background legal rules. According to a consultant’s report for a multi-county resort area in Colorado, “Revenue-sharing agreements are most commonly used to resolve or present annexation disputes.” See BBC RESEARCH, LOCAL REVENUE-SHARING METHODOLOGIES 22, available at http://www.hmccolorado.org/Revenuesharingfinalreport.pdf (2001). In fact, all of the case studies they presented involved revenue sharing to forestall annexation battles.

\(^{217}\) See Gillette, supra note 18, at 235, for a description of the Meadowlands revenue sharing program.

\(^{218}\) See id. for a description of the Louisville-Jefferson County revenue sharing compact. Under state law, some Kentucky cities and their counties are required to share occupational license fees, see KY. REV. STAT. ANN. §§ 79.310–.330 (Mitchie1996).
Intergovernmental Cooperation

maximizing\textsuperscript{219} he describes in his discussion of voluntary intergovernmental cooperation. For central urban areas, into whose lap many of the costs of these agreements have fallen, the efficiency maximizing regionalization of services was accompanied by a more sinister reverse “burden sharing,” which transferred wealth from urban core to peripheral suburb.\textsuperscript{220} As that imposition of costs continues unabated, voluntary burden sharing (of both the formal and informal types) can never be more than a trickle against a flood of efficiency maximizing.

In sum, though Professor Gillette’s underlying premise that voluntary agreements are preferable to coercive state-mandated redistribution has some appeal, his proposal highlights one of the intractable problems with self-contained local governments in a metropolitan region. While he notes that a purely self-interested, affluent municipality should be willing to engage in metropolitan redistribution, on the basis that overall regional health will be enhanced when the less favored quarters of the region prosper, he attributes the paucity of examples of this self-interested behavior to legal doctrine and contracting costs.\textsuperscript{221} Yet his proposal relies on the preservation of the very legal rules that have already imposed significant costs on the less favored segments of the metropolitan region. To reject reformulation of the rules that have benefited the favored quarter, on the grounds that it would be too costly for that privileged segment of metropolitan regions, seems to value preservation of privilege more than reduction of inequality.

B. The New Regionalist Critique of Intergovernmental Cooperation

Commentators have identified ways in which state laws governing municipal incorporation, annexation, and local taxation facilitate the preservation of local autonomy and intra-regional inequality.\textsuperscript{222} In

\textsuperscript{219} See Gillette, supra note 18, at 231–32. Gillette does not use the term efficiency maximizing, but it is an accurate description of the contrast he makes between, on the one hand, intergovernmental burden sharing agreements, and, on the other hand, those that involve “regional allocation functions,” id. at 231, and are characterized as “agreements for coordination.” Id. at 232.

\textsuperscript{220} See supra notes 93 and 203–04 and accompanying text.

\textsuperscript{221} See Gillette, supra note 18, at 212–31.

\textsuperscript{222} See, e.g., WEIHER, supra note 40, at 176–88. Weiher’s study concluded that local government formation in the United States is “peculiarly ‘anti-government.’” Id. at 165. He explained: Incorporation is an act of anti-government in the sense that it permits residents to escape the collective political demands of larger, more inclusive governmental units. Government becomes not the agency for brokering the legitimate interests of a diverse society, but an instrument for
addition, they have noted the prevalence of intergovernmental cooperation in metropolitan regions and have described the ways in which these collaborative efforts focus on the provision of services, rather than on the elimination of inequalities. Most have not, however, examined whether the existing pattern of regionalization, in which efficiencies are purportedly achieved without loss of governmental autonomy, may ultimately have a negative impact on overall regional welfare. To the extent that the New Regionalist critique has been applied to intergovernmental cooperation, it has praised this "governance over government" solution as a way for local governments to provide services they are unable to produce, and as a voluntary collaborative approach to metropolitan area problems.

Id. at 182.

Weiber includes several examples, including incorporations done by private corporations to avoid taxation and regulation, as well as incorporations by wealthy citizens to escape from financing services for low income individuals, id. at 184. See also Miller, supra note 39, at 34–62; Downs, supra note 1 at 19–22; Jackson, supra note 5, at 139; Briffault, supra note 26, at 72–81; Cashin, supra note 5, at 1992.

223. See, e.g., Cashin, supra note 5, at 2030–33; Briffault, supra note 29, at 374–82; Rusk, supra note 2, at 122–23.

224. In contrast to most current thinking about regionalism, Professor Frug’s recent study of European Union structures and their potential relevance for American metropolitan regions reaches the same conclusion as the one posited here—that the current array of voluntary intergovernmental cooperative agreements “are not stepping stones toward comprehensive regional solutions but successful methods of avoiding them.” See Frug, supra note 15, at 1787.

225. See Savitch & Vogel, supra note 9, at 161.

226. Local governments may separate their deliberations over local public services into two discrete components—first, the decision whether to provide a particular service; second, the decision as to which government entity should produce that service. The flexibility afforded to most general purpose local governments under current state laws allow them to enter a wide range of voluntary agreements whereby another entity actually produces the service to be provided by the receiving entity in the contract. See Briffault, supra note 2, at 1144–45, in which he quotes two urban economists who distinguish the production and provision aspects of local service delivery. Whereas the provision decision results from “collective public choice processes to determine how much of each service to provide and how to pay for it, . . . the production of public services refers to ‘the technical processes of combining resources to . . . render a service.’”) Id. at 1145 n.148, (quoting Roger B. Parks & Ronald J. Oakerson, Metropolitan Organization and Governance: A Local Public Economy Approach, 25 Urb. Aff. Q. 18, 20, 21 (1989)). The public choice analysis makes the normative claim that local governments’ primary purpose is to determine the appropriate range and quality of local services; under this view, the production source of the services is irrelevant. See Briffault supra note 2, at 1145.

227. See, e.g., Cashin, supra note 5, at 1989; ACIR, supra note 20, at 98 (suggesting that intergovernmental cooperation appeals to local governments because it allows for better service provision without structural governmental reorganization). Professor Kathryn Foster views the ability to execute interlocal service agreements as a factor that correlates positively with heightened
Intergovernmental Cooperation

Voluntary intergovernmental efforts may in fact provide a benefit to all segments of a metropolitan area. Presumably, for instance, central cities and suburbs alike are better off with a regional transportation system than without one. On the surface, then, intergovernmental cooperation may appear to establish a win-win opportunity for enhancing regional welfare. Its appeal is multi-faceted: it is based on cooperation rather than coercion; it preserves the pre-existing local government entities; and it enhances the overall regional welfare by providing better or more efficient services to the constituents of various governments in the region. Missing from the analysis, however, is careful consideration whether the intergovernmental cooperation approach to metropolitan issues has in fact facilitated the preservation of widespread regional disparities.

To answer that question, the inquiry must go beyond the limited focus on whether the central urban core is better off with a particular regional service than without one. More relevant to the New Regionalist analysis is the question whether the central city, and ultimately the entire region, is as well off with the current regime of voluntary cooperation as it would have been if metropolitan area governments had adopted, not only regional transportation and sewerage systems, but also regional policies for things such as housing, job creation, tax revenue distribution, and schools. So long as the favored quarter is not required to participate more broadly in regionwide governance, no empirical evidence will be available to answer that question. And because of the rules of intergovernmental cooperation, the central city is in a “take it or leave it” situation. Like the favored quarter, the city can choose to participate in regional governance when it would advance its own interest, but, unlike the affluent quarter, the regional governance efforts it most needs are responses to problems that are more heavily concentrated in non-favored quarter segments of the metropolitan area. As a result, because the New Regionalism’s primary goal is to correct the socio-economic disparities

---

228. See Foster, supra note 11, at 93.

in metropolitan regions, the voluntary intergovernmental cooperation approach to regionalism is likely to leave untouched the root sources of the very disparity it seeks to remedy. The end result is a selective regionalism, for which the metropolitan area is “all in it together” when regional action benefits the favored quarter, but for which the “it’s your problem” response can be given to central city proposals to correct inadequacies in city infrastructure or services.

IV. CONCLUSION: MOVING TOWARDS REGIONAL EQUITY

A. Recognizing the Anti-Regional Impacts of Intergovernmental Cooperation

Intergovernmental cooperative endeavors are flourishing in metropolitan areas. Most recently, this trend has received the endorsement of commentators who predict that continued focus on voluntary regionalization efforts, when seen through the self-interested lens of the favored quarter suburbs, will lead to increased metropolitan equity.229 Yet at the same time, the gap between central city and suburbia has widened. Though the correlation does not necessarily establish a causal connection, this Article has suggested that the two phenomena are related, and that more than mere coincidence is involved. By identifying several ways in which intergovernmental cooperation may both preserve and exacerbate intra-regional inequality, this Article hopes to encourage reexamination of the New Regionalist agenda, at least in so far as it praises intergovernmental cooperation as a realistic tool for narrowing the urban-suburban gap. At a minimum, the analysis offered here should temper the enthusiasm with which the New Regionalist embraces intergovernmental cooperation.

Because of the wide variation in terms of the participants, the services, and the relationships among and between the governmental entities involved, general pronouncements about the anti-regional impact of intergovernmental cooperation would be overstated and unhelpful. Some of the examples discussed in this Article, in fact, suggest that metropolitan equity at times could be enhanced by some voluntary cooperative efforts.230 At the same time, though, an uncritical

229. See supra notes 76–80 and 206–211 and accompanying text.
230. See supra note 189 (discussing Barnes v. Dep’t of Housing & Urban Dev., 341 N.W.2d 766 (Iowa 1984), in which the court invalidated actions taken by a voluntarily created regional housing authority). This Article’s discussion of joint services agreements has also suggested ways in which some of those agreements could have positive impacts on regional welfare. See Part III.A.2.
Intergovernmental Cooperation

endorsement of widespread intergovernmental cooperation appears misguided. The possible anti-regional impacts are at least three-fold. First, to the extent that intergovernmental cooperative efforts result in the establishment of a separate governmental entity, they remove difficult funding decisions from the general purpose government and ultimately skew the revenue allocation in favor of those regional services, at the expense of what remains to be funded out of the local government’s general revenues, most notably spending for social services.\textsuperscript{231} In addition, intergovernmental cooperation may extract a benefit for wealthier suburban areas while it imposes costs on the central urban areas.\textsuperscript{232} And finally, the “pick and choose” phenomenon of intergovernmental cooperation leaves affluent, autonomous local governments in a privileged position to select those aspects of metropolitan policy for which their cooperation is a desirable and interest-maximizing possibility, freeing them from participation in regional policies that would redistribute revenues to the rest of the metropolitan area. The combined impact of these phenomena tips the balance against the urban core, which is left without recourse to state or local laws to correct the imbalance.

B. Refining the New Regionalist Agenda

Though the “New Regionalism” has yet to result in a cohesive doctrinal approach to metropolitan area problems, it appears to have been the catalyst for a growing coalescence around the urgent need for government action to correct metropolitan area inequities. Having come to that position from a variety of ideological and analytical perspectives, commentators and policy makers must now evaluate whether current urban strategies are in fact consistent with their goal. Recognizing the perhaps unintended anti-regional costs of voluntary intergovernmental efforts, both past and ongoing, is itself an important step; articulating alternatives to solve the problem, however, is infinitely more challenging.

Surely, it is too late to envision a metropolitan area without regional special districts and other intergovernmental cooperative endeavors. Not

\textsuperscript{231} This phenomenon of “full line forcing” is discussed \textit{supra} at note 197. If a service is provided at a regional level by a single purpose government, the absence of full line forcing will result in a disproportionate amount of money for regional services (usually infrastructure intensive services such as waste treatment or transportation), at the expense of what remains to be funded out of the government’s general revenues, most notably spending for social services.

\textsuperscript{232} This claim appears to find clearest support in several studies of funding for regional transportation systems and highway spending. \textit{See supra} note 204.
only are they entrenched in the metropolitan landscape, they also undoubtedly confer many benefits on all segments of metropolitan areas. The favored quarter, however, has received disproportionate benefits from widespread intergovernmental cooperation. Not only has it been subsidized by the rest of the region in the provision of some essential services, it has also solidified its privilege with its ability to choose to participate in regional governance only when it would be to its own advantage. That double advantage requires offsetting recompense and action to counter the misallocation of resources in metropolitan regions. Leaving the elimination of the enormous, and ever-growing, urban-suburban gap in the hands of autonomous local governments implausibly hopes that those who have benefited from the current trends will decide to reverse them.

Central to the New Regionalist analysis is the elimination of the socio-economic disparities that plague all major metropolitan areas. Many scholars and politicians have noted the increasing schism and have presented proposals for change. And not surprisingly, current regionalism proposals vary widely in terms of the structures they propose, the implementation mechanisms they suggest, and the extent of state and local law revision they require. It is not for lack of analysis and proposals to reconfigure metropolitan areas that intra-regional inequality has intensified and that metropolitan regions show increasing levels of fragmentation and overlap of governmental units. Moving beyond voluntary intergovernmental efforts, however, leads to the inevitable conclusion that true regionalism efforts will impose financial and political costs on the favored quarter’s local government units. As a result, the political barriers are substantial. In the final analysis, however, it is the responsibility of the state governments that have facilitated the problem to solve it.233

Though the similarities of metropolitan area demographics are substantial, it would be a mistake to suggest that the New Regionalism can develop a standard issue, one size fits all solution to metropolitan area disparities. Restoration of the suburban-urban balance can be

233. Federal policies and incentives have also played a role in shaping the current anti-regional landscape in metropolitan areas, but they are beyond the scope of this Article. See, e.g., FOSTER, supra note 24, at 111 (noting how federal funding and federal tax laws have often encouraged the proliferation of regional special districts); Mark Alan Hughes, Federal Roadblocks to Regional Cooperation: The Administrative Geography of Federal Programs in Larger Metropolitan Areas, in URBAN-SUBURBAN INTERDEPENDENCIES, supra 5 at 161 (explaining that housing authorities, county welfare agencies, job training programs all based on the artificial territorial jurisdiction of city and county, in large part because of federal policies); Savitch & Vogel, supra note 9, at 164 (noting that most regional plan commissions were established solely in order to qualify for federal transportation funds).
pursued through many mechanisms, ranging from complete governmental reorganization at the regional level, to the establishment of more limited regional units with control over regional problems such as schools and affordable housing, to more modest redistributive measures, such as tax sharing or restructuring of the regional tax base itself. And, as the few successful efforts to consolidate municipal and county governments in metropolitan areas can teach us, the legal and political context is extremely important. Existing state statutes, the local political climate, the history of the relationships among existing units of local governments, the strengths and interests of minority groups, and the influence of major political parties, are all extremely important determinants of the shape any regional action will take. The suitability of a proposed regional reform in any particular situation will depend on the interplay of these and many other factors. Notwithstanding the wide range of options, however, the failures of voluntary intergovernmental efforts suggests that the one common necessary ingredient is state commitment to the goal of eliminating the suburban gap. Metropolitan area redistribution makes sense, not only because it may be in the favored quarter’s own long-term self-interest or because it may now be politically feasible through the forging of metropolitan area alliances among the “unfavored” quarters in the metropolitan area, but also because it is “the right thing to do.”

234. States might consider, for instance, something like a metropolitan area benefits tax, to be levied on favored quarter participants in all voluntary intergovernmental efforts. This tax would impose distinct and immediate costs on intergovernmental cooperation. It would recognize that although intergovernmental cooperation may produce tangible benefits for some segments of a metropolitan region, it is also likely to impose costs on other areas. It would also work against the “pick and choose” phenomenon of intergovernmental cooperation, by conditioning favored quarter ability to extract a benefit from its presence in a metropolitan area on its willingness to contribute to narrowing the gap with the other segments of the metropolitan area, without which the intergovernmental cooperation would be impossible. It is of course unlikely that this tax would alone close the gap between favored quarter and the rest of the region. Some supra-regional redistribution is likely to be essential to the realization of metropolitan area equity. See Anita A. Summers, supra note 73, at 190–92; Rusk, supra note 2, at 107. A metropolitan area benefits tax, however, would correctly tax the wealth produced for the favored quarter by intergovernmental cooperative agreements.

235. See Valente et al., supra note 29, at 448–60 for fuller discussion of the consolidation phenomenon and the factors that shape it.

236. See, e.g., Gillette, supra note 18; see also supra notes 85–88 and accompanying text the discussion of the so-called “interdependence imperative.”

237. See generally Orfield, Metropolitics, supra note 10; Cashin, supra note 5. This justification for regionalism depends on the ability of central city and surrounding, declining inner ring suburbs to create a new political majority.

238. See supra note 94.
As this Article joins the chorus in support of the New Regionalists’ call for fairer allocation of resources and opportunities in metropolitan America, it has suggested that intergovernmental cooperation, at least as currently structured, is not a viable tool for the realization of that goal. True regional equity requires a reformulation of the legal rules and structures that preserve and possibly exacerbate the widening disparity between urban core and other metropolitan area constituents. With the anti-regional effects of intergovernmental cooperation identified, one intuitively appealing and popular solution has been taken off the regionalists’ table. What remains are proposals that are all likely to produce strong resistance from favored quarter governments. If this Article is correct in its claim that intergovernmental cooperative efforts are facilitating the preservation of the urban-suburban gap, however, it is time to reexamine those alternative, though undoubtedly more controversial, proposals in the elusive search for regional equity.
SO THE ARMY HIRED AN AX-MURDERER: THE ASSAULT AND BATTERY EXCEPTION TO THE FEDERAL TORT CLAIMS ACT DOES NOT BAR SUITS FOR NEGLIGENT HIRING, RETENTION AND SUPERVISION

Rebecca L. Andrews

Abstract: The Federal Tort Claims Act (FTCA) waives the federal government’s sovereign immunity as to claims for injuries caused by an act or omission of a government employee within his or her scope of duty. However, this waiver is not absolute and the government has retained immunity for many claims, including those arising out of an assault or battery. The federal circuit courts are split regarding whether this exception applies to claims for the negligent hiring, retention and supervision of federal employees who commit an assault or battery. While the U.S. Supreme Court has left the question unanswered, the Ninth Circuit Court of Appeals has stated that such claims are allowed, while most other circuits have taken the opposite view. This Comment argues that claims for the negligent hiring, retention and supervision of federal employees are not barred by the assault and battery exception to the FTCA. The legislative history and intent of the FTCA urge such a reading, as does recent Supreme Court jurisprudence. Finally, any danger that barred claims may be disguised as claims for negligent hiring, retention and supervision can be avoided through the use of the Federal Rules of Civil Procedure.

INTRODUCTION

Marvin, a recruiter for the United States Army, visited the home of a young potential enlistee.\(^1\) While in her home as a representative of the United States government, he assaulted and raped her. When Marvin had applied to enlist in the Army he failed to note on his application that he had served time in a state prison. Disregarding its own procedures, the Army did not investigate whether Marvin had a criminal record. Such an investigation would have revealed that Marvin had previously been convicted of forcibly raping a minor female. This conviction may have rendered Marvin ineligible to join the Army. Because of its own negligence in hiring Marvin, the government not only approved of Marvin’s enlistment but placed him in the recruitment office where he would have frequent contact with young women. While the government’s negligence in hiring, retaining and supervising Marvin was a proximate cause of the victim’s injury in this case and the claim for compensation arises out of this negligence as well as the assault and

\(^1\) Hypothetical created by the author.
battery, many federal courts would not allow the victim to recover against the United States government.\(^2\)

The Federal Tort Claims Act (FTCA) is a waiver of sovereign immunity that allows individuals to sue the federal government for injuries suffered as a result of government negligence.\(^3\) The FTCA is not a complete waiver because Congress has created several exceptions to it.\(^4\) One such exception is labeled the “intentional tort exception” or the “assault and battery exception.”\(^5\) This exception retains the federal government’s sovereign immunity for all claims arising “out of assault, battery, false imprisonment, false arrest, abuse of process, or malicious prosecution.”\(^6\) However, it does not cover all intentional torts, including the intentional infliction of emotional distress or rape.\(^7\)

Courts vary in their interpretation of the assault and battery exception. The U.S. Supreme Court has held that the assault and battery exception does not bar claims for government negligence that proximately causes an assault or battery by a federal employee, when there is an independent duty owed by the government to the plaintiff.\(^8\) Some courts interpret this exception to allow only claims for antecedent government negligence in performing a duty owed the plaintiff unrelated to the government’s employment relationship with the assailant.\(^9\) These courts have held that the assault and battery exception bars claims based on the negligent hiring, retention and supervision of government employees,\(^10\) sometimes arguing that such claims are respondeat superior in disguise.\(^11\) The

\(^2\) The Second, Fourth, Fifth and Tenth Circuits, as well as a federal district court for the district of Maine, have all either held, or implied, that they would not allow this hypothetical victim to recover. See infra notes 98–104 and 170–209 and accompanying text.
\(^4\) Id. § 2680.
\(^5\) See Sheridan v. United States, 487 U.S. 392, 400 (1988); Senger v. United States, 103 F.3d 1437, 1438 (9th Cir. 1996). Because this Comment deals exclusively with assault and battery, the exception will be labeled the “assault and battery exception” throughout.
\(^6\) 28 U.S.C. § 2680(h).
\(^7\) Id.
\(^8\) Sheridan, 487 U.S. at 403.
\(^10\) See, e.g., Lilly, 141 F. Supp. 2d at 628; La Francis, 66 F. Supp. 2d at 342; Guccione, 878 F.2d at 33.
Supreme Court has not, however, required this reading and the Ninth Circuit has refused to adopt it. Instead, the Ninth Circuit has read the assault and battery exception to allow federal district courts jurisdiction over claims of negligence directly related to the employment relationship between the government and the assailant, including claims for negligent hiring, retention and supervision.

This Comment argues that the assault and battery exception to the FTCA is not a bar to claims against the United States government for the negligent hiring, retention and supervision of government employees who commit assault and battery outside of the scope of their employment. Part I outlines the distinction between claims based on respondeat superior and those based on negligent hiring, retention and supervision. Part II reviews the statutory language and legislative history of the FTCA and the assault and battery exception. Part III traces the Supreme Court’s jurisprudence on the assault and battery exception. Part IV details the lower courts’ reactions to the Supreme Court’s decisions. Finally, Part V argues that the FTCA grants federal district court jurisdiction over claims for the negligent hiring, retention and supervision of government employees who commit assault and battery outside of their scope of duty.

I. RESPONDEAT SUPERIOR AND NEGLIGENT HIRING, RETENTION AND SUPERVISION ARE PRACTICALLY AND ANALYTICALLY DISTINCT

According to tort law, both today and at the time of the FTCA’s passage in 1946, an employer may be liable for an employee’s tort under at least two theories: (1) respondeat superior and (2) negligent hiring, retention or supervision. Under respondeat superior, an employer is liable for an employee’s tort regardless of fault. Under negligent hiring, retention or supervision, if the employer knew or should have known that the employee would likely subject third parties to an unreasonable risk of
harm an employer can be held liable for an employee’s intentional tort.\textsuperscript{17} Under a traditional tort law analysis these two theories of liability are analytically distinct.\textsuperscript{18} Negligent hiring, retention and supervision requires a negligent act or acts on the part of a supervising employee while respondeat superior has no such requirement.\textsuperscript{19}

A. Respondeat Superior

The doctrine of respondeat superior is based on the vicarious liability of an employer for an employee’s torts committed within the scope of employment.\textsuperscript{20} Although the employer may not be at fault, the employer may still be liable under state tort law for injuries caused by its employee acting within his or her scope of duty.\textsuperscript{21} The policy rationales behind respondeat superior include compensating the plaintiff, preventing future tortious conduct, and allocating risk to those most able to pay and to those who benefit from the employee’s actions.\textsuperscript{22}

The meaning of “scope of duty” is vague, and at the time of the FTCA’s passage it could be decided by a number of factors.\textsuperscript{23} Such factors included the time, place, and purpose of the act, whether it was a task normally undertaken by servants, whether it was the kind of task the employee was employed to perform, and whether it was motivated by the desire to further the employer’s purposes.\textsuperscript{24} The fact that an employee’s action is improper or prohibited does not automatically exempt it from the employee’s scope of duty if it satisfies some of the other factors listed above.\textsuperscript{25} The U.S. Supreme Court, state courts and a leading commentator have all noted that an employer may be vicariously liable for its employees’ intentional torts committed while acting within the

\begin{itemize}
\item \textsuperscript{17} See 27 AM. JUR. 2d Employment Relationship \textsection 472 (1996).
\item \textsuperscript{18} See id.; 22 AM. JUR. 2d Damages \textsection 788 (1996).
\item \textsuperscript{19} See 27 AM. JUR. 2d Employment Relationship \textsection 472 (1996); 22 AM. JUR. 2d Damages \textsection 788 (1996). Because an employer as an incorporeal entity cannot act without the action of its employees, both of these torts are forms of vicarious liability. For the purposes of this article, the actions of a supervising employee will be ascribed to the employer while those of an employee acting outside of his or her scope of duty will not.
\item \textsuperscript{20} 22 AM. JUR. 2d Damages \textsection 788 (1996).
\item \textsuperscript{21} Id.
\item \textsuperscript{22} 27 AM. JUR. 2d Employment Relationship \textsection 472 (1996).
\item \textsuperscript{23} See PROSSER, supra note 16, at 476.
\item \textsuperscript{24} See RESTATEMENT (SECOND) OF AGENCY \textsection 228, at 504 (1958).
\item \textsuperscript{25} See PROSSER, supra note 16, at 476.
\end{itemize}
Federal Tort Claims Act

scope of duty. In these situations, the employment relationship is enough to create the liability without any independent action attributable to the employer. Because the employer’s liability does not depend upon the employer’s independent negligent or wrongful conduct, respondeat superior does not create an independent cause of action, but merely extends liability for damages from the employee to the employer.

B. Negligent Hiring, Retention and Supervision

In contrast to the doctrine of respondeat superior, negligent hiring, retention and supervision requires a negligent act or omission on the part of the employer. While the plaintiff pleading respondeat superior does not need to demonstrate that the employer acted negligently, a plaintiff pleading negligent hiring, retention and supervision must demonstrate that the employer was negligent. At the time of the FTCA’s passage, the doctrine of negligent hiring, retention and supervision was still in its infancy and would have fallen under the duty to exercise control over the actions of others. According to commentator William J. Prosser it had been recognized in 1941 that employers owe a duty to prevent their employees from harming others even in the absence of a special relationship to the victim. Today, a contemporary plaintiff arguing that an employer violated this duty must prove that the employer knew or should have known that the employee would likely subject third parties to an unreasonable risk of harm.

26. See, e.g., Burlington Indus., Inc. v. Ellerth, 524 U.S. 742, 756 (1998); Hechinger Co. v. Johnson, 761 A.2d 15, 24–25 (D.C. 2000) (stating that if employee acts in part to serve employer's interest, employer will be held liable for intentional torts of employee even if prompted partially by personal motives, such as revenge); Clark v. Pangan, 998 P.2d 268, 270–71 (Utah 2000) (noting Utah has long held an employer can be vicariously liable for intentional torts of employees acting within scope of duty); Jordan v. Cates, 935 P.2d 289, 292 (Okla. 1997) (noting requirements to establish employer liability for intentional tort of employee); PROSSER, supra note 16, at 476 (noting an employer may be liable for the intentional torts of its employee in certain circumstances).

27. 22 AM. JUR. 2d Damages § 788 (1996).

28. See id.

29. 27 AM. JUR. 2d Employment Relationship § 472 (1996). See RESTATEMENT (SECOND) OF AGENCY § 219, at 481 (1958) (stating that masters are not liable for the tort of a servant if that servant was “acting outside of the scope of employment unless the master (a) intended the conduct or the consequences, or (b) the master was negligent or reckless . . . .” (emphasis added)).


32. Id.

33. 27 AM. JUR. 2d Employment Relationship § 472 (1996).
The policy rationales behind respondeat superior mentioned above highlight the differences between respondeat superior and negligent hiring, retention and supervision. In the case of respondeat superior there is a deliberate decision by the court to place liability with the employer even though the only cause of the injury is the employee’s action.\textsuperscript{34} This decision is based on the policies of compensating the plaintiff, preventing future tortious conduct, and allocating risk to those most able to pay and to those who benefit from the employee’s actions.\textsuperscript{35} In contrast, in the case of negligent hiring, retention and supervision there are at least two causes of the injury: negligence attributed to the employer and the employee’s tortious action.\textsuperscript{36}

In sum, the tort of negligent hiring, retention and supervision requires an affirmative, negligent action on the part of the employer. This same action is not a requirement of respondeat superior. Thus, these torts are different and distinct.

\textbf{II. THE FEDERAL TORT CLAIMS ACT}

The FTCA grants jurisdiction to federal district courts to hear claims against the United States government for injuries caused by the negligent or wrongful acts of its employees.\textsuperscript{37} The FTCA was passed in 1946 to provide an effective means of resolving citizens’ tort claims against the federal government.\textsuperscript{38} Without the FTCA, courts would not have jurisdiction to hear these claims because the federal government, as sovereign, would be immune from suit.\textsuperscript{39} If a court finds at any time that the alleged tort does not fall within the parameters of the FTCA, it must dismiss the claim immediately as required by Federal Rule of Civil Procedure 12(h)(3).\textsuperscript{40}

There are many exceptions to the FTCA, since Congress did not waive immunity for all torts.\textsuperscript{41} One such exception is the assault and battery exception that retains governmental immunity for any claim that “arises

\begin{itemize}
  \item \textsuperscript{34} See id.
  \item \textsuperscript{35} See id.
  \item \textsuperscript{36} 22 AM. JUR. 2d Damages § 788 (1996).
  \item \textsuperscript{38} Dalehite v. United States, 346 U.S. 15, 25 (1953).
  \item \textsuperscript{39} United States v. Sherwood, 312 U.S. 584, 586 (1941).
  \item \textsuperscript{40} FED. R. CIV. P. 12(h)(3).
  \item \textsuperscript{41} 28 U.S.C. § 2680 (2000).
\end{itemize}
Federal Tort Claims Act

out of” an assault or battery. Although there is little legislative history for this particular exception, the limited history suggests that the government was concerned about the financial danger of liability for the intentionally tortious acts of its employees.

A. Legislative History and Procedural Context of the FTCA

Before the passage of the FTCA, the federal government was immune from suit for torts committed by its employees by virtue of the doctrine of sovereign immunity. The source of this doctrine is unclear; some say it springs from the Constitution while others insist that it is a child of English common law. Regardless of its source, sovereign immunity has been widely criticized as unjust and the federal government has slowly been picking away at the concept by enacting legislation that waives it in certain circumstances. One such piece of legislation is the FTCA.

In response to a need for a simple way to compensate victims, in 1946 Congress passed the FTCA, 28 U.S.C. §§ 1346, 2671–2680, which provides federal district court jurisdiction over claims against the United States Government for personal injuries caused by certain types of government negligence. The FTCA specifically allows for the recovery of damages for injury “caused by the negligent or wrongful act or omission of any employee of the Government while acting within the scope of his office or employment.” Prior to the FTCA’s passage the only redress for these injuries was a private bill passed by Congress, a process which was cumbersome and provided compensation for only a fraction of the claims made. The FTCA was passed both to cure this

42. Id. § 2680(h).
43. See Panella v. United States, 216 F.2d 622, 625 (2d Cir. 1954) (describing legislative history of § 2680(h) as “meager”).
46. See, e.g., Blue Fox, Inc., 525 U.S. at 260.
48. Id. § 1346.
49. Id.
50. See Dalehite v. United States, 346 U.S. 15, 25, 25 n.9 (1953); Irvin M. Gottlieb, The Federal Tort Claims Act—A Statutory Interpretation, 35 GEO. L.J. 1, 4 (1946) (noting that “more than 2,000 private claim bills are introduced in each Congress, a substantial percentage of which are claims for property damage or personal injury . . . . [T]he cost of passing a single private claim bill amounted to
perennial Congressional headache and to address the need for access to the courts for the mass of citizens injured by government negligence that were not served by the private bill procedure.\textsuperscript{51}

A claim can be dismissed for lack of subject matter jurisdiction at any time if it is excepted from the FTCA.\textsuperscript{52} Federal Rule of Civil Procedure 12(h)(3) states that when a lack of subject matter jurisdiction becomes apparent to the court, the court must dismiss the claim.\textsuperscript{53} Because the FTCA is a grant of jurisdiction, claims that are barred by one of its exceptions must be dismissed at once.\textsuperscript{54} Additionally, the government has only provided for bench trials in cases asserting tort liability of the federal government.\textsuperscript{55} Thus, if a respondeat superior claim reaches trial in the guise of a negligent hiring, retention or supervision claim, the subtle distinction between these two doctrines will not be decided by an untrained jury.\textsuperscript{56}

\textbf{B. The Assault and Battery Exception to the FTCA}

The waiver of sovereign immunity contained in the FTCA is not absolute and the government has retained immunity from suit for many intentional torts, including claims for assault and battery by federal employees.\textsuperscript{57} 28 U.S.C. § 2680(h) provides that the broad grant of jurisdiction in the FTCA\textsuperscript{58} shall not apply to any claims "arising out of assault [and] battery."\textsuperscript{59} While this Comment deals exclusively with the assault and battery exception, the analysis herein may be applicable to one or more of the other exceptions.

The legislative history of the assault and battery exception to the FTCA, although sparse, implies that the government was concerned

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{51}See \textit{Indian Towing v. United States}, 350 U.S. 61, 68–69 (1955).
\item \textsuperscript{52}See \textit{FED. R. CIV. P. 12(h)(3)} (mandating Federal Courts to dismiss cases for which they do not have subject matter jurisdiction).
\item \textsuperscript{53}See id.
\item \textsuperscript{54}See \textit{id.}; 28 U.S.C. § 1346 (2000).
\item \textsuperscript{55}See 28 U.S.C. § 2402; \textit{see also} \textit{GREGORY C. SISK, LITIGATION WITH THE FEDERAL GOVERNMENT: CASES & MATERIALS} 72–79 (2000).
\item \textsuperscript{56}See 28 U.S.C. § 2402.
\item \textsuperscript{57}Id. § 2680. For example, Section 2680(h) contains exceptions for claims arising out of discretionary functions, the loss, miscarriage or negligent transmission of mail, and an exception for many intentional torts, including assault and battery. \textit{Id.} §§ 2680(a), 2680(b), 2680(h).
\item \textsuperscript{58}Id. § 1346.
\item \textsuperscript{59}Id. § 2680(h).
\end{itemize}
\end{footnotesize}
about the danger of liability for intentional torts that could be exaggerated and easily proven. 60 In Senate Committee hearings in the 76th Congress on an immunity-waiving bill similar to the FTCA, the Justice Department’s representative noted that the bill’s assault and battery exception corollary excluded those types of torts that “would be difficult to make a defense against, and which are easily exaggerated.” 61 Thus, it appears Congress did not want the government to be automatically liable each time one of the federal government’s employees committed an assault and battery.

The paucity of discussion of the FTCA does not provide a basis for any firm conclusions as to Congress’ intent regarding intentional torts caused by negligence. 62 As one commentator noted, the very sparseness of Congress’ discussion of this exception implies that Congress simply divided the world of torts into those caused by negligence and those by intention and did not pause to ponder those intentional torts that were caused by antecedent governmental negligence. 63 In the course of Congressional hearings on a similar bill, the representative from the Department of Justice urging passage of the bill was questioned as follows:

Mr. Robsion. On that point of deliberate assault that is where some agent of the Government gets in a fight with some fellow?

Mr. Shea. Yes.

Mr. Robsion. And socks him?

Mr. Shea. That is right . . . .

Mr. Cravens. This refers to a deliberate assault?

Mr. Shea. That is right.

Mr. Cravens. If he hit someone deliberately?

Mr. Shea. That is right.

Mr. Cravens. It is not intended to exclude negligent assaults?

60. See Tort Claims Against the United States: Hearings Before a Subcommittee of the Committee on the Judiciary United States Senate, 76th Cong. 39 (1942).

61. See id.


63. See Zolensky, supra note 62, at 810.
Mr. Shea. No. An injury caused by negligence could be considered under the bill.64

This exchange suggests that Congress considered torts to be either intentional or caused by negligence. Based on this limited legislative history, it is unclear whether Congress contemplated or intended to bar assaults and batteries caused by the negligent hiring, retention or supervision of government employees.

In sum, the FTCA was clearly intended to provide a remedy for individuals harmed by the negligence of the federal government. The exceptions to the FTCA appear to have been intended to shield the government from unwieldy claims founded on intentional torts. Unfortunately, Congress’ intent with regard to intentional torts proximately caused by governmental negligence largely remains a mystery.

III. INTERPRETING THE SCOPE OF THE FEDERAL TORT CLAIMS ACT

Throughout the history of the assault and battery exception to the FTCA, courts have conflicted in their interpretations of its breadth.65 Some courts, following a plurality of the U.S. Supreme Court in United States v. Shearer,66 adopted a but-for analysis, reasoning that if the plaintiff’s injury would not have been inflicted but-for an assault and battery by a government employee, then the claim arose out of an assault and battery and was barred.67 This framework, however, conflicted with the Second Circuit’s employee/non-employee analysis, which held that the assault and battery exception only applied to assaults and batteries committed by government employees, and not to assaults and batteries committed by non-employees.68 Ultimately, the Supreme Court adopted this employee/non-employee framework in Sheridan v. United States69

---

65. Compare Thigpen v. United States, 800 F.2d 393, 394 (4th Cir. 1986) (holding that the assault and battery exception barred any claim involving an assault and battery), with Bennett v. United States, 803 F.2d 1502, 1503–05 (9th Cir. 1986) (allowing a claim to proceed which alleged negligent hiring and retention of a federal employee who assaulted his students).
67. Thigpen, 800 F.2d at 394.
68. See Panella v. United States, 216 F.2d 622, 624–25 (2d Cir. 1954).
Federal Tort Claims Act

and modified it to include a distinction between employees acting within the scope of duty and those acting outside of the scope of duty. 70 This modification meant that if an employee committed an assault and battery outside of his or her scope of duty, and that assault and battery was also caused by some type of government negligence predicated on a duty owed the victim, the government could be held liable. 71

The Sheridan Court declined to decide whether the federal government could ever be held liable for negligently hiring, retaining or supervising employees who commit assaults and batteries outside of their scope of duty. 72 Although most circuits have read the assault and battery exception to exclude claims for the negligent hiring, retention and supervision of federal employees who commit assaults and batteries, 73 the Ninth Circuit has maintained that such claims are allowed. 74

A. A Broad Reading of the Assault and Battery Exception: United States v. Shearer and the But-For Analysis

Prior to the United States Supreme Court’s decision in Sheridan, many courts broadly interpreted the FTCA’s assault and battery exception. 75 This broad reading was in response to the opinion of a plurality of the Supreme Court in United States v. Shearer. 76 The Shearer plurality advanced a but-for analysis to the assault and battery exception. 77 The plurality reasoned that the exception’s “arising out of” language meant that any claim which would not exist but-for an assault and battery was barred, regardless of the government’s antecedent negligence or any independent duty owed to the victim. 78

70. Id.
71. See infra notes 129–42 and accompanying text.
72. Sheridan, 487 U.S. at 403 n.8.
73. See Jared M. Viders, Comment, Negligent Hiring, Supervision and Training—The Scope of the Assault and Battery Exception: Senger v. United States, 39 B.C. L. REV. 452, 452 n.6 (1998).
74. See, e.g., Senger v. United States, 103 F.3d 1437, 1442 (9th Cir. 1996); Brock v. United States, 64 F.3d 1421, 1425 (9th Cir. 1995); Bennett v. United States, 803 F.2d 1502, 1504 (9th Cir. 1986).
75. See Hoot v. United States, 790 F.2d 836, 838 (10th Cir. 1986); Thigpen v. United States, 800 F.2d 393, 395 (4th Cir. 1985); Johnson by Johnson v. United States, 788 F.2d 845, 851 (2d Cir. 1986).
76. 473 U.S. 52 (1986); see Hoot, 790 F.2d at 838; Thigpen, 800 F.2d at 395; Johnson by Johnson, 788 F.2d at 850.
77. See Shearer, 473 U.S. at 55.
78. Id. at 54–57.
In Shearer, the mother of an off-duty serviceman sued the government under the FTCA for negligent retention when another off-duty serviceman, Private Heard, killed her son.\textsuperscript{79} Prior to this killing, Private Heard was convicted of manslaughter and sentenced to four years in prison while assigned to an army base in Germany.\textsuperscript{80} The government was aware of Private Heard’s prior conviction and of his violent history. Additionally, three of his superior officers had recommended that he be removed from service, but the government did nothing.\textsuperscript{81} Further, the army did not evaluate Private Heard’s mental state until after he had murdered the plaintiff’s son.\textsuperscript{82}

While the plurality focused on the FTCA’s assault and battery exception, a majority of the U.S. Supreme Court decided Shearer based on the Feres doctrine.\textsuperscript{83} The Feres doctrine stems from the Court’s decision in Feres v. United States\textsuperscript{84} and stands for the principle “that a soldier may not recover under the Federal Torts Claim Act for injuries which ‘arise out of or are in the course of activity incident to service.’”\textsuperscript{85} The majority of the Court stated that the Feres doctrine applied to the facts of the Shearer case and reasoned that finding otherwise would require courts to second-guess the decisions of commanding officers.\textsuperscript{86}

A four-member plurality of the Court maintained that the claim was barred by the FTCA’s assault and battery exception as well as the Feres doctrine.\textsuperscript{87} The plurality read the assault and battery exception in Section 2680(h) broadly: “[the assault and battery exception] does not merely bar claims for assault or battery; in sweeping language it excludes any claim arising out of assault or battery. We read this provision to cover claims like respondent’s that sound in negligence but stem from a battery committed by a Government employee.”\textsuperscript{88} In reaching this conclusion, the plurality interpreted the legislative history of the assault and battery exception to indicate that Congress expected that the federal government

\textsuperscript{79} Id. at 53.
\textsuperscript{80} Id. at 53–54.
\textsuperscript{81} United States v. Shearer, 723 F.2d 1102, 1104 (3d Cir. 1983) rev’d, 473 U.S. 52 (1985).
\textsuperscript{82} Id.
\textsuperscript{83} Shearer, 473 U.S. at 57–59.
\textsuperscript{84} Feres v. United States, 340 U.S. 135 (1950).
\textsuperscript{85} Shearer, 473 U.S. at 57 (quoting Feres, 340 U.S. at 146 (1950)).
\textsuperscript{86} Id. at 52.
\textsuperscript{87} Id. at 54–56.
\textsuperscript{88} Id. at 55 (emphasis in original).
would not be liable for the intentional torts of its employees. The plurality reasoned this reading of the assault and battery exception was confirmed by an amendment to the FTCA passed in 1974 that waived sovereign immunity for claims arising out of the assaults and batteries of law enforcement officers. This legislation did not mention that the government would be liable for the negligent hiring, retention or supervision of a government employee. Therefore, the plurality reasoned, the premise of this legislation was that unamended, the FTCA does not provide for government liability for these torts regardless of antecedent government negligence.

The Shearer plurality acknowledged Panella v. United States, a Second Circuit case which held that the assault and battery exception does not apply to the assaults and batteries of non-employees. By acknowledging Panella, the plurality implicitly accepted the reasoning that the federal government is liable for a tort committed by a non-employee that is proximately caused by the government’s negligence, although it is not liable for the same tort committed by a government employee preceded by identical negligence. In essence, the plurality adopted a but-for test. If the plaintiff would not have a claim but-for the federal employee committing an assault and battery, the government is immune from suit regardless of the government’s negligence.

In the wake of Shearer, some courts followed the plurality’s lead and adopted a but-for test. The Fourth Circuit adopted the but-for test in Thigpen v. United States. The Thigpen court relied on Shearer and previous Fourth Circuit precedent to hold that the assault and battery exception “bars any claim that depends on the existence of an assault and battery.” The Second Circuit agreed in Johnson by Johnson v. United

89. Id.
91. Shearer, 473 U.S. at 55–56.
92. Id. at 56.
93. 216 F.2d 622 (2d Cir. 1954).
94. Shearer, 473 U.S. at 56–57 (citing Panella, 216 F.2d at 626).
95. Id.
96. See id. at 55.
97. See id.
98. See Hoot v. United States, 790 F.2d 836, 838 (10th Cir. 1986); Thigpen v. United States, 800 F.2d 393, 395 (4th Cir. 1986); Johnson by Johnson v. United States, 788 F.2d 845, 851 (2d Cir. 1986).
99. 800 F.2d 393 (4th Cir. 1986).
100. Id. at 395.
States, stating that: “[t]he statute’s plain language, ‘arising out of,’ reflects an intent by Congress to bar a suit against the government for injuries caused by a government employee’s commission of an assault and battery.” Finally, the Tenth Circuit embraced the but-for test in Hoot v. United States, and opined that “[a]bsent the assault and battery perpetrated on Hoot by Firth, there would be no claim.”

As set forth above, the Shearer plurality adopted a reading of the assault and battery exception that denies government liability if a plaintiff’s claim is in any way premised on an assault and battery. However, the Shearer plurality acknowledged that the government was liable for the negligently caused assaults and batteries of non-employees. In response to the Shearer plurality’s reasoning, the Second, Fourth and Tenth Circuits all adopted but-for tests for FTCA claims.

B. A Narrower Exception: Sheridan v. United States and the Search for an Independent Duty

At the same time that the but-for test was developed, some courts interpreted the assault and battery exception more narrowly. These courts concluded that the plaintiff had a cause of action if the government owed the victim an independent duty and the assailant was not an employee or was acting outside of the scope of duty. This interpretation of the assault and battery exception reads it in light of the scope of the FTCA’s waiver of immunity. Because the FTCA only waives immunity for the negligent acts of government employees acting within their scope of duty, only those claims which would provide a cause of action under this waiver are subject to the exception. Therefore, the assault and battery exception does not apply to claims

101. 788 F.2d 845 (2d Cir. 1986).
102. Id. at 850 (emphasis in original).
103. 790 F.2d 836 (10th Cir. 1986).
104. Id. at 839.
109. See Panella, 216 F.2d at 624–25.
Federal Tort Claims Act

stemming from an assault and battery committed by a non-employee or by a government employee acting outside of his or her scope of duty.110

1. The Origins of the Independent Duty Test: Panella v. United States

The door to the independent duty test was first opened by the Second Circuit in Panella v. United States when the court acknowledged that the assault and battery exception does not apply to assaults and batteries by non-employees.111 While “[i]t is true that [the assault and battery exception] can literally be read to apply to assaults committed by persons other than government employees,” the circuit court found this reading inconsistent with the entire the FTCA.112 The Second Circuit recognized that the waiver of sovereign immunity only applies to negligent or wrongful acts or omissions “of any employee of the Government while acting within the scope of his office or employment.”113

The Second Circuit reasoned that an intentional tort committed by a non-employee does not itself give rise to government liability, because the FTCA only waives sovereign immunity for acts committed by government employees acting within their scope of duty.114 Because the FTCA only grants federal district courts jurisdiction over the negligent acts of government employees, the assault and battery exception cannot apply to cases where the assault and battery is committed by a non-employee.115 In such cases, the antecedent negligence of the government that proximately caused the assault and battery is the only remedy available because the FTCA does not provide a remedy, barred or otherwise.116 Similarly, because there is no claim available for the assault and battery itself, the root of the claim against the government is based in negligence, not in assault and battery.117

111. See 216 F.2d at 624. The court relied on the language of 28 U.S.C. § 1346(b)(1), which grants federal district courts jurisdiction over torts committed by Federal Government employees. Id.
112. Id. at 623–25.
114. Panella, 216 F.2d at 624.
115. See id.
116. See id.
117. Id. at 623.
2. The Independent Duty Test Becomes the Law of the Land: Sheridan v. United States

The Second Circuit’s reasoning in *Panella* was adopted by the United States Supreme Court in *Sheridan v. United States*.\(^\text{118}\) In *Sheridan*, a heavily intoxicated off-duty serviceman, Carr, was found by several naval corpsmen in a hospital building owned by the United States Navy.\(^\text{119}\) The corpsmen attempted to take Carr to the emergency room, but fled when they discovered that Carr was carrying a rifle, and did not report the incident, or Carr’s presence, to the proper authorities.\(^\text{120}\) Later that evening Carr fired his rifle into the plaintiffs’ automobile, injuring one of the plaintiffs and causing property damage.\(^\text{121}\) Recognizing that it was facing a circuit split regarding the scope of the assault and battery exception,\(^\text{122}\) the Court began its analysis by looking at the statute’s language and acknowledged that the words “arising out of” could be broad enough to bar any claim if it was based in any way on an assault and battery.\(^\text{123}\)

However, citing *United States v. Muniz*,\(^\text{124}\) the *Sheridan* Court noted that there are instances where the existence of an assault or battery does not bar a claim against the federal government for negligence.\(^\text{125}\) In *Muniz*, a federal prisoner was severely beaten by other inmates: he suffered a fractured skull and lost sight in his right eye.\(^\text{126}\) Although the *Muniz* Court reversed the appellate court’s dismissal of the claim, it did not specifically address the assault and battery exception.\(^\text{127}\) In *Sheridan*, the Court held that *Muniz* stood for the proposition that the existence of an assault and battery is not an absolute bar to federal district court jurisdiction over a claim against the government for negligently allowing an assault and battery to occur.\(^\text{128}\)

---

119. *Id.* at 395.
120. *Id.*
121. *Id.*
122. *Id.* at 398.
123. *Id.*
127. *Id.* at 153.
128. See *Sheridan*, 487 U.S. at 399.
The Sheridan Court suggested two possible interpretations of Muniz. The narrower interpretation, which the Court adopted, reasoned that an assault and battery by a non-employee does not fall under the assault and battery exception. The broader theory, which the Court neither adopted nor rejected, assumed that because Muniz had alleged an independent basis for government liability, namely the government’s negligence, his claim did not arise solely out of the assault and battery but also out of the negligence of the prison officials. By adopting the first theory, the Sheridan Court relied on the Panella court’s distinction between assaults and batteries committed by employees and those committed by non-employees and moved the emphasis in the FTCA from “employee” to “acting within the scope of his office or employment.”

Applying this interpretation to the facts of Sheridan, the Supreme Court found that the assault and battery exception was inapplicable to the case at bar because Carr was not acting within the scope of his duties when he fired the shots into the plaintiffs’ automobile. Consequently, his actions did not independently give rise to government liability. The Court also held, however, that the government could be liable for the independent negligence of the corpsmen who found Carr and did not report his presence to the appropriate authorities. The Court observed that the government owed a duty to the plaintiffs independent of its employment relationship with the assailant. This duty was based on the regulations prohibiting possession of firearms on base and on the corpsmen’s voluntarily undertaking to care for an obviously intoxicated and armed person.

In Sheridan, the Court also employed a traditional tort law analysis of duty and causation. Although the Court did not discuss it explicitly, the corpsmen who found Carr in the hall were acting within the scope of their duty when they failed to report their discovery to the authorities,
thus opening the government up to liability under the FTCA.\textsuperscript{139} Despite the fact that the claim would not exist but-for the assault and battery, the Court did not find that the assault and battery exception barred the claim.\textsuperscript{140} The Court reasoned that the government would be liable if the assault and battery had been committed by a non-employee and that it would be absurd to find no liability simply because the assailant was a government employee.\textsuperscript{141} Since the \textit{Sheridan} Court found that Carr’s employment status was irrelevant to its decision, it declined to decide whether negligent hiring, supervision or retention might give rise to government liability for the injuries resulting from a foreseeable assault and battery.\textsuperscript{142}

Justice Kennedy concurred in the judgment of the Court in \textit{Sheridan}, but concluded that relying on the assailant’s employment status or scope of duty was a flaw in the majority’s reasoning.\textsuperscript{143} Instead, according to Justice O’Connor’s dissent, Justice Kennedy focused on the acts or omissions of the federal government.\textsuperscript{144} Justice Kennedy observed that it is a basic precept of tort law that one injury can arise from multiple wrongful acts.\textsuperscript{145} Justice Kennedy criticized the dissent’s view, which advocated the \textit{Shearer} plurality’s but-for analysis,\textsuperscript{146} arguing that this rationale would obliterate the multi-causal nature of many torts and give no legal significance to negligent acts that may have preceded the assault and battery.\textsuperscript{147}

However, Justice Kennedy amputated this rationale as it applies to preceding acts of negligent hiring, retention or supervision.\textsuperscript{148} He reasoned that in many cases it would be easy to assign blame for the intentional torts of employees to the prior negligence of supervisors and therefore circumvent the purpose of the exception.\textsuperscript{149} Although he did not use the term “respondeat superior” in his concurrence, Justice Kennedy appeared to conflate that concept with negligent hiring, retention and

\begin{enumerate}
\item\textsuperscript{139} See \textit{id}.
\item\textsuperscript{140} \textit{Id.} at 403.
\item\textsuperscript{141} \textit{Id}.
\item\textsuperscript{142} \textit{Id.} at 403 n.8.
\item\textsuperscript{143} \textit{Id} at 405 (Kennedy, J., concurring).
\item\textsuperscript{144} \textit{Id.} at 404 (O’Connor, J., dissenting).
\item\textsuperscript{145} \textit{Id.} at 405 (Kennedy, J., concurring).
\item\textsuperscript{146} \textit{Id.} at 406.
\item\textsuperscript{147} \textit{Id}.
\item\textsuperscript{148} \textit{Id.} at 406–07.
\item\textsuperscript{149} \textit{Id.} at 407.
\end{enumerate}
supervision because he did not explain his assertion that proving negligent hiring, retention or supervision is substantially easier than proving any other type of negligence.\textsuperscript{150} Instead, he simply asserted this proposition, implying that he was referring to respondeat superior, which is substantially easier to prove than negligence because it does not require a showing of duty, causation, breach and injury.\textsuperscript{151} Although Justice Kennedy recognized that negligence and assault and battery could be distinct causes of a single injury, he was unwilling to make a distinction between negligent hiring, retention and supervision and respondeat superior.\textsuperscript{152}

To summarize, the U.S. Supreme Court began its examination of the assault and battery exception to the FTCA with a plurality opinion narrowly drawing the boundaries on government liability. The Court has since loosened the strictures around the exception by recognizing that the government may be liable for the injuries caused by a federal government employee’s assault and battery under certain circumstances. While the Court has not stated that negligent hiring, retention and supervision are some of those circumstances, neither have it closed the door on that possibility.

IV. THE CIRCUIT COURTS POST-SHERIDAN V. UNITED STATES: DISAGREEING WHETHER HIRING, RETENTION, AND SUPERVISION ARE INDEPENDENT DUTIES

In the wake of Sheridan, the majority of courts have adopted an independent duty test that excludes claims based on the government’s negligent hiring, retention and supervision of its employees.\textsuperscript{153} Under this test, the government is only liable if it owes a duty to the victim of the assault and battery, independent of its employment relationship with the assailant, and if the government’s breach of that duty is a proximate cause of the victim’s injuries.\textsuperscript{154} This doctrine does not extend to include jurisdiction over claims for the negligent hiring, retention or supervision of employees who commit assaults and batteries outside of their scope of duty.\textsuperscript{155} The Ninth Circuit, however, has recognized government liability

\textsuperscript{150} Id.
\textsuperscript{151} See 22 AM. JUR. 2d Damages § 788 (1996).
\textsuperscript{152} Sheridan, 487 U.S. at 404–08.
\textsuperscript{153} See Viders, supra note 73, at 452.
\textsuperscript{154} See, e.g., Harris v. United States, 797 F. Supp. 91, 95 (D.P.R. 1992).
for the negligent hiring, retention and supervision of government employees where the government knew or should have known that the employee would likely commit an assault or battery and negligently hired, retained or supervised the employee.  

A. A Duty Independent of the Employment Relationship Avoids the Assault and Battery Exception

After the Sheridan decision, some district courts fashioned an independent duty test employing the Sheridan Court’s reasoning. Under this test, courts considered whether the government owed the plaintiff an independent duty unconnected to the government’s employment relationship with the assailant. If the government had some type of special relationship to the plaintiff, then, these courts held, the assault and battery exception was not a bar to a cause of action.  

The District of Columbia Circuit adopted this test in Bembenista v. United States. In Bembenista, the plaintiff was sexually assaulted by a federal medical technician while in an unconscious or semi-conscious state. Six months prior to the Sheridan decision, the district court hearing the claim dismissed it as barred by the assault and battery exception. But, after Sheridan, an appellate court held the plaintiff could proceed under a theory that the government had breached its “special obligation of protective care” to the plaintiff. The D.C. circuit court found that under District of Columbia law, the relationship between a patient and a hospital gave rise to an independent duty to protect patients from the foreseeable, injurious acts of third persons. Further, the court stated the breach of this duty of care to Mrs. Bembenista was unrelated to the medical technician’s employment status: no matter who assaulted Mrs. Bembenista, the government would be liable. The Bembenista court did not consider whether the defendant was acting

156. See, e.g., Bennett v. United States, 803 F.2d 1502, 1503-05 (9th Cir. 1986).
158. See id. at 339.
159. Id.
160. 866 F.2d 493 (D.C. Cir. 1989).
161. Id. at 495.
162. Id. at 497.
163. Id.
164. Id. at 498.
165. Id.
within the scope of his duty when he assaulted Mrs. Bembenista, perhaps judging that the answer was self-evident or irrelevant to the resolution of the issues.\footnote{166} Other courts have found that an independent duty exists where children were mistreated by their federally employed teacher,\footnote{167} infants were injured in a federal hospital,\footnote{168} and customers were sexually assaulted by a Postmaster while in the post office.\footnote{169}

**B. Negligent Hiring, Retention and Supervision Claims Barred by the Assault and Battery Exception Under the Independent Duty Framework**

Courts that have adopted the independent duty test have held that it stops short of allowing district court jurisdiction over claims for negligent hiring, retention and supervision.\footnote{170} Although the Sheridan Court did not preclude courts from finding jurisdiction based on negligent hiring, retention or supervision,\footnote{171} many courts have refused to extend the logic of Sheridan this far.\footnote{172} These courts have restricted their jurisdiction to those cases where a clear duty, independent of the employment relationship, exists between the United States and the victim of the assault and battery.\footnote{173}

In *Leleux v. United States*, the Fifth Circuit employed a narrow interpretation of the Sheridan Court’s statement that government liability requires that the breach of an independent duty owed by the government to the victim was a proximate cause of the injury.\footnote{174} In *Leleux*, a high-school age navy recruit was seduced by an enlisted petty officer, Sistrunk, who worked in the recruitment office.\footnote{175} As a result of this encounter, the plaintiff acquired the genital herpes virus and sued the

\footnote{166}{Id. at 498–99.}
\footnote{167}{Harris v. United States, 797 F. Supp. 91, 93 (D.P.R. 1992).}
\footnote{168}{Gess v. United States, 952 F. Supp. 1529, 1532 (M.D. Ala. 1996).}
\footnote{169}{Strange v. United States, 114 F.3d 1189, 1997 WL 295589, at *1 (6th Cir. 1997).}
\footnote{170}{See, e.g., Leleux v. United States, 178 F.3d 750, 757 (5th Cir. 1999); Guccione v. United States, 878 F.2d 32, 33 (2d Cir. 1989); Miami N., Inc. v. United States Dep’t of Labor, 939 F. Supp. 53, 56 (D. Me. 1996).}
\footnote{171}{Sheridan v. United States, 487 U.S. 392, 403 n.8 (1988).}
\footnote{172}{See, e.g., Leleux, 178 F.3d at 756; Guccione, 878 F.2d at 33; Miami N., Inc., 939 F. Supp. at 56.}
\footnote{173}{See, e.g., Leleux, 178 F.3d at 757; Guccione, 878 F.2d at 33; Miami N., Inc., 939 F. Supp. at 56.}
\footnote{174}{178 F.3d 750, 757 (5th Cir. 1999).}
\footnote{175}{Id. at 753.}
federal government under the FTCA. 176 Although a compelling argument could be made that the government owes a duty of care to the young people it actively recruits, the circuit court determined that this was an employee-third party relationship with no special duty owed to the plaintiff. 177 Further, the Fifth Circuit found that any government negligence that existed was directly related to the government’s employment relationship with Sistrunk and therefore barred by the assault and battery exception. 178 Thus, the circuit court held that the plaintiff’s case was barred by the assault and battery exception because the government did not owe an independent duty to the plaintiff. 179

Similarly, in Franklin v. United States, 180 the Tenth Circuit found that the lack of an independent duty could bar a claim under the FTCA. 181 In Franklin, a patient at a VA hospital in Oklahoma died after an operation was performed on him without his consent. 182 Analyzing general and Oklahoma tort law, the circuit court found that the performance of an operation without consent sounds in medical battery and not negligence. 183 Although the court cited Sheridan, it found that any duty owed the plaintiff was a direct result of the hospital employees’ employment status. 184 Therefore, there was no duty owed the plaintiff independent of the hospital workers employment relationship with the federal government and Sheridan was inapplicable. 185 While the court found that the assault and battery exception would normally provide a bar to this case, a narrow exception authorizing government liability for torts committed by VA personnel allowed the case to go forward. 186

Two other courts have also found that the Sheridan rationale does not extend to claims for negligent hiring, retention or supervision. In Bajkowski v. United States 187 a district court in the Eastern District of North Carolina held that the assault and battery exception barred a claim for negligent retention and supervision of an employee of the United

176. Id.
177. Id. at 758.
178. Id. at 757–58.
179. Id. at 759.
180. 992 F.2d 1492 (10th Cir. 1993).
181. See id. at 1498–99.
182. Id. at 1495.
183. Id. at 1497.
184. Id. at 1499.
185. Id.
186. Id. at 1499–1502.
States Army. The army re-enlisted the assailant despite his criminal record. Three weeks before he attacked the plaintiff, the Army released the assailant pending a trial for rape. The court held that if the assailant were not a member of the armed forces at the time of the attack, the government would have no duty to supervise him or monitor his behavior. Therefore, the case did not fall within the scope of Sheridan and was barred by the assault and battery exception to the FTCA.

Similarly, in Malone v. United States, a district court in the Southern District of Georgia found that a claim for negligent supervision of an employee of the United States Army was barred by the assault and battery exception. In Malone, the employee, Woods, had already violently raped another woman when he attacked the plaintiff. After he was arrested for this crime, his commanding officers placed him under restricted status. However, they transferred Woods to another location and the commanding officer in charge of Woods’ restriction did not tell anyone who might have been able to stop Woods from leaving the base about his restricted status. Woods violated this restriction, left the base, met the plaintiff in a bar and forcibly raped and sodomized her. Despite this lax supervision of a known rapist, the court determined that a claim for negligent supervision was barred by the assault and battery exception to the FTCA because there was no affirmative, independent duty owed to the plaintiff outside of the federal government’s employment relationship with the assailant.

In sum, many courts have refused to extend the logic of Shearer to negligent hiring, retention and supervision. These courts have restricted their jurisdiction to those cases where a clear duty, independent of the employment relationship, exists between the United States and the victim of the assault and battery.

188. Id. at 540.
189. Id.
190. Id.
191. Id.
192. Id. at 541–42.
194. Id. at 1380–82 (holding that plaintiff’s claims also barred by both Georgia immunity and the Mindes rule that judicial review of military decisions should be avoided).
195. Id. at 1374.
196. Id. at 1376.
197. Id.
198. Id. at 1376–77.
199. Id. at 1380–81.
C. Negligent Hiring, Retention, and Supervision Claims are Actionable Under the FTCA in the Ninth Circuit

Even before Sheridan, the Ninth Circuit had interpreted the assault and battery exception to be inapplicable to cases involving negligent hiring, supervision or retention of a government employee.\(^{200}\) In Bennett v. United States,\(^{201}\) an early Ninth Circuit case, the court allowed a suit to go forward for negligent hiring and retention of a government employee.\(^{202}\) In Bennett, the Bureau of Indian Affairs (BIA) hired a teacher for one of its boarding schools who had previously been arrested and charged with child molestation.\(^{203}\) While he was at the BIA boarding school, the teacher kidnapped, assaulted and raped several students.\(^{204}\) The teacher had admitted on his application that he had been arrested and charged with violating an Oklahoma statute outlawing “Outrage to Public Decency,” but the BIA did not investigate this admission.\(^{205}\) Such an investigation would have revealed that the prior charges were for molestation similar to what had occurred at the BIA boarding school.\(^{206}\) The government admitted that hiring and retaining the teacher was negligent, but argued that the claim was barred by the assault and battery exception to the FTCA.\(^{207}\) The Ninth Circuit disagreed and held that the assault and battery exception was not a bar to liability.\(^{208}\) Although the court noted the Shearer plurality’s conclusion that the assault and battery exception bars claims for negligent supervision of an employee who commits an intentional tort, it declined to follow this rationale.\(^{209}\)

The Bennett court had several rationales for deciding that the assault and battery exception does not bar a claim for negligent supervision of a government employee who commits an assault and battery.\(^{210}\) The Bennett court relied on Ninth Circuit precedent in Jablonski v. United States,\(^{211}\) in which the court held that the assault and battery exception

\(^{200}\) See, e.g., Bennett v. United States, 803 F.2d 1502, 1503–05 (9th Cir. 1986).
\(^{201}\) 803 F.2d 1502 (9th Cir. 1986).
\(^{202}\) Id. at 1504–05.
\(^{203}\) Id. at 1503.
\(^{204}\) Id.
\(^{205}\) Id.
\(^{206}\) Id.
\(^{207}\) Id.
\(^{208}\) Id. at 1505.
\(^{209}\) See id. at 1503.
\(^{210}\) See id.
\(^{211}\) 712 F.2d 391 (9th Cir. 1983).
did not excuse government negligence. In Jablonski, the court noted that one of the major policy reasons behind the intentional tort exceptions to the FTCA was to prevent the government from having to defend lawsuits for acts it was powerless to prevent. Applying this rationale to the Bennett case, the court determined that the government should have known that the teacher posed a serious risk to the children at the BIA boarding school and therefore liability was not based on respondeat superior, and therefore an employee’s act it was powerless to prevent, but on the government’s own independent negligence. Thus, the problem of being forced to defend lawsuits for acts beyond its power was not an issue.

Further, the Ninth Circuit dismissed the Shearer plurality’s notion that the government’s immunity is waived only if the tortfeasor is a non-employee. The Bennett court ruled that this conclusion was anything but certain. The court discerned no evidence in the congressional record that Congress meant to open the government up to liability for negligently supervising non-employees but not its own employees. Additionally, the court recognized a clear distinction between claims for respondeat superior, which are barred by the assault and battery exception, and claims based on the government’s own negligence. Finally, the court noted that the broad immunity suggested by the Shearer plurality was inconsistent with the overarching purpose of the FTCA: to provide redress for those who have been injured as a proximate result of government negligence.

The Ninth Circuit continued to rely on this analysis after the Sheridan opinion. In Brock v. United States, the Ninth Circuit held that the assault and battery exception did not bar a claim by a U.S. Forest Service employee who was sexually harassed and raped by her supervisor, McKinney, on assignment in the field. After she was raped, she

212. See Bennett, 803 F.2d at 1503–05.
213. See id. at 1503–04.
214. See id. at 1503 (citing Jablonski ex rel. Pahls v. United States, 712 F.2d 391 (9th Cir.1983)).
215. See id.
216. See id. at 1504.
217. See id.
218. Id.
219. Id.
220. Id.
221. 64 F.3d 1421 (9th Cir. 1995).
222. Id. at 1425.
refused to go back into the field with McKinney and was transferred to a
desk job where sexual harassment and unwanted touching by McKinney
continued.\textsuperscript{223} Eventually, the plaintiff was transferred to another
department to avoid further contact with McKinney.\textsuperscript{224} After filing a
claim against the government for McKinney’s behavior and for
government negligence in failing to properly supervise him, although
they knew of his behavior, she was subject to torment by her
colleagues.\textsuperscript{225} The Ninth Circuit refused to hold that the claims for
negligent supervision were barred by the assault and battery exception to
the FTCA.\textsuperscript{226} The court recognized that the question remained open in the
wake of \textit{Sheridan}\textsuperscript{227} and relied on Ninth Circuit precedent in \textit{Bennett} to
hold that claims for negligent hiring and supervision of employees who
commit intentional torts are not barred by the assault and battery
exception.\textsuperscript{228}

The Ninth Circuit reaffirmed this conclusion most recently in \textit{Senger v. United States.}\textsuperscript{229} In \textit{Senger}, a tow truck operator was attempting to tow
a U.S. Postal Service employee’s truck from the parking lot of the Main
Post Office in Portland at the request of the Postal Service.\textsuperscript{230} While
doing this, he was assaulted by the car’s owner, a Postal Service
employee.\textsuperscript{231} The court held that the plaintiff’s claim that the post office
inadequately supervised an employee for a business invitee was not
barred by the assault and battery exception.\textsuperscript{232} The Ninth Circuit
reaffirmed the distinction between claims for respondeat superior, which
are barred by the assault and battery exception, and claims for negligent
supervision, which are not barred.\textsuperscript{233}

As set forth above, the Ninth Circuit continues to hold that the assault
and battery exception is not a bar to claims for negligent hiring, retention,
and supervision despite opinions to the contrary. Many of the
contrary opinions come from courts that have held that their jurisdiction

\begin{itemize}
\item 223. \textit{Id.} at 1422.
\item 224. \textit{Id.}
\item 225. \textit{Id.}
\item 226. \textit{Id.} at 1425.
\item 227. \textit{Id.} (citing \textit{Sheridan v. United States}, 487 U.S. 392, 403 n.8 (1988)).
\item 228. \textit{Id.} at 1425.
\item 229. 103 F.3d 1437 (9th Cir. 1996).
\item 230. \textit{Id.} at 1438.
\item 231. \textit{Id.}
\item 232. \textit{Id.} at 1442–43.
\item 233. \textit{Id.} at 1441.
\end{itemize}
is constrained to cases where a clear duty, independent of the employment relationship, exists between the government and the plaintiff. The Ninth Circuit reached the opposite conclusion by ruling that the intent behind the FTCA was to provide redress to citizens injured by governmental negligence, and therefore the assault and battery exception is not a bar to claims for negligent hiring, retention, and supervision. Further, the Ninth Circuit has distinguished between those claims, which are rooted in negligence, and those that are based solely on respondeat superior and that negligent hiring, retention, and supervision claims are allowed under the FTCA.

V. THE ASSAULT AND BATTERY EXCEPTION TO THE FEDERAL TORTS CLAIM ACT DOES NOT BAR CLAIMS FOR NEGLIGENT HIRING, RETENTION OR SUPERVISION

Synthesizing the language and history of the FTCA, the substance of a claim for negligent hiring, retention or supervision, and past precedent, reveals that the assault and battery exception does not deny federal district courts jurisdiction to hear such claims. Reading the assault and battery exception to bar federal district courts from hearing negligent hiring, retention, and supervision claims against the federal government substantially frustrates the overarching purpose of the FTCA to provide a remedy for the negligent acts of government officials. Further, the negligent acts of federal employees in the hiring, retention and supervision of other federal employees are readily distinguishable from the intentional assaults and batteries themselves. Because virtually all FTCA claims are heard before a judge, claims that truly are barred can be dismissed through the normal channels of federal civil procedure without the danger that a jury will confuse a claim for respondeat superior with negligence. Finally, the United States Supreme Court’s use of traditional tort law in Sheridan urges a reading of the assault and battery exception which allows federal district courts jurisdiction over claims based on a government employee’s negligence in hiring, retaining or supervising another federal employee.

234. See supra notes 48–51 and accompanying text.
235. See supra Part I.
236. See supra note 55.
237. See supra notes 138–42 and accompanying text.
A. **Interpreting the Assault and Battery Exception to Bar only Respondeat Superior Claims is Consistent with its Legislative History and Language**

Interpreting the assault and battery exception in the context of its legislative history and the FTCA as a whole, it is clear that the exception’s scope should be limited to claims for respondeat superior. The assault and battery exception’s legislative history indicates that Congress intended to provide a mechanism for redress for citizens injured by the negligent acts of government employees. Additionally, the FTCA only provides liability for the negligent acts of employees acting within their scope of duty. Once employees act outside of that scope of duty, the FTCA no longer provides a remedy. If the FTCA does not provide a remedy, then the assault and battery exception is irrelevant. However, the FTCA does provide a remedy for the negligence of a federal employee in hiring, retaining and supervising another federal employee who commits an assault and battery. This is because such a claim is predicated on the government’s negligence, not the assault and battery of the employee, for which the FTCA does not provide a remedy regardless of the assault and battery exception. Therefore, the government is liable for the negligent acts of federal employees who supervise employees who commit an assault and battery outside of their scope of duty. Finally, the ultimate purpose of the FTCA is to provide a remedy for citizens wronged by the negligence of government officials and interpreting the assault and battery exception too broadly undermines this purpose.

The sparse legislative history associated with the assault and battery exception to the FTCA indicates that Congress intended to waive the government’s immunity for negligent acts or omissions of the Federal Government. One of the few mentioned underlying concerns driving the passage of the assault and battery exception was that the government did not want to have to litigate intentional torts that are difficult to

---

238. *See supra* notes 60–64 and accompanying text.
239. *See supra* notes 129–34 and accompanying text.
241. *See supra* notes 133–34 and accompanying text.
243. *See supra* note 134 and accompanying text.
244. *See supra* note 51 and accompanying text.
245. *See supra* notes 60–64 and accompanying text.
defend and easily exaggerated.246 Negligent hiring, retention and supervision claims, on the other hand, are not more difficult to defend than other types of negligence claims for which the government has waived immunity.247 Congress’ concern, paired with the relative ease of defending negligence claims, exposes Congress’s intent that the assault and battery exception is only a bar to claims for vicarious liability for the intentional torts of government employees. Claims for negligent hiring, retention and supervision require the plaintiff to prove the same elements of duty, breach, causation and harm as in any other negligence claim.248 If Congress was concerned about the government’s liability for more easily proven claims, then they were most likely referring to claims for respondeat superior liability for intentional torts which merely require a showing that an intentional tort was committed by an employee of the federal government within the employee’s scope of employment.249

The later Congressional Amendment to the assault and battery exception allowing for government liability when federal law enforcement officers commit an intentional assault and battery does not alter the fundamental analysis of the exception. While the Shearer plurality reasoned that this amendment to the assault and battery exception demonstrated Congress’ belief that the government would not be liable in any way when a federal employee commits an assault and battery, the plurality did not recognize that these assaults and batteries are the very few that will most likely be committed in the scope of duty.250 This amendment to the assault and battery exception allows for liability when federal investigative or law enforcement officers are accused of assault, battery, false imprisonment, false arrest, abuse of process, or malicious prosecution.251 Congress detailed part of these officers’ job description in the amendment, noting that law enforcement and investigative officers have the power to execute searches, to seize evidence, and to make arrests for violations of Federal law.252 This inclusion evidences Congress’ acknowledgment that these types of

246. See Zolensky, supra note 63, at 811 n.43.
247. See id. at 811–12 (noting that while the element of intent in an intentional assault and battery is easily proven, proving that the negligence of an employee caused that assault and battery is not).
249. See RESTATEMENT (SECOND) OF AGENCY § 228, at 504 (1958); Zolensky, supra note 63, at 811.
250. See supra notes 90–92 and accompanying text.
252. Id.
officers will be accused of on-the-job assaults and batteries.\textsuperscript{253} Because these assaults and batteries are committed within the employee’s scope of duty they are exempt from this Comment’s analysis: these assaults and batteries are within the scope of the FTCA and were, prior to the amendment, barred by the assault and battery exception.\textsuperscript{254} In contrast, assaults and batteries committed \textit{outside} of the scope of duty are not within the ambit of the FTCA and therefore antecedent negligence that caused these assaults and batteries may subject the government to liability under the FTCA.\textsuperscript{255} This amendment does not support the \textit{Shearer} plurality’s conclusion that the assault and battery exception bars claims for the negligent hiring, retention and supervision of government employees acting \textit{outside} of their scope of duty.

Analyzing the assault and battery exception as part of a larger piece of legislation, it is clear that it only applies to respondeat superior claims for intentional torts committed within the scope of the employee’s duty. As the U.S. Supreme Court noted in \textit{Sheridan}, the FTCA only waives sovereign immunity for injuries caused by an “employee of the Government while acting within the scope of his office or employment.”\textsuperscript{256} Therefore, according to the Court, the assault and battery exception does not apply to assaults and batteries committed by employees acting outside of their scope of employment.\textsuperscript{257} The \textit{Sheridan} Court concluded that the assault and battery exception did not apply to the assault and battery in that case because the assailant was acting outside of his scope of duty.\textsuperscript{258} The Court then determined that the government had breached a duty owed to the plaintiff to report persons who had unauthorized weapons on the base and that this breach was the proximate cause of the plaintiffs’ injuries.\textsuperscript{259} Similarly, if the government’s negligent hiring, retention or supervision of an employee is the proximate cause of an assault and battery by a federal employee acting outside of his or her scope of duty, the assault and battery exception is irrelevant.\textsuperscript{260} Once the assault and battery exception has been removed from the equation, the district courts have jurisdiction because

\begin{footnotes}
\footnotetext{253}{See id.}
\footnotetext{254}{See supra notes 132–34 and accompanying text.}
\footnotetext{255}{See supra notes 129–42.}
\footnotetext{257}{Id.}
\footnotetext{258}{Id. at 401–02.}
\footnotetext{259}{See id. at 402.}
\footnotetext{260}{See supra notes 132–34.}
\end{footnotes}
the claim is predicated on the government’s negligence, not the employee’s assault and battery. To conclude otherwise would deny the premise relied on by the Supreme Court in *Sheridan* that the assault and battery exception does not apply to claims where the government employee acted outside of his or her scope of duty.

Viewed in light of the overarching purpose of the FTCA, the ambiguous language of the assault and battery exception should be construed to allow district courts jurisdiction to hear claims for negligent hiring, retention and supervision. While the arising out of language could literally be read to deny federal district courts jurisdiction for any claim related to an assault or battery, this language becomes ambiguous when an injury is proximately caused by both an intentional tort and negligence. When this happens the injury could be said to arise out of negligence or the intentional tort and it is necessary to consider the intent of the entire FTCA to properly interpret the exception. While the main purpose of the FTCA was to replace the arduous private bill process used at the time to compensate plaintiffs for injury at the hands of the federal government, a second and almost equally important purpose was to compensate plaintiffs who had been injured by the federal government’s negligence. This second purpose makes clear that the FTCA waives sovereign immunity for claims arising out of negligence but preserves it for claims arising out of intentional torts that were not caused by the independent negligence of the federal government. Taken as a whole, the FTCA and the legislative history of the assault and battery exception suggest that the exception should not be interpreted to deny government liability for negligent hiring, retention and supervision claims.

B. The Assault and Battery Exception is Not a Bar to Actions Claiming Negligent Hiring, Retention and Supervision of Government Employees Because Such Torts are Analytically Distinct from the Assault and Battery by an Employee

Claims founded on negligent hiring, retention and supervision of employees who act outside of their scope of duty should not be barred by

---

261. See supra notes 132–34.
262. See *Sheridan*, 487 U.S. at 400.
265. See supra note 40 and accompanying text.
the assault and battery exception to the FTCA because they are distinct from claims based solely on the intentionally tortious conduct of the employee. Negligent hiring, retention and supervision claims are based on the independent negligence of the employer, in this case the federal government, and thus do not arise solely out of the assault and battery of the employee. While courts purport to fear that claims that are actually based on vicarious liability will be couched in terms of negligent hiring, retention and supervision, the two torts are analytically distinct and any hidden respondeat superior claims can be dispensed with through the normal channels of the Federal Rules of Civil Procedure.

1. Negligent Hiring, Retention and Supervision Claims are based on the Independent Negligence of the Employer and are Readily Distinguished from Claims for Respondeat Superior

Although the injury involved in a negligent hiring, retention and supervision claim can be the result of an intentional assault and battery, such claims are founded on the employer’s independent negligence and not solely on the employee’s assault and battery. In most jurisdictions, in order to prove negligent hiring, retention or supervision, the plaintiff cannot just prove that the employee committed an intentional tort, but must prove that the employer knew, or reasonably should have known, that the employee was dangerous. Further, as with any independent negligence claim, the plaintiff must prove that the employer’s negligence proximately caused the injury. Therefore, a claim for negligent hiring, retention and supervision is rooted in the independent negligence of the employer and arises out of that negligent act or omission. Since the assault and battery exception only bars those claims which arise out of intentional torts, this exception should not bar claims based on the independent negligence of a federal government employee acting within

268. Id.
272. Id.
his or her scope of duty when hiring, retaining or supervising another federal employee.274

Respondeat superior claims, which are based on vicarious liability and are barred by the assault and battery exception, are readily distinguishable from claims based on negligent hiring, retention and supervision.275 Respondeat superior claims are founded entirely on the employee’s intentionally tortious act and do not require any proof of negligence on the part of the employer.276 Claims based on the theory of respondeat superior posit that the employer should be liable for the wrongful acts of his or her employee if that employee was acting on the employer’s behalf, that is, within the employee’s scope of duty.277 Such claims arise out of the intentionally tortious conduct of the employee and have no basis in the employer’s independent negligence.278 In contrast, negligent hiring, retention and supervision claims are not based solely on the employee committing an assault and battery but have separate and distinct roots in the employer’s negligence.279

2. Courts Can Dismiss Negligent Hiring, Retention and Supervision Claims that are Revealed to be Claims for Respondeat Superior

The risk that plaintiffs will successfully disguise their respondeat superior claims with allegations of negligent supervision is overstated. The assault and battery exception does not merely prevent a cause of action from arising, but denies the federal district courts subject matter jurisdiction for claims arising out of assault and battery.280 Under Federal Rule of Civil Procedure 12(h)(3), respondeat superior claims that are disguised as negligent hiring, retention and supervision claims must be dismissed as soon as their true nature becomes clear.281 Although the notice pleading requirements of the Federal Rules of Civil Procedure only require notice of the claims that the plaintiff is alleging, such pleadings still require the plaintiff to inform the defendant of the

274. Bennett v. United States, 803 F.2d 1502, 1504 (9th Cir. 1986).
276. Id.
278. See id.
allegations.\textsuperscript{282} As a result, the lack of subject matter jurisdiction is often evident at the pleading stage of litigation if the pleadings do not allege any circumstances beyond the employee’s tortious conduct.\textsuperscript{283} Even if the litigation moves beyond the pleading stage, the government may make a motion for summary judgment to uncover whether any factual basis exists for the negligence claim.\textsuperscript{284} Utilizing the rules of civil procedure would preserve the purpose of the assault and battery exception, to avoid litigating a claim arising out of assault and battery. It also would advance the overall purpose of the FTCA: to provide a forum for meritorious claims against the government for its own negligence.

Although one could argue that the fact finder could confuse the concepts of negligent hiring, retention and supervision with respondeat superior, this concern is practically non-existent in the context of FTCA litigation. While the distinctions between respondeat superior and negligent hiring, retention and supervision are clear in theory, it is true that the distinction may begin to blur for a jury faced with a sympathetic plaintiff, an employee who has done something wrong, and an employer with deep pockets. However, almost all FTCA cases are argued before a judge,\textsuperscript{285} and therefore the danger of confusing the two doctrines is greatly reduced.

In sum, the negligent hiring, retention and supervision claims are not barred by the assault and battery exception to the FTCA. Such claims are analytically distinct from the federal employee’s intentionally tortious actions and therefore cannot be disguised as a claim for respondeat superior. Any risk that these respondeat superior claims are masked as negligent hiring, retention, or supervision is negated by the fact-finding role played by the federal judge in FTCA claims. Because a judge and not a jury will determine the nature of the claim asserted, the judge can dismiss any meritless claims through the normal channels of civil procedure.

\textsuperscript{282} Jack H. Friedenthal et al., Civil Procedure § 5.7, at 258 (3d ed. 1999).
\textsuperscript{283} See Bryson v. United States, 463 F. Supp. 908, 912 (E.D. Penn. 1978) (noting that courts can and have examined the facts of a case to see if there is a claim evident beyond the assault and battery of the employee. If there is not, the courts simply dismiss the claim).
\textsuperscript{284} See Friedenthal et al., supra note 282, at § 9.1, 452.
\textsuperscript{285} See 28 U.S.C. § 2402 (2000); see also Sisk, supra note 55, at 72–79.
Federal Tort Claims Act

C. Allowing Negligent Hiring, Retention and Supervision Claims when an Employee Acts Outside of his or her Scope of Employment is Consistent with Sheridan v. United States and Ninth Circuit Precedent

The U.S. Supreme Court in *Sheridan* refused to interpret the “arising out-of” language of the assault and battery exception to bar any claim that would not exist but-for the intentional tort.\(^{286}\) Instead, the Court adopted an interpretation based on principles of basic tort law.\(^{287}\) The Court held that the government can be held liable for the breach of an independent duty owed to the plaintiff that causes an assault and battery by a government employee acting outside of his or her scope of duty.\(^{288}\) The Court’s refusal to adopt this rationale implies that the FTCA’s assault and battery exception does not apply to federal employees’ negligent acts of hiring, retention or supervision of other employees who commit intentional torts outside of their scope of duty.

The Supreme Court’s traditional tort law analysis in *Sheridan* requires a reading of the assault and battery exception to grant jurisdiction to federal district courts to hear claims for the negligent hiring, retention and supervision of federal employees. The *Sheridan* Court determined that the assault and battery exception did not bar claims based on an employee’s intentional assault and battery while acting outside the scope of duty, if the government owed a duty to the victim that was independent of the employment relationship.\(^{289}\) The Court concluded that because the employee in *Sheridan* was not acting within his scope of duty when he shot at the plaintiffs’ car, the FTCA would not have provided a remedy regardless of the assault and battery exception.\(^{290}\) The *Sheridan* Court thus rejected the *Shearer* plurality’s insistence on a reading of the assault and battery exception that would bar any cause of action connected to an intentional assault and battery, and instead held that the plaintiff could proceed against the government based on its independent negligence.\(^{291}\)

Although the *Sheridan* Court couched its analysis in terms of the employee acting outside the scope of duty, the Court ultimately held that

\(^{287}\) See id.
\(^{288}\) See id. at 403.
\(^{289}\) See id. at 402–03.
\(^{290}\) See id. at 401.
\(^{291}\) Id. at 402–03.
the exception’s arising out of language does not apply where the claim is
based on a negligent act of the government independent of the intentional
tort of the assailant.292 Employing this analysis, it does not matter
whether that duty was based on the employment relationship or on some
other independent duty owed. If the federal government owes a duty to
the plaintiff under the law of the forum state, then an injury caused by a
breach of that duty should be actionable under the FTCA. The FTCA
provides a remedy for the negligent acts of government employees293
while the assault and battery exception exempts the government from
liability for the solely intentional torts of its employees.294 This exception
does not apply if the assault and battery was made possible by the
independent, negligent act of a government employee within the ambit of
the FTCA.295 Distinguishing between distinct duties owed based on
whether they are independent of the employment relationship is arbitrary
and relies on the faulty logic of the but-for analysis rejected by the
Sheridan court.296

Finally, the Ninth Circuit’s rationale in *Bennett v. United States*
is a
persuasive reason for allowing suits for the negligent hiring, retention
and supervision of government employees who commit assaults and
batteries outside of their scope of duty. First, these types of suits do not
implicate the major policy concern motivating the assault and battery
exception; they do not make the government liable for suits that are
difficult to defend and for actions that it was powerless to prevent.297 The
premise behind the tort of negligent hiring, retention and supervision is
that the employer knew or should have known that its employee was a
threat to third parties and therefore could have prevented the employer’s
tortious act.298 Second, the Ninth Circuit noted that there is no evidence
in the legislative history of the FTCA that Congress intended to accept
liability for the negligent supervision of non-employees but not its own
employees.299 If the government is liable for negligent supervision for
one group of individuals, it is logically inconsistent to exempt the
government from liability for the identical negligent supervision of a

---

292. *Id.* at 401.
294. *Id.* § 2680(h).
296. *Id.*
298. *See supra* note 30 and accompanying text.
299. *See supra* notes 216–18 and accompanying text.
different group of individuals. Third, as the Ninth Circuit noted, negligent hiring, retention and supervision claims are distinct from respondeat superior.\textsuperscript{300} Finally, the court recognized that a broad interpretation of the assault and battery exception is inconsistent with the FTCA’s purpose to provide a remedy to individuals injured by government negligence.\textsuperscript{301} Ultimately, both Supreme Court and Ninth Circuit precedent urge a reading of the assault and battery exception that makes the government liable for the negligent hiring, retention, and supervision of employees who commit assault and battery.

V. CONCLUSION

The FTCA’s assault and battery exception should not be interpreted to bar federal courts from hearing claims for the negligent hiring, retention and supervision of government employees. Such a reading substantially frustrates one of Congress’ central purposes in enacting the FTCA: providing a remedy for injuries caused by the negligent acts of government officials. Claims for negligent hiring, retention and supervision are rooted in the negligence of government officials. Although some courts may fear that allowing such claims would allow barred respondeat superior claims to slip through the cracks, this fear is unfounded. The federal district court judges who hear negligent hiring, retention and supervision claims can dismiss those that are revealed to be respondeat superior claims. Finally, allowing these claims is a natural extension of U.S. Supreme Court’s logic in \textit{Sheridan v. United States}. The Court recognized that the negligent acts of government employees open the government up to liability under the FTCA regardless of whether the injury was caused by an assault and battery by a government employee acting outside of his scope of duty.

---

300. \textit{See supra} note 219 and accompanying text.
301. \textit{See supra} note 220 and accompanying text.
BACK TO PRIMA PAINT CORP. V. FLOOD & CONKLIN MANUFACTURING CO.: TO CHALLENGE AN ARBITRATION AGREEMENT YOU MUST CHALLENGE THE ARBITRATION AGREEMENT

Andre V. Egle

Abstract: The Federal Arbitration Act (FAA) requires courts to order parties in a dispute arising out of a commercial contract containing an arbitration provision to proceed to arbitration unless the formation or performance of the arbitration agreement itself is at issue. In 1967, the U.S. Supreme Court held in Prima Paint Corp. v. Flood & Conklin Manufacturing Co. that under the FAA, courts, instead of arbitrators, should resolve claims for fraudulent inducement of arbitration agreements. However, courts were not permitted to resolve claims for fraud in the inducement of the underlying commercial contracts. The Court also held that when deciding whether to enforce an arbitration agreement, a court should only consider the issues related to making and performing that agreement. The federal circuit courts have applied Prima Paint in two ways. The Third, Ninth, and Eleventh Circuits have held that a court may consider a claim that a commercial contract containing an arbitration agreement is void, even if the party has not alleged that the arbitration agreement is invalid. In contrast, the Fifth and Sixth Circuits have held that alleging that a contract is void is not enough to put the contract’s arbitration agreement at issue under the FAA and Prima Paint. This comment argues that to put an arbitration agreement at issue a party should specifically plead that it is invalid. A mere allegation that the underlying commercial contract is void is insufficient because federal law encourages arbitration and treats arbitration agreements as severable from the contracts in which they are included. Only if a court finds that an arbitration clause is merely a part of the underlying commercial contract should the court resolve a claim that the contract is void.

In 1925, Congress passed the United States Arbitration Law codified as the Federal Arbitration Act (FAA). The FAA was Congress’ response to the reluctance of federal courts to enforce arbitration agreements. The FAA requires courts to order arbitration of a dispute arising out of a contract containing an arbitration provision unless the formation or performance of the arbitration agreement is not in issue. Since then, arbitration of controversies arising out of maritime and commercial transactions has become an increasingly popular means of dispute resolution. However, it is common for the parties to arbitration agreements to challenge the validity of the agreements in a judicial forum. The challenges are especially common when a commercial or

2. See Julius H. Cohen & Kenneth Dayton, The New Federal Arbitration Law, 12 VA. L. REV. 265, 265 (1926) (noting that the FAA “reversed the hoary doctrine that agreements for arbitration are revocable at will and are unenforceable”).
maritime contract contains broad boilerplate language mandating arbitration of “any controversy or claim arising out of or relating to [the] contract.” In many cases, the party that wants to avoid arbitration asserts that the arbitration agreement is invalid merely because the underlying contract is deficient. Lower federal courts have reached different conclusions as to whether a federal court or an arbitrator should resolve these challenges in light of the relevant provisions of the FAA and the U.S. Supreme Court’s decision in Prima Paint Corp. v. Flood & Conklin Manufacturing Co.

The FAA is a powerful statute that governs enforceability of the majority of commercial arbitration agreements throughout the nation and preempts any additional requirements for arbitration agreements imposed by states. The FAA mandates that federal courts enforce an arbitration provision “in any maritime transaction or a contract evidencing a transaction involving commerce” if the provision satisfies three conditions. First, it must be in writing. Second, the arbitration provision must relate to a maritime transaction or a transaction involving interstate commerce. Third, the arbitration agreement must be valid and able to withstand any legal or equitable grounds for the revocation of any contract.

Many courts have addressed the FAA’s requirement that, before a federal court may enforce an agreement to arbitrate, the agreement must be as valid as any other contract. More specifically, courts have reached different conclusions regarding whether a court or an arbitrator should resolve challenges to the validity of an arbitration clause that is a part of a larger commercial contract. A partial answer to this question was provided by the U.S. Supreme Court in 1967 in Prima Paint. There, the

---

7. See Southland Corp. v. Keating, 465 U.S. 1, 10–15 (1984) (explaining that the FAA has created a substantive federal law that preempts conflicting state law provisions as to enforceability of arbitration agreements involving interstate commerce and maritime transactions).
8. 9 U.S.C. § 2 defines “commerce” for FAA purposes as “commerce among the several States or with foreign nations.” Section 1 further specifically excludes from “commercial contracts” under the FAA “contracts of employment of seamen, railroad employees, or any other class of workers engaged in foreign or interstate commerce.” Id. § 1.
10. See id.
11. See id. § 2.
12. Id.
Challenging Arbitration Agreements Under the FAA

Court held that when determining the validity of an arbitration agreement a federal court can only consider issues related to the formation and performance of the arbitration agreement itself.\(^\text{14}\) The Court also stated that if a party challenges the arbitration clause itself as fraudulently induced, courts may proceed to resolve the challenge.\(^\text{15}\) However, the Court did not settle the question of whether a claim that the entire contract, and thus an arbitration clause contained therein, is void and unenforceable should be resolved in court or in arbitration.

*Prima Paint* has prompted conflicting decisions in the federal Circuit Courts regarding the validity of arbitration agreements under the FAA.\(^\text{16}\) The main point of controversy among the Circuit Courts relates to a situation where a party challenges the arbitration agreement by alleging that the underlying commercial contract is void, instead of voidable. A contract is voidable if it is generally valid, but one or more parties to the contract has the power to avoid the contractual relationship on grounds such as fraud in inducement of the contract, duress, or mistake.\(^\text{17}\) In contrast, a contract is void if one of its essential elements, such as mutual assent, is missing.\(^\text{18}\) A void contract is not a contract at all; it is a “promise” or “agreement” that is void of legal effect.\(^\text{19}\)

Circuit Courts for the Third, Ninth, and Eleventh Circuits have held that *Prima Paint*’s holding—that courts should only resolve allegations of the invalidity of an arbitration agreement itself—does not apply when the validity of the entire contract is challenged.\(^\text{20}\) Therefore, these courts hold that judges—not arbitrators—should consider the contract’s validity. In contrast, the appellate courts for the Fifth and Sixth Circuits have taken the position that even if a party to the contract claims that the contract is void, the party must specifically allege that the arbitration provision is invalid before the court may consider the challenges to the contract.\(^\text{21}\)

\(^{14}\) Id. at 404.

\(^{15}\) Id.

\(^{16}\) See Burden v. Check Into Cash of Ky., LLC, 267 F.3d 483, 488 (6th Cir. 2001) (summarizing approaches of Courts of Appeals to interpretation of *Prima Paint*).

\(^{17}\) See RESTATEMENT (SECOND) OF CONTRACTS § 7 cmt. b (1981).

\(^{18}\) See id. cmt. a.

\(^{19}\) See id.


This Comment argues that under *Prima Paint*, federal courts should determine the enforceability of an arbitration provision in a commercial or maritime contract only if a party to the contract specifically alleges that the provision should be revoked as invalid at law or in equity. A specific challenge to an arbitration clause is required even in situations where a party to the contract alleges that the entire contract is void. As an alternative, a party may challenge an arbitration clause by asserting the invalidity of the entire contract only if the arbitration clause is not severable from the contract. Part I explains that the FAA promotes enforcement of arbitration agreements and that, in most cases, the FAA and the courts treat arbitration agreements as severable from the rest of the contract. Part II discusses the U.S. Supreme Court’s opinion in *Prima Paint*. Part III discusses Circuit Courts’ interpretation of *Prima Paint* and describes the two major approaches used in the different Circuits. Part IV argues that the FAA and *Prima Paint* require parties who ask federal courts to resolve disputes arising under contracts containing an arbitration provision to specifically allege and prove that the provision is invalid.

I. THE FAA PROMOTES ENFORCING ARBITRATION AGREEMENTS AND DIRECTS FEDERAL COURTS TO ENFORCE THEM AS SEVERABLE FROM THE UNDERLYING CONTRACTS

Congress adopted the FAA to create an enforcement mechanism for otherwise valid arbitration agreements in commercial and maritime transactions.\(^{22}\) The FAA was intended to further arbitration as a means of alternative dispute resolution that would eliminate the delays associated with judicial proceedings, avoid the expense of litigation, and promote decisions regarded as just in the business world.\(^{23}\) Federal courts had in the past treated arbitration agreements as not binding and revocable at will.\(^{24}\) However, Congress strongly encouraged courts to enforce arbitration agreements by explicitly stating in the FAA that written arbitration provisions in maritime transactions and transactions involving interstate commerce should be “valid, irrevocable, and enforceable.”\(^{25}\)

---

23. Id. at 269.
24. Id. at 265.

202
The U.S. Supreme Court has also favored enforcing arbitration agreements in commercial contracts. In Moses H. Cone Memorial Hospital v. Mercury Construction Corp., the Court determined that Congress declared a liberal policy favoring arbitration agreements in the FAA and created "a body of federal substantive law of arbitrability, applicable to any arbitration agreement within the coverage of the Act." Further, the Court stated that any doubts regarding the scope of arbitrable issues, such as construction of the contractual language or the defenses to arbitration, should be resolved in favor of arbitration.

The FAA is a powerful substantive and procedural statute that preempts all inconsistent state laws and promotes uniformity in enforcement of arbitration agreements. The Act applies to any arbitration agreements relating to interstate commerce or maritime transactions. Significantly, the FAA treats arbitration agreements as severable from the contracts that contain them.

A. The FAA Promotes Arbitration

Prior to the FAA’s enactment, American courts acknowledged arbitration as an option but adhered to the traditional view that arbitration agreements are revocable at will and should not be enforced by courts. As a result, a party that wanted to avoid an arbitration agreement only had to refuse to proceed and the court would not enforce the agreement. Suing for damages arising out of the breach of the arbitration agreement generally could not adequately redress the aggrieved party’s inability to arbitrate. Thus, no meaningful remedy existed for the intentional breach of arbitration agreements. Congress passed the FAA to remedy courts’ reluctance to enforce arbitration agreements.

27. Id. at 24 (interpreting section 2 of the FAA codified as 9 U.S.C. § 2).
28. Id. at 24–25.
29. See infra Part I.B.
30. See infra Part I.B.
31. See infra Part I.C.
32. See Cohen & Dayton, supra note 2, at 270.
33. See id.
34. Id.
35. See Dean Witter Reynolds Inc. v. Byrd, 470 U.S. 213, 219 (1985) (“The legislative history of the Act establishes that the purpose behind its passage was to ensure judicial enforcement of privately made agreements to arbitrate.”).
Section 2 of the FAA, passed in 1925, states that “[a] written provision in any maritime transaction or a contract evidencing a transaction involving commerce to settle by arbitration a controversy thereafter arising out of such contract or transaction . . . shall be valid, irrevocable, and enforceable.”36 According to the statute, arbitration agreements may be enforced in two ways. First, under section 3 of the FAA, a party wishing to enforce an arbitration agreement related to ongoing litigation can apply for a stay of the litigation. 37 In this situation, the court where the law suit is pending, “upon being satisfied that the issue involved in such suit or proceeding is referable to arbitration . . . shall . . . stay the trial of the action until such arbitration has been had in accordance with the terms of the agreement.”38 Second, under section 4 of the FAA, a party purporting to arbitrate a dispute before the other party files a law suit may petition the court for an order directing the parties to proceed with arbitration under the agreement. 39 Here, the court “upon being satisfied that the making of the agreement for arbitration . . . is not in issue . . . shall make an order directing the parties to proceed to arbitration in accordance with the terms of the agreement.”40

Congress adopted the FAA to encourage the arbitration of contract disputes and to provide an enforcement mechanism for private arbitration agreements.41 However, the foremost purpose of the FAA was to encourage courts to enforce arbitration agreements by ordering specific performance in situations where a party refuses to comply with the agreement’s terms.42 No such remedy existed in federal courts prior to enactment of the FAA.43 In addition, Julius Henry Cohen, the American Bar Association’s draftsman of the FAA, suggested that the purpose

37. See id.
38. Id.
39. See id. § 4.
40. Id.
41. See generally Cohen & Dayton, supra note 2; see also Dean Witter Reynolds Inc. v. Byrd, 470 U.S. 213, 219 (1985) (“The preeminent concern of Congress in passing the Act was to enforce private [arbitration] agreements into which parties had entered and that concern requires that [federal courts] rigorously enforce agreements to arbitrate, even if the result is ‘piecemeal’ litigation, at least absent a countervailing policy manifested in another federal statute.”).
42. See Cohen & Dayton, supra note 2, at 271–72.
43. See id. at 276 (explaining that prior to enactment of the FAA federal courts recognized the existence and validity of arbitration agreements, but refused to enforce them by way of specific performance); see also Byrd, 470 U.S. at 219–20 (pointing out that Congress adopted the FAA “to overrule the judiciary’s longstanding refusal to enforce agreements to arbitrate”).
behind the Act was in line with the threefold purpose of arbitration as a means of alternative dispute resolution. 44 First, arbitration would effectively eliminate the long delays usually incident to court proceedings. 45 Second, arbitration would help parties to avoid the expense of litigation. 46 Third, arbitration, rather than regular judicial proceedings, would provide a better means of reaching a decision regarded as just in the business world. 47 The third purpose is particularly important because courts mainly apply general rules that may not fit a particular commercial dispute. 48 Further, in an ordinary jury trial, a dispute may not receive an adequate analysis because of the jurors’ lack of expertise in commercial matters, whereas an experienced commercial arbitrator can skillfully scrutinize a complex dispute. 49

Consistent with arbitration’s purpose to prevent long delays in court proceedings, the U.S. Supreme Court has stated on several occasions that the FAA’s ultimate goal is to “move the parties to an arbitrable dispute out of court and into arbitration as quickly and easily as possible.” 50 For example, in Moses H. Cone, the Court explained that the liberal federal policy favoring the enforcement of arbitration agreements stems directly from section 2 of the FAA. 51 The Court declared that under the FAA, any doubts as to the scope of issues subject to arbitration should be resolved in favor of arbitration. 52 The lower federal courts have followed the Court’s pro-arbitration policy. 53

B. The FAA Preempts Inconsistent State Law

The FAA has both substantive and procedural components. It is substantive because it sets uniform requirements for the enforceability of

44. See Cohen & Dayton, supra note 2, at 269.
45. Id.
46. Id.
47. Id.
48. Id.
49. See id.
51. See supra note 36 and accompanying text; Moses H. Cone, 460 U.S. at 24.
any arbitration agreements relating to certain types of transactions. For example, section 2 of the FAA states that a written arbitration provision “shall be valid, irrevocable, and enforceable, save upon such grounds as exist at law or in equity for the revocation of any contract.” However, the FAA is also a procedural statute because it spells out the procedures that federal courts must follow when enforcing valid arbitration agreements. Under the FAA, courts must “make an order directing the parties to proceed to arbitration” upon satisfaction that “the making of the agreement for arbitration or the failure to comply therewith is not in issue.” This procedural scheme ensures that valid arbitration agreements are properly enforced.

The U.S. Supreme Court has interpreted the FAA to have created federal substantive arbitration law that preempts all state laws that set additional requirements and limitations on the enforceability of arbitration agreements. The Court has held that the FAA governs the enforceability of arbitration agreements in both federal and state courts “notwithstanding any state substantive or procedural policies to the contrary.” Therefore, the FAA applies to any arbitration agreements relating to interstate commerce or maritime transactions, regardless of whether a party seeks to enforce the arbitration agreement in a federal or state court.

54. See Moses H. Cone, 460 U.S. at 24.
56. Id. § 4.
57. See Moses H. Cone, 460 U.S. at 25 n.32; Southland Corp. v. Keating, 465 U.S. 1, 11 (1984) (stating that “the FAA rests on the authority of Congress to enact substantive rules under the Commerce clause of § 8 of Article I of the Constitution,” which implies that the substantive rules of the Act should bind both state and federal courts). See also Robert Lawrence Co. v. Devonshire Fabrics, Inc., 271 F.2d 402, 407 (2d Cir. 1959) (stating that Congress used its Article I, § 8, clause 3 power to enact § 2 of the FAA which is a “declaration of national [substantive arbitration] law equally applicable in state or federal courts”). But see Keating, 465 U.S. at 21–36 (O’Connor, J., dissenting) (arguing that the FAA is a merely procedural statute that Congress adopted through exercise of its power to control jurisdiction of lower federal courts under Article III of the Constitution, which grants to Congress no power to control proceedings in state courts); Cohen & Dayton, supra note 2, at 266 (noting that the FAA was intended to reverse “the hoary doctrine that agreements for arbitration are revocable at will and are unenforceable”).
60. See id. at 14–15. The Court has emphasized, however, that the FAA does not create independent federal question jurisdiction under 28 U.S.C. § 1331. Id. at 25 n.32. The necessity of an independent ground for federal subject matter jurisdiction in suits to enforce arbitration agreements under the FAA is implicit in language of section 4 of the FAA providing that “any United States district court which, save for such agreement, would have jurisdiction under Title 28, in a civil action or in admiralty of the subject matter of a suit arising out of the controversy between the
C. Arbitration Clauses are Severable Under the FAA

According to section 2 of the FAA, arbitration clauses are severable from the contracts that contain them. Therefore, section 2 states that a written arbitration provision in a contract “shall be valid, irrevocable, and enforceable.” Thus, the FAA provides that an arbitration clause may be separately enforced unless there are any legal or equitable grounds for its revocation. This conclusion stems from the statute’s specific focus on arbitration provisions as separate contractual units. Indeed, section 2 of the FAA specifically mentions the enforceability of an arbitration provision “in any maritime transaction or contract” as independent from the enforceability of the contract itself. Thus, at least one circuit has held that the statute “does not purport to affect the contract as a whole.”

In addition, federal and state courts have always treated arbitration agreements as independent contracts. For example, in Hamilton v. Home Insurance Co., the U.S. Supreme Court explicitly stated that “it is . . . well settled that the agreement [to arbitrate] . . . is collateral and independent; and that a breach of this agreement, while it will support a separate action, cannot be pleaded in bar to an action on the principal contract.” Further, courts’ treatment of arbitration agreements as severable contractual units has not been affected by the passage of the FAA. In the years following the adoption of the Act, courts have held that the illegality of part of the contract does not operate to nullify an agreement to arbitrate. Finally, the Court in Prima Paint affirmed the position that, under the FAA, arbitration agreements are severable from the contracts in which they are embedded.
In sum, the FAA treats arbitration agreements as severable from commercial contracts that contain them and promotes their enforcement as long as the agreements satisfy the basic elements of any valid contract. The U.S. Supreme Court has interpreted the FAA as creating a federal policy favoring arbitration. Both substantive and procedural aspects of the FAA further the statute’s main goal—to make specific performance of arbitration agreements available to aggrieved parties.

II. IN PRIMA PAINT, THE COURT HELD THAT UNDER THE FAA, A COURT MAY CONSIDER CLAIMS FOR FRAUD IN THE INDUCEMENT OF AN ARBITRATION CLAUSE, BUT NOT FRAUD IN THE INDUCEMENT OF THE ENTIRE CONTRACT

In Prima Paint the U.S. Supreme Court interpreted section 4 of the FAA and held that when a person alleges that a contract containing an arbitration clause is invalid on grounds of fraudulent inducement, a federal court can only resolve challenges of fraudulent inducement concerning the arbitration clause itself. Moreover, the Court emphasized that when ruling upon an application for stay of the action pending arbitration, “a federal court may consider only issues relating to the making and performance of the agreement to arbitrate.” According to the Court, challenges of fraud in inducement of the underlying contract should be reserved for an arbitrator. Given the specific factual setting of the case, the Court did not decide how to allocate the respective authorities of a court and an arbitrator in cases where a party asserts that the contract containing an arbitration clause is void from its

holding that federal courts should consider only issues relating to the making and performance of arbitration agreements “has come to be known as the severability doctrine”;

Stevens/Leinweber/Sullens, Inc. v. Holm Dev. & Mgmt, Inc., 795 P.2d 1308, 1312–13 (Ariz. App. 1990) (stating that under the concept of separability endorsed by Prima Paint, “an arbitration provision is considered to be an independent and separate agreement between the parties to the underlying contract”).

72. See Moses H. Cone, 460 U.S. at 24.
73. See supra notes 41–43 and accompanying text.
74. Prima Paint, 388 U.S. at 403–04.
75. Id. at 404.
76. Id.
77. The plaintiff in Prima Paint alleged that the contract containing the arbitration provision was a result of fraudulent inducement, i.e., that it was voidable. See id. at 398.
very inception and argues that the arbitration clause is automatically invalid as well.

*Prima Paint* was the first and only Supreme Court case to interpret the FAA in the context of the allocation of the respective powers of a federal court and an arbitrator. The ultimate issue in *Prima Paint* was whether a federal court or an arbitrator should resolve a claim of fraud in the inducement of a contract containing an arbitration clause. The Court held that a federal court deciding whether to stay an action and order arbitration under section 3 of the FAA “may consider only issues relating to the making and performance of the agreement to arbitrate.” More specifically, the Court determined that if a party to the contract asserts a claim of fraud in the inducement of the arbitration clause itself, a federal court can adjudicate such a claim. However, the FAA does not authorize a federal court to resolve claims of fraud in the inducement of the contract as a whole.

*Prima Paint* involved a consulting agreement between two corporations. This agreement followed the formation of a contract under which the plaintiff purchased the defendant’s paint business. The consulting agreement contained a broad arbitration clause providing that “any controversy or claim arising out of or relating to this Agreement, or breach thereof, shall be settled in arbitration in the City of New York.” After the defendant had filed a bankruptcy petition, the plaintiff refused to perform the consulting agreement and filed suit in a federal court seeking rescission of the agreement on the ground that the agreement was fraudulently induced by the defendant’s misrepresentation of its solvency. The district court granted the defendant’s motion to stay the action pending arbitration, holding that an allegation of fraud in the inducement of a contract containing an arbitration clause was a question

---

78. See Sandvik AB v. Advent Int’l Corp., 220 F.3d 99, 105 (3d Cir. 2000) (“[T]he Supreme Court in *Prima Paint* did not grapple with what is to be done when a party contends not that the underlying contract is merely voidable, but rather that no contract ever existed.”).
80. *Id.* at 404 (emphasis added).
81. *Id.* at 403–04.
82. *Id.* at 404.
83. *Id.* at 397.
84. *Id.*
85. *Id.* at 398.
86. *Id.*
87. *Id.* at 399.
for the arbitrator and not for the court. The U.S. Supreme Court affirmed.

The Court emphasized that section 4 of the FAA allows federal courts to resolve only claims of fraud in inducement of an arbitration clause itself, not claims of fraud in the inducement of the underlying contract. According to the Court, Congress’s explicitly mandated such scheme. This section 4 axiom prompted the Court to acknowledge the Second Circuit’s position in Robert Lawrence Co. v. Devonshire Fabrics, Inc. that as a matter of federal law arbitration clauses are separable from contracts where they are included. Consequently, the Prima Paint Court agreed with the lower court’s holding that the arbitration agreement was severable from the commercial contract for the purpose of evaluating claims of fraudulent inducement of the underlying contracts.

Because of the case’s factual setting, the Prima Paint Court focused solely on situations where a party challenges the entire contract as fraudulently induced, i.e., where the party asserts that the contract is voidable. A contract is voidable even if it is generally valid, but one or more parties to the contract have the power to avoid the contractual relationship on grounds such as fraud in inducement of the contract, duress, or mistake. The Court did not consider how the authority of a court and an arbitrator would be allocated in a case where a party asserts that a contract containing the arbitration clause was fraudulently executed, i.e., that the contract is void from its very inception.

---

88. Id. (noting that the arbitration clause in the agreement gave the arbitrator very broad authorities).
89. Id. at 407.
90. Id. at 403–04.
91. Id. at 403.
92. 271 F.2d 402 (2d Cir. 1959).
93. See id. at 409–10.
94. See Doctor’s Assocs., Inc. v. Distajo, 66 F.3d 438, 452 (2d Cir. 1995) (noting that the U.S. Supreme Court in Prima Paint endorsed the Second Circuit’s position that an arbitration clause is a separable part of the contract); see also Burden v. Check Into Cash of Ky., LLC, 267 F.3d 483, 488 (6th Cir. 2001) (stating that “[t]he Court in Prima Paint found that arbitration clauses were ‘separable’ from the contracts in which they were included”).
95. Fraud in inducement occurs when there is a genuine mutual assent to the contract but one of the parties misrepresents certain facts, such as the quality of goods. See E. ALAN FARNsworth, CONTRACTS § 4.10, at 243–44 (3d ed. 1999).
96. See Restatement (Second) Of Contracts § 7 cmt. b (1981).
97. As opposed to fraud in inducement of a contract, fraud in execution, also know as fraud in factum, occurs when the misrepresentation goes “to the very character of the proposed contract itself, as when one party induces the other to sign a document by falsely stating that it has no legal effect.” See Farnsworth, supra note 95, at 243. If the other party neither knows nor has reason to
A contract is void if one of its essential elements, such as mutual assent, is missing; a void contract is not a contract but merely a “promise” or “agreement” that is void of legal effect. Nevertheless, in *Prima Paint*, the Court made it clear that a federal court should order arbitration once it is satisfied that “the making of the agreement for arbitration . . . is not in issue.”

In *Prima Paint*, the U.S. Supreme Court confirmed the FAA’s mandate that a federal court should decide the validity of an arbitration agreement only if a party specifically puts the arbitration agreement at issue. Under the Court’s ruling in *Prima Paint*, an arbitrator should rule upon a party’s claim for fraud in inducement of the underlying commercial contract, but a court should rule upon a claim of fraud in inducement of the arbitration clause. The Court did not decide whether a court or an arbitrator should resolve a party’s claim of fraud in the execution of the contract. Significantly, the Court treated the arbitration agreement and the commercial contract as separate and independent of one another.

### III. FEDERAL COURTS ARE DIVIDED AS TO HOW THE *PRIMA PAINT* RULE APPLIES WHERE A PARTY DOES NOT SPECIFICALLY CHALLENGE AN ARBITRATION PROVISION BUT ASSERTS FRAUD IN EXECUTION OF THE CONTRACT

The lower federal courts have applied *Prima Paint* in different ways. The Third, Ninth, and Eleventh Circuits have limited *Prima Paint*’s application to claims of fraud in the inducement of a contract containing an arbitration clause, i.e., claims where a party to a valid commercial contract seeks to avoid or rescind the contract by arguing that it is voidable. These courts have held that if a party alleges that a contract
with an arbitration clause was never legally formed because it lacked mutual assent, a court, not an arbitrator, should decide whether the contract is valid.104 Thus, by alleging that the entire contract is void, a party may successfully avoid arbitration. In contrast, the Fifth and Sixth Circuits have held that a party purporting to avoid arbitration of disputes arising out of a commercial contract must specifically challenge the validity of the arbitration clause.105 Thus, a party in the Fifth or Sixth Circuits cannot avoid arbitration by merely alleging that the entire contract is void.

A. The Third, Ninth, and Eleventh Circuits Have Limited Prima Paint to Claims of Fraudulent Inducement

The Third, Ninth, and Eleventh Circuits have held that Prima Paint does not apply to contracts that are fraudulently executed, such as in situations where a party asserts that the arbitration clause is invalid because the underlying contract is void from its very inception. The courts reason that a party alleging that it never assented to the contract with an arbitration provision challenges “the very existence of any agreement, including the existence of an agreement to arbitrate.”106 The Ninth Circuit’s decision in Three Valleys Municipal Water District v. E.F. Hutton & Co.107 took the lead in developing this theory. The Ninth Circuit held that Prima Paint applies to arbitration provisions in voidable contracts, but does not apply to arbitration agreements in contracts void from their very inception.108 Since Three Valleys, the Third and Eleventh Circuits have followed the Ninth Circuit’s approach.109

In Three Valleys, the Ninth Circuit held that the Prima Paint rule is limited to claims of avoidance or rescission of a contract.110 Three

---

107. 925 F.2d 1136 (9th Cir. 1991).
108. See id. at 1140.
109. See, e.g., Chastain, 957 F.2d at 855; Sandvik, 220 F.3d at 106–07.
110. Three Valleys, 925 F.2d at 1140.
Valleys involved arbitration clauses in securities accounts agreements. The plaintiffs alleged that the district court could not compel arbitration because the defendant’s representative who signed the agreements had no authority to do so. Thus, the plaintiffs challenged the agreements as void from their inception based on a lack of the plaintiffs’ assent. The Ninth Circuit agreed.

Focusing on Prima Paint’s factual setting, the Ninth Circuit reasoned that the ruling in Prima Paint authorized arbitrators to consider only claims for fraudulent inducement of a contract, as opposed to claims for fraud in the execution of a contract. Thus, the Ninth Circuit limited the scope of Prima Paint’s rule to challenges “seeking to avoid or rescind the contract.” The court reasoned that if a party challenges the very existence of a contract in which the arbitration clause is embedded, a district court should resolve such a claim because the challenging party may have never agreed to the authority of an arbitrator.

Similarly, in Chastain v. Robinson-Humphrey Co., the Eleventh Circuit held that Prima Paint does not apply when a party contends that the contract containing an arbitration provision is void. In Chastain, the plaintiff claimed that she never assented to either the arbitration agreement or the underlying contract. The plaintiff’s father opened a securities trading account with the defendant, executing two customer agreements in connection with the account. One of the agreements bore the plaintiff’s name, but the plaintiff did not personally sign the agreement. The second agreement bore the defendant’s name only. The plaintiff alleged that she had never authorized her father to open the

111. Id. at 1137.
112. Id. at 1138.
113. Id. at 1139–41.
114. See supra notes 83–88 and accompanying text.
115. Three Valleys, 925 F.2d at 1140.
116. Id. (emphasis in original).
117. Id. at 1137.
118. See id. at 1140.
119. 957 F.2d 851 (11th Cir. 1992).
120. Id. at 855.
121. Id. at 853.
122. Id.
123. Id.
124. Id.
125. Id.
account in her name.126 Thus, the plaintiff alleged that both the arbitration agreement and the underlying contract were void from their very inception because the plaintiff’s signature was forged. Chastain presented a factual situation upon which the U.S. Supreme Court in Prima Paint had not ruled: where a party to the commercial contract containing an arbitration clause asserts that the contract was void from its very inception.

The Eleventh Circuit concluded in Chastain that if a party disputes signing the contract that requires arbitration, then that party may not have agreed to proceed with arbitration at all.127 In the court’s view, alleging a lack of assent precluded resolving the dispute by arbitration.128 The court considered the Prima Paint ruling to be distinguishable because it did not require “arbitrators to adjudicate a party’s contention . . . that a contract never existed at all.”129 The court reasoned that alleging that a contract was invalid was enough to cast doubt on the validity of the arbitration agreement.130

In line with Three Valleys and Chastain, the Third Circuit held in Sandvik AB v. Advent International Corp.,131 that under the Court’s ruling in Prima Paint, an arbitrator’s jurisdiction should be limited to resolution of claims of inducement of the contract containing an arbitration clause.132 In Sandvik, the plaintiff claimed breach of a joint venture agreement containing an arbitration clause.133 The defendant responded that the agreement was void because its agent lacked authority to sign the agreement.134 At the same time, the defendant moved to compel arbitration under the FAA.135 The defendant explained that its dispute of the existence of the contract does not automatically assume the dispute of the arbitration clause contained therein.136 The district court

126. Id.
127. Id. at 854.
128. Id. (citing Cancanon v. Smith Barney, Harris, Upham & Co., 805 F.2d 998, 1000 (11th Cir. 1986)).
129. Id. at 855.
130. See id.
131. 220 F.3d 99 (3d Cir. 2000).
132. Id. at 106–07.
133. The arbitration clause provided that “[a]ny dispute arising out of or in connection with this agreement . . . shall . . . be finally settled by arbitration in accordance with the rules of the Netherlands Arbitration Institute.” Id. at 101.
134. Id. at 100.
135. Id.
136. See id.
denied the defendant’s motion to compel, reasoning that the challenge to
the existence of the underlying contract was sufficient to put the validity
of the arbitration agreement in dispute.137 The Third Circuit affirmed.138

The court noted that Prima Paint did not consider the situation in
which a party asserts that a contract underlying the arbitration agreement
is not merely voidable, but non-existent.139 Further, the Third Circuit
reasoned that determining whether an arbitration agreement exists under
the FAA may depend on the validity of the underlying contract.140 Thus,
before a district court compels arbitration it must first determine whether
the underlying contract was valid.141

The Third Circuit noted, however, that although the defendant failed
to show that the arbitration agreement was valid and severable from the
underlying contract, an arbitration agreement can exist separately from a
larger contract if such agreement independently meets the conditions of
contract formation.142 Further, the court acknowledged that in most cases
where there is no defect in signatory power of a party, if one party
promises to arbitrate in exchange for the other party’s promise to
arbitrate, each promise forms a sufficient consideration for the other.143

Thus, the Third, Ninth, and Eleventh Circuits recognize a distinction
under Prima Paint between “void/fraud in execution” challenges and
“voidable/fraud in inducement” challenges to a contract containing an
arbitration agreement. In these Circuits, an allegation that a contract is
void is sufficient to assert that the arbitration agreement is invalid,
thereby authorizing a court under Prima Paint to rule on the contract’s
validity. In contrast, alleging that a contract is merely voidable fails to
put the arbitration agreement “in issue” under Prima Paint and an
arbitrator will resolve the merits of the allegation.

137. See id.
138. Id.
139. Id. at 105.
140. Id. at 106 (citing Par-Knit Mills, Inc. v. Stockbridge Fabrics Co., 636 F.2d 51 (3d Cir.
1980)).
141. “[W]e conclude that the doctrine of severability presumes an underlying, existent,
agreement. Such an agreement exists, under the Prima Paint doctrine, even if one of the parties
seeks to rescind it on the basis of fraud in the inducement. [Such agreement] . . . does not [exist] if
no contract ever existed.” Id.
142. Id. at 108. Formation of a valid contract is accomplished when the parties have expressed
mutual assent to the contract’s terms supported by adequate consideration. See RESTATEMENT
143. Id. (citing Sauer-Getriebe KG v. White Hydraulics, 715 F.2d 348, 350 (7th Cir. 1983)).
B. The Fifth and Sixth Circuits Have Held that Prima Paint Requires a Specific Allegation of Invalidity of an Arbitration Agreement to Prevent its Enforcement

The Fifth and Sixth Circuits have held that, although *Prima Paint* was decided in the context of fraud in the inducement of a contract, it pronounced a general rule that a party resisting arbitration must specifically challenge the arbitration provision in a commercial contract even if the party alleges that the entire contract is void. The Circuit Courts have reasoned that the rule of *Prima Paint* should apply to allegations of fraud in the execution of the contract because the fraud in the inducement alleged in *Prima Paint* would pervade the entire contract containing the arbitration clause just as much as the fraud in the execution of the contract. The most current representative cases supporting this approach are the Fifth Circuit’s decision in *Lawrence v. Comprehensive Business Services* and the Sixth Circuit’s decision in *Burden v. Check Into Cash of Kentucky, LLC*.

1. Lawrence v. Comprehensive Business Services: An Arbitration Agreement May be Enforceable Even if the Underlying Contract is Illegal

In *Lawrence*, the Fifth Circuit held that an arbitration agreement may be enforceable under the FAA even if the underlying contract is illegal. The court opined that in order for a party to obtain a court ruling on the validity of an arbitration agreement, that party must assert the illegality of the arbitration clause itself. Importantly, the court emphasized that the question of the validity of an arbitration agreement, as separate from the underlying commercial contract, is a matter of the federal law under section 2 of the FAA.

---

146. *Id.*
147. 267 F.3d 483 (6th Cir. 2001).
148. See *id.* at 1161–62.
149. *Id.* at 1162.
150. *Id.* (discussing preemption aspects of the FAA under Southland Corp. v. Keating, 465 U.S. 1, 10–13 (1984)).
The plaintiffs in Lawrence challenged an agreement containing an arbitration clause as illegal and therefore void.\textsuperscript{151} Plaintiff Robert Lawrence, a Texas certified public accountant, entered into a franchise agreement with the defendant, allowing him to use the defendant’s trade name.\textsuperscript{152} The agreement required Lawrence to make periodic royalty payments to the defendant and contained an arbitration clause.\textsuperscript{153} After signing the contract, the plaintiff learned that the Texas State Board of Public Accountancy had prohibited other franchisees from operating an accounting practice under the defendant’s name.\textsuperscript{154} The plaintiff advised the defendant that he could not carry out the contract because if he did, he would lose his license.\textsuperscript{155} The plaintiff sued the defendant, seeking a judgment declaring the contract illegal and unenforceable.\textsuperscript{156} In response, the defendant moved to stay the litigation and compel arbitration pursuant to the arbitration clause in the contract.\textsuperscript{157} The plaintiff argued that ordering arbitration under the arbitration clause in an illegal contract is improper.\textsuperscript{158} The Fifth Circuit rejected this argument and held that under the command of Prima Paint, an arbitrator should decide whether the franchise agreement between the parties was valid and legal because the plaintiff did not challenge the arbitration clause itself.\textsuperscript{159}

The Fifth Circuit declined the plaintiff’s invitation to recognize the difference between the Prima Paint plaintiff’s claim of fraud in inducement of the contract\textsuperscript{160} and the Lawrence plaintiffs’ fraud in execution\textsuperscript{161} argument, because fraud in inducement and fraud in execution have the same pervasive effect on the contract.\textsuperscript{162} Moreover, the court stated that it read the Prima Paint decision as mandating courts to order arbitration even if a party asserts that the contract containing an arbitration clause fails to comply with certain state regulations and is

\begin{itemize}
  \item \textsuperscript{151} See id. at 1161.
  \item \textsuperscript{152} Id. at 1160.
  \item \textsuperscript{153} Id.
  \item \textsuperscript{154} Id.
  \item \textsuperscript{155} Id.
  \item \textsuperscript{156} Id. at 1161.
  \item \textsuperscript{157} Id.
  \item \textsuperscript{158} Id.
  \item \textsuperscript{159} Id. at 1162.
  \item \textsuperscript{160} See supra note 95 and accompanying text.
  \item \textsuperscript{161} See supra note 97 and accompanying text.
  \item \textsuperscript{162} Lawrence, 833 F.2d at 1162.
\end{itemize}
therefore void.\textsuperscript{163} Finally, the Fifth Circuit stated that the FAA established that, “as a matter of federal law, any doubts concerning the scope of arbitrable issues should be resolved in favor of arbitration.”\textsuperscript{164}

2. \textit{Burden v. Check Into Cash of Kentucky, LLC: Prima Paint Requires Specifically Challenging the Validity of the Arbitration Agreement}

In \textit{Burden}, the Sixth Circuit strongly suggested in dicta that the FAA, as interpreted by the U.S. Supreme Court in \textit{Prima Paint}, requires a specific challenge to the validity of the arbitration provision in the contract.\textsuperscript{165} The plaintiffs in \textit{Burden} were trustees for four bankruptcy estates and other numerous residents of Kentucky.\textsuperscript{166} The main defendant was a creditor of the bankruptcy estates.\textsuperscript{167} Plaintiffs alleged that the defendant violated Kentucky usury laws by loaning money to Kentucky residents at allegedly usurious interest rates.\textsuperscript{168} The reverse side of each loan agreement at issue contained an arbitration clause providing that all claims, demands, or disputes “arising under this Agreement or the transaction in connection with which this Agreement has been executed” should be resolved by arbitration.\textsuperscript{169} The plaintiffs contended that prior to December of 1997, the loan agreements had no arbitration clause on the reverse side of the form.\textsuperscript{170} Further, the plaintiffs alleged that the defendants never informed them about the arbitration clause, and that the plaintiffs learned about the clause only when the defendant attached its copy to their motion to compel arbitration.\textsuperscript{171} The plaintiffs claimed that the loan agreements containing arbitration clauses were invalid.\textsuperscript{172} Finally, the plaintiffs relied on the theory that the court, not an arbitrator,

\begin{itemize}
\item \textsuperscript{163} \textit{Id.} (citing Mesa Operating Ltd. P’ship v. La. Intrastate Gas Corp., 797 F.2d 238, 244 (5th Cir. 1986)).
\item \textsuperscript{164} \textit{Id.} at 1164 (citing Moses H. Cone Mem’l Hosp. v. Mercury Constr. Corp., 460 U.S. 1, 24-25 (1983)).
\item \textsuperscript{165} \textit{Id.} at 491.
\item \textsuperscript{166} \textit{Id.} at 486.
\item \textsuperscript{167} \textit{Id.}
\item \textsuperscript{168} \textit{Id.}
\item \textsuperscript{169} \textit{Id.} at 487.
\item \textsuperscript{170} \textit{Id.}
\item \textsuperscript{171} \textit{Id.}
\item \textsuperscript{172} \textit{Id.} at 489.
\end{itemize}
must consider their allegations that the loan agreements containing the arbitration clause were void from their very inception.173

The Sixth Circuit ruled that the district court should have compelled arbitration under the Prima Paint decision because the plaintiffs failed to identify any misrepresentation particular to the arbitration agreements and separate from the loan agreements.174 The court noted that the Prima Paint rule commands a court, rather than an arbitrator, to adjudicate claims of fraud in the inducement only if such claims concern the arbitration clause itself.175 The Sixth Circuit concluded that the plaintiffs’ allegations that the arbitration agreements were part of the defendants’ “fraudulent scheme” were arbitrable under Prima Paint.176 In reaching that result, the court relied on the FAA’s policy favoring enforcement of arbitration agreements.177

Although the Burden court acknowledged the Ninth and Third Circuits’ void/voidable distinction, the court found that such distinction was improper under Prima Paint and its own precedent.178 The court reasoned that the only question that a court should resolve when determining an arbitrator’s authority is whether the issues in dispute involve “the making or the performance of the section 3 arbitration clause itself.”179 In light of the language of section 3 of the FAA,180 the court decided to adhere to Prima Paint’s mandate that a court can only adjudicate an arbitration dispute if the claim of fraud relates to the making of the arbitration agreement.181

173. Id.
174. Id. at 491; see also Showden v. Checkpoint Check Cashing; Elite Fin. Serv., Inc., 290 F.3d 631, 637–38 (4th Cir. 2002) (stating that the Burden court properly denied the plaintiffs’ challenge to enforcement of arbitration clauses in loan agreements because the plaintiffs failed to specifically allege lack of their assent to the arbitration clauses, and instead challenged the substance of the loan agreements); Bess v. Check Express, 294 F.3d 1298, 1306 (11th Cir. 2002) (noting that the Burden court concluded that the plaintiffs’ allegations that the loan agreements were void as illegal “constituted challenge to substance of loan agreements and should thus be decided by arbitrator rather than by court”).
175. Id. at 488.
176. Id. at 491.
177. “As a matter of federal law, any doubts concerning the scope of arbitrable issues should be resolved in favor of arbitration.” Id. at 488 (quoting Wilson Elec. Contractors, Inc. v. Minotte Contracting Corp., 878 F.2d 167, 169 (6th Cir. 1989)).
178. See id. at 489–91 (citing C.B.S. Employees Fed. Credit Union v. Donaldson, Lufkin & Jenrette Sec. Corp., 912 F.2d 1563, 1566 (6th Cir. 1990)).
179. Id. at 489 (citing C.B.S. Employees, 912 F.2d at 1567–68).
181. Burden, 267 F.3d at 489 (citing C.B.S. Employees, 912 F.2d at 1566).
Thus, the Circuit Courts’ interpretation of the *Prima Paint* decision is split into two different approaches. The Third, Ninth, and Eleventh Circuits emphasize that because *Prima Paint* concerned fraud in inducement of the contract containing an arbitration clause, its holding should be limited to situations in which a contract is valid, but voidable.\(^{182}\) On the other hand, the Fifth and Sixth Circuits have adhered to the Supreme Court’s admonition in *Prima Paint* that a judicial forum can only resolve the validity of an arbitration clause if a party specifically challenges making or performing the clause itself.\(^{183}\)

### IV. PARTIES MUST SPECIFICALLY ALLEGED THAT AN ARBITRATION AGREEMENT IS INVALID TO BRING THE DISPUTE ARISING OUT OF THE CONTRACT WITHIN A COURT’S JURISDICTION

The Third, Ninth, and Eleventh Circuit Courts’ rulings that the *Prima Paint* decision does not apply when an underlying contract may be void\(^{184}\) ignore the FAA section 2’s clear mandate that federal courts should enforce arbitration agreements independently from the contracts containing them.\(^{185}\) Consequently, a court should not disturb an arbitrator’s authority unless the validity of the arbitration agreement itself is at issue.\(^{186}\) Because the FAA treats arbitration clauses as severable from the commercial contracts containing them,\(^{187}\) a mere allegation that the underlying contract is void or unenforceable should not be sufficient to put the arbitration clause at issue under section 4 of the FAA\(^{188}\) and deprive an arbitrator of jurisdiction over the contractual dispute.

Alleging that a contract containing an arbitration agreement is void or invalid should not be enough to put the validity of the arbitration agreement at issue under section 4 of the FAA. By holding that a claim of fraud in the inducement of a contract is not a claim directed at the “making” of the agreement to arbitrate, the U.S. Supreme Court, in

---

\(^{182}\) *See, e.g.*, Chastain v. Robinson-Humphrey Co., 957 F.2d 851, 855 (11th Cir. 1992).
\(^{183}\) *See C.B.S. Employees*, 912 F.3d at 1567–68.
\(^{184}\) *See supra* Part III.A.
\(^{185}\) *See supra* Part I.C.
\(^{187}\) *See supra* Part I.C.
\(^{188}\) *See* 9 U.S.C. § 4 (2000); *Prima Paint*, 388 U.S. at 404 (holding that if a party moves for stay pending arbitration, a federal court “may consider only issues relating to the making and performance of the agreement to arbitrate”.

220
Challenging Arbitration Agreements Under the FAA

*Prima Paint* did not distinguish claims of fraud in the inducement from any other challenges to commercial contracts.189 Instead, the Court emphasized that the FAA requires federal courts to consider only issues related to the making and performance of arbitration agreements.190 The Third, Ninth, and Eleventh Circuits’ position that alleging a contract is void is enough to put an arbitration agreement at issue misinterprets the rule set forth in *Prima Paint*.191 Although these Circuits’ classification of challenges as “void/fraud in execution” and “voidable/fraud in inducement” may help litigants that have failed to properly plead the invalidity of an arbitration clause, this distinction should be abandoned because it contradicts the clear language of the FAA and the strong federal policy favoring arbitration agreements.192 The distinction should only be used to aid courts in determining whether a party has made a direct challenge to an arbitration agreement.

A.  The FAA Requires Parties to Specifically Allege that an Arbitration Agreement is Invalid

Congress adopted the FAA to create an enforcement mechanism for arbitration agreements in commercial and maritime transactions by directing the courts to order specific performance of the agreements.193 The FAA was intended to further arbitration as a speedy and efficient way to resolve commercial disputes.194 Based on that congressional intent, the U.S. Supreme Court has declared that any doubts as to the scope of issues subject to arbitration should be resolved in favor of arbitration.195

Under the FAA, a written arbitration provision in a commercial or maritime contract is valid and enforceable unless it could be revoked on legal or equitable grounds as any other contract.196 The FAA’s language treats arbitration agreements as separate from the contracts in which they are included. This is evidenced by the FAA’s provision that an

189.  See *Prima Paint*, 388 U.S. at 403–04.
190.  See id. at 404.
191.  See supra Part II.
192.  See supra Part I.A.
193.  See Cohen & Dayton, supra note 2, at 276–78.
194.  See id. at 269.
arbitration provision contained within a commercial contract is independent from the underlying contract. Consistent with this treatment, section 4 of the FAA provides that courts should order parties to proceed to arbitration as soon as the court is satisfied that “the making of the agreement for arbitration or the failure to comply therewith is not in issue.” Thus, for example, if a party claims that the arbitration agreement was fraudulently induced, the court may adjudicate that claim under section 4 of the FAA.

The FAA’s focus on the arbitration clause as a separate contractual unit strongly suggests that a party seeking to challenge the validity of an arbitration clause should directly challenge the arbitration clause. It should not be enough for a party to a commercial contract to refuse to proceed with arbitration merely because the party asserts that the contract was void and therefore its every term, including the arbitration provision, is not binding. To the contrary, a party should specifically assert that an arbitration provision, as a separate agreement, is invalid on some legal or equitable ground. Thus, a mere allegation that a contract is invalid should not remove the dispute from an arbitrator’s jurisdiction unless the arbitration provision is not severable from the underlying contract.

B. The Prima Paint Rule Should Apply to Allegations that the Contract Containing an Arbitration Agreement is Void

A strong federal policy favoring arbitration supports the separate treatment of arbitration clauses from the contracts in which they are included. In enacting the FAA, Congress created federal substantive arbitration law encouraging the enforcement of arbitration agreements. The drafters of the FAA acknowledged that substituting judicial resolution of commercial disputes with arbitration would effectively eliminate the long delays incident to court proceedings, help parties to avoid the expense of litigation, and provide a better means of reaching a decision regarded as just in the business world. In line with these

201. Id.
202. See supra Part I.
203. See supra Part I.B.
204. See supra notes 44–49 and accompanying text.
goals, the U.S. Supreme Court ruled in *Prima Paint* that an arbitrator should decide issues related to fraud in inducement of a contract containing an arbitration clause.\(^{205}\)

Consistent with the FAA’s meaning and policy, in *Prima Paint* the Court authorized federal courts to adjudicate allegations related only to invalidity of arbitration clauses.\(^{206}\) However, the Court did not authorize the lower courts to consider claims that contracts containing arbitration clauses are generally void as illegal or as lacking mutual assent. Such claims should be normally arbitrated pursuant to the contract\(^{207}\) because the arbitration clauses are in most cases severable from the commercial contracts in which they are embedded.\(^{208}\) Thus, a party’s claim that it did not assent to the contract does not necessarily mean that it did not assent to the arbitration clause. Consequently, by alleging that a contract containing an arbitration clause is void, a party in most cases fails to question the authority of an arbitrator and thereby fails to effectively put “the making of an arbitration agreement” at issue within the meaning of section 4 of the FAA.\(^{209}\)

Courts should not consider allegations that an entire contract is void simply because the arbitration clauses are often embedded in contracts in a way that the parties fail to notice them. In cases where a person believes in good faith that he or she did not assent to an arbitration clause, the person may simply plead that the clause itself, not the underlying contract, is void. However, parties attacking arbitration clauses usually have no firm arguments that the arbitration clause should be revoked under the requirements of section 2 of the FAA.\(^{210}\) Such parties attempt to get into court “through the backdoor,” by alleging that the entire contract containing the arbitration clause is invalid.\(^{211}\) Under the *Prima Paint* decision such a challenge should not be sufficient to

---


206. *Id.*; see also *Snowden v. Checkpoint Cashing; Elite Fin. Servs., Inc.*, 290 F.3d 631, 637 (4th Cir. 2002) (noting that in *Prima Paint*, the Court held that section 4 of the FAA, and by implication, section 3 of the FAA, limits a federal court’s jurisdiction to challenges to the arbitration clause itself”).


208. *See supra* Part I.C.


210. *See, e.g.*, *Lawrence*, 833 F.2d at 1162–65 (declining the plaintiffs’ direct challenge of the arbitration clause in the franchise agreement).

211. *See, e.g.*, *id.* (The plaintiffs were unable to effectively challenge the arbitration clause itself, that is why they attempted to challenge the entire contract containing the clause as void).
require a court to resolve whether an arbitration agreement is valid under section 4 of the FAA. 212


The Sixth Circuit in Burden correctly concluded that the Prima Paint rule applies in situations where a party contends that a contract containing an arbitration clause is void because it allegedly violates certain statutory provisions. 213 Likewise, in Lawrence the Fifth Circuit appropriately held that under Prima Paint, an arbitrator should decide whether the contract between the parties was valid and legal because the plaintiff failed to successfully challenge the validity of the arbitration clause. 214 Although Prima Paint did not specifically address this situation, the plain language of section 4 of the FAA 215 coupled with the principle of severability of arbitration clauses, 216 provide strong support for the conclusion that parties must specifically plead that the arbitration agreement is invalid. Where a party to a commercial or maritime transaction alleges that a contract containing an arbitration clause is void, the court should not consider the validity of the arbitration agreement unless a party specifically asserts that the arbitration clause is invalid as a separately standing agreement. 217 The Burden and Lawrence courts properly focused on the fact that the plaintiffs refusing to arbitrate had failed to challenge the validity of the arbitration agreement. 218 As the court pointed out, and as sections 2 and 4 of the FAA require, such a challenge is essential to overcome the FAA’s presumption that arbitration agreements are enforceable in commercial contracts. 219

The Sixth Circuit’s approach in Burden and the Fifth Circuit’s approach in Lawrence is sound because alleging that the underlying commercial contract is void does not necessarily assume that the

213. See supra Part III.B.2.
214. See supra Part III.B.1.
216. See supra Part I.C.
217. See Prima Paint, 388 U.S. at 403–04.
219. The existence of such a presumption follows from the structure of section 2 of the FAA providing that “[a]n arbitration provision . . . shall be valid, irrevocable, and enforceable, save upon . . . [any legal or equitable] grounds . . . for the revocation of any contract.” See 9 U.S.C. § 2.
arbitration agreement is likewise void for at least two reasons. First, the FAA, *Prima Paint*, and other cases decided after the adoption of the FAA treat arbitration clauses as separate contracts. Second, parties to a commercial transaction containing an arbitration clause likely understand that they will forego a judicial forum if any dispute arises out of the contract.

In contrast, the Third, Ninth, and Eleventh Circuits’ conclusion that a court should consider a party’s claim that a contract containing an arbitration clause is void is questionable because it automatically assumes that arbitration clauses are merely a part of the main commercial contract. The Circuits adhering to this approach have never explicitly considered the severability of arbitration clauses. By holding that alleging that a contract is void is sufficient to put an arbitration provision at issue within the meaning of section 4 of the FAA, the courts have erroneously implicitly assumed that arbitration clauses are not severable from the underlying contracts. This assumption that an arbitration clause is merely a part of the underlying contract is flawed for two reasons. First, courts adhering to this approach assume a fact that is worth an independent determination—whether an arbitration provision stands separately from the contract. Second, the assumption that an arbitration provision is merely a part of the underlying contract flies in the face of the FAA’s treatment of arbitration provisions as separate agreements, which the Court in *Prima Paint* implicitly recognized. Thus, the Third, Ninth, and Eleventh Circuits should reconsider their approach to claims that a commercial contract containing an arbitration clause is void, and allow courts to proceed to resolve such claims only upon an express finding that the arbitration clause is not severable from the allegedly void contract.

220. See *id.*; see also Robert Lawrence Co. v. Devonshire Fabrics, Inc., 271 F.2d 402, 409–10 (2d Cir. 1959) (summarizing the cases that treat arbitration clauses as separate agreements unaffected by invalidity of the underlying commercial contracts).


222. See supra Part I.C.


D. The Void/Voidable Distinction is Only Appropriate if an Arbitration Provision is Not Severable from the Underlying Commercial Contract

It is true that the void/voidable distinction can be useful in determining whether a party has in fact assented to an arbitration agreement. Federal courts, when faced with an allegation that a contract containing an arbitration clause is void from its very inception, should first ascertain whether the arbitration clause is separable from the contract. Given that the FAA treats arbitration provisions as severable from the underlying contracts, courts should presume that an arbitration provision is severable unless a party proves the opposite.

If a court finds that the arbitration clause is merely a part of the underlying contract, the court should resolve the claim whether the contract is void. This would be consistent with the position of the Third, Ninth, and Eleventh Circuits that “a party who contests the making of a contract containing an arbitration provision cannot be compelled to arbitrate the threshold issue of the existence of an agreement to arbitrate.” However, if the court determines that the arbitration clause is severable from the allegedly void contract, the court should consider the enforceability of the arbitration clause only if a party puts the clause itself in issue. This approach would help promote a uniform application of Prima Paint’s rule and avoid confusion related to the specificity of factual situations in which the challenges to the contract containing arbitration clauses arise. Further, it would be consistent with the plain meaning of the FAA and a federal policy favoring enforcement of arbitration agreements.

V. CONCLUSION

In a dispute over the validity or performance of a contract containing an arbitration clause, a federal court, upon being satisfied that the subject matter of the dispute falls within the scope of arbitrable issues, should order the parties to proceed to arbitration unless any party to the contract specifically asserts that the arbitration provision is not enforceable. The

225. Id.
228. See supra Part I.A.
U.S. Supreme Court in *Prima Paint* expressed no opinion as to which forum, a court or an arbitrator, should adjudicate a party’s allegation that an entire contract containing an arbitration clause is void. However, the Court made it clear that, under section 4 of the FAA, a specific challenge to an arbitration agreement is necessary in order to avoid its enforcement.

A federal court considering whether a contract containing an arbitration clause is void from its very inception should first ascertain whether the arbitration clause is separable from the contract. In view of the FAA’s language and policy favoring enforcement of arbitration agreements, the court should presume severability of an arbitration provision unless there is proof to the contrary. Only if the court finds that an arbitration clause is merely a part of the underlying commercial contract should the court resolve a claim that the contract is void. This approach would promote uniform application of the *Prima Paint* rule to different factual settings. Further, it is consistent with the plain requirements of sections 2, 3 and 4 of the FAA and the strong federal policy favoring enforcement of arbitration agreements in interstate and international commerce.
RESEARCHER LIABILITY FOR NEGLIGENCE IN HUMAN SUBJECT RESEARCH: INFORMED CONSENT AND RESEARCHER MALPRACTICE ACTIONS

Roger L. Jansson, M.P.H., M.P.A.

Abstract: Two sets of federal regulations, the “Common Rule” and Food and Drug Administration (FDA) regulations, govern human subject research that is either federally-funded or involves FDA regulated products. These regulations require, inter alia, that: (1) researchers obtain informed consent from human subjects, and (2) that an Institutional Review Board (IRB) independently review and approve the research protocol. Although the federal regulations do not provide an express cause of action against researchers, research subjects should be able to bring informed consent and malpractice actions against researchers by establishing a duty of care and standard of care. Researchers owe human subjects a duty of care analogous to the special relationship between physicians and patients. The federal regulations should provide the minimum standard of care for informed consent in human subject research, and complying with them should be a partial defense. In contrast, expert testimony should establish the standard of care for researcher malpractice, and IRB approval should be a partial defense.

In 1993, the Kennedy Krieger Institute (KKI), a prestigious research institute associated with Johns Hopkins University, initiated a two-year research program in Baltimore to study the effectiveness of methods used to reduce lead exposure in children. The researchers measured the extent to which the children’s blood became contaminated with lead, and compared those blood levels with the measurements of lead dust in the children’s homes over the same period of time. Two outraged parents filed separate negligence actions when their children developed elevated lead levels in their blood while participating in the research program. The parents alleged that they were not fully informed of the study’s risks, and that the researchers violated a duty to warn them when the children’s blood-lead levels became elevated.

In *Grimes v. Kennedy Krieger Institute*, the Circuit Court for Baltimore City granted KKI’s motion for summary judgment on the ground that the researchers and institution had no legal duty to the

2. *Id.* at 812.
3. *Id.* at 825–29.
4. *Id.*
5. 782 A.2d 807, 832 (Md. 2001).
plaintiffs. Both plaintiffs appealed, and Maryland’s high court held that there are several potential sources for a duty of care that researchers owe to human subjects, and that the breach of such duties may result in negligence. The *Grimes* court also held that a parent cannot consent to a child’s participation in nontherapeutic research that poses “any risk” to the subject, a standard of care seemingly higher than that provided by the federal regulations. The court later clarified that “any risk” means “any articulable risk beyond the minimal kind . . . inherent in any endeavor.” It remains unclear whether this definition sets a higher standard of care than the federal regulations, or merely interprets allowable risk under the regulations.

The *Grimes* decision is part of a larger trend in which injured research subjects have filed lawsuits against researchers. These high-profile lawsuits have generated intense media attention. This period of increased legal and public scrutiny of the research enterprise will likely lead to a dramatic rise in lawsuits by human subjects. However, few cases have generated reported decisions; thus the case law interpreting tort liability of researchers is scarce. Yet, in the near future many state courts will be faced with the task of developing an approach for dealing with negligence claims in human subject research.

---

6. *Id.* at 858.
7. *Id.*
10. See infra notes 162–63.
When human subjects are harmed, the most likely source of liability for researchers is based in negligence. There are two initial considerations when determining whether human research subjects can bring negligence actions against researchers. First, is there a duty of care owed to human research subjects? Second, if so, what is the standard of care?

This Comment argues that state courts should recognize that researchers owe human subjects a duty of care rooted in informed consent and researcher malpractice. Part I describes the existing federal regulations and the elements of negligence. Part II explores the special relationship between a researcher and subject as a potential source for the duty of care researchers owe to human subjects. Part III examines the standard of care in negligence actions by research subjects for informed consent and researcher malpractice. Part IV argues that researchers owe a duty of care to human subjects, and that plaintiffs can establish a standard of care in negligence actions against researchers for informed consent and researcher malpractice actions. The federal regulations governing human subject research provide the best standard of care for informed consent, and complying with them should provide a partial defense. Further, expert testimony under the medical malpractice model provides the best standard of care for researcher malpractice, although Institutional Review Board (IRB) approval should provide a partial defense.

I. HUMAN SUBJECT RESEARCH AND NEGLIGENCE

Two sets of federal regulations govern human subject research: the Common Rule and Food and Drug Administration (FDA) regulations. The Common Rule applies to research funded by federal departments

13. Other potential sources of liability including battery, contract-related claims, fraud and misrepresentation, strict liability in product liability cases, and constitutional claims are beyond the scope of this Comment. For a discussion of other bases of liability, see Jay M. Zitter, Annotation, Recovery for Nonconsensual Human Medical Experimentation, 2002 A.L.R. 5th 11 (2002). In addition, many injuries to human subjects may simply be adverse events, which are also beyond the scope of this Comment. For a discussion of other compensation systems for human subjects, see President’s Commission for the Study of Ethical Problems in Medicine & Biomedical & Behavioral Research, Compensating for Research Injuries: The Ethical & Legal Implications of Programs to Redress Injured Subjects (1982).


15. The Common Rule is a set of federal regulations issued to protect human subjects. The Common Rule is explained in detail at infra notes 31–51 and accompanying text.
and agencies that conduct, support, and regulate human subjects research.\textsuperscript{16} In contrast, the FDA regulations apply to research involving experimental drugs, biological products, and medical devices that are subject to FDA approval.\textsuperscript{17} In order for injured subjects to successfully bring negligence actions against researchers, they must first prove that the researcher owed a duty to the subject, and that the duty was breached.\textsuperscript{18} The federal regulations governing human subject research do not provide an express cause of action for negligence. Therefore, an independent duty must exist at common law in order for researchers to be held negligent.

\textbf{A. The Common Rule and FDA Regulations}

The federal regulations currently governing human subject research are rooted in the international guidelines and codes that arose after human subjects were abused in the Twentieth Century.\textsuperscript{19} After World War II and Nazi medical experiments, American judges developed the Nuremberg Code, setting forth ten ethical principles that provided a foundation for the protection of human subjects.\textsuperscript{20} Later, doctors and scientists of the World Medical Association adopted the Declaration of Helsinki and set international guidelines for biomedical research by physicians.\textsuperscript{21} In 1966, the Public Health Service developed internal policy guidelines requiring peer review of all research involving human subjects.\textsuperscript{22}

\begin{itemize}
\item \textsuperscript{16} 45 C.F.R. § 46.101 (2001).
\item \textsuperscript{17} 21 C.F.R. §§ 50.1, 56.101 (2001).
\item \textsuperscript{18} See \textit{Keeton et al.}, supra note 14, § 30 at 164–65.
\end{itemize}
By the mid-1970s, several highly publicized American research abuses led Congress to enact the National Research Act of 1974, which created the National Commission for the Protection of Human Subjects in Biomedical and Behavioral Research (National Commission). In 1978, the National Commission published The Belmont Report, setting forth three ethical principles—respect for persons, beneficence, and justice—that later served as the foundation for the federal regulations governing human subject research. Each principle translated into a particular application to protect human subjects: “respect for persons” translated into “informed consent,” “beneficence” into “risk-benefit assessment,” and “justice” into “selection of subjects.” The Belmont Report deliberately set a high standard of disclosure for informed consent in the research setting, stating that “the research subject, being in essence a volunteer, may wish to know considerably more about risks gratuitously undertaken than do patients who deliver themselves into the hands of a clinician for needed care.”

The Federal Policy for the Protection of Human Subjects, known as the Common Rule, is a set of federal regulations that incorporates the ethical principles and guidelines of The Belmont Report. With minor variations, it applies to all federal departments and agencies that conduct, support, and regulate human subject research. The Common Rule was developed under the Public Health Service Act’s mandate “to protect the

---

23. See, e.g., Henry K. Beecher, Ethics and Clinical Research, 274 NEW ENG. J. MED. 1354, 1354–60 (1966) (publicizing 22 examples of unethical research protocols); JAMES H. JONES, BAD BLOOD: THE TUSKEGEE SYPHILIS EXPERIMENT (1993) (recounting the infamous study in which hundreds of African-American men with syphilis were left untreated for decades to gain scientific knowledge about progression of the disease).


25. Id. at § 203 (charging the National Commission with identifying the basic ethical principles underlying all research involving human subjects and developing research guidelines from those principles).


27. Id. at 4.

28. Id. at 4–20.

29. Id.


31. See ADVISORY COMMITTEE ON HUMAN RADIATION EXPERIMENTS, supra note 22, at 425–38 (explaining minor variations of the Common Rule across federal departments and agencies).
The regulation that formed the basis for the Common Rule was originally adopted in 1981 by the Department of Health, Education, and Welfare (DHEW), a predecessor to the Department of Health and Human Services (HHS). In 1991, this regulation was adopted by other federal departments and agencies as the Common Rule. The Common Rule applies to federally-funded research on human subjects, and to research conducted at institutions that have contractually agreed through a Federal Wide Assurance (FWA) to apply the Common Rule to all research regardless of the funding source. The HHS Office for Human Research Protections (OHRP) enforces institutional compliance with the Common Rule.

The Common Rule places the responsibility of obtaining informed consent directly on the researcher, and states that the researcher must minimize the possibility of coercion and that the research must provide the prospective subject with sufficient opportunity to consider whether to participate in the study. The elements of consent include an explanation of the purposes and procedures of the research, a description of any reasonably foreseeable risks and benefits, and a disclosure of alternatives. The consent process must also explain the confidentiality policy, state that participation is voluntary, explain whether any compensation is available if injury occurs, and provide information on whom to contact with additional questions. The informed consent

33. See ADVISORY COMMITTEE ON HUMAN RADIATION EXPERIMENTS, supra note 22, at 425.
34. Each department and agency has codified the Common Rule in different parts of the Code of Federal Regulations. When citing the Common Rule, this Comment always refers to the Health and Human Services (HHS) regulations at 45 C.F.R. § 46.101–24.
35. See 45 C.F.R. § 46.103 (setting forth the assurance system, which requires all institutions receiving federal funding to conduct research to provide formal assurance to Office for Human Research Protections (OHRP) (or other applicable agency designee) that it will develop a system to protect human subjects). In December 2000, OHRP developed a Federal Wide Assurance (FWA) to streamline the assurance process. See http://ohrp.osophs.dhhs.gov/irbasur.htm (last visited Jan. 3, 2003).
36. OHRP, located in the office of the Secretary of Health and Human Services since June 2000, was formerly known as the Office for the Protection from Research Risks (OPRR), which was located in the National Institutes of Health (NIH). See 65 Fed. Reg. 37,136 (June 13, 2000). Even though the Food and Drug Administration (FDA) is part of HHS, the FDA (not OHRP) oversees compliance with 21 C.F.R. sections 50 and 56. See 21 C.F.R. § 56.120-124.
37. 45 C.F.R. § 46.116.
38. Id. § 46.116(a).
39. Id.
cannot include any exculpatory language that waives “any of the subject’s legal rights, or releases or appears to release the investigator, the sponsor, the institution or its agents from liability for negligence.”

The Common Rule also sets forth circumstances where the requirements for informed consent can be altered or waived. In addition, the Common Rule states that these informed consent requirements do not preempt any federal, state, or local laws which require additional information to be disclosed for informed consent.

The Common Rule also requires IRBs to independently review research protocols that involve human subjects, and sets forth the requirements for the review of research and the criteria for its approval. The IRB must review the informed consent form for compliance with federal requirements, and must perform continuing reviews at least once a year. The criteria for IRB approval of research include a determination that: (1) the risks to subjects are minimized, (2) the risks to subjects are reasonable in relation to anticipated benefits, (3) the selection of subjects is equitable, and (4) informed consent is sought and documented. In addition, the Common Rule allows expedited review for certain kinds of research involving minimal risk, and for minor changes in approved research. Thus, the Common Rule places the responsibility on IRBs to review and approve research protocols, and


40. Id. § 46.116; see also 21 C.F.R. § 50.20 (2001) (prohibiting exculpatory agreements concerning negligence in clinical investigations regulated by the FDA).
41. 45 C.F.R. §§ 46.116(c)–(d) (listing circumstances where the IRB may allow the elements of consent to be altered or waived).
42. Id. § 46.116(c).
43. Id. § 46.107 (requiring the IRB to include at least five members who possess the professional competence necessary to review specific research activities, including at least one member who is not affiliated with the institution).
44. Id. § 46.102(d) (“Research’ means a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge.”).
45. Id. § 46.102(f) (“Human subject’ means a living individual about whom an investigator (whether professional or student) conducting research obtains (1) Data through intervention or interaction with the individual, or (2) Identifiable private information.”).
46. Id. § 46.109.
47. Id. § 46.111.
48. Id. §§ 46.109(b)–(c).
49. Id. § 46.109(e).
50. Id. § 46.111(a).
51. Id. § 46.110.
merely requires researchers to obtain informed consent of subjects and IRB approval.

The FDA developed a similar set of federal regulations in 1981 to oversee research involving experimental drugs, biological products and medical devices subject to FDA approval, even if it is privately funded. The FDA regulations also require informed consent, IRB review and approval. The most notable difference between the FDA regulations and the Common Rule is that the FDA regulations provide fewer exceptions to the informed consent requirements, and fewer waivers. Together, the Common Rule and the FDA regulations serve as guidelines for human subject research.

B. Negligence Actions Against Researchers Generally

The Common Rule does not provide an express cause of action for negligence. Instead, it merely states that violating the federal regulations may result in a loss of federal funding and thus suspension or termination of research. The OHRP has recently increased enforcement and temporarily suspended all human subject research at several prominent institutions for violations of the federal regulations.

However, researchers may be liable to injured subjects for their negligent actions. There may be a special relationship between the researcher and human subject that imposes a duty. In order to prevail in a negligence action against a researcher, an injured subject must prove that: (1) a duty is owed by the defendant; (2) the duty was breached; (3)
Researcher Negligence in Human Subject Research

the breach caused injury to the subject; and (4) a cognizable injury. These alternate causes of action correspond to two distinct interests of patients: self-determination and competent care. Thus, a jury’s exoneration of a physician from liability for medical malpractice does not forestall a plaintiff’s claim for failure to obtain informed consent and vice versa. Researcher liability may also be based on these alternate causes of action, resulting in two distinct types of researcher negligence: informed consent and researcher malpractice.

Physicians’ obligation to obtain patients’ informed consent before providing medical treatment has become a general duty owed by physicians to patients. To establish negligence in medical informed consent litigation, the plaintiff must prove: (1) that the physician owed a duty to disclose information to the patient; (2) the physician breached the duty under the appropriate standard of disclosure; (3) the plaintiff was injured; (4) the injury was the result of an undisclosed outcome or risk; and (5) had the plaintiff been informed of the outcome or risk, the plaintiff would not have consented. Before providing treatment, the physician must obtain consent by disclosing and explaining all information that is necessary for the patient’s decision, which typically includes the nature and purpose of the treatment, expected benefits, foreseeable risks, reasonable alternatives, and foreseeable risks of forgoing treatment. Lack of informed consent is generally treated as professional negligence, and is based on a physician’s professional duty

60. See KEETON ET AL., supra note 14, § 30 at 164–65. Proving causation and cognizable injury is beyond the scope of this Comment.


65. Id. at 29.

66. See, e.g., Truman v. Thomas, 611 P.2d 902, 905 (Cal. 1980) (en banc).

to provide patients with the appropriate information before they consent to treatment.\textsuperscript{68} Courts considering informed consent actions against researchers must determine how to establish the appropriate standard of care for informed consent in human subject research.

Professional negligence, or malpractice, is a special type of negligence in which professional standards of care have been developed for persons possessing or claiming to possess special knowledge or skill.\textsuperscript{69} Medical malpractice is the type of professional negligence most analogous to researcher malpractice, and is generally defined as a failure to exercise the required degree of care, skill and diligence ordinarily possessed by a reasonable and prudent physician in the same medical specialty acting under the same or similar circumstances.\textsuperscript{70} In medical malpractice, the standard of care is usually derived from expert testimony by the medical profession.\textsuperscript{71} Courts considering researcher malpractice actions must determine how to establish the appropriate standard of care.

Because negligence actions against researchers have been rarely reported, there are two important issues that remain unexplored. First, do researchers owe a duty of care to human subjects? Second, if they do, how can subjects establish standards of care to assess whether researchers have breached that duty for informed consent and researcher malpractice?

II. DUTY OF CARE

Courts have historically treated researcher negligence actions as part of the medical malpractice framework.\textsuperscript{72} Recently courts have begun to develop a specialized analysis for negligence causes of action in human subject research.\textsuperscript{73} To bring a successful negligence claim against researchers,\textsuperscript{74} a plaintiff must first establish that the defendant owed a

\begin{thebibliography}{99}
\bibitem{68} Faden & Beauchamp, \emph{supra} note 64, at 29.
\bibitem{69} \emph{Id}.
\bibitem{70} See \textsc{BARRY R. FURROW ET AL.}, \textsc{Health Law} 269 (2d ed. 2000).
\bibitem{71} \emph{Id} at 270; see Moringa v. Vue, 85 Wash. App. 822, 832, 935 P.2d 637, 642 (1997) (stating that expert testimony is required unless evidence is observable by lay persons and describable without medical training).
\bibitem{72} See \textit{infra} notes 77–79 and accompanying text.
\bibitem{73} See \textit{infra} notes 80–94 and accompanying text.
Researcher Negligence in Human Subject Research

duty of care. Although many possible sources of this duty exist, this Comment focuses on the special relationship between researchers and subjects, and analogizes it to the special relationship between physicians and patients.

A. Historical Treatment of Researcher Negligence Actions

Historically, courts have been reluctant to allow clinical research that deviates from accepted medical practice. Courts have relied on the medical malpractice framework to analyze negligence actions in human subject research. Early experimental deviations from medical standards of care were automatically considered malpractice, without regard to the level of care used or the experimental nature of the treatment. Courts did not accept clinical research as a necessary endeavor until the 1930s.

Analyzing human subject research as a separate negligence cause of action, apart from medical malpractice actions, is a relatively recent development. Halushka v. University of Saskatchewan, a 1965 Canadian case, was one of the first North American decisions to distinguish between medical practice and medical research. In the 1970s,

mandated that the hospital assume the duty of obtaining informed consent.”). There is no precedent for holding IRB members liable for negligence. However, there is split authority on whether state peer review statutes designed to encourage peer review by providing physicians who serve on committees with immunity for decisions to discipline their peers will protect IRB members from liability. Compare Konrady v. Oesterling, 149 F.R.D. 592, 598 (D. Minn. 1993) (holding that an IRB was not a peer review committee under the Minnesota peer review statute, and that the records were not protected), with Doe v. Ill. Masonic Med. Ctr., 696 N.E.2d 707, 711 (III. App. Ct. 1998) (holding that documents used by hospital IRB in the course of a medical study come within the scope of privilege of the Illinois peer review statute). See also VA. CODE ANN. § 8.01-44.1 (Michie 2001) (explicitly providing civil immunity to IRB members in Virginia).

75. See KEETON ET AL., supra note 14, at 164–65.

76. See Grimes v. Kennedy Krieger Inst., Inc., 782 A.2d 807, 858 (Md. 2001) (stating that potential sources for the duty the researchers owe to subjects include: the special relationship between researcher and subject, the informed consent quasi-contract, an implied duty from the federal regulations, and duties from international codes).


78. See, e.g., Carpenter v. Blake, 60 Barb. N.Y. 488, 523–24, rev’d 50 N.Y. 696 (N.Y. Gen. Term 1871) (holding that an innovative treatment was negligent because it deviated from standard practice).

79. See Fortner v. Koch, 261 N.W. 762, 765 (Mich. 1935) (“[w]e recognize the fact that, if the general practice of medicine and surgery is to progress, there must be a certain amount of experimentation carried on.”).

80. [1965] D.L.R.2d 436, 443–44 (“The duty imposed upon those engaged in medical research to those who offer themselves as subjects for experimentation is at least as great as, if not greater than, the duty owed by the ordinary physician or surgeon to his patient.”).
there was a transition period in which courts blurred the distinction between causes of action based on traditional medical malpractice and those based on research. The distinction between clinical practice and research has also been a difficult and complex problem in regulatory and ethical codes. Early discussions of research ethics and regulation often distinguished between “therapeutic” and “nontherapeutic” research, although many commentators have argued that this distinction is illogical and confusing. The Common Rule abandoned this distinction, simply referring to “research,” but some courts continue to refer to “nontherapeutic research” in their opinions.

Recently, courts have begun to develop a specialized analysis for negligence causes of action in human subject research. In 1986, a federal district court in the Middle District of North Carolina became the first American court to carefully address the duty of care and standard of care for negligence actions based on informed consent under the Common Rule. In Whitlock v. Duke University, the court held that under the Common Rule there is a heightened duty for disclosure of foreseeable risks that differs from the medical context. The Whitlock court adopted the Common Rule as the standard of care for informed consent in human subject research, but it did not reach the question of whether a duty of


82. See ROBERT J. LEVINE, ETHICS AND REGULATION OF CLINICAL RESEARCH 3 (2d ed. 1986) (defining “research” as “a class of activities designed to develop or contribute to generalizable knowledge,” whereas “practice of medicine” refers to “a class of activities designed solely to enhance the well-being of an individual patient or client”).

83. Id. at 8 (explaining that the original Declaration of Helsinki distinguished “nontherapeutic” non-clinical research from “therapeutic” clinical research).

84. Id. at 8–10 (explaining that many types of research cannot be defined as therapeutic or nontherapeutic, such as placebo-controlled, double-blind drug trials in which nobody knows who is receiving placebo).


87. 637 F. Supp. 1463 (M.D.N.C. 1986), aff’d, 829 F.2d 1340 (4th Cir. 1987) (involving an experienced diver in a simulated deep sea diving experiment to research high pressure nervous syndrome).

88. Id. at 1471. In dicta, the court also considered standards of disclosure under the Nuremberg Code and the Declaration of Helsinki. Id. at 1470–71.

89. See id. at 1471.
care is implied by the Common Rule because it found that the Common Rule’s standard of care was not breached. 90

In 2001, Maryland’s high court held in *Grimes* that there are several potential sources of duty that researchers owe to human subjects. 91 These include the special relationship between researcher and subject, the informed consent quasi-contract, implied duties from the federal regulations, and duties from international codes. 92 Although the *Grimes* court concluded that researchers may owe human subjects a duty of care, it did not clearly articulate which of these sources the duty arises from. 93 However, *Grimes* was the first case to explicitly state that “the very nature of nontherapeutic scientific research on human subjects can, and normally will, create a special relationship out of which duties arise.” 94

**B. A Source for the Duty of Care: The Special Relationship Between Researchers and Subjects**

Although there are a number of potential sources of duty that researchers owe to human subjects, the special relationship between researchers and human subjects offers the most likely choice. In general, a person has no duty to aid someone unless he or she placed that person in danger or had a “special relationship” with the person that created a duty. 95 A duty to aid or protect someone is typically found in relationships involving dependence or mutual dependence. 96 The relationship between a researcher and human subjects is analogous to the special, fiduciary relationship between physicians and patients. 97

---

90. *Id.* at 1475.
91. *Grimes*, 782 A.2d at 858.
92. *Id.*
93. *See id.*
94. *Id.* at 835–36.
95. *See RESTATEMENT (SECOND) OF TORTS § 314A (1965) (describing four commonly recognized special relationships in which an actor is under a duty to another: (1) common carrier-passenger; (2) innkeeper-guest; (3) landowner-invitee; and (4) certain custodial relationships).*
96. *See id. cmt. b.*
97. *See, e.g., Angela R. Holder, Do Researchers and Subjects Have a Fiduciary Relationship?, 4 IRB: ETHICS & HUMAN RESEARCH 6–7 (1982).*
1. **Physician-Patient Relationship**

In many states, the existence of a special physician-patient relationship is essential to negligence actions against physicians. Physicians have been regarded as fiduciaries of their patients and as such are expected to act in their patients’ best interests. This fiduciary duty arises out of the trust patients must place in a physician’s skill, learning, and experience. Once the physician gains this trust, the physician assumes the duty of informing patients of the nature and hazards of their disease and treatment.

Courts have also viewed the special relationship between physicians and patients as express or implied contracts. Although a physician has no duty to treat a patient until entering into a consensual transaction, once the physician enters into this relationship the physician is obligated to treat the patient at a certain standard of care. Consequently, medical malpractice law has included a mixture of contract and tort influences for over a century.

2. **Researcher-Subject Relationship**

If a medical researcher is also the subject’s treating physician, courts may view the dual physician/researcher-patient/subject relationship as “special” under traditional principles of medical malpractice. However, it is less clear whether courts would view researchers who are not a subject’s treating physician as owing a special duty to the research subject. *Grimes* was the first case to explicitly hold that the researcher-subject relationship itself constitutes a “special relationship” similar to

---

98. See, e.g., Anderson v. Houser, 523 S.E.2d 342, 345 (Ga. Ct. App. 1999) (reasoning that doctor-patient privity establishes the legal duty to conform to a standard of conduct and is essential to medical malpractice claims).

99. See, e.g., Carson v. Fine, 123 Wash. 2d 206, 218, 867 P.2d 610, 617 (1994) (en banc) (recognizing that the physician-patient relationship is a fiduciary one).

100. See, e.g., Canterbury v. Spence, 464 F. 2d 772, 782 (D.C. Cir. 1972) (stating that the doctor’s fiduciary duty of disclosure requires informed consent).

101. See, e.g., Moldoff, supra note 67, at § 1.

102. See FURROW, supra note 70, at 265.


105. Holder, supra note 97, at 6.
that between a physician and patient. The Grimes court stated that “the very nature of nontherapeutic research can, and normally will, create special relationships out of which duties arise.”

Although the Grimes court did not cite any authority for holding that a special relationship exists between researchers and subjects, commentators have noted that there are several similarities between the research-subject and physician-patient relationship. These commentators have argued that the researcher-subject relationship is fiduciary, because a researcher’s specialized knowledge makes the subject dependent on the researcher. Similar to the medical malpractice context, the risk that researcher negligence could cause a subject bodily injury might lead a court to impose special duties of care and disclosure on the researcher.

In addition, a researcher enters into a quasi-contractual relationship with the subject by obtaining the subject’s written informed consent. This quasi-contractual relationship might serve as the basis for a special relationship similar to that of physician and patient. In Dahl v. Hem Pharmaceuticals Corp., the Ninth Circuit held that a consent form in a research protocol formed a unilateral contract. The Grimes court took this contract analysis a step further and held that the informed consent agreements in research protocols may constitute bilateral contracts that create special relationships between researchers and subjects. Thus, there are significant similarities between the physician-patient and researcher-subject relationship that may impose similar duties of care.

107. Id. at 835–36.
108. See Holder, supra note 97, at 6–7; see also Delgado & Leskovac, supra note 62, at 107–12 (arguing that the researcher has a fiduciary duty with the human subject, similar to that in the typical physician-patient relationship).
110. See Grimes, 782 A.2d at 846 (stating that recruitment of otherwise healthy subjects into potentially hazardous study conditions for the purpose of creating statistics for testing scientific hypotheses “would normally warrant or create such special relationships as a matter of law”).
111. See 45 C.F.R. § 46.117 (2001) (requiring a written informed consent form unless the IRB waives the requirement in exceptional circumstances).
112. 7 F.3d 1399 (9th Cir. 1993).
113. Id. at 1404–05 (allowing plaintiff to continue receiving the study drug even after the sponsor had ended the study, based on a reliance argument in contract law).
III. STANDARD OF CARE

Once a plaintiff in a negligence action establishes that the defendant owes the plaintiff a duty of care, the court must determine whether the defendant met the standard of care.115 Courts considering informed consent and malpractice actions against researchers must determine the standard of care that researchers owe subjects in each action. In both informed consent and researcher malpractice actions, there are two potential sources for the standard of care: federal regulations and expert testimony.

A. Standard of Care for Informed Consent in Research

Physicians’ obligation to obtain patients’ informed consent before providing medical treatment has become a general duty owed by physicians to patients.116 The informed consent requirement in medicine is based on a respect for the patient’s autonomy and right to self-determination.117 As early as 1914, Justice Cardozo articulated this principle in an oft-quoted opinion: “Every human being of adult years and sound mind has a right to determine what shall be done with his own body . . . .”118 In order to meet the standard of care for informed consent, the physician must obtain, disclose, and explain all of the information necessary for a patient to decide.119 This typically includes discussing the nature and purpose of the treatment, its expected benefits, foreseeable risks, and reasonable alternatives, as well as the foreseeable risks of forgoing treatment.120 Courts faced with determining the standard of care in researcher informed consent actions will most likely consider the federal regulations and expert testimony. Typically, courts have adopted the federal regulations as the standard of care for informed consent actions against researchers121 and relied on expert testimony to determine whether the standard of care was met.122

115. See Keeton et al., supra note 14, at 164–65.
116. See generally King, supra note 64, at 114–50.
119. See, e.g., Truman v. Thomas, 611 P.2d 902, 905 (Cal. 1980) (en banc).
120. See generally John H. Derrick, Annotation, Medical Malpractice: Liability for Failure of Physician to Inform Patient of Alternative Modes of Diagnosis or Treatment, 38 A.L.R. 4th 900 (1985); Moldoff, supra note 67.
121. See infra Part III.A.1.
122. See infra Part III.A.2.
Researcher Negligence in Human Subject Research

1. Using the Federal Regulations to Establish the Standard of Care in Researcher Informed Consent Actions

Courts have held that a statute or administrative regulation can provide the standard of care if a duty of care exists at common law.\(^{123}\) Violations of federal statutes and regulations have served as the basis for negligence actions in many areas of law. These include violating safety standards for drugs, medical devices, and pesticides,\(^ {124}\) not complying with environmental regulations on toxic waste,\(^ {125}\) and violating the Occupational Safety and Health Act (OSHA) regulations.\(^ {126}\) Courts can adopt an administrative regulation as the standard of care if the administrative regulation’s purpose is: (1) to protect a class of persons including the person whose interest was invaded, (2) to protect the particular interest invaded, (3) to protect that interest against the kind of harm which has resulted, and (4) to protect that interest against the particular hazard from which the harm results.\(^ {127}\) In most jurisdictions, an unexcused violation of a relevant federal statute or regulation constitutes negligence per se as a matter of law.\(^ {128}\) A minority of jurisdictions treat violations of statutes or regulations as evidence of negligence, creating a rebuttable presumption of negligence.\(^ {129}\)

Although state courts are not required to adopt federal regulations as the standard of care in tort litigation, commentators have argued that state courts should give federal standards more weight.\(^ {130}\) As early as

\(^{123}\) See Restatement (Second) of Torts § 285 (1965).


\(^{126}\) See John P. Ludington, Annotation, Violation of OSHA Regulation As Affecting Tort Liability, 79 A.L.R. 3d 962 (1977).

\(^{127}\) Restatement (Second) of Torts § 286 (1965).

\(^{128}\) See, e.g., Sanchez v. Galey, 733 P.2d 1234, 1242 (Idaho 1986) (reasoning that violation of OSHA regulations constituted negligence per se); see also Steagall v. Dot Mfg. Corp., 446 S.W.2d 515, 518 (Tenn. 1969) (stating that violation of the Federal Hazardous Substances Act may result in negligence per se).

\(^{129}\) Sherman, supra note 124, at 877–84 (comparing approaches to negligence per se in various states).

\(^{130}\) See Lars Noah, Rewarding Regulatory Compliance: The Pursuit of Symmetry in Products Liability, 88 Geo. L.J. 2147, 2152 (2000) ("Although not formally obligated to adopt relevant federal regulations as particularizing the standard of care in tort litigation, at least absent preemption, courts should take them more seriously than they do at present."); see also Richard Ausness, The Case for a "Strong" Regulatory Compliance Defense, 55 Md. L. Rev. 1210, 1253–54 (1996)
1960, one commentator proposed that the standard of care for human subject research should be at least partly based on applicable statutes and regulations. In 1997, a California court took the lead and held in *Daum v. SpineCare Medical Group* that the federal regulations designed to protect human research subjects provided the standard of care for informed consent in a negligence action against a researcher. The *Daum* court held that the human subject could successfully use the FDA regulations to prove the elements of negligence per se. Further, the court held that the trial court should have instructed the jury on negligence per se based on the defendant’s violation of the federal regulations on informed consent. Thus, the *Daum* court derived the standard of care for informed consent from the federal regulations, and limited the role of expert testimony to whether the standard was met.

Several other courts have relied on the federal regulations to establish the standard of care for informed consent in human subject research. The federal district court in *Whitlock* held that the Common Rule established the standard of care for an informed consent claim against researchers in North Carolina, and stated that there is a heightened standard of disclosure of foreseeable risks in research than in the medical context. Further, the Supreme Court of Washington adopted the Common Rule as the standard of care for informed consent in *Vodopest v. MacGregor*, holding that University of Washington researchers violated the Common Rule by requiring subjects to waive their legal rights to bring negligence actions in the consent form, a requirement that violates the Common (proposing a regulatory compliance defense but permitting plaintiffs to overcome it by producing clear and convincing evidence that the regulation was inadequate).

---

133. Id. at 264 (involving spinal surgery using an experimental fixation device).
134. Id. at 273.
135. See id. at 279 (“Instead of being able to try the case based on the statutes and regulations governing clinical trials, the Daums were forced to engage in a battle of experts over the duty of disclosure.”).
136. Id.
137. 637 F. Supp. 1463, 1475 (D.N.C. 1986), aff’d, 829 F.2d 1340 (4th Cir. 1987) (holding that the defendant did not violate the standard of care provided by the Common Rule, because the risk of organic brain damage was not foreseeable); see supra notes 87–90 and accompanying text.
138. Id. at 1471.
139. 128 Wash. 2d 840, 913 P.2d 779 (1996) (involving research on the effects of high altitude on subjects).
Rule’s informed consent provision. Similarly, courts have held that the FDA regulations set the standard of care for obtaining a patient’s informed consent before implanting experimental medical devices. Thus, there appears to be an emerging trend among courts to use the federal regulations as the standard of care for informed consent in human subject research.

Defendants have argued that complying with federal safety standards should shield them from state tort liability. These defendants invoke the federal preemption doctrine, arguing that the primacy of federal law under the Supremacy Clause overrides state tort law. Defendants have successfully employed the preemption doctrine to overcome claims by injured consumers involving federally regulated cigarette labels, medical device labels and designs, and motor vehicle designs. In addition, many defendants have argued that compliance with federal regulations should be a complete defense to state tort liability. However, few jurisdictions view regulatory compliance as a complete defense. Most states allow juries to take a defendant’s compliance with federal regulations into account, but instruct juries to treat federal safety standards as only a minimum floor of acceptable conduct. While state courts have often been willing to treat violation of federal statutes or regulations as negligence per se, they have been reluctant to treat compliance as a complete defense.

Although no courts have explicitly addressed whether complying with the federal requirements for informed consent is a complete defense in researcher negligence actions, several courts have granted summary judgment to defendants for compliance with the federal regulation.

140. Id. at 857–62, 913 P.2d at 787–89.
142. See Ausness, supra note 130, at 1225–37.
143. U.S. CONST. art. VI, cl. 2.
144. See Ausness, supra note 130, at 1225–37 (reviewing federal preemption of state product liability claims).
145. See id. at 1239–47 (reviewing regulatory compliance as a defense).
146. Id.
147. See Paul Dueffert, The Role of Regulatory Compliance in Tort Actions, 26 HARV. J. ON LEGIS. 175, 176, 186–88 (1989); see also RESTATEMENT (SECOND) OF TORTS § 288C (1965) (“Compliance with a legislative enactment or an administrative regulation does not prevent a finding of negligence where a reasonable man would take additional precautions.”).
148. See Noah, supra note 130, at 2151–52.
requirements for disclosure of risks. In *Slater v. Optical Radiation Corp.*, Judge Posner, writing for the Seventh Circuit Court of Appeals, observed that when the risks are adequately explained to the research subject, “he cannot complain that the risks materialized.” Further, Judge Posner cautioned that “[i]f experimental procedures are subject to hindsight evaluation by juries, so that failed experiments threaten to impose enormous tort liability on the experimenter, there will be fewer experimental treatments, and patients will suffer.” Thus, under *Slater*, defendants could argue regulatory compliance as a defense.

However, complying with the federal requirements for informed consent should not be viewed as a complete defense because the federal regulations explicitly state that they do not affect state or local laws in general. Further, the federal regulations specifically state that the informed consent requirements do not preempt any federal, state, or local laws which require additional information to be disclosed for informed consent. Thus, states can set higher standards of care for informed consent in human subject research without violating the federal regulations. Several states have specific statutes or regulations addressing human subject research that could provide a standard of care for informed consent. Although these laws usually adopt the identical requirements as the Common Rule, some states have enacted statutes or regulations providing additional protections to certain groups deemed vulnerable when participating as human subjects, such as prisoners, the mentally ill, and developmentally disabled persons.

150. 961 F.2d 1330 (7th Cir. 1992).
151. Id. at 1333–34.
152. Id.
154. Id. § 46.116(e).
At least two courts have imposed higher standards of care than those in the federal regulations for informed consent in human subject research without relying on state statutes or regulations governing human subject research. The Supreme Court of California, in *Moore v. Regents of the University of California*, 157 held that medical researchers must disclose conflicts of interest under state law on informed consent, 158 even though the federal regulations do not mention conflicts of interest. Further, the Maryland high court in *Grimes* arguably set a higher standard of care than the Common Rule for informed consent. 159 After holding that researchers may owe duties to human subjects, the *Grimes* court further held that parents cannot consent to their children’s participation in nontherapeutic research that involves “any articulable risk beyond the minimal kind of risk that is inherent in any endeavor.” 160 The federal regulations, however, allow research that poses greater than minimal risks for children in certain situations. 161 Thus, it is unclear whether this ancillary holding established a higher standard of care for parental consent in children’s research, 162 or was merely an attempt to interpret the level of risk allowed under the federal regulations. 163

In sum, there are few state statutes, regulations, or judicial decisions that set higher standards of care for human subject research than those in the federal regulations. Although the federal regulations allow states to provide additional protections to human subjects, the trend at this point appears to be for courts to adopt the federal regulations as the standard of care for informed consent.

---

158. Id. at 485.
160. Id. at 862.
162. See Lainie Friedman Ross, *In Defense of the Hopkins Lead Abatement Studies*, 30 J.L. MED. & ETHICS 50, 54 (2002) (“At trial, the legal and moral authority of parents to enroll their children in certain types of nontherapeutic research should be reaffirmed.”).
2. Using Expert Testimony to Evaluate the Standard of Care for Informed Consent

Even if the federal regulations provide the standard of care for informed consent, expert testimony may be required to determine whether researchers breached the standard of care. The federal regulations require researchers to disclose certain elements of consent, including an explanation of the purposes and procedures of the research, a description of any reasonably foreseeable risks and benefits, and a disclosure of alternatives. Courts must develop tests to decide whether researchers adequately met these federal requirements for disclosure. Currently, there are two different tests for establishing the appropriate standard of disclosure in medical informed consent actions: (1) the professional practice standard; and (2) what the “reasonable person” would want to know. The professional practice standard, adopted in over twenty-five states, requires expert testimony to establish the scope of a reasonable disclosure, based on what a reasonable practitioner would have disclosed in a similar situation. The “reasonable person” standard, growing in popularity and approaching a majority position, bases the appropriate scope of disclosure on what a “reasonable person” would want to know, regardless of professional practice. The trier of fact may determine what a “reasonable person” would want to know without expert testimony, but expert testimony may still be needed to clarify the nature of the treatment and its probability of risk.

Thus, even if the federal regulations are used to establish the standard of care for informed consent in human subject research, expert witnesses will most likely be required to help factfinders determine whether the researcher complied with the standard of care. For instance, in Stewart v. Cleveland Clinic Foundation, expert testimony was used to determine whether the standard of care set by the federal regulations was met. The defendant argued that the informed consent for a clinical trial

164. 45 C.F.R. § 46.116(a); see supra notes 37–42 and accompanying text.
165. See King, supra note 64, at 30–34 (also noting that a third “subjective” standard has been proposed in legal commentary, based on what the particular patient wants to know).
166. See FURROW, supra note 70, at 318.
167. Id. at 318.
168. Id. at 318–19.
169. Id. at 319.
170. Id.
172. Id. at 501.
complied with the Common Rule, while the plaintiff provided a contradictory expert report. The Stewart court held that a genuine issue of material fact existed as to whether the informed consent form complied with the federal requirements, and allowed the jury to decide. In sum, courts have generally adopted the federal regulations as the standard of care for informed consent, and have used expert testimony to determine whether the standard was breached.

B. Standard of Care for Researcher Malpractice

Medical malpractice is the type of professional negligence most analogous to researcher malpractice, and is generally defined as a failure to exercise the required degree of care, skill and diligence ordinarily possessed by a reasonable and prudent physician in the same medical specialty acting under the same or similar circumstances. The liability of health care providers for medical malpractice is governed by (1) state statutes and (2) expert testimony under medical malpractice law that varies from state to state. Although courts could determine the standard of care for researcher malpractice actions by using the federal regulations, no court has taken this approach because the federal regulations only impose general requirements on conducting research. At least one court has relied on expert testimony to establish the standard of care in researcher malpractice actions.

1. Using the Federal Regulations to Establish the Standard of Care for Researcher Malpractice

No courts have used the federal regulations to establish the standard of care for researcher malpractice outside the context of informed consent claims. Although the federal regulations are intended to protect the rights and safety of human subjects, they do not impose specific requirements on researchers beyond obtaining the subjects’ informed

173. Id. at 495.
174. Id. at 497.
175. Id. at 501.
176. See FURROW, supra note 70, at 269.
177. Id.
178. Heinrich v. Sweet, 308 F.3d 48, 70 (1st Cir. 2002).
179. See supra note 32 and accompanying text.
consent and IRB approval. The federal regulations place the responsibility of weighing risks and approving research on IRBs, not on researchers. Thus, unlike in informed consent, the federal regulations do not provide a specific standard of care for researcher malpractice actions.

However, it is possible that IRB approval could provide a partial defense in researcher malpractice claims, even though no courts have explicitly addressed this issue. IRB approval requires that an independent group of IRB members with professional competence determine, inter alia, that the research had scientific merit, risks were minimized, and the benefits outweighed the risks. Obtaining IRB review may therefore provide evidence of due care in researcher malpractice actions. But, it should be noted that the *Grimes* court allowed a negligence action against the researchers to proceed even though the researchers had obtained IRB approval. Indeed, the *Grimes* court accused the IRB of abdicating its responsibility by helping the researchers to “miscast the characteristics of the study” and by “aid[ing] researchers in getting around federal regulations.” Thus, not all courts have found IRB approval to be a partial defense to researcher malpractice actions.

2. Using Expert Testimony to Establish the Standard of Care for Researcher Malpractice

Most negligence actions against researchers to date have been based on informed consent claims rather than researcher malpractice, so there is little case law to determine the standard of care for researcher malpractice. Consequently, most commentators have focused on informed consent issues in human subject research, and have ignored or rejected researcher malpractice actions based on expert testimony.

181. *Id.* § 46.111.
182. *See supra* notes 43–51 and accompanying text.
183. *See supra* notes 43–51 and accompanying text.
186. *Id.* at 814.
187. *See supra* notes 132–41 and accompanying text.
188. *See Morin, supra* note 81; *see also* Delgado & Leskovac, *supra* note 62.
Nevertheless, expert testimony could be used to establish the standard of care for a reasonable researcher in the same research specialty acting under the same or similar circumstances, similar to the medical malpractice model. In the medical malpractice model, the standard of care is usually derived from expert testimony by members of the medical profession. The appropriate standard of care arises out of the complex interactions of professional leaders, journals, peer discussions and meetings. Thus, standards of care emerge gradually through the interplay of numerous comments from various professional sources once a practice becomes generally accepted. Once the standard of care is established by expert testimony, courts merely enforce the standard in tort suits rather than determining what the standard “should” be. To prove medical malpractice, the plaintiff must normally show that the defendant health care provider violated the standard of care through expert testimony. Very few cases have allowed patients to recover when the defendant health care provider complied with the standard of practice. 

Eugenics: Relying on Innovative Tort Doctrine To Provide Relief When Gene Therapy Fails, 35 GA. L. REV. 1277, 1303–08 (2001) (arguing that the medical malpractice model should be used to establish the standard of care for gene therapy research, similar to how it has been used to establish standards of care for innovative medical treatments that lack customary standards of care).

190. See Furrow, supra note 70, at 269.
191. Id. at 270.
192. Id. at 271–76.
193. In determining the standard of care for medical specialists, most jurisdictions have moved from a locality rule to a national standard. See Jay M. Zitter, Annotation, Standard of Care Owed to Patient By Medical Specialist as Determined by Local, “Like Community,” State, National, or Other Standards, 18 A.L.R. 4th 603 § 3 (1982) (listing cases that recognize that a specialist has the duty to possess and exercise that degree of skill and care ordinarily employed, under similar circumstances, by the members of his specialty in good standing, located in the same locality or community) and 18 A.L.R. 4th 603 § 6 (listing medical malpractice cases in which courts adopt standards of care based on national or nongeographic standards of similar specialists). The standard of care for general practitioners, interns, and residents is usually still based on the local or similar community standard. See, e.g., Bahr v. Harper-Grace Hosps., 497 N.W.2d 526 (Mich. Ct. App. 1993).
194. See, e.g., Osborn v. Irwin Mem’l Blood Bank, 7 Cal. Rptr. 2d 101, 128 (Cal. Ct. App. 1992) (stating that “professional prudence is defined by actual or accepted practice within a profession, rather than theories about what ‘should’ have been done”).
conforming with customary practice is generally a conclusive defense for a health care professional. 197

Recently in *Heinrich v. Sweet*, 198 the Court of Appeals for the First Circuit adapted the medical malpractice framework to human subject research. 199 The First Circuit vacated a jury verdict for the plaintiffs on the negligence claim and reversed in favor of the defendants, noting that evidence that the physician had breached the standard of care was flawed. 200 The *Heinrich* court relied on expert testimony to determine the standard of care in researcher malpractice, and held that the plaintiffs’ expert testimony lacked adequate foundation to show that the research violated the standard of care at the time it was performed, in 1960–61. 201 Although the research predated the modern federal regulations, the *Heinrich* court noted that the research received three levels of review by administrative committees at the hospital, and that “approval by these various committees is very compelling evidence” that the trials complied with the standard of care. 202 In addition, the *Heinrich* court observed that adequate informed consent is a partial defense to researcher malpractice claims, but admitted that there are situations in which deviating from the research protocol described in informed consent could lead to independent negligence. 203 Thus, the *Heinrich* court adapted the medical malpractice model to establish the standard of care in researcher malpractice.

In sum, although there is a trend for courts to adopt the federal regulations as the standard of care for informed consent actions against

197. Although the medical profession is in the unusual position of setting its own legal standards of conduct, merely by adopting its own medical practices as the standard, this privilege is based in part on the reasoning that physicians set those standards with “primary regard to protection of the public rather than to such considerations as increased profitability.” See Rosell v. Volkswagen of Am., 709 P.2d 517, 522, 523 (Ariz. 1985); see also Richard N. Pearson, *The Role of Custom in Medical Malpractice Cases*, 51 IND. L.J. 528, 536–37 (1976). But see Philip G. Peters, *The Quiet Demise of Deference to Custom: Malpractice Law at the Millenium*, 57 WASH. & LEE L. REV. 163, 170 (2000) (observing that judicial deference to physician customs is gradually eroding, with twelve states expressly rejecting deference to physician customs and another nine adopting a standard of care based on the reasonable physician).

198. 308 F.3d 48 (1st Cir. 2002).

199. Id. at 52–53 (explaining that the medical malpractice case was brought in 1995 for the deaths of two patients given an experimental treatment for brain cancer over four decades ago, in medical trials only recently made public).

200. Id. at 70–71.

201. Id. at 65–67.

202. Id. at 69.

203. Id. at 70.
researchers, limiting the scope of expert testimony to determining whether the federally-established standard was breached, no courts have adopted the federal regulations as the standard of care for researcher malpractice. Rather, courts such as the First Circuit in Heinrich have used expert testimony to establish the standard of care in researcher malpractice actions and viewed IRB approval as a partial defense.

IV. INJURED RESEARCH SUBJECTS SHOULD BE ABLE TO BRING INFORMED CONSENT AND RESEARCHER MALPRACTICE ACTIONS AGAINST RESEARCHERS

An injured research subject should be able to successfully bring a negligence action against a researcher by establishing that the researcher owes the subject a duty of care based on the special relationship between a researcher and subject. Once the injured research subject has demonstrated that a duty of care exists, courts must determine whether the researcher has violated the standard of care. A plaintiff should be able to establish that the federal regulations are the standard of care for informed consent in human subject research, using expert testimony to determine if the federal requirements for informed consent have been met. Violating the federal requirements for informed consent should result in a finding of negligence, and compliance with the federal regulations should provide a partial defense. Further, a plaintiff should be able to establish the standard of care for other researcher malpractice actions by using expert testimony, similar to the medical malpractice model. Failure to obtain IRB approval before proceeding with human subject research protocols should result in a finding of negligence, and IRB approval should provide a partial defense.

A. Establishing the Duty of Care For Human Subject Research

A plaintiff who participates in a research study should be able to establish a duty of care against researchers. Although the duty arising from the researcher-subject relationship has rarely been addressed by the courts, plaintiffs should be able to successfully establish that researchers owe a duty of care to human subjects that resembles the special physician-patient relationship.204 Grimes is the first case to explicitly hold that researchers owe duties to human subjects based on special

204. See supra notes 97–101 and accompanying text.
relationships. The Grimes court stated that “the very nature of nontherapeutic scientific research on human subjects can, and normally will, create special relationships out of which duties arise,” and noted that no laws preclude “the parties to a scientific study” from entering into “special relationships with the subjects of the study that can create duties.” Plaintiffs should be able to rely on Grimes as persuasive authority that the special relationship between researchers and subjects exists even outside of Maryland.

Further, plaintiffs should be able to establish that researchers owe subjects a duty of care by emphasizing the similarities between the researcher-subject relationship and the physician-patient relationship. Plaintiffs should argue first that negligence in human subject research may result in bodily injury creating a duty similar to that imposed in medical malpractice. Second, the researcher-subject relationship is fiduciary in nature because the subject is dependent on the researcher’s specialized knowledge. Third, the researcher-subject relationship is initiated by an informed consent quasi-contract, similar to the patient-physician relationship. Drawing on these similarities, injured research subjects should be able to successfully establish that researchers owe an affirmative duty of care to human subjects based on the special relationship that exists between researchers and subjects.

Although it can be argued that the researcher-subject relationship is distinguishable from the physician-patient relationship because researchers can potentially have little to no personal interaction with human subjects, the researcher-subject relationship arguably imposes a higher duty of care. Research often involves unforeseeable and unknown risks, increasing the subject’s dependence on the researcher’s specialized knowledge. There is often less accepted scientific evidence and experience to establish what is safe in human subject research than in

---

206. Id. at 834–35.
207. Id.
208. See supra Part II.B.
209. See Grimes, 782 A.2d at 846 (stating that recruitment of otherwise healthy subjects into potentially hazardous study conditions for the purpose of testing scientific hypotheses “would normally warrant or create such special relationships as a matter of law”).
210. See Holder, supra note 97, at 6–7; Delgado & Leskovac, supra note 62, at 107–12.
211. See supra notes 111–14 and accompanying text.
212. See supra notes 102–04 and accompanying text.
213. See Delgado & Leskovac, supra note 62, at 69.
medical practice. Further, the research subject often has less chance of receiving a direct benefit than medical patients, and is often acting out of altruism rather than self-interest. Whereas physicians are primarily attempting to help patients, researchers and subjects may have conflicting interests. Indeed, the federal requirements for written informed consent prior to research have been held to be more stringent than the unwritten consent implied in the physician-patient relationship, and to create a higher standard of disclosure of foreseeable risks. Thus, courts should find that researchers owe a duty of care to their subjects.

B. Establishing the Standard of Care for Human Subject Research

Plaintiffs should be able to bring two types of professional negligence actions against researchers: informed consent, and researcher malpractice. The federal regulations governing human subject research should provide the minimum standard of care for informed consent. Violating the federal requirements for informed consent should result be deemed negligence, and researchers’ compliance with them should be a partial defense. In contrast, plaintiffs should be able to establish a standard of care for researcher malpractice actions through expert testimony, and researchers who obtain IRB approval should be allowed a partial defense.

1. Establishing the Standard of Care for Informed Consent in Research

A standard of care may be established by administrative regulations, and violation of federal regulations has served as the basis for negligence actions in many areas of law. Courts may adopt the standard of care from federal regulations that protect: (1) a class of persons whose interest is invaded, (2) the particular interest that is invaded, (3) against the harm

---

214. See id.
215. Id.; see National Commission, supra note 26 and accompanying text.
216. See, e.g., Rossell v. Volkswagen of Am., 709 P.2d 517, 522–23 (Ariz. 1985); see also supra note 197.
217. See Morin, supra note 81, at 213; see also Delgado & Leskovac, supra note 62, at 69.
221. See supra notes 124–26 and accompanying text.
which has resulted, and (4) against the particular hazard from which the harm results.222 The Daum court held that the federal regulations provide the standard of care for informed consent in human subject research because they meet these four elements of negligence per se.223 In particular, the Daum court found that the federal regulations are “intended to protect the rights and safety of subjects”224 and that human subjects are “within the class of persons for whose protection the [federal regulations] were enacted.”225

Several courts have used the federal regulations to determine the standard of care for informed consent in human subject research. The Whitlock court adopted the Common Rule as the standard of care for informed consent, and held that there is a heightened standard of disclosure for informed consent in the research context.226 Similarly, the Daum court held that violations of the federal regulation’s requirements for informed consent result in negligence per se, and limited the role of expert testimony to whether the federal standard was met.227 Courts have found that the federal regulations provide detailed, clear guidelines for courts and juries to assess whether a subject’s informed consent conformed to the higher standard of care owed by researchers.228

In addition, other courts have relied on the FDA regulations to set the standard of care for obtaining a patient’s informed consent before implanting experimental medical devices.229 In Vodopest v. MacGregor,230 the Supreme Court of Washington held that researchers violated the Common Rule by requiring subjects to waive their legal rights to bring negligence actions in the consent form,231 because the

222. RESTATEMENT (SECOND) OF TORTS § 286 (1965).
224. Id. at 273.
225. Id.
227. Daum, 61 Cal. Rptr. 2d at 273; see supra notes 132–35 and accompanying text.
228. See, e.g., United States v. Najarian, 915 F. Supp. 1460, 1473 n.21 (D. Minn. 1996) (observing that “the requisites of ‘informed consent’ are detailed in the FDA’s Regulations” and holding, inter alia, that 21 C.F.R. §§ 50.20–50.27 is not void for vagueness).
231. Id. at 857–62, 913 P.2d at 787–89.
Common Rule prohibits such exculpatory agreements.\textsuperscript{232} Courts have thus recognized that both the Common Rule and FDA regulations provide a workable standard of care by which to measure a research subject’s informed consent.

Even though the federal regulations provide a detailed standard of care for informed consent, expert testimony will most likely be necessary to determine whether a particular researcher complied with the standard of care. The Common Rule sets forth specific elements of informed consent, including explaining the purposes and procedures of the research, describing the reasonably foreseeable risks and benefits, and disclosing the alternatives.\textsuperscript{233} Courts currently apply two different tests for establishing the appropriate scope of disclosure in informed consent actions: (1) the professional practice standard; and (2) what the “reasonable person” would want to know.\textsuperscript{234} Jurisdictions using the professional practice standard should rely on expert testimony to determine the scope of disclosure for each of the elements of informed consent, based on what a reasonable researcher would have disclosed in a similar situation.\textsuperscript{235} Jurisdictions using the “reasonable person” standard should allow the trier of fact to determine the scope of disclosure for each of the elements of informed consent that a “reasonable person” would want to know, but should use expert testimony to clarify the purposes, procedures, and foreseeable risks of the research.\textsuperscript{236} For instance, in \textit{Stewart v. Cleveland Clinic Foundation},\textsuperscript{237} the defendant clinic argued that the informed consent form for a clinical trial complied with the Common Rule,\textsuperscript{238} while the plaintiff provided contradictory expert testimony.\textsuperscript{239} The court held that a genuine issue of material fact existed as to whether the informed consent complied with the federal requirements, and allowed the jury to consider expert testimony in determining whether the researchers complied with the federal regulations.\textsuperscript{240} Similarly, the \textit{Daum} court limited the use of expert

\textsuperscript{232} See supra note 40 and accompanying text.
\textsuperscript{233} 45 C.F.R. § 46.116(a); see supra notes 38–40 and accompanying text.
\textsuperscript{234} See King, supra note 65, at 29.
\textsuperscript{235} See \textit{Furrow}, supra note 70, at 318.
\textsuperscript{236} \textit{Id}. at 319.
\textsuperscript{237} 736 N.E.2d 491 (Ohio Ct. App. 1999).
\textsuperscript{238} \textit{Id}. at 495.
\textsuperscript{239} \textit{Id}. at 497.
\textsuperscript{240} \textit{Id}. at 501.
testimony to determine whether the standard of care provided by the federal regulations was met.241

However, complying with the federal regulations’ informed consent requirements should only be a partial defense for researchers. Although several courts have granted summary judgment to defendants for complying with the federal requirements for disclosure of risks,242 there are several reasons why such compliance should not be a complete defense. First, expert testimony will usually be required to determine whether the informed consent was adequate under the federal regulations, thereby precluding summary judgment.243 Second, the federal regulations explicitly state that the informed consent requirements do not preempt applicable federal, state, or local laws that require additional information to be disclosed.244 Thus, courts must consider whether state laws set higher standards for informed consent and whether the researcher’s actions violated them. For example, the Supreme Court of California held that researchers must disclose conflicts of interest under state law, even though the Common Rule does not require it.245 Assuming that state or local laws do not require higher standards for informed consent, plaintiffs should be able to establish the Common Rule as the standard of care for informed consent in human subject research. Violating the federal requirements should result in a finding of negligence, and compliance should only be a partial defense.

2. Establishing the Standard of Care for Researcher Malpractice

The federal regulations require researchers to obtain IRB review246 and approval247 before obtaining informed consent and conducting research on human subjects. The federal regulations do not, however, provide a specific standard of care for the conduct of research. The criteria for IRB approval generally requires that the IRB determine, inter alia, that “risks to subjects are minimized”248 and “reasonable in relation

242. See supra notes 149–52 and accompanying text.
243. See, e.g., supra notes 171–75 and accompanying text.
244. 45 C.F.R. § 46.116(e) (2001).
247. Id. § 46.111.
248. Id. § 46.111(a)(1).
Researcher Negligence in Human Subject Research

to anticipated benefits, if any, to subjects and the importance of the knowledge that may reasonably be expected to result."249 The IRB also must make technical decisions about the scientific merit of research protocols before approving the research.250 Thus, the federal regulations place the primary responsibility of risk-benefit analysis and approval of research on IRBs, and do not provide detailed standards of care for the actual conduct of research in researcher malpractice. However, obtaining IRB approval could become part of the standard of care for researcher malpractice, in which case conducting research without IRB approval could result in researcher malpractice.

Setting the federal regulations aside, plaintiffs should be able to successfully establish the standard of care in researcher malpractice actions through expert testimony, similar to the medical malpractice model. The standard of care in medical malpractice is generally established by expert testimony, and is defined as a failure to exercise the required degree of care, skill, and diligence ordinarily possessed by a reasonable and prudent physician in the same medical specialty acting under the same or similar circumstances.251 Early cases treated research that deviated from medical standards of care as negligent, without developing a specialized standard of care for research.252 Courts should adapt the medical malpractice model to establish a specialized standard of care in researcher malpractice actions, using expert testimony to determine the degree of care, skill and diligence ordinarily possessed by a reasonable and prudent researcher in the same research specialty acting under the same or similar circumstances. The First Circuit adopted this approach in *Heinrich* using expert testimony to establish the standard of care for researcher malpractice, and holding that the plaintiffs failed to show that the applicable standard of care was breached.253

Although the federal regulations do not provide a specific standard of care for researcher malpractice, obtaining IRB approval should only be a partial defense. IRB approval demonstrates that an independent group of IRB members with professional competence determined, inter alia, that the research had scientific merit, the benefits outweighed the risks, and the informed consent form was in compliance with federal

249. Id. § 46.111(a)(2); see supra notes 43–51 and accompanying text.
250. Id. § 46.111(a)(1) (stating IRBs must determine that research protocols use "procedures which are consistent with sound research design").
251. See FURROW, supra note 70, at 269.
252. See supra notes 77–79 and accompanying text.
253. See supra notes 198–203 and accompanying text.
requirements.254 Although the Heinrich court observed that “the bar to be surmounted in litigation over current charges of [researcher] malpractice is a demanding one,”255 the circuit court found that approval by committees similar to IRBs was a partial defense.256 If the researcher complied with the federal regulations and obtained IRB approval, it should at least be evidence of meeting the standard of care for researcher malpractice.

There are several reasons why IRB approval should only create a rebuttable presumption of due care in research malpractice actions, rather than serving as a complete defense. First, complying with the federal requirement of obtaining IRB approval will not necessarily prevent a finding of researcher malpractice if a reasonable person would have taken additional precautions.257 Second, the federal regulations explicitly allow higher standards of care set by federal, state, or local laws.258 Consequently, courts must evaluate whether other laws set higher standards of care than those required to obtain IRB approval. Third, IRB approval itself may be flawed if the researcher provided inaccurate information about the risks and benefits of the research, or failed to provide information about adverse events for annual continuing review.259 The IRB relies on the researcher’s training and honesty when reviewing and approving research protocols. It is always possible that this reliance could be misplaced.

Further, the IRB itself may not have adequately performed the IRB review. For instance, the Grimes court criticized the IRB’s performance in approving the research, and allowed a negligence action against researchers to proceed because the IRB approval was flawed.260 Finally, even if the researcher obtained proper IRB approval before conducting the research, it is possible that the research was negligently performed,261 or that the researcher deviated from the IRB-approved protocol. Juries should therefore consider IRB approval as evidence that the standard of

254. See supra notes 43–51 and accompanying text.
255. Heinrich v. Sweet, 308 F.3d 48, 71 (1st Cir. 2002).
256. See supra note 202 and accompanying text.
257. RESTATEMENT (SECOND) OF TORTS § 288C (1965).
259. Id. § 46.109(c); see note 49 and accompanying text.
260. See supra notes 185–86 and accompanying text.
care for researcher malpractice was met, but should ultimately assess the value of that evidence in the context of expert testimony.

Thus, courts should adopt the federal regulations as the standard of care for informed consent in human subject research, using expert testimony to establish whether the standard of care for informed consent was breached. Researchers who violate the federal requirements for informed consent should be held liable, although researchers who comply with these requirements should be granted a partial defense. Because the federal regulations do not provide detailed standards of care for researchers conducting research, plaintiffs should be able to establish the standard of care in researcher malpractice through expert testimony. Researchers who proceed without IRB approval should be held liable, while researchers who obtain IRB approval should only be entitled to a partial defense.

CONCLUSION

Injured human subjects are becoming more likely to sue researchers. Although the case law on human subject research is just beginning to emerge, plaintiffs should be able to successfully establish negligence claims against researchers by establishing a duty of care based on the special relationship between researchers and subjects. Having established that researchers owe human subjects a duty of care, injured research subjects should be able to bring negligence actions against researchers based on informed consent and researcher malpractice. Research subjects should be able to establish the federal regulations as the minimum standard of care in informed consent, and use expert witnesses to determine whether the federal requirements for informed consent were violated. Violating the standard of care for informed consent in the federal regulations should result in a finding of negligence, and researchers who comply with the regulations should be allowed a partial defense. Although the federal regulations do not provide a specific standard of care for researcher malpractice, research subjects should be able to establish specialized standards of care for researcher malpractice through expert testimony, similar to the traditional medical malpractice model. Failure to obtain IRB approval should result in a finding of negligence, and IRB approval should only be a partial defense.
IS ASSENT STILL A PREREQUISITE FOR CONTRACT FORMATION IN TODAY’S E-CONOMY?

Melissa Robertson

Abstract: A browse-wrap agreement is an online contract that governs the use of a Web site but does not require users of the site to affirmatively agree to the terms and conditions of the contract. The terms of a browse-wrap agreement are accessible to the user only by clicking on an Internet link, often inconspicuously located at the bottom of a Web page, marked “Terms and Conditions.” Browse-wrap agreements purport to bind users to these terms and conditions when the user merely performs a function of the Web site, such as submitting a query on the site’s database or downloading software. Despite the prevalent use of browse-wrap agreements, courts are just beginning to consider their enforceability. To date, four federal district courts have addressed the issue. Each court has approached the issue of contract formation differently and has reached a different result. Courts should refuse to enforce browse-wrap agreements. Users do not always have adequate notice that using a Web site binds them to the terms and conditions of a browse-wrap. Even if the Web site does provide notice to users that such terms and conditions exist, users are not given the opportunity to adequately manifest their assent to such terms. Although courts must be flexible as contracts evolve to accommodate electronic commerce, browse-wrap agreements stray too far from the basic contractual principles of notice and assent. Accordingly, courts should not enforce them.

Imagine you are browsing the Internet from your home computer.¹ You are searching for the latest version of a popular software program that allows you to download and play music from the Internet. You find the software and proceed to download it. Soon after, you learn that the software has infected your computer with a virus that has virtually destroyed your computer’s hard drive. You sue the software company for damages. The company moves to stay the proceedings and compel arbitration in Florida pursuant to an arbitration clause contained in the terms and conditions of its Web site. You do not remember clicking on an icon that says “I Agree” or otherwise forming a contract on the Web site. In fact, you never even saw a list of terms and conditions on the Web site. You learn that the terms and conditions of the site are hidden behind a link at the bottom of the Web site, far below the icon which allowed you to download the software. Next to this otherwise unmarked link, there is a small statement that says, “By using this Web site you agree to be bound by our Terms and Conditions.” But you never assented to arbitration in Florida. Surely a court of law would never bind you to the terms and conditions of a Web site when you did not even know the terms existed—would it? It just might.

¹. Hypothetical created by the author.
Courts should enforce online contracts only where the consumer has adequate notice of the terms and conditions, the user affirmatively assents to be bound by such terms, and the terms being invoked are conscionable. Since the advent of the Internet, parties have been contracting online. Because the validity of online contracts is unsettled, many online businesses require users to affirmatively agree to express contracts called “click-wrap” agreements to govern the use of their Web sites. However, many other businesses rely on online contracts called “browse-wrap” agreements that purport to bind users even though they do not require users of the site to perform an affirmative act, or even know about the contract. Web sites with browse-wrap agreements usually display a notice on the site that states that using the Web site binds users to the terms and conditions of the site. This announcement is usually followed by a link to the site’s full text of terms and conditions, which may or may not be on the same Web page as the notice. Despite the prevalent use of browse-wrap agreements, courts have just recently begun to consider the enforceability of such contracts.

The law regarding the enforceability of browse-wrap agreements is unsettled. To date, four federal district courts have addressed the enforceability of browse-wrap agreements. Prior to 2000, no court had


3. See Mark H. Wildasin, Shrink Wrap, Click Wrap, and Now Browse Wrap: Did You Just Make a Contract?, METROPOLITAN CORPORATE COUNSEL, Oct. 2001, at 13. Click-wrap agreements are online contracts that contain a Web site’s terms and conditions. A user must click an icon that states “I Agree” or a similar phrase of agreement to indicate assent to the terms and conditions of the site.

4. Browse-wrap agreements are also called “web-wrap” or “browse-through” agreements.

5. See id, supra note 3, at 13.

6. See id.

7. Id.

8. See id.

9. See id.

addressed the issue. Initially, courts did not definitively hold whether terms within browse-wrap agreements were enforceable. One court noted that users of a Web site with a browse-wrap agreement may or may not understand that they are bound by the terms and conditions of the site merely by visiting it. However, that court refused to declare the browse-wrap agreement unenforceable. The apparent enforceability of analogous contracts prevalent in the software industry, such as shrink-wrap and click-wrap licenses, suggested that courts would enforce terms within browse-wrap agreements as long as they were not unconscionable. In Register.com, Inc. v. Verio, Inc., a federal district court confirmed this assumption by enforcing a term within a browse-wrap agreement. However, a recent federal district court opinion, Specht v. Netscape Communications Corp., places the enforceability of browse-wrap agreements into question. In Specht, the court held browse-wrap agreements to be unenforceable because they do not require users to affirmatively assent to the terms and conditions; and without assent, no valid contract exists.

This Comment argues that courts should not enforce browse-wrap agreements. Rather, courts should enforce online contracts only where users have adequate notice of the terms and conditions and affirmatively agree to be bound by such terms. Accordingly, assuming that the terms

---

14. Id. at 982.
15. A shrink-wrap agreement is an unsigned license agreement commonly used in the software industry. Shrink-wrap agreements generally come inside a software product’s packaging or are displayed on a user’s computer screen when the user installs the software on his or her computer. See infra Part II.A.
16. See infra Part II.B.
17. In the context of software contracts, the terms “license,” “contract,” and “agreement” are often used interchangeably. See, e.g., Jane K. Winn & Benjamin Wright, The Law of Electronic Commerce, § 6.02[A], at 6-3–6-7 (4th ed. 2001).
18. See Winn, supra note 17, § 6.01, at 6-2. (“It seems likely that courts will uphold clickwrap and webwrap contracts based on the Internet display of a standardized form contract by a vendor and some act on the part of the accepting party indicating acceptance of the offered terms. It will be more difficult to predict whether each provision of such a standard form contract will be enforced, however.”).
20. See id. at 248.
invoked are conscionable, courts should enforce click-wrap agreements because they provide users with adequate notice and require them to affirmatively assent to the terms. Part I of this Comment establishes contract law principles that are relevant to determining whether browse-wrap agreements are enforceable. Part II examines the enforceability of analogous contracts such as shrink-wrap and click-wrap agreements, as well as online contracts under the Uniform Computer Information Transactions Act. Part III of this Comment analyzes the four federal district court cases that have addressed the enforceability of browse-wrap agreements. Part IV argues that courts should not enforce browse-wrap agreements because they do not provide users with adequate notice of the terms and conditions of the agreement or require users to adequately manifest assent. Further, Part IV argues that courts should enforce click-wrap agreements because they do provide users with adequate notice and require users to manifest assent. This Comment concludes that click-wrap agreements strike an essential balance between facilitating online business and ensuring that users are not bound by contracts of which they had no knowledge.

I. CONTRACT LAW PRINCIPLES RELEVANT TO THE ENFORCEABILITY OF ONLINE CONTRACTS

A contract, in its simplest terms, is the binding promise of one party to another. Contract law generally requires three elements to create a binding contract: offer, acceptance, and consideration. The Restatement (Second) of Contracts defines a contract as “a promise or a set of promises for the breach of which the law gives a remedy, or the performance of which the law in some way recognizes as a duty.”

common law, a contract was formed when one party accepted another party’s offer.\textsuperscript{28} The \textit{Restatement} provides that “the manifestation of mutual assent to an exchange ordinarily takes the form of an offer or proposal by one party followed by an acceptance by the other party or parties.”\textsuperscript{29} The offeree’s acceptance, which may take the form of a return promise or act, creates a binding contract.\textsuperscript{30}

Contract law has evolved over the years to accommodate modern business practices.\textsuperscript{31} For example, Article 2 of the Uniform Commercial Code (UCC) loosens the requirements of contract formation such that a contract may be formed “in any manner sufficient to show agreement.”\textsuperscript{32} However, the terms of a contract will not be enforced if a court finds them to be unconscionable, regardless of the manner in which the contract was formed.\textsuperscript{33} For example, courts will generally enforce standardized form contracts because parties have a duty to read the contract before agreeing to its terms,\textsuperscript{34} but they will invalidate the terms of the contract if they are unconscionable.\textsuperscript{35}

Because the UCC substantially loosens the requirements necessary to form a valid contract, parties may manifest their assent electronically.\textsuperscript{36} UCC § 2-204 provides that a contract may be made “in any manner sufficient to show agreement, including conduct by both parties which recognizes the existence of such a contract.”\textsuperscript{37} Further, UCC § 2-206 provides, “an offer to make a contract shall be construed as inviting acceptance in any manner and by any medium reasonable in the circumstances.”\textsuperscript{38} Courts have subsequently interpreted UCC §§ 2-204 and 2-206 to allow parties to create enforceable contracts over the Internet.\textsuperscript{39} However, there are many issues regarding the enforceability of online contracts. For example, what conduct is required by the parties to manifest their assent to online contracts? How do procedural and substantive unconscionability affect the enforceability of terms within

\begin{itemize}
\item \textsuperscript{28} See FARNSWORTH, supra note 25, § 3.3, at 113.
\item \textsuperscript{29} See \textit{FARNSWORTH, supra} note 25, § 3.3, at 113.
\item \textsuperscript{30} See \textit{FARNSWORTH, supra} note 25, § 3.3, at 113.
\item \textsuperscript{31} See \textit{Wildasin, supra} note 3, at 13.
\item \textsuperscript{32} Id.
\item \textsuperscript{33} U.C.C. § 2-302 (2000).
\item \textsuperscript{34} See FARNSWORTH, supra note 25, § 4.26, at 297.
\item \textsuperscript{35} See \textit{Williams v. Walker-Thomas Furniture Co.}, 350 F.2d 445, 448–49 (D.C. Cir. 1965).
\item \textsuperscript{36} See WINN, supra note 17, § 6.02, at 6-3.
\item \textsuperscript{37} U.C.C. § 2-204(1) (2000).
\item \textsuperscript{38} Id. § 2-206(1)(a).
\item \textsuperscript{39} See GEORGE B. DELTA & JEFFREY H. MATSUURA, LAW OF THE INTERNET § 10.07 (2d ed. 2002).
\end{itemize}
online contracts? And how does the enforceability of standardized form contracts and the duty to read apply to online contracts? This Comment will now discuss how the issues of assent, unconscionability, the enforceability of standardized form contracts and the duty to read relate to the enforceability of online contracts.

A. Assent

To create an enforceable contract, a party must assent to its terms. However, it is not entirely clear what conduct is sufficient to adequately manifest a user’s assent to the terms and conditions posted on a Web site. Several courts have expressed concern over a user’s ability to adequately manifest assent to online contracts. According to E. Allan Farnsworth, “[s]ince it is difficult for a workable system of contract law to take account of assent unless there has been an overt expression of it, courts have required that assent to the formation of a contract be manifested in some way, by words or other conduct, if it is to be effective.” Courts, however, are just now determining what conduct is required from a user to create an enforceable online contract. Must a user click on an icon that says “I Agree?” Or is it enough that a user downloads software or submits a query on a Web site that contains a notice that says “By using this site you agree to be bound by our terms and conditions?” Must the user actually read the terms and conditions of the site? And does the user even have to see the link to the terms and conditions? These questions remain largely unanswered.

Whether a court finds that a user manifested assent to an online contract may depend largely on whether the court applies a subjective or objective theory of assent. Under the subjective theory, a court examines the actual intentions of the parties and requires, as often stated, a “meeting of the minds.” If a party did not specifically intend to assent to a particular term, then that term is not enforceable. However, under the objective theory of assent, a court looks only to “the external or objective appearance of the parties’ intentions as manifested by their intentions.”

40. Farnsworth, supra note 25, § 3.1, at 110.
42. Farnsworth, supra note 25, § 3.1, at 110.
43. See Hillman, supra note 22, at 488–89.
44. See Farnsworth, supra note 25, § 3.6, at 116–17.
45. See id.
Therefore, even if a party did not intend to assent to a particular term, a court will enforce it if a reasonable person would conclude that the party intended to assent by their words or conduct. Although the subjective theory of assent is reflected in the Restatement of Contracts, the objective theory is most commonly employed by the courts. Generally speaking, there is no contract without assent, but once the objective manifestations of assent are present, the author is bound. There are many unanswered questions regarding assent and the enforceability of online contracts.

B. Unconscionability

Because online contracts are unilaterally imposed on the user, these contracts may include terms that are materially unfavorable or unfair to the user. Depending on the circumstances, a court may find such terms to be unconscionable. The equitable doctrine of unconscionability was codified in UCC § 2-302. It provides that if a court finds a contract clause to be unconscionable, the court can refuse to enforce the contract entirely, refuse to enforce only the unconscionable clause, or limit the application of the clause so as to avoid an unconscionable result.

Courts generally recognize two kinds of unconscionability: "procedural" and "substantive." Procedural unconscionability relates to how a term becomes part of a contract. For example, procedural unconscionability may involve inconspicuous or unintelligible print, a lack of opportunity to review the terms, or an inability to ask questions regarding the terms and meanings of specific clauses. Procedural unconscionability may also involve a unilaterally imposed standardized form contract by a party with far greater bargaining power. However,
while disparity in bargaining power or a standardized form contract alone will not render terms within the contract unconscionable, these factors may be significant when considered together with the substantive clauses of the contract.\textsuperscript{58}

Substantive unconscionability refers to contracts or terms that are oppressive or overly harsh.\textsuperscript{59} In determining whether a contract is substantively unconscionable, courts have looked at several factors.\textsuperscript{60} These include provisions that deprive one party of the benefits he or she is entitled to receive under the contract and provisions that bear no reasonable relation to the business risk involved.\textsuperscript{61} Further, courts have considered whether there is an unreasonable disparity in the cost and the selling price and whether the contract creates an unreasonable advantage for one party without producing an adequate benefit to the disadvantaged party.\textsuperscript{62} Although the presence of one of these factors does not necessarily mean that a court will invalidate the contract as unconscionable, courts will generally invalidate the contract if both procedural and substantive unconscionability exist.\textsuperscript{63}

\section*{C. Standardized Form Contracts and the Duty to Read}

Although contracts scholars have argued that standardized form contracts are procedurally unconscionable, courts routinely enforce such contracts because parties have a common law duty to read.\textsuperscript{64} Standardized form contracts are regularly upheld by courts even if one party did not read or understand the terms of the contract.\textsuperscript{65} Courts will generally enforce such contracts even if the terms of the contract are preprinted by one party and presented to the other party on a “take it or leave it” basis.\textsuperscript{66} Courts have viewed online contracts as a novel type of standardized form contract.\textsuperscript{67}

\begin{itemize}
\item \textsuperscript{58} See id. at 312.
\item \textsuperscript{59} See \textsc{John D. Calamari & Joseph M. Perillo, The Law of Contracts} § 9.40, at 373 (4th ed. 1998); \textit{see also Dawson}, supra note 55, at 694.
\item \textsuperscript{60} See \textit{Farnsworth}, supra note 25, § 4.28, at 311.
\item \textsuperscript{61} See id.
\item \textsuperscript{62} See \textsc{Calamari & Perillo, supra note 59, § 9.40, at 373.}
\item \textsuperscript{63} Id.
\item \textsuperscript{64} See id. § 9.42, at 377.
\item \textsuperscript{65} See id.
\item \textsuperscript{66} See id.; \textit{see also Carnival Cruise Lines, Inc. v. Shute, 499 U.S. 585, 594–95 (1991)} (holding a forum selection clause printed on the back of a cruise ship ticket reasonable and enforceable because such terms reduce litigation costs, insurance costs, and passenger fairs, despite the fact enforcement of the clause effectively denied plaintiff’s day in court). For a discussion of the enforceability of forum selection clauses within online contracts, see Kaustuv M. Das, Comment, \textit{Forum-Selection

272
The enforceability of form contracts is typically justified by the common law rule that a party has a duty to read a contract. Failing to read the terms of a contract does not excuse a party from being legally obligated to fulfill the contract. The exception to this rule arises when the disputed terms are not sufficiently called to the attention of the adhering party. Determining whether the terms were sufficiently called to the attention of the adhering party depends on whether a reasonable person under the circumstances would understand that the disputed terms were part of the binding contract.

Although courts will generally enforce terms within standardized form contracts, they have made exceptions to the rule in limited circumstances where the terms of the contract were unfair under the circumstances. Courts can invalidate such contracts in whole or in part. Three grounds exist for courts to invalidate terms within a standardized contract: (1) the adhering party did not truly assent to a particular term; (2) the term contravenes public policy and is therefore void; and (3) the term is unconscionable.

In Williams v. Walker-Thomas Furniture Co., the federal Court of Appeals for the D.C. Circuit considered the enforceability of a standardized form contract and concluded that, because the adhering party likely did not know of the grossly unfair terms within the preprinted form contract, the trial court erred in failing to determine whether the doctrine of unconscionability was applicable. The court explained that when a party of little bargaining power signs an unconscionable contract without having any knowledge of its terms, the party has not truly assented to be bound by such terms. Under such circumstances, the court stated, the traditional duty to read rule should be

---

Clauses in Consumer Clickwrap and Browsewrap Agreements and the “Reasonably Communicated” Test, 77 WASH. L. REV. 481 (2002).

67. WINN, supra note 17, § 6.02, at 6–3.
68. See CALAMARI & PERILLO, supra note 59, § 9.41, at 377.
69. See id.
70. See id. § 9.42, at 378.
71. See id.
72. See id. § 9.41, at 377.
73. See FARNSWORTH, supra note 25, § 4.26, at 309.
74. See CALAMARI & PERILLO, supra note 59, § 9.43, at 382–83.
75. 350 F.2d 445, 449 (D.C. Cir. 1965) (remanding to trial court to determine whether the doctrine of unconscionability applied where a clause in an installment sales agreement resulted in repossession of all items purchased from the store when the purchaser fell behind on payment of one item).
76. See id.
77. See id.
abandoned to avoid an unconscionable result. Additionally, a court may find a term buried in a form contract to be unenforceable if a reasonable person who has carefully read the contract could not be expected to understand the term. The odds of unenforceability are greater if the problematic term is in an inconspicuous place, such as the back of a paper form or in fine print.

If a court finds that a party drafting a contract has included terms that are egregiously one-sided, the court may construe the term in the manner most favorable to the adhering party and most unfavorable to the drafting party, without actually invalidating the term. Courts may also follow the Restatement (Second) of Contracts § 211(3), which suggests invalidating a specific term of the contract when “the other party has reason to believe that the party manifesting such assent would not do so if he knew that the writing contained a particular term . . . .”

Although standardized form contracts raise a number of issues regarding assent and fairness, their efficiency and convenience have made them a significant and inescapable aspect of modern life. Many scholars have argued that a party never truly assents to all the terms within standardized form contracts. Most notably, Karl Llewellyn, a principal drafter of the UCC, argued that a party may give a “blanket assent” to reasonable terms within a standardized form contract, but that the party cannot truly assent to anything but the few dickered terms of the contract. Llewellyn concluded that as long as the terms contained in a standardized form contract are not unfair in presentation or substance, courts should enforce them if the parties gave a blanket assent to be bound by such terms.

Another contracts scholar, Todd D. Rakoff, has argued that because parties do not truly assent to standardized form terms, proponents of such forms can not argue that standardized form contracts must be enforced to
Notice, Assent & Browse-Wrap Agreements

uphold the principle of “freedom of contract.”87 He contended that “[o]nce it is recognized that contracts of adhesion arise from the matrix of organizational hierarchy, the argument for enforcement of form terms as a recognition of ‘freedom of contract’ in its usual sense is unsupportable.”88 Regardless, courts have the tools to keep the drafters of standardized form contracts in check.

II. THE ENFORCEABILITY OF SHRINK-WRAP AND CLICK-WRAP AGREEMENTS

When presented with the question of whether terms within browse-wrap agreements are enforceable, courts have found an absence of authority directly on point. Searching for a useful analogy, courts have looked to cases addressing the enforceability of shrink-wrap and click-wrap agreements for guidance.89 Courts generally enforce terms in both shrink-wrap90 and click-wrap agreements.91 The Uniform Computer Information Transaction Act (UCITA), a model law intending to provide guidance on the enforceability of electronic contracts, provides that shrink-wrap and click-wrap agreements are enforceable as long as the user has a reasonable opportunity to review the terms and manifests his or her assent in a manner sufficient to show agreement.92

A. Shrink-Wrap Agreements

A shrink-wrap agreement is an unsigned license agreement commonly used in the software industry.93 Shrink-wrap agreements generally come in two forms. The first includes the terms and conditions of a product inside the sealed plastic of the software’s packaging; hence the name, “shrink-wrap agreement.”94 The second displays the terms and conditions on a user’s computer screen when the user installs the software program.95 Both types contain a notice to the user that using the product

87. Rakoff, supra note 84, at 1237.
88. Id.
90. See Wildasin, supra note 3, at 13.
91. See WINN, supra note 17, § 6.02[A], at 6-7–6-8.
92. See DELTA, supra note 39, § 10.07, at 10-64.1–10-64.2.
93. See Wildasin, supra note 3, at 13.
94. See id.
95. See id.
or keeping the product beyond a certain time period binds the user to the
terms and conditions of the product.96

In the early 1990s, courts considering the enforceability of shrink-
wrap agreements declined to enforce them on the grounds that they
constituted “additional terms” not part of the original contract.97 Under
UCC § 2-207, such additional terms required express consent beyond
merely opening the package and retaining the product for longer than a
specified time.98 Accordingly, a user merely opening a package or
viewing terms when starting up a software program did not constitute
express assent to the terms of the shrink-wrap.99 Without assent, no valid
contract existed.

However, a new trend emerged in enforcing shrink-wrap agreements
in 1996 with the seminal case ProCD, Inc. v. Zeidenberg.100 In ProCD,
the Seventh Circuit held that shrink-wrap agreements are enforceable
unless their terms are objectionable on grounds applicable to contracts in
general, such as if they violate a rule of positive law or are
unconscionable.101 The shrink-wrap agreement at issue in ProCD came
inside the product’s packaging with a notice on the outside of the
package requiring the user to review all terms and conditions within the
agreement.102 The notice further stated that if the buyer did not wish to be
bound by the terms, the buyer may return the product for a full refund.
If the buyer did not return the product, the buyer presumably agreed to be
bound by the terms of the shrink-wrap agreement.103 In holding such
terms to be enforceable, the ProCD court applied UCC § 2-204, which
allows a contract for the sale of goods to be made in any manner
sufficient to show agreement, including conduct by both parties.104 The
Seventh Circuit concluded that opening a package and retaining the
product constituted acts sufficient to demonstrate assent.105 Since the
ProCD decision, many courts, both federal and state, have relied on the

96. See id.
97. See WINN, supra note 17, § 6.02[A], at 6-5.
further discussion of UCC § 2-207 as applied to shrink-wrap licenses, see William H. Danne, Jr.,
Annotation, What are Additional Terms Materially Altering Contract within the Meaning of UCC
100. 86 F.3d 1447 (7th Cir. 1996).
101. Id. at 1449.
102. Id. at 1450.
103. See id. at 1452–53.
104. See id. at 1452.
105. See id. at 1452–53.
Seventh Circuit’s reasoning in ProCD to enforce similar shrink-wrap agreements.106

B. Click-Wrap Agreements

Courts have relied on the same line of reasoning in ProCD to enforce click-wrap agreements.107 Click-wrap agreements, which acquired their name from their similarity to shrink-wrap agreements, are electronic contracts that contain a Web site’s terms and conditions.108 To form a click-wrap agreement, a user “clicks” on an icon that states “I Agree” or a similar phrase of assent.109 The user may then proceed with the desired action on the Web site, such as downloading software or searching the site’s database.110

Although this area of law is not yet entirely settled, the few courts that have addressed the validity of click-wrap agreements have found them to be enforceable when the user had an opportunity to review the terms and the terms were not unconscionable.111 According to the UCC, a user may agree to be bound by a contract in any manner sufficient to show agreement, and courts have generally found that clicking on an icon that says “I Agree” or a similar phrase of assent is sufficient to demonstrate agreement to be bound by the terms.112

106. See, e.g., Hill v. Gateway 2000, Inc., 105 F.3d 1147, 1150 (7th Cir. 1997) (holding that, because the plaintiffs retained a computer beyond the thirty day period specified on the computer’s box, they assented to the terms contained inside the box, including the arbitration clause at issue); Mortenson Co. v. Timberline Software Corp., 140 Wash. 2d 568, 583–84, 998 P.2d 305, 313 (2000) (holding that a limitation of liability in a license agreement contained in software packaging was conscionable and enforceable even though it denied plaintiff from recovering consequential damages); Brower v. Gateway 2000, Inc., 676 N.Y.S.2d 569, 571 (N.Y. App. Div. 1998) (holding that plaintiffs manifested their acceptance to the terms contained inside a computer’s box when they retained the computer beyond the specified thirty day period). But see Klocek v. Gateway, Inc., 104 F. Supp. 1332, 1341 (D. Kan. 2000) (denying defendant Gateway’s motion to dismiss because the court found that retaining a computer beyond the specified five day period was not sufficient to demonstrate assent to defendant’s arbitration clause, which was in defendant’s shrink-wrap agreement).
108. See id.
109. See id.
110. See id.
111. See DELTA, supra note 39, § 10.05, at 10-57–10-58.
112. See, e.g., CompuServe, Inc. v. Patterson, 89 F.3d 1257, 1260 (6th Cir. 1996) (finding the click-wrap agreement between plaintiff, an Ohio-based Internet service provider, and defendant, a customer from Texas, to be a valid contract; therefore, defendant had sufficient minimum contacts with Ohio for an Ohio district court to assert personal jurisdiction); In re RealNetworks, Inc., Privacy Litig., No. 00 C 1366, 2000 U.S. Dist. LEXIS 6584, at *21 (N.D. Ill. May 8, 2000) (rejecting an intervenor’s arguments that an arbitration clause within a click-wrap agreement is
C. The Enforceability of Online Contracts Under UCITA

Recognizing the lack of uniformity in the case law regarding transactions in the computer industry, including the enforceability of software licenses and online contracts, the National Conference of Commissioners on Uniform State Laws (NCCUSL) formed a committee in March of 1994 to draft a new UCC Article 2B to govern these transactions. The committee met numerous times between 1994 and 1999, attempting to meet the divergent requests of both consumer groups and copyright industries. After failing to finalize a new Article 2B, the NCCUSL created a model law, the UCITA in 1999. According to one commentator, “[t]he essence of UCITA . . . is that it is a commercial contract law that provides a framework for forming software [and online] contracts, much as the UCC provided the framework for contracts to purchase and sell goods.” Although UCITA has only been adopted in a few jurisdictions, it is important to consider because courts may still look to it for guidance.

UCITA contains provisions regarding the enforceability of electronic contracts. It champions a “freedom of contract” approach to electronic contracting and generally favors the drafters of standardized form contracts. The Act reflects the same line of reasoning employed by the Seventh Circuit in ProCD. UCITA maintains that shrink-wrap and click-wrap agreements are generally enforceable, as long as the agreements meet basic requirements regarding the user’s opportunity to review the terms and conditions and manifest their assent to be bound. Although UCITA does not specifically consider browse-wrap agreements, it lays


113. See Harrison, supra note 26, at 929.
114. See id.; see also WINN, supra note 17, § 6.01, at 6-2.
115. See WINN, supra note 17, § 6.01, at 6-2.
116. See DELTA, supra note 39, § 10.07, at 10-64.1.
118. See WINN, supra note 17, § 6.01, at 6-2.
119. See id.
120. See id.
121. DELTA, supra note 39, § 10.07, at 10-64.1–10-64.2.
out basic requirements for the enforceability of online contracts. UCITA explicitly contrasts its example of an enforceable online agreement with an online contract "that places the terms and conditions of the agreement behind jump-links labeled 'terms and conditions' or 'legal' and is tucked unobtrusively at the bottom of the [Web] page, where they are unlikely to be noticed by any but the most cautious or dilatory user."122

UCITA requires that electronic contracts provide users with an opportunity to review the terms and conditions of the contract.123 According to a comment to UCITA, "[a]n opportunity to review requires that the record be made available in a manner that ought to call it to the attention of a reasonable person and in a form that readily permits review."124 Winn and Wright contrast a format that meets this requirement, e.g., clearly marked terms and conditions prominently displayed on a Web site, with a Web interface that hides terms and conditions behind hyperlinks at the bottom of a Web page that the user must scroll down to see.125 Many terms and conditions, such as warranty disclaimers, must be displayed conspicuously.126 However, the user does not actually have to review the terms and conditions to create an enforceable contract.127 This rule is consistent with the duty to read standardized form contracts.

UCITA also contains provisions intended to clarify what constitutes manifestation of assent to online contracts.128 With respect to the enforceability of click-wrap agreements, UCITA established a bright-line rule:129 if a user must click on an icon that states "I Agree" two times to

122. WINN, supra note 17, § 6.02[A], at 6-8, n.28.
123. DELTA, supra note 39, § 10.07, at 10-64.2.
124. U.C.I.T.A. § 112, cited in WINN, supra note 17, § 6.02[A], at 6-8–6-9, n. 28.
125. WINN, supra note 17, § 6.02[A], at 6-8–6-9, n. 28.
126. U.C.I.T.A. § 102(a)(14), reprinted in WINN, supra note 17, § 6.04[A], at 6-27–6-28, n.127:

Conspicuous terms include the following:
(A) with respect to a person:
(i) a heading in capitals in a size equal to or greater than, or in contrasting type, font, or color to, the surrounding text;
(ii) language in the body of a record or display in larger or other contrasting type, font or color or set off from the surrounding text by symbols or other marks that draw attention to the language; and
(iii) a term prominently referenced in an electronic record or display which is readily accessible or reviewable from the record or display . . . .

Id.
127. WINN, supra note 17, § 6.02[A], at 6-3.
128. See id. § 5.08[C], at 5-55.
129. A critic of this rule has argued that creating a bright-line rule for click-wraps is premature at best and misleading at worst. See id.
get through a contract formation interface, then the user has manifested assent to the terms and conditions of the click-wrap. For other software contracts, UCITA established the rule that a user has manifested assent if that person either: (1) “authenticates a record”; or (2) intentionally engages in conduct with reason to know that the other party will infer assent from that conduct, provided that person has knowledge of the contents of the record, or was given a reasonable opportunity to review it. In sum, courts have enforced shrink-wrap and click-wrap agreements and UCITA provides that these contracts should be enforced as long as users have a reasonable opportunity to review the terms and conditions of the contracts and manifest their assent in a manner sufficient to show agreement.

III. FEDERAL COURTS HAVE REACHED CONFLICTING CONCLUSIONS REGARDING THE ENFORCEABILITY OF BROWSE-WRAP AGREEMENTS

Although four federal district courts addressing the enforceability of browse-wrap agreements have turned to the Seventh Circuit’s decision in ProCD for guidance, they have not all reached the same conclusion. Two courts reasoned that browse-wrap agreements should be enforced as long as the user has demonstrated assent. For example, a user could indicate sufficient intent to be bound by downloading software from the Web site, searching the Web site’s database, or by accessing a site that

130. See id.; see also Groff v. America Online, No. PC 97-0331, 1998 R.I. Super. LEXIS 46, at *13 (R.I. Super. Ct. May 27, 1998) (applying this “two clicks” rule prior to the enactment of UCITA to conclude that the plaintiff user had formed a valid contract and was therefore bound by defendant’s forum-selection clause).


132. UCITA § 112(a), reprinted in WINN, supra note 17, § 5.08[C], at 5-55 (emphasis added).

133. See Wildasin, supra note 3, at 13.

contained language such as “By using this site you agree to be bound by the terms and conditions contained herein.” In contrast, the other two courts concluded that the user must explicitly assent to the terms—not merely use the site—to indicate assent. To better understand the courts’ reasoning, each decision will be discussed separately and in chronological order.

A. Ticketmaster Corp. v. Tickets.com, Inc.: Refusing to Enforce a Browse-Wrap Agreement

In *Ticketmaster Corp. v. Tickets.com, Inc.*, an unpublished opinion, a federal district court in the Central District of California refused to enforce the plaintiff’s browse-wrap agreement because the defendant had not assented to be bound by its terms and conditions. The plaintiff, Ticketmaster Corp., maintained a Web site where users could receive information about upcoming events and purchase tickets online. Defendant Tickets.com used automated software to extract information from the plaintiff’s Web site about upcoming concerts and ticket vendor locations, change the data’s font and format, and post the data on its own Web site. The primary source of information for the defendant’s online ticket clearing house was Ticketmaster’s Web site. Ticketmaster filed suit alleging that the defendant’s actions were a breach of contract because, among other things, Ticketmaster's browse-wrap agreement prohibited using any of the information obtained from its site for commercial use.

Ticketmaster’s browse-wrap agreement consisted of a general statement that use of the site bound users to its terms and conditions. Next to this statement was a link to the full text of the Web site’s terms and conditions. Ticketmaster argued that the defendant should be

138. *Id.* at *18.
139. *Id.* at *4–5.
140. *Id.* at *8–9.
141. *See id.*
142. *See id.* at *11.
143. *Id.* at *6.
144. *Id.*
bound by the terms contained therein. The court disagreed, stating that Ticketmaster “lacks sufficient proof of agreement by defendant to be taken seriously . . . .” However, the court did not provide further explanation or analysis. The court noted its unfamiliarity with concepts such as automated software. The court also noted that a user on plaintiff’s Web site was not required to click an “I agree” icon before using the site, but it did not elaborate on the significance of such an act.

B. Pollstar v. Gigmania, Ltd.: A Browse-Wrap Agreement May Be Enforceable

Although the facts of Pollstar v. Gigmania, Ltd. were similar to those in Ticketmaster, the federal district court in the Eastern District of California considering the case was hesitant to reach the same result. Plaintiff Pollstar alleged that defendant Gigmania breached its browse-wrap agreement when Gigmania accessed Pollstar’s Web site, copied information, and posted it on its own Web site. Pollstar's Web site contained a notice stating that if users accessed any of the information on the site, they agreed to be bound by the terms and conditions posted on the site. This notice, which was in small gray text on a light gray background, provided a link to the full text of the Web site's terms and conditions. However, the link was not underlined, which is a common method of indicating an Internet link within a Web site’s text. Like Ticketmaster’s browse-wrap agreement, Pollstar’s browse-wrap

145. See id. at *7–8.
146. Id. at *18.
147. Id. at *8 (“[Tickets.com] does not obtain the information in the same way as does the public (that is, by opening up an interior Web page and reading the information off the screen), but rather by a sophisticated computer method of monitoring the thousands of interior [Ticketmaster] Web pages electronically by the use of a mysterious (to the court) device[] know[n] as [] 'webcrawlers' or 'spiders.'”).
148. Id. at *6.
149. The court provided a more detailed analysis of Ticketmaster’s copyright claim, which the court concluded was not valid because factual information may not be copyrighted. See id. at *9–10.
151. Id. at 982.
152. Id. at 976.
153. Id. at 977.
154. Id. at 980–81.
155. Id. at 981.
prohibited use of information obtained from the site for commercial purposes.\textsuperscript{156}

Gigmania moved for dismissal, arguing that the browse-wrap agreement was unenforceable because users of Pollstar’s Web site did not have adequate notice of its terms and conditions.\textsuperscript{157} The court denied Gigmania’s motion and stated that it found the reasoning of \textit{ProCD} and \textit{Hill v. Gateway 2000, Inc.}\textsuperscript{158} persuasive.\textsuperscript{159} In both \textit{ProCD} and \textit{Hill}, courts enforced shrink-wrap agreements based on the proposition in UCC § 2-204 that parties may form an enforceable contract in any manner sufficient to show agreement.\textsuperscript{160} Expressly reserving judgment on the enforceability of browse-wrap agreements, the \textit{Gigmania} court stated that it hesitated to declare browse-wrap agreements unenforceable because decisions within the Seventh Circuit, such as \textit{ProCD} and \textit{Hill}, demonstrated that users can be bound by terms they did not see.\textsuperscript{161} Although the court denied Gigmania’s motion to dismiss and recognized the potential validity of browse-wrap agreements, it expressed concern that users of Pollstar’s Web site may not be aware of the terms because notice of the license agreement was in small gray text on a light gray background and the link to the terms was not underlined.\textsuperscript{162}

\textbf{C. Register.com, Inc. v. Verio, Inc.: Enforcing a Term Within a Browse-Wrap Agreement}

In \textit{Register.com, Inc. v. Verio, Inc.},\textsuperscript{163} a district court in the Southern District of New York was first to declare that browse-wrap agreements are enforceable as long as users receive clear notice of the terms.\textsuperscript{164} Plaintiff Register.com, a registrar of Internet domain names, sought an injunction against defendant Verio, a competitor and Internet service provider,\textsuperscript{165} for using automated software to access information on its Web site about its customers and compiling the information for mass

\begin{footnotesize}
\begin{itemize}
\item[\textsuperscript{156}] Id. at 976 n.1.
\item[\textsuperscript{157}] Id. at 980–81.
\item[\textsuperscript{158}] 105 F.3d 1147 (7th Cir. 1997).
\item[\textsuperscript{159}] \textit{Pollstar}, 170 F. Supp. 2d at 981.
\item[\textsuperscript{160}] See supra notes 100–106 and accompanying text.
\item[\textsuperscript{161}] \textit{Pollstar}, 170 F. Supp. 2d at 982.
\item[\textsuperscript{162}] Id. at 981.
\item[\textsuperscript{163}] 126 F. Supp. 2d 238 (S.D.N.Y. 2000).
\item[\textsuperscript{164}] See id. at 248.
\item[\textsuperscript{165}] Id. at 241.
\end{itemize}
\end{footnotesize}
marketing purposes, such as sending “spam.” Register.com alleged that this process constituted a breach of contract because its Web site contained a browse-wrap agreement that prohibited commercial use of any information obtained from its site. Like the Web sites at issue in Ticketmaster and Pollstar, Register.com’s Web site stated that using the site demonstrated assent to be bound by the terms and conditions of the agreement.

Register.com argued that the defendant had formed a contract when it submitted a query on Register.com’s database. The defendant dismissed this argument by maintaining that merely submitting a query did not adequately indicate assent to the terms; an independent act demonstrating agreement to be bound by the terms of the browse-wrap was required. The court agreed with Register.com that the “terms of use are clearly posted on [the] Web site. The conclusion of the terms paragraph states, ‘By submitting this query, you agree to abide by these terms.’” Therefore, the court concluded, “there can be no question that by proceeding to submit a . . . query, Verio manifested its assent to be bound by Register.com’s terms of use, and a contract was formed and subsequently breached.”

D. Specht v. Netscape Communications Corp.: Refusing to Enforce an Arbitration Clause in a Browse-Wrap Agreement

In Specht v. Netscape Communications Corp., a different federal district court in the Southern District of New York refused to enforce an arbitration clause within a browse-wrap agreement. The plaintiffs, several individual consumers, filed a claim against defendants, Netscape and its parent company, America Online. The consumers alleged that the defendants’ SmartDownload software transmitted private information about their Internet activity to the defendants in violation of the Electronic Communications and Privacy Act and the Computer Fraud

166. See id. at 244. “Spam” refers to unsolicited emails, the online equivalent to junk mail. For a discussion of spam, see WINN, supra note 17, § 2.06, at 2-37–2-40.
168. Id. at 248.
169. Id. at 245–46.
170. Id. at 248.
171. Id.
172. Id.
174. See id. at 596.
175. Id. at 587 n.2, 587–88.
Notice, Assent & Browse-Wrap Agreements

and Abuse Act.\textsuperscript{176} Pursuant to the arbitration clause in the browse-wrap agreement, the defendants sought to stay the proceedings and compel arbitration.\textsuperscript{177} The court held that the plaintiffs were not bound by the terms of the browse-wrap agreement because they never affirmatively assented to its terms.\textsuperscript{178}

The software at issue in \textit{Specht} was available on the defendants’ Web site, where users like the plaintiffs could download it for free by clicking on an icon that said, “Download.”\textsuperscript{179} At the bottom of the Web page was the statement “Please review and agree to the terms of the Netscape SmartDownload software license agreement before downloading and using the software.”\textsuperscript{180} This statement could only be seen by scrolling down the Web page past the “Download” icon.\textsuperscript{181} The statement contained an underlined hyperlink to the defendants’ full text of terms and conditions.\textsuperscript{182} The full text contained a notice in all capital letters stating that, if users did not agree to be bound by the terms, they must not download the software.\textsuperscript{183}

Distinguishing the enforceability of click-wrap agreements from the defendants’ browse-wrap agreement, the court stated that the defendants’ terms would have been enforceable merely by requiring users to click an “I Assent” icon or something similar before they may download the software.\textsuperscript{184} Unlike browse-wraps, the court noted, click-wraps \textit{require} users to click their assent to the terms before they may proceed with an activity on a Web site.\textsuperscript{185} The court rejected the defendants’ argument that clicking “Download” indicated assent in the same way that clicking “I Assent” does.\textsuperscript{186} The court explained that downloading is “hardly an unambiguous indication of assent.”\textsuperscript{187} Rather, the purpose of downloading is to obtain a product.\textsuperscript{188} However, “clicking on an icon stating ‘I assent’ has no meaning or purpose other than to indicate such
assent.”189 The court concluded that “Netscape’s failure to require users of SmartDownload to indicate assent to its license as a precondition to downloading and using its software is fatal to its argument that a contract has been formed.”190

The Specht court was also troubled that users were not required to link to the page displaying the full text of terms and conditions before proceeding with the download.191 Citing ProCD,192 the defendants argued that users who downloaded their software were bound by their terms regardless of whether they actually read them; just as those who buy products containing shrink-wrap agreements are bound by the terms of the shrink-wrap agreements regardless of whether they actually read them.193 The court rejected this argument, distinguishing defendants’ terms from terms within shrink-wraps, because “shrink-wrap agreement[s] . . . require users to perform an affirmative action unambiguously expressing assent before they may use the software.”194 Accordingly, the court held that users are not bound by electronic contracts unless they affirmatively indicate their assent.195

IV. COURTS SHOULD ONLY ENFORCE ONLINE CONTRACTS THAT PROVIDE USERS WITH ADEQUATE NOTICE AND REQUIRE AFFIRMATIVE ASSENT

Considering the countless number of Web sites that are purportedly governed by browse-wrap agreements,196 courts must come to a consensus regarding their enforceability. Courts should not enforce browse-wrap agreements because they do not satisfy the requirements necessary to create a valid contract, such as notice and assent. Online contracts should only be enforced where users have adequate notice of the terms and conditions of the contract and affirmatively agree to be bound by such terms. Accordingly, courts should enforce click-wrap agreements, assuming that the terms invoked are conscionable, because they provide users with adequate notice and require them to affirmatively assent to their terms.

189. Id.
190. Id.
191. Id. at 596.
192. 86 F.3d 1447 (7th Cir. 1996).
194. Id. at 595.
195. Id.
196. See Wildasin, supra note 3, at 13.
A. A Uniform Consensus Regarding the Enforceability of Browse-Wrap Agreements is Needed

The four courts to date that have addressed the enforceability of browse-wrap agreements have not provided consistent guidance on the issue, much less established a cohesive body of law. Both Ticketmaster and Pollstar fail to provide useful precedent on the enforceability of browse-wrap agreements. Although the Specht court and the Register.com court applied similar reasoning to determine whether the parties were able to adequately manifest their assent, they reached conflicting conclusions. Because browse-wrap agreements are used to govern the use of countless Web sites, businesses as well as consumers should know whether terms in such agreements are enforceable.

1. Ticketmaster Corp. v. Tickets.com and Pollstar v. Gigmania Fail to Provide Useful Precedent

Neither Ticketmaster nor Pollstar provide helpful precedent because the Ticketmaster court does not supply explicit analysis and the Pollstar court never reaches a definite conclusion regarding the enforceability of browse-wrap agreements. The Ticketmaster court addressed whether the browse-wrap agreement was a valid contract in one sentence: “The contract theory lacks sufficient proof of agreement by defendant to be taken seriously as a ground for preliminary injunction.” The court failed to explain its reasoning or provide any further guidance on the issue. The court’s admitted unfamiliarity with concepts such as automated software most likely contributed to its apparent issue-dodging. Additionally, the court may have been satisfied with its conclusion based on copyright law that factual information may not be copyrighted. Therefore, even if the defendant was bound by Ticketmaster’s agreement, Ticketmaster still lacked a claim sufficient for a preliminary injunction because defendant obtained and posted factual information about upcoming concerts. Regardless of the court’s reasons for skirting the browse-wrap enforceability issue, the single

198. However, the court did state that it did not intend to make any significant legal pronouncements. Id. at *4. Rather, the court stated, directives regarding areas of unsettled law such as the enforceability of browse-wrap agreements should come from the courts of appeals. Id.
199. Id. at *8; see supra note 147 and accompanying text.
201. Id. at *9.
sentence that the court devotes to the issue does not establish useful authority.

Although the *Pollstar* court analyzed the enforceability of browse-wrap agreements more thoroughly than the *Ticketmaster* court, it reached an equally unsatisfying conclusion.\(^{(202)}\) The *Pollstar* court merely stated that a browse-wrap agreement “may” be enforceable.\(^{(203)}\) The *Pollstar* court’s conclusion is most likely explained by the procedural posture of the case. Defendant Gigmania had moved for dismissal and the court was reluctant to dispose of Pollstar’s claims on an issue about which it was uncertain.\(^{(204)}\) Although the court expressly stated its concern that users may not have adequate notice of a browse-wrap agreement, the court did not make any pronouncements regarding their enforceability generally. The court explained its denial of defendant’s motion for dismissal by suggesting that plaintiff’s browse-wrap *may* be enforceable without actually deciding whether browse-wrap agreements are enforceable.\(^{(205)}\) Accordingly, *Pollstar*, like *Ticketmaster*, fails to provide useful authority.


Although both *Specht* and *Register.com* were decided by federal district courts in the Southern District of New York, the courts reached opposite conclusions. The courts’ significantly different decisions are most likely explained by the cases’ different facts. In *Register.com*, the term at issue prohibited users like the defendant from using information obtained from the Web site for the purposes of sending spam.\(^{(206)}\) This type of term is not inherently troubling. In fact, such a term is most likely appreciated by Register.com’s customers. In contrast, the term at issue in *Specht* was a mandatory arbitration clause.\(^{(207)}\) Courts are generally more cautious about enforcing clauses that waive adhering parties’ legal right to sue.\(^{(208)}\) Although such waivers are valid if executed appropriately, courts are reluctant to enforce them if it is questionable whether the


\(^{(203)}\) *Id.*

\(^{(204)}\) *Id.*

\(^{(205)}\) *Id.*


\(^{(208)}\) See DAWSON, supra note 55, at 186.

288
adhering party agreed to be bound by such a clause. The term at issue in Register.com protected the public while the term at issue in Specht deprived the public of a legal right. This difference may explain the different outcomes by two federal district courts within the same district.

Further, the parties in Specht and Register.com were not equally sympathetic. In Specht, the parties arguing against enforcement were individual consumers filing suit against a major corporation that was trying to deny them their day in court. In Register.com, the defendant was an online business that was allegedly surreptitiously obtaining e-mail addresses of unwitting online consumers from the database of an unwilling competitor for the purpose of clogging the consumers’ inboxes with spam. Thus, although neither court noted these factual differences as the reason for its decision, these differences likely played an implicit role in the courts’ analysis.

Both the Specht court and the Register.com court based their holdings on the adequacy of the parties’ manifestation of assent. The Specht court justified its contrary conclusion by distinguishing the facts of Register.com. But this attempted distinction is tenuous. The Register.com court found that clicking on the “Submit Query” button was an adequate indication of assent, whereas the Specht court found that clicking on the “Download” button was not an adequate indication of assent. It should not matter that the buttons used different language, nor should it matter that the Specht court applied California law while the Register.com court applied New York law. Both opinions rest on whether a contract was formed between the parties and both states require mutual assent to form an enforceable contract. Arguably, Register.com’s Web site contained clearer notice to its users that submitting a query bound them to the Web site’s terms and conditions. Yet it is doubtful that the Specht court would have enforced Netscape’s arbitration clause if its Web site had contained the same language as Register.com’s Web site.

209. See id.
211. Register.com, 126 F. Supp. 2d at 241.
214. See Casamiquela, supra note 212, at 484–85.
215. See id.
216. Register.com’s website stated that using the site bound the user to the terms and conditions of the site. See Register.com, 126 F. Supp. 2d at 242–43. By contrast, Netscape’s notice asked the user to “please review” the terms and conditions. See Specht, 150 F. Supp. 2d at 588.
because the *Specht* court demanded explicit affirmative assent.\textsuperscript{217} Merely submitting a query as the defendant did in *Register.com* does not meet this demand.\textsuperscript{218} In sum, these four federal district court opinions—one of which is unpublished—do not provide useful guidance regarding the enforceability of browse-wrap agreements because they have distinguishable facts that arguably explain their conflicting conclusions.

### B. Courts Should Not Enforce Browse-Wrap Agreements

Courts should refuse to enforce browse-wrap agreements because they do not satisfy the requirements necessary to create a valid contract.\textsuperscript{219} Browse-wrap agreements do not necessarily give users adequate notice that merely using a Web site binds them to the terms and conditions of that site.\textsuperscript{220} Further, even if the Web site provides notice that such terms and conditions exist, users are not given the opportunity to adequately manifest their assent.\textsuperscript{221} The *ProCD* court and other courts’ conclusion that shrink-wrap and click-wrap agreements are generally enforceable is not applicable to browse-wrap agreements because browse-wrap agreements do not require users to affirmatively assent. Further, the manifestation of assent purportedly acceptable to create a browse-wrap agreement does not satisfy the standards for adequate manifestation of assent promulgated by UCITA. Finally, declaring browse-wrap agreements unenforceable will benefit online businesses by encouraging them to use other, more enforceable online contracts.

#### 1. Browse-Wrap Agreements Fail to Provide Users with Sufficient Notice to Create an Enforceable Contract

Users often do not receive adequate notice that they are entering into a binding contract when they perform an act purportedly sufficient to bind them to a browse-wrap agreement.\textsuperscript{222} The average consumer may not even realize that such terms governing Web sites exist.\textsuperscript{223} The notices intended to alert consumers that using a Web site binds them to the terms and conditions of the site are often located at the bottom of a Web page

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{217} *Specht*, 150 F. Supp. 2d at 595.
\item \textsuperscript{218} *See id.*
\item \textsuperscript{219} *See infra*, Part II.
\item \textsuperscript{220} *See Specht*, 150 F. Supp. 2d at 595.
\item \textsuperscript{221} *See id.*
\item \textsuperscript{222} *See Specht*, 150 F. Supp. 2d at 594 (citing Pollstar v. Gigmania, 170 F. Supp. 2d 974, 980–82 (E.D. Cal. 2000)).
\item \textsuperscript{223} *See Specht*, 150 F. Supp. 2d at 594.
\end{itemize}
\end{footnotesize}
where a user has to scroll down to see them. The notice may also be written in a color similar to the background color, like the Web site in *Pollstar.* In *Specht,* Netscape’s notice was problematic because its statement asking the user to “please review” the terms of the software license before downloading the software was too mild an invitation to notify the user that a binding contract was being formed. A user may perform an act that the notice claims will bind the user to the site without actually ever seeing the notice, much less the text of the agreement’s actual terms and conditions. There is nothing about submitting a query on a database or downloading free software that inherently alerts a user that a binding contract is being formed. Absent clear notice, users have no way of knowing that they are entering into a contract.

Browse-wrap agreements do not provide users with adequate notice of their terms and conditions, and therefore should not be enforced. The common law requires that the terms of a contract be clear, unambiguous, and plainly visible. Binding a party to a contract of which the party had no knowledge contravenes traditional principles of contract law. Courts apply this principle to cases involving standardized form contracts. As in traditional contract law, there is an exception to the duty to read rule for terms that are not adequately called to the attention of the user. Such terms are not enforceable. Because it is unlikely that the terms and conditions of a browse-wrap agreement are adequately called to the attention of the user, or “plainly visible” as required to create an enforceable contract, the browse-wrap agreement should not be enforced.

Further, browse-wrap agreements fail to meet even the standards promulgated by UCITA regarding adequate notice of online terms and conditions. UCITA requires that users be given an opportunity to review the terms and conditions of an online contract before being bound. UCITA explicitly contrasts its example of an enforceable online agreement with an online contract “that places the terms and conditions

225. See *Specht,* 150 F. Supp. 2d at 595–96.
226. See *id.*
227. See *id.*
228. See FARNSWORTH, supra note 25, § 7.8, at 454–55; see also CALAMARI & PERILLO, supra note 59, § 9.42, at 378–79.
229. See *infra* Part II.
231. See *id.*
232. See *id.*
233. U.C.I.T.A. § 112, reprinted in *WINN,* supra note 17, § 6.02[A], at 6–8, n. 27.
of the agreement behind jumplinks labeled ‘terms and conditions’ or ‘legal’ and tucked unobtrusively at the bottom of the [web] page, where they are unlikely to be noticed by any but the most cautious or dilatory user.”

While UCITA does not mention browse-wrap agreements by name, the example UCITA provides of an online contract that does not give users adequate notice is indeed a browse-wrap agreement.

2. **Browse-Wrap Agreements Do Not Require Users to Adequately Manifest Their Assent**

Even if a Web site provides adequate notice of its terms and conditions, users can not adequately manifest their assent to a browse-wrap agreement. It is possible to create a Web site that provides ample notice to users that use of the site binds them to the terms and conditions governing the site. It is also possible to create clear, explicit interfaces that allow users to review the terms and conditions of an agreement. But adequate notice alone does not create an enforceable contract. Users must also manifest their assent to be bound by a browse-wrap agreement. Browse-wrap agreements are not enforceable because users are not required to affirmatively indicate their assent.

An adhering party must manifest assent to the terms and conditions of an agreement to create an enforceable contract. Although there are exceptions to this rule, they are generally not applicable to the creation of an enforceable standardized form contract. Failing to manifest assent to the terms and conditions of a standardized form contract could result in the invalidation of its terms. The *Specht* court, applying this principle to a browse-wrap agreement, stated:

> The case law on software licensing has not eroded the importance of assent in contract formation. Mutual assent is the bedrock of any agreement to which the law will give force. Defendants’ position [that downloading software when a notice states that such an act constitutes assent to the terms of the browse-wrap], if accepted,

---

234. **WINN, supra note 17, § 6.02[A], at 6-8, n.28.**
235. The term “browse-wrap” was not adopted until after UCITA was finalized by the NCCUSL in 1999. *See, e.g.*, Specht v. Netscape Communications Corp. 150 F. Supp. 2d 585, 594 (S.D.N.Y. 2001) (E.D. Cal. 2000) (noting that the *Pollstar* court was the first court to use the term “browse-wrap” when it used the term on Oct. 27, 2000).
236. For example, a large window with the entire text of the terms and conditions inside may provide users with a reasonable opportunity to review the terms and conditions of the browse-wrap.
237. **See FARNSWORTH, supra note 25, § 3.1, at 110.**
238. **See CALAMARI & PERILLO, supra note 59, § 9.43, at 382-83.**
239. **See id.**
would so expand the definition of assent as to render it meaningless.\textsuperscript{240}

According to the \textit{Specht} court, assent to a browse-wrap agreement must be affirmative.\textsuperscript{241}

Although a user may subjectively intend to manifest assent to the terms and conditions of a browse-wrap agreement by downloading software or submitting a query, it does not follow that such an action objectively manifests assent. Users can not manifest assent to a browse-wrap agreement because there is no way for them to affirmatively assent. Click-wrap agreements, by contrast, require users to affirmatively manifest assent. Courts should enforce conscionable terms and conditions contained in click-wrap agreements. Such contracts meet the contractual principles of notice and assent. Notice is inherently provided by the window that displays the “I Accept” and “I Do Not Accept” icons to the user. Click-wrap agreements require users to affirmatively indicate their assent by clicking on an icon. While such an act does not guarantee that users have read and understood all of the terms and conditions contained in the contract, the act is analogous to a user signing a paper standardized form contract.\textsuperscript{242} Courts should enforce terms within click-wrap agreements because, like signing a paper contract, the user is notified that they are entering into a binding agreement and must indicate their assent before a contract is formed.

Users do not have an opportunity to manifest assent to browse-wrap agreements in the same way that they may manifest assent to shrink-wrap and click-wrap agreements. Although the \textit{ProCD} court’s reasoning is arguably applicable to click-wrap agreements, it is not equally applicable to browse-wrap agreements. Users’ ability to assent to terms of shrink-wrap agreements is central to \textit{ProCD}’s holding that such contracts are enforceable.\textsuperscript{243} Similarly, users’ ability to assent to click-wrap agreements is central to courts holding that such contracts are enforceable.\textsuperscript{244} Users are not given the same opportunity to assent to browse-wrap agreements.\textsuperscript{245} The court in \textit{Specht v. Netscape} explained that “downloading is hardly an unambiguous indication of assent. The

\begin{itemize}
\item \textsuperscript{240} \textit{Specht v. Netscape Communication Corp.}, 150 F. Supp. 2d 585, 596 (S.D.N.Y. 2000).
\item \textsuperscript{241} \textit{See id}.
\item \textsuperscript{242} \textit{See id}.
\item \textsuperscript{243} The issue of whether such an act must be authenticated to ensure that an unauthorized person is not forming a contract on a user’s computer is beyond the scope of this Comment.
\item \textsuperscript{244} \textit{ProCD v. Zeidenberg}, 86 F.3d 1447, 1452 (7th Cir. 1996).
\item \textsuperscript{245} \textit{See Specht}, 150 F. Supp. 2d at 595.
\end{itemize}
primary purpose of downloading is to obtain a product, not to assent to an agreement. 246 Although a user may form a contract in any manner sufficient to show agreement, performing a Web site function does not objectively meet this requirement. Therefore, courts should not apply the reasoning of the ProCD court to hold that browse-wrap agreements enforceable because such contracts do not allow users to adequately manifest their assent.

In addition to failing to meet traditional contract law requirements for manifesting assent, browse-wrap agreements also fail to meet UCITA’s requirements for manifesting assent. For software contracts other than click-wraps, UCITA provides that a user has manifested assent “if that person either [1] authenticates a record or [2] intentionally engages in conduct with reason to know that the other party will infer assent from that conduct, provided that person has knowledge of the contents of the record, or was given a reasonable opportunity to review it.” 247 Accordingly, a user may meet the first method of the UCITA test by electronically signing an online contract or engaging in some kind of intentional activity to manifest assent. A browse-wrap agreement does not require or even allow users to sign it. If it did, it would no longer be a browse-wrap agreement because it would involve affirmative assent.

In sum, under UCITA, a user must engage in intentional activity to manifest assent to a browse-wrap agreement. However, even if a court concludes that a user intentionally entered into a browse-wrap agreement by downloading software, for example, and therefore satisfied the first prong of the UCITA test, the user is still not bound unless he or she also satisfies the second prong of the UCITA test by either knowing the terms and conditions of the agreement or by having a reasonable opportunity to review them. Further, browse-wrap agreements do not provide users with a reasonable opportunity to review terms and conditions because a user does not necessarily even know that the terms are there. Therefore, the browse-wrap agreement should not be enforced under contract law principles and UCITA’s requirements for manifestation of assent.

246. Id.
247. U.C.I.T.A. § 112(a), (emphasis added), reprinted in WINN, supra note 17, § 5.08[C], at 5-55.
3. **Declaring Browse-Wrap Agreements Unenforceable Will Encourage Online Businesses to Use Other, More Enforceable Online Contracts and Protect Consumers**

If courts refuse to enforce browse-wrap agreements, the result will benefit online businesses as well as consumers. The current state of the law gives online businesses with browse-wrap agreements a fifty-fifty chance that their terms will be enforced. These odds are not high enough to justify such a large number of online businesses relying on browse-wrap agreements to uphold their terms and conditions. Online businesses would benefit from a definitive standard regarding the enforceability of such agreements. 

Online businesses that currently employ browse-wrap agreements are at risk that courts will refuse to enforce their terms and conditions.\(^{248}\) Even if courts generally enforced terms and conditions within browse-wrap agreements, online businesses could still be at risk that certain terms would be invalidated on the grounds of unconscionability. Courts generally do not invalidate terms on such grounds unless they find both procedural and substantive unconscionability.\(^{249}\) Because users may not have had adequate notice of the terms of a browse-wrap, they could make a strong argument that the browse-wrap was procedurally unconscionable. Therefore, users would have a strong argument that a term denying their right to sue that they were unaware of is substantively unconscionable.\(^{250}\) Although businesses are always subject to a court invalidating a term on the grounds of unconscionability, the odds of that happening are much smaller if the parties had clear notice of the terms and agreed to them.\(^{251}\)

If courts refuse to enforce browse-wrap agreements, it will encourage online businesses to replace browse-wrap agreements with online contracts that require users to adequately manifest their assent. By using these contracts, online businesses will benefit because it will be more likely that the terms and conditions of their agreements will be enforced.

It can be argued that courts’ refusal to enforce browse-wrap agreements could hurt online businesses. Studies show that the fewer steps Web sites require users to complete in a transaction, the more transactions that will be completed.\(^{252}\) While this may be true, it is also

---

248. See Specht, 150 F. Supp. 2d at 596.
249. See Farnsworth, supra note 25, § 4.28, at 312.
250. See, e.g., Specht, 150 F. Supp. 2d at 596.
252. See Ruthenberg, supra note 2, at 14 n.6.
true that online contracts formed without the user’s clear assent could be invalidated, thereby eliminating the protections afforded to online businesses by such terms.253

Further, if courts refuse to enforce browse-wrap agreements, consumers will be protected from being bound by online terms and conditions of which they had no notice. Although it can be argued that adhering parties never truly assent to the terms and conditions of standardized form contracts, most standardized form contracts require users to at least take some affirmative action that recognizes the formation of a contract, such as signing a document.254 While parties may not explicitly assent to each term in a standardized form contract, they are at least aware that they are entering into a binding agreement. Courts should apply this same principle to online contracts.

VII. CONCLUSION

Enforcing the terms and conditions of a browse-wrap agreement against a consumer offends traditional principles of contract law, as well as modern principles that require parties to assent to standardized form contracts. Under contract law and UCITA, courts must ensure that users assent to online contracts. While, as Karl Llewellyn suggests, there may not be “true assent” to standardized form contracts, consumers must still assent to be bound; even if it is a “blanket assent” to the terms of a standardized form contract. Browse-wrap agreements stray too far from the basic contractual principles of notice and assent. Accordingly, they must not be enforced.

253. See FARNSWORTH, supra note 25, § 4.28, at 312.
254. See CALAMARI, supra note 59, § 9.43, at 383.
BEggars can’t be voters: Why Washington’s felon re-enfranchisement law violates the equal protection clause

jill E. simmons

Abstract: The Washington State constitution denies persons convicted of felonies the right to vote until their civil rights have been restored. Civil rights are restored when offenders complete all aspects of their sentence, including paying the legal-financial obligations imposed at sentencing. Payment of legal-financial obligations presents a significant hurdle to offenders trying to reclaim their right to vote. According to the Washington Department of Corrections, roughly 46,500 offenders in Washington have not had their right to vote restored solely because of unpaid legal-financial obligations. The right to vote is a fundamental right secured by the United States Constitution, yet the United States Supreme Court has affirmed that states have the right, under the Fourteenth Amendment, to disenfranchise persons convicted of crimes. While the constitutional requirements of felon disenfranchisement are settled, the requirements of felon re-enfranchisement are an open question. This Comment argues that felon re-enfranchisement laws must not discriminate in ways that violate the traditional voting rights requirements of the Equal Protection Clause. As one requirement, the U.S. Supreme Court has held that states cannot require the payment of money as a qualification for voting. Therefore, Washington’s requirement that offenders pay their legal-financial obligations before re-enfranchisement violates the Equal Protection Clause because it conditions the fundamental right to vote on the payment of money to the state.

John was released last year from a Washington State prison after serving time for a felony theft conviction. Although he has a job earning $7.50 an hour, John has not been able to pay the $1,200 in monetary sanctions—officially called “legal-financial obligations”—that were assessed during sentencing. Suppose that under Washington State law, John does not have the right to free speech because he has not finished

1. Hypothetical created by the author for illustrative purposes.
   [A] sum of money that is ordered by a superior court of the state of Washington . . . which may include restitution to the victim, statutorily imposed crime victims’ compensation fees as assessed pursuant to RCW 7.68.035, court costs, county or local drug funds, court-appointed attorneys' fees, and costs of defense, fines, and any other financial obligation that is assessed to the offender as a result of a felony conviction. Upon conviction for vehicular assault while under the influence of intoxicating liquor or any drug, RCW 46.61.522(1)(b), or vehicular homicide while under the influence of intoxicating liquor or any drug, RCW 46.61.520(1)(a), legal-financial obligations may also include payment to a public agency of the expense of an emergency response to the incident resulting in the conviction, subject to RCW 38.52.430.

Although the statute does not hyphenate the phrase “legal-financial obligations,” the two words are hyphenated throughout this Comment for clarity and to emphasize the connection between the legal and financial aspects of the obligation.
paying his legal-financial obligations. Under this law, Washington
requires offenders to complete all aspects of their sentence, including
payment of legal-financial obligations, before restoring their free speech
rights. Angered by a recent ballot initiative, John wants to express his
views by passing out flyers on a street corner, but he is prevented from
doing so until he pays his legal-financial obligations. Across town, Sarah
is on the sidewalk handing out flyers on the same ballot initiative. She
also was released from jail last year, but her family could afford to pay
her $1,500 of legal-financial obligations. As a consequence, her free
speech rights were fully restored upon payment.

The scenario described above seems preposterous. Surely no court
would uphold a state law that restricted offenders’ free speech rights
following release from prison until they paid their legal-financial
obligations. Yet, while no state restricts First Amendment rights in this
way, the scenario reflects the law in many states regarding another
fundamental right: voting. Some states, including Washington, deny
offenders the right to vote until they have paid their legal-financial
obligations. As a result, while John can pass out flyers expressing his
feelings on the upcoming initiative, he cannot vote on it. Sarah, on the
other hand, can vote. The only difference between Sarah and John is their
ability to pay.

In the United States, an estimated 3.9 million U.S. citizens are
disenfranchised because of criminal convictions, including over one
million who have fully completed their sentences. Forty-eight states

3. The author has invented this state law for illustrative purposes. No law in Washington State
requires that offenders pay their legal financial obligations before regaining free speech rights.

4. No state has enacted a law that restricts offenders’ free speech rights following release from
prison. However, in Turner v. Safely, the U.S. Supreme Court stated that restrictions on prisoners’
fundamental rights are permissible only if necessary to advance legitimate penological interests. 482
U.S. 78, 89 (1987). It is difficult to imagine a scenario where restrictions on released offenders’ free
speech rights would be necessary to serve legitimate penological interests.

5. Washington is among more than a dozen states that deny felons voting rights until they have
satisfied all conditions of their sentence, including payment of legal-financial obligations. See
UNITED STATES DEPARTMENT OF JUSTICE, CIVIL RIGHTS DIVISION, RESTORING YOUR RIGHT TO
(summarizing what offenders must do to restore their voting rights in each state.); Jamie Fellner &
Marc Mauer, Losing the Vote: The Impact of Felony Disenfranchisement Laws in the United States 4
(Human Rights Watch and The Sentencing Project, October 1998). For example, Alaska (ALASKA
STAT. § 15.05.030 (Michie 2002)), Connecticut (CONN. GEN. STAT. §§ 9-46, 9-46a (2001)),
Minnesota (MINN. STAT. §§ 201.014, 609.165 (2002)), and Texas (TEX. ELEC. CODE ANN. § 11.002
(Vernon 2002)) deny felons the right to vote until they have completed all aspects of their sentence.

restrict offenders’ voting rights to some extent. State restrictions range from disenfranchisement only during incarceration to permanent denial of the right to vote. In Washington, persons convicted of felonies are disenfranchised until they have received a “discharge,” which they will receive after they have served any prison sentence, completed community placement, and paid all legal-financial obligations imposed during sentencing. Felons may bypass this process only if the governor grants them a pardon. As of 1998, more than 150,000 Washington residents did not have the right to vote because of felony convictions. In addition, the Washington Department of Corrections (DOC) estimates that as of December 2001, 46,500 offenders remained disenfranchised solely because of pending legal-financial obligations.

The United States Constitution does not expressly guarantee the right to vote, but the U.S. Supreme Court has nonetheless held that once states grant citizens the right to vote, voting becomes a fundamental right protected by the Constitution. According to the Court, the right to vote is a “fundamental matter in a free and democratic society,” and as such any infringement on that right must be “meticulously scrutinized.” In the last fifty years, the Court has struck down a number of state laws that restricted citizens’ voting rights, including laws requiring payment of money to the state. However, the Court has upheld state laws that

9. Under Washington’s Sentencing Reform Act, offenders are sentenced to periods of community placement rather than parole and probation. WASH. REV. CODE § 9.94A.700 (2002). While under community placement, the offender lives in the community but remains under the supervision of the Department of Corrections. Id.
10. WASH. CONST. art. VI, § 3 (denying the right to vote for all persons convicted of “infamous crimes” unless they have had their civil rights restored); WASH. REV. CODE § 9.96.050 (2002) (providing for the restoration of civil rights upon discharge).
13. DEPARTMENT OF CORRECTIONS, AGENCY FISCAL NOTE FOR SENATE BILL 6519 (2002).
16. Id.
disenfranchise persons convicted of crimes. In *Richardson v. Ramirez*,\(^\text{18}\) the Court held that, in spite of the heightened protection of voting rights, Section 2 of the Fourteenth Amendment allows states to distinguish between persons convicted of crimes and all other citizens when granting the right to vote.\(^\text{19}\) Since *Ramirez*, the Court has continued to uphold a state’s ability to deny felons the right to vote upon conviction, but has struck down state disenfranchisement laws that are otherwise discriminatory.\(^\text{20}\)

This Comment does not challenge Washington’s right under Section 2 of the Fourteenth Amendment to disenfranchise persons convicted of crimes. Rather, it argues that Section 2, as interpreted in *Ramirez*, allows states to distinguish between offenders and other citizens for voting rights purposes but not among offenders in ways that violate traditional voting rights principles. This Comment contends that when reinstating offenders’ voting rights states must comply with the voting rights requirements of the Equal Protection Clause and, therefore, may not impose qualifications that discriminate against certain offenders. Specifically, Washington may not require felons to pay money to the state as a prerequisite for re-granting voting rights.\(^\text{21}\) Currently, Washington conditions the grant of offenders’ voting rights on the payment of legal-financial obligations. Thus, Washington gives voting rights to some offenders but not others—even offenders convicted of the same crime—based on their economic circumstances.\(^\text{22}\) By requiring payment of money to the state without advancing a compelling state interest, Washington’s re-enfranchisement system violates the Equal Protection Clause.

The U.S. Supreme Court has not determined the constitutional requirements of felon re-enfranchisement. Consequently, this Comment draws on analogous precedent in cases protecting voting rights, cases identifying the scope of offenders’ rights, and cases addressing other fundamental rights to demonstrate why Washington’s re-enfranchisement requirement violates the Equal Protection Clause. Part I of this Comment describes Washington’s felon disenfranchisement and re-enfranchisement laws and the effects of conditioning offenders’ voting

---

19. Id. at 54.
rights on the payment of legal-financial obligations. Part II provides an overview of the U.S. Supreme Court’s voting rights doctrine. Part III acknowledges that states may disenfranchise persons convicted of crimes, but notes that the Court has not addressed re-enfranchisement of offenders. Part IV specifically highlights the Court’s treatment of laws that condition voting on the payment of money and examines the extent to which other fundamental rights may be restricted based on a failure to pay legal-financial obligations. Finally, Part V argues that Washington’s re-enfranchisement of offenders, only after paying legal-financial obligations, violates the Equal Protection Clause because it conditions the fundamental right to vote on the payment of money to the state without advancing a compelling state interest.

I. WASHINGTON CONDITIONS RESTORATION OF FELONS’ VOTING RIGHTS ON THE PAYMENT OF LEGAL-FINANCIAL OBLIGATIONS

The Washington State Constitution denies to all persons convicted of felonies the right to vote unless they have had their civil rights restored. Convicted felons’ civil rights are restored if they complete all the requirements of their sentence, including jail time, community placement, and payment of all legal-financial obligations. The civil rights restoration process reinstates all rights “lost by operation of law upon conviction.” The rights “lost by operation of law” include only political rights and the right to bear arms. Moreover, the right to vote is the only fundamental right lost upon conviction. Many offenders have a difficult time satisfying their legal-financial obligations, resulting sometimes in permanent denial of the right to vote.

---

23. WASH. CONST. art. VI, § 3 (“All persons convicted of infamous crime unless restored to their civil rights . . . are excluded from the elective franchise.”).
25. Id. § 9.94A.637(3).
26. The right to vote, WASH. CONST. art. VI, § 3; WASH. REV. CODE § 29.10.097 (2002), and the right to serve on a jury, id. § 2.36.070, are the only political rights expressly lost upon conviction. However, a number of other political rights are contingent on being an eligible voter, including the right to run for office and to sponsor a voter’s initiative. Id. §§ 29.15.010; 29.79.010.
27. Id. § 9.41.040(1)(b)(i).
28. See infra footnotes 39–47 and accompanying text.
29. See infra footnotes 68–72 and accompanying text.
A. Washington Denies Offenders the Right to Vote Unless Their Civil Rights Have Been Restored, Which Only Occurs Upon Payment of All Legal-Financial Obligations

The Washington State Constitution denies persons convicted of “infamous crimes” the right to vote unless their civil rights have been restored. Washington law defines “infamous crime” as “a crime punishable by death in the state penitentiary or imprisonment in a state correctional facility.” All felony crimes in Washington are captured under this definition. Convicted felons’ civil rights are restored if they complete all the requirements of their sentence, including paying legal-financial obligations, or if they receive a pardon from the governor. When offenders complete all aspects of their sentence, the sentencing court issues a certificate of discharge. A discharge automatically restores an offender’s civil rights, and the certificate of discharge must state that the offender’s civil rights have been restored.

Washington’s civil rights restoration process reinstates all rights, including voting, “lost by operation of law upon conviction.” Besides the right to vote, the only rights “lost by operation of law” are the right to serve on a jury and the right to bear arms. The U.S. Supreme Court has established that voting is a fundamental right protected by the U.S. Constitution. However, the Court has not construed the right to serve on

30. WASH. CONST. art. VI, § 3.
32. Id. § 9A.20.021(10).
33. Id. §§ 9.96.050; 9.94A.637.
34. Id. § 9.96.010. Section 9.96.020 provides that the governor may restore felons’ civil rights by signing a civil rights restoration certificate. However, although the statute remains valid, the civil rights restoration by the governor was essentially supplanted by the Sentencing Reform Act’s process of civil rights restoration upon discharge. Id. § 9.94A.637. Personal correspondence with Everett Billingslea, Counsel to Governor Gary Locke. (On file with the author). Consequently, current Governor Gary Locke does not restore civil rights for Washington offenders, but instead refers them to the sentencing court. Id.
35. Id.
37. Id.
38. WASH. CONST. art. VI, § 3; WASH. REV. CODE § 29.10.097 (2002).
39. WASH. REV. CODE § 2.36.070.
40. Id. § 9.41.040(1)(b)(i).
Felon Re-Enfranchisement

a jury or the right of individuals to bear arms as fundamental rights. Likewise, the Washington Supreme Court has not discussed the right to serve on a jury and has stopped short of applying fundamental status to the Washington State Constitution’s right to bear arms. Fundamental rights other than voting, such as the right to free speech and the right to be free from unreasonable search and seizure, are not entirely lost upon conviction, and their restoration is not contingent on payment of legal-financial obligations. Instead, other fundamental rights are only restricted during incarceration and parole when necessary to advance “legitimate penological interests.” Consequently, the only fundamental right restored through the civil rights restoration process, and thus the only fundamental right conditioned on the payment of legal-financial obligations, is an offender’s right to vote.

Washington’s constitutional history does not provide a rationale for denying offenders of the fundamental right to vote, nor does the legislative record provide an explanation for requiring felons to pay legal-financial obligations prior to regaining the franchise. Other courts

42. United States v. Miller, 307 U.S. 174, 178 (1939) (holding that the Second Amendment is a right held by the states and does not protect the possession of a weapon by private citizens).
44. Rights protected by the Constitution’s Bill of Rights are fundamental. JOHN E. NOWAK & RONALD D. ROTUNDA, CONSTITUTIONAL LAW 1002 (6th ed. 2000).
46. Id. at 89.
47. Even though the right to bear arms is not fundamental, most offenders’ right to possess firearms may be restored regardless of whether they have paid their legal-financial obligations. WASH. REV. CODE § 9.41.040(4)(b)(i) (2002). With or without payment of legal-financial obligations, most offenders can petition the sentencing court for the restoration of their right to possess a firearm if they have lived in the community for five or more consecutive years without being convicted of another crime. Id. Thus, offenders who have not paid their legal-financial obligations may have their right to possess a firearm restored, yet their right to vote—a fundamental right—will be denied because their sentence has not been discharged.
48. Washington’s first disenfranchisement law was passed in 1866, when Washington was a territory. State v. Collin, 69 Wash. 268, 270 (1912). This law was incorporated into the Washington Constitution as Article VI, section 3 at the Constitutional Convention of 1889. Id. However, the delegates added the phrase “unless restored to civil rights” upon a separate motion. BEVERLY PAULIK ROSENOW, THE JOURNAL OF THE WASHINGTON STATE CONSTITUTIONAL CONVENTION 1889, 638–39 (1962). The limited history that does exist from the Constitutional Convention does not describe the discussions surrounding felon disenfranchisement or the addition of the civil rights restoration provision. Id.; see also ROBERT F. UTTER & HUGH SPITZER, THE WASHINGTON STATE CONSTITUTION: A REFERENCE GUIDE 122–23 (2002).
49. The requirement that offenders satisfy all conditions of their sentence, including payment of legal-financial obligations, before restoration of civil rights is part of the Sentencing Reform Act of
and commentators have attributed state felon disenfranchisement laws to a state’s interest in promoting responsible use of the ballot.\textsuperscript{50} However, no published court opinion or commentator has discussed states’ rationales for requiring felons to pay their legal-financial obligations before regaining the right to vote.

\textbf{B. Washington Courts Assess Legal-Financial Obligations at Sentencing, and Failure to Pay Legal-Financial Obligations May Result in Civil or Criminal Sanctions}

When a person is convicted of a felony in Washington, the court must order the payment of legal-financial obligations as a part of the offender’s sentence.\textsuperscript{51} There are many different kinds of legal-financial obligations imposed in Washington, but the most common\textsuperscript{52} are restitution,\textsuperscript{53} a victim penalty assessment,\textsuperscript{54} trial court costs,\textsuperscript{55} fines,\textsuperscript{56} and local drug penalties.\textsuperscript{57} The sentencing court must impose a five hundred-

\textsuperscript{50} See Shepard v. Trevino, 575 F.2d 1110, 1115 (5th Cir. 1978) (noting that felons, “like insane persons[,] have raised questions about their ability to vote responsibly”); Green v. Bd. of Elections, 380 F.2d 445, 451–52 (2d Cir. 1967) (stating that allowing felons to vote may lead to political corruption); Roger Clegg, \textit{Who Should Vote?}, 6 TEX. REV. L. & POL. 159, 172 (2001) (arguing that felons should be denied the right to vote because their criminal acts call into question their trustworthiness and loyalty); Note, The Disenfranchisement of Ex-felons: Citizenship, Criminality, and “the Purity of the Ballot Box,” 102 HARV. L. REV. 1300, 1305–09 (1989) (surveying the justifications for felon disenfranchisement); ALEXANDER KEYSSAR, \textit{The Right to Vote: The Contested History of Democracy in the United States} 162–63 (2000) (discussing historical justifications for felon disenfranchisement).


\textsuperscript{54} Id. \S 7.68.035.

\textsuperscript{55} Id. \S 10.01.160.

\textsuperscript{56} Id. \S 9.94A.550.

dollar victim penalty assessment in all felony cases, and restitution is mandatory where the crime results in harm to person or property. Most other legal-financial obligations may be ordered at the discretion of the court or are mandatory only to the extent that the offender is able to pay. Consequently, regardless of financial condition, every person convicted of a felony in Washington is assessed at least five hundred dollars in legal-financial obligations.

At sentencing, the court must establish a payment schedule based on the offender’s financial resources. The offender must abide by the payment schedule to avoid further sanctions. From the date of sentencing, offenders are charged interest on their unpaid legal-financial obligations at a minimum rate of twelve percent per year. If offenders fall behind in their payments, the state may pursue a number of debt collection devices, including garnishment of wages. Additionally, the sentencing court and the DOC may impose sanctions, such as jail time or community service, if offenders willfully fail to pay legal-financial obligations.

C. Legal-Financial Obligations Present a Substantial Hurdle to the Restoration of Offenders’ Voting Rights

Many offenders have a difficult time meeting their legal-financial obligation requirements. Indigent offenders who are never able to pay are

58. State v. Curry, 118 Wash. 2d 911, 917, 829 P.2d 166, 186 (1992) (holding that the victim penalty assessment is mandatory because there is no statutory provision for waiver).
59. Wash. Rev. Code § 9.94A.753 (2002). Restitution must be ordered by the court in cases of damage to property or injury to person “unless extraordinary circumstances exist which make restitution inappropriate in the court’s judgment.” Id.
60. For example, trial court costs and fines are assessed at the discretion of the court. Id. § 10.01.160. Also, the Department of Corrections (DOC) supervision fees are mandatory but may be waived if the offender is unable to pay because of unemployment, handicap, or other extenuating circumstances that make payment difficult. Id. § 9.94A.780 (2002).
61. Id. § 9.94A.760(1). If the court does not set a payment schedule, the DOC must set one. Id.
62. Id. §§ 9.94A.634(3)(a)(i); 9.94A.634(3)(c).
63. Interest on criminal judgments accrues at the same rate as that required for civil judgments, and accrues at the highest rate permitted by law. See id. §§ 10.82.090; 4.56.110. The maximum interested rate permitted by law is the higher of: (1) 12% or (2) the rate of a 26-week T-bill plus four percent. Id. § 19.52.020.
64. Id. § 9.94A.760(3).
65. Id. §§ 9.94A.634(3)(a)(i); 9.94A.634(3)(c).
permanently ineligible for re-enfranchisement.\textsuperscript{66} While the DOC and county courts track overall collection rates, they do not record how much the average offender is required to pay or how many offenders have difficulty paying their legal-financial obligations.\textsuperscript{67} The limited evidence available regarding offenders’ ability to pay suggests that thousands of offenders in Washington are currently unable to receive a discharge, and thus are ineligible for re-enfranchisement, because they cannot pay their legal-financial obligations. In a fiscal note to a 2002 Washington State Senate bill, the DOC estimated that as of December 31, 2001, about 46,500 offenders had completed all the requirements of their sentences except for paying their legal-financial obligations.\textsuperscript{68} Additionally, approximately ninety percent of the offenders who appear before the sentencing court for failure to pay legal-financial obligations qualify for a public defender,\textsuperscript{69} which means that they are at or below 125 percent of the federal poverty line.\textsuperscript{70} The high rate of indigency among offenders who are delinquent in making their legal-financial obligation payments suggests that they simply lack the financial resources to meet their payment schedule. Data on collection rates also supports this conclusion. The DOC reports that in the year 2000 it collected only twenty eight percent of the assessed legal-financial obligations.\textsuperscript{71} Of the mandatory victim penalty assessments imposed between 1995 and 2000, the DOC reports that it collected just 25.4 percent.\textsuperscript{72}

In sum, convicted felons in Washington cannot regain voting rights until they have paid all legal-financial obligations imposed at sentencing. Therefore, although Washington has a process for restoring offenders’ voting rights, that process is foreclosed to many offenders because they lack the financial resources to pay their debts. The U.S. Supreme Court’s voting rights case law,\textsuperscript{73} along with its opinions regarding other

\begin{footnotesize}
\textsuperscript{66} If offenders do not pay their legal financial obligations they are ineligible for a sentence discharge; a discharge is a pre-requisite for re-enfranchisement. WASH. REV. CODE §9.94A.637 (2002).
\textsuperscript{67} Telephone Interview with Barbara Miner, Director of King County Clerk’s Office (July 10, 2002).
\textsuperscript{68} DEPARTMENT OF CORRECTIONS, AGENCY FISCAL NOTE FOR SENATE BILL 6519 (2002).
\textsuperscript{69} DEPARTMENT OF COMMUNITY, TRADE, AND ECONOMIC DEVELOPMENT, LOCAL GOVERNMENT FISCAL NOTE FOR HOUSE BILL 2712 (2002).
\textsuperscript{71} COMPREHENSIVE REVIEW, supra note 52, at 56–57.
\textsuperscript{72} Id.
\textsuperscript{73} See infra Section II.
\end{footnotesize}
fundamental rights,\textsuperscript{74} suggest Washington’s bar on voting for failure to pay legal-financial obligations violates the Equal Protection Clause.\textsuperscript{75} The following sections investigate this possibility.

II. VOTING IS A FUNDAMENTAL RIGHT SECURED BY THE EQUAL PROTECTION CLAUSE

Although the right to vote is not explicitly guaranteed in the text of the U.S. Constitution, the U.S. Supreme Court has held that once states grant citizens the right to vote, it becomes a fundamental right protected by the Constitution.\textsuperscript{76} As a fundamental right, state laws alleged to restrict voting rights must survive strict scrutiny under the Equal Protection Clause.\textsuperscript{77} To satisfy strict scrutiny, state laws must treat all citizens alike when providing the right to vote, unless disparate treatment is necessary to advance a compelling governmental interest.\textsuperscript{78}

A. Once Granted by the State, the Right to Vote is Fundamental

Neither the text of the U.S. Constitution nor any constitutional amendment contains an affirmative grant of the right to vote,\textsuperscript{79} and until the mid-twentieth century voting was not considered a fundamental right.\textsuperscript{80} The Constitution scarcely mentions voting,\textsuperscript{81} and the amendments that address voting rights do so only in negative terms, providing that the right to vote “shall not be abridged or denied” on account of race,\textsuperscript{82} sex,\textsuperscript{83} age,\textsuperscript{84} or payment of a poll tax.\textsuperscript{85} During the nineteenth century, the U.S. Supreme Court held that “the Constitution of the United States has not

\textsuperscript{74}. See infra Section IV.
\textsuperscript{75}. See infra Section V.
\textsuperscript{76}. Id.
\textsuperscript{78}. Wesberry v. Sanders, 376 U.S. 1, 18 (1964).
\textsuperscript{80}. Id.
\textsuperscript{81}. The only mention of citizen electoral rights is in Article I, section 2 which states “The House of Representatives shall be . . . chosen every second Year by the People of the several States . . . .”
\textsuperscript{82}. U.S. CONST. amend. XV.
\textsuperscript{83}. U.S. CONST. amend. XIX.
\textsuperscript{84}. U.S. CONST. amend. XXVI.
\textsuperscript{85}. U.S. CONST. amend. XXIV.
conferred the right to vote upon any one.” Nonetheless, following the Civil War the Court began to recognize that voting was one of a citizen’s most important political rights. In *Yick Wo v. Hopkins*, the Court stated that, “[t]hough not regarded strictly as a natural right, but as a privilege merely conceded by society . . . [voting] is regarded as a fundamental political right, because [it is] preservative of all rights.” During the first half of the twentieth century, the Court did not address the constitutional status of voting rights but instead focused on the problem of disenfranchisement of African Americans. Then, in a series of cases between 1959 and 1969, the Court expanded the scope of the right to vote and declared voting a “fundamental matter in a free and democratic society.”

According to the U.S. Supreme Court, the fundamental right to vote is implied in the U.S. Constitution’s structure and arises once states provide citizens with the opportunity to vote. In *Wesberry v. Sanders*, the Court stated that “our Constitution leaves no room for classification of people in a way that unnecessarily abridges” the right to vote because “[n]o right is more precious in a free country than having a voice in the election of those who make the laws . . . . Other rights, even the most basic, are illusory if the right to vote is undermined.” The Court has acknowledged that states may determine the conditions under which the right to vote can be exercised, but it has held that once states grant voting rights to their citizens they must not do so in a discriminatory manner. Upon granting the franchise, states must treat all eligible voters similarly, unless disparate treatment is necessary to advance a compelling interest.

---

87. 118 U.S. 356 (1886).
88. Id. at 370.
89. IDSACHAROFF ET AL., supra note 79, at 46.
93. Id. at 17–18.
B. State Laws Infringing on the Fundamental Right to Vote Are Subject to Strict Scrutiny

The Equal Protection Clause forbids states from treating similarly situated people differently, but there is no universal standard for determining whether a state law inappropriately imposes disparate treatment. Instead, the U.S. Supreme Court employs three standards of review for Equal Protection challenges: strict scrutiny, intermediate scrutiny, and rational basis scrutiny. Courts reviewing an Equal Protection challenge must choose between these standards based on the answers to two initial questions. First, on what basis does the law classify groups of people for disparate treatment? Second, what right is at stake?

Laws challenged under the Equal Protection Clause receive strict scrutiny—the most intensive level of review—if the law involves disparate treatment based on suspect classifications such as race or if the disparate treatment threatens a fundamental right. Under strict scrutiny, courts will strike down state laws unless they are narrowly tailored to advance a compelling governmental interest. Laws receive

98. See Dunn, 405 U.S. at 335 (highlighting the three considerations encompassed within these two questions).
99. NOWAK & ROTUNDA, supra note 44, at 639.
100. Id.
101. Id.
102. Id.
103. The U.S. Supreme Court has recognized race, alienage, and national origin as suspect classifications requiring strict scrutiny. City of Cleburne v. Cleburne Living Ctr., Inc., 473 U.S. 432, 440 (1985). See also NOWAK & ROTUNDA, supra note 44, at 640. The Court has declined to extend the suspect classification label to other groups of people, including the disabled and the poor. See Cleburne, 473 U.S. at 439 (people with disabilities); San Antonio Indep. Sch. Dist. v. Rodriguez, 411 U.S. 1, 28–29 (1973) (families with low incomes).
104. LAWRENCE TRIBE, AMERICAN CONSTITUTIONAL LAW 1454 (2d ed. 1988). Fundamental rights which invoke strict scrutiny include, inter alia, the right to vote, the right to marry, the right to travel, and those rights explicitly protected in the Bill of Rights. See Dunn, 405 U.S. at 336 (right to vote); Zablocki v. Redhail, 434 U.S. 374, 383 (1978) (right to marry); Shapiro v. Thompson, 394 U.S. 618, 634 (1969) (right to travel); NOWAK & ROTUNDA, supra note 44, at 1002 (rights protected in the Bill of Rights).
105. Miller v. Johnson, 515 U.S. 900, 921 (1995). At times, the U.S. Supreme Court has used the word “necessary” rather than “narrowly tailored” in its articulations of the strict scrutiny test. See Dunn, 405 U.S. at 337. Both expressions of the strict scrutiny test capture the spirit of the Court’s standard—that the restriction must not merely advance the compelling state interest but must be essential to its achievement—and are used interchangeably throughout this Comment.
intermediate scrutiny if they involve classifications that require heightened judicial review but are not “suspect” and do not implicate a fundamental right.\textsuperscript{106} To survive intermediate scrutiny, state laws must substantially further an important governmental interest.\textsuperscript{107} All other challenges receive rational basis scrutiny, which requires state laws to be rationally related to a legitimate governmental purpose.\textsuperscript{108}

Regardless of the basis for classification, voting laws receive strict scrutiny because they implicate a fundamental right.\textsuperscript{109} Under the strict scrutiny standard, laws that impose voter qualifications will be struck down unless they are necessary to advance a compelling governmental interest.\textsuperscript{110} In \textit{Reynolds v. Sims},\textsuperscript{111} the U.S. Supreme Court recognized voting as a fundamental right and directed courts to apply strict scrutiny, stating “any alleged infringement of the right of citizens to vote must be carefully and meticulously scrutinized.”\textsuperscript{112} Further, in \textit{Buckley v. Valeo},\textsuperscript{113} the Court stated that restrictions on the electoral process must survive “exacting scrutiny”\textsuperscript{114} and will only be sustained if they further a “vital governmental interest.”\textsuperscript{115} The Court has struck down a number of voter qualifications on Equal Protection grounds because the qualifications imposed conditions that were not narrowly tailored to advance compelling governmental interests.\textsuperscript{116} For example, the Court has invalidated voter qualifications that established a requirement of wealth,\textsuperscript{117} required residents to pay property taxes in order to vote in school board elections,\textsuperscript{118} excluded residents based on military status,\textsuperscript{119} or imposed lengthy residency requirements.\textsuperscript{120}

\begin{itemize}
\item \textsuperscript{106} TRIBE, \textit{supra} note 104, at 1613. The U.S. Supreme Court has recognized gender and illegitimacy as classifications that require intermediate scrutiny. \textit{Id.} at 1614.
\item \textsuperscript{107} Clark v. Jeter, 486 U.S. 456, 461 (1988).
\item \textsuperscript{109} ISSACHAROFF ET AL., \textit{supra} note 79, at 303–05.
\item \textsuperscript{110} Dunn, 405 U.S. at 337.
\item \textsuperscript{111} 377 U.S. 533 (1964).
\item \textsuperscript{112} \textit{Id.}
\item \textsuperscript{113} 424 U.S. 1 (1974).
\item \textsuperscript{114} \textit{Id.} at 93–94.
\item \textsuperscript{115} \textit{Id.}
\item \textsuperscript{116} Gordon v. Lance, 403 U.S. 1, 5 (1971).
\item \textsuperscript{118} Kramer v. Union Free Sch. Dist., 395 U.S. 621, 622 (1969).
\end{itemize}
In sum, since the mid-twentieth century, the U.S. Supreme Court has upheld citizens’ fundamental right to vote and permitted only those voter qualifications that are narrowly tailored to advance a compelling governmental interest. Most state voting rights restrictions have been struck down as unnecessary infringements on a fundamental right. However, as the next section illustrates, the Court has permitted one exception to the rule that states cannot treat citizens differently with respect to voting rights: states can exclude from the franchise citizens convicted of crimes.121

III. STATES CAN DISENFRANCHISE OFFENDERS UPON CONVICTION, BUT THE U.S. SUPREME COURT HAS NOT YET DECIDED THE CONSTITUTIONAL REQUIREMENTS FOR FELON RE-ENFRANCHISEMENT

Although the right to vote is fundamental, states may restrict offenders’ right to vote without running afoul of the Equal Protection Clause if disenfranchisement laws are not otherwise discriminatory.122 However, the U.S. Supreme Court has never addressed the Equal Protection Clause’s requirements for felon re-enfranchisement laws. Interpreting Section 2 of the Fourteenth Amendment, the Court in Ramirez permitted states to deny the right to vote to persons convicted of crimes.123 Under Ramirez, states do not violate the Equal Protection Clause when they distinguish for voting rights purposes between citizens who have committed crimes and those who have not.124 However, in Hunter v. Underwood,125 the Court clarified that felon disenfranchisement laws are not entirely immune from all Equal Protection challenges.126 State disenfranchisement laws violate the Equal Protection Clause if the laws discriminate on a basis other than offender

---

119. Carrington v. Rush, 380 U.S. 89, 96–97 (1965) (striking down a Texas constitutional provision prohibiting members of the armed forces who move to Texas during the course of military duty from voting while they remain in the military).
121. See infra Section III.A.
124. Id.
126. Id. at 233.
status. The Court has not decided whether states, when re-enfranchising offenders, are permitted to make distinctions using classifications such as race or economic status that cannot be used when granting the initial right to vote.

A. The U.S. Constitution Permits States to Deny Voting Rights to Citizens Convicted of Crimes

A state’s ability to deny felons the right to vote is derived from Section 2 of the Fourteenth Amendment. According to Section 2, “representatives will be apportioned among the several States according to their respective numbers.” States that deny the right to vote to any eligible male “except for participation in rebellion, or other crime” will have their representation reduced in proportion to the number of persons denied the right to vote. Based on the explicit exception of criminal convictions for apportionment purposes, the U.S. Supreme Court concluded in Ramirez that Section 2 of the Fourteenth Amendment implicitly authorizes states to deny the right to vote to persons convicted of crimes.

In Ramirez, three California residents convicted of felonies sued their county clerk for refusing to register them to vote. They argued that California’s felon disenfranchisement law violated the Equal Protection Clause because it denied them the right vote while granting it to all other residents. Their claim rested on prior U.S. Supreme Court voting rights cases, which held that states could not impose voter qualifications that distinguish between different groups of eligible voters based on irrelevant factors. However, the Supreme Court’s opinion in Ramirez, relying on the text and original intent of Section 2 of the Fourteenth

127. Id.
128. U.S. CONST. amend XIV, § 2 (emphasis added).
129. Id.
131. Id. at 31–33.
132. Id. at 33. The plaintiffs also claimed that California’s felon disenfranchisement laws violated the Equal Protection Clause because county clerks throughout the state were inconsistent in their application of the felon disenfranchisement law, thereby arbitrarily granting some felons but not others voting rights. Id. Because the California Supreme Court had not considered this claim, the U.S. Supreme Court remanded the case for consideration of whether the counties’ implementation of California’s disenfranchisement law violated the Equal Protection Clause. Id. at 56.
133. Id.
Amendment, held that states do not violate the Equal Protection Clause when they disenfranchise felons based on their status as offenders.\(^{134}\)

**B. Felon Disenfranchisement Laws May Violate the Equal Protection Clause When They Are Otherwise Discriminatory**

A decade after *Ramirez*, the U.S. Supreme Court clarified that state disenfranchisement laws can violate the Equal Protection Clause if the laws disenfranchised offenders on a basis other than criminal conviction.\(^{135}\) In *Hunter v. Underwood*, citizens convicted of misdemeanor crimes challenged an Alabama constitutional provision that disenfranchised persons convicted of crimes “involving moral turpitude.”\(^{136}\) Alabama’s definition of crimes of “moral turpitude” included misdemeanor offenses such as presenting a worthless check and petty larceny, but excluded more serious offenses such as second-degree manslaughter and assault on a police officer.\(^{137}\) The plaintiffs argued that the selection of crimes “involving moral turpitude” was racially discriminatory, intending to disenfranchise blacks more often than whites.\(^{138}\) The Court struck down Alabama’s law, citing the unusual selection of crimes\(^{139}\) and legislative history demonstrating that the constitutional provision was enacted with the purpose of disenfranchising blacks.\(^{140}\) The *Hunter* Court held that felon disenfranchisement laws that are enacted with “the desire to discriminate” against persons on account of race and that in fact have a discriminatory effect violate the Equal Protection Clause.\(^{141}\)

The explicit holding of the Court in *Hunter*—that felon disenfranchisement laws violate the Equal Protection Clause if they are

\(^{134}\) Id. at 54.

\(^{135}\) 471 U.S. 222, 233 (1985). In *Ramirez*, the Court suggested, but did not decide, that felon disenfranchisement laws violate the Equal Protection Clause if otherwise discriminatory when it remanded the case to the California Supreme Court for consideration of whether California’s implementation of its disenfranchisement law was arbitrary and therefore violated the Equal Protection Clause. 418 U.S. at 56. See also *supra* note 132.

\(^{136}\) *Hunter*, 471 U.S. at 223 (quoting the Ala. Const. § 182 (1901)).

\(^{137}\) Id. at 226–27.

\(^{138}\) Id. at 224.

\(^{139}\) Id. at 227.

\(^{140}\) Id. at 229–30.

\(^{141}\) Id. at 233.
racially discriminatory—suggests that where felon disenfranchisement laws are discriminatory on any basis other than offender status, they must survive a traditional Equal Protection analysis. To resolve the plaintiffs’ Fourteenth Amendment claim, the Hunter Court did not first discuss Ramirez or disenfranchisement laws’ immunity from Equal Protection challenges, but instead began with a traditional Equal Protection analysis. Furthermore, at the end of its analysis, the Court noted that its decision was entirely consistent with the Ramirez decision because “§ 2 was not designed to permit . . . purposeful racial discrimination . . . .”144 The Hunter Court’s willingness to entertain an Equal Protection challenge based on racial discrimination without first considering the specific requirements of Section 2 of the Fourteenth Amendment demonstrates that felon disenfranchisement laws are not entirely immune from Equal Protection scrutiny.145

C. The U.S. Supreme Court Has Not Considered the Constitutional Requirements of Felon Re-Enfranchisement Laws

The constitutional requirements of felon re-enfranchisement remain an unsettled area of law because both Ramirez and Hunter exclusively addressed felon disenfranchisement. Ramirez considered whether states, under the crime exception in Section 2 of the Fourteenth Amendment, could deny felons the right to vote while granting it to all other citizens.146 Hunter dealt with a state’s initial decision about which crimes would fall under Section 2’s crime exception and clarified that the state may not make those determinations for discriminatory reasons.147 Ramirez and Hunter leave unanswered the central constitutional question of felon re-enfranchisement: whether states, when re-granting voting rights, can distinguish among offenders by imposing qualifications that would otherwise violate the Equal Protection Clause’s voting rights requirements.

Although no U.S. Supreme Court opinion has addressed the constitutionality of felon re-enfranchisement requirements, recently in

142. Id.
143. Id. at 227–28.
144. Id. at 233.
145. Id. at 227–28.
**Felon Re-Enfranchisement**

*Johnson v. Bush*, 148 a federal district court in the Southern District of Florida considered whether a requirement that offenders pay restitution before becoming eligible for re-enfranchisement amounted to an unconstitutional poll tax. 149 Plaintiffs challenged that the restitution requirement was a poll tax equivalent, and therefore violated the Fourteenth and Twenty-Fourth Amendments, as well as the Voting Rights Act of 1965. 150 The federal district court did not consider the three bases of the plaintiffs’ challenge separately, and merely held that Florida was entitled to summary judgment on the plaintiff’s poll tax claims. 151 To support its grant of summary judgment, the *Johnson* court noted that Florida’s restitution requirement was unlike traditional poll taxes because it was a part of Florida’s civil rights restoration process rather than a requirement for voter registration or a restraint on exercise of the franchise. 152 In addition, according to the *Johnson* court, payment of restitution is related to an offender’s “readiness to return to the electorate” and therefore is not an unconstitutional poll tax. 153 Moreover, the court suggested that upon conviction felons lose all stake in the right to vote, and consequently any decision by Florida to restore voting rights does not invoke constitutional protections. 154

---


149. Id. at 1342–43. The *Johnson* court has been the only court to consider whether a requirement that offenders pay money as a prerequisite to the restoration of voting rights violates the Equal Protection Clause. The Second Circuit, in a case opinion pre-dating *Hunter* and *Ramirez*, did consider whether offenders could challenge the constitutionality of state laws which require the payment of a fee for restoration of voting rights, but did not decide whether such requirements actually violate the Equal Protection Clause. See *Bynum v. Conn. Comm’n on Forfeited Rights*, 410 F.2d 173, 175–76 (2d Cir. 1969). In holding that the plaintiff’s challenge was valid, the Second Circuit reasoned “[t]he focal question is whether [a state], once having agreed to permit ex-felons to regain their vote and having established administrative machinery for this purpose, can deny access to this relief solely because one is too poor to pay the required fee.” Id. The Second Circuit referred the case to a three-judge panel, which never produced an opinion on the validity of Connecticut’s re-enfranchisement requirements. Id. at 177. The Connecticut Legislature later repealed the re-enfranchisement statute that required payment of a fee. See *Conn. Gen. Stat.* § 9-48 (2001). More recently, the Fourth Circuit, in an unpublished opinion, rejected without much discussion the plaintiff’s claim that Virginia’s requirement that offenders pay money to apply for the restoration of their voting rights violated the Twenty-Fourth Amendment’s prohibition against poll taxes. See *Howard v. Gilmore*, No. 99-2285, 2000 U.S. App. LEXIS 2680, at *4 (4th Cir. Feb. 23, 2000). The Fourth Circuit did not examine the re-enfranchisement requirements under the Fourteenth Amendment’s Equal Protection Clause. Id.


151. Id. at 1342.

152. Id. at 1343.

153. Id.

154. Id.
Though the Johnson court held that Florida’s restitution requirement was not a poll tax, the court, like the U.S. Supreme Court, did not directly decide the central question of felon re-enfranchisement. Can states, in their re-enfranchisement laws, distinguish between groups of felons in ways that otherwise violate the voting rights requirements of the Equal Protection Clause? The U.S. Supreme Court’s opinions in analogous areas of law, explored in the next section, are helpful in answering this question.

IV. STATES CANNOT RESTRICT FUNDAMENTAL RIGHTS BASED ON AN INABILITY TO PAY MONEY TO THE STATE WITHOUT ADVANCING A COMPELLING INTEREST

Although the U.S. Supreme Court has not decided the requirements for felon re-enfranchisement laws, it has addressed restrictions of other fundamental rights based on an inability to pay. First, with respect to voting rights, the Court has held that states cannot condition voting on financial resources or payment of money to the state.155 Second, outside the context of voting, the Court has held that states may not condition offenders’ fundamental rights on the payment of legal-financial obligations and court costs.156 Third, addressing fundamental rights more generally, the Court has held that states may not restrict citizens’ fundamental rights because of a failure to pay pre-existing debt.157 In addition, restricting fundamental rights to serve the state’s “debt collection” interests is unconstitutional where the state has equally effective, yet less restrictive options available.158

A. States May Not Impose Voter Qualifications Based on Payment of Money to the State or an Individual’s Affluence

Voter qualifications that require citizens to pay money to the state, most often identified with a poll tax requirement, violate the Equal Protection Clause.159 Poll tax requirements are taxes, or the functional equivalent, that burden the exercise of an individual’s right to vote. Poll

158. Id. at 389–90; Williams, 399 U.S. at 245.
taxes violate the Fourteenth \(^{160}\) and Twenty-Fourth Amendments \(^{161}\) of the U.S. Constitution, as well as the Voting Rights Act of 1965. \(^{162}\) Moreover, the U.S. Supreme Court has stated that the Fourteenth Amendment’s voting rights requirements prohibit more than explicit poll taxes, extending to any state requirement that introduces affluence as a qualification of voting. \(^{163}\)

In *Harper v. Virginia Board of Elections*, \(^{164}\) the U.S. Supreme Court struck down a Virginia law requiring voters to pay a $1.50 poll tax in order to vote in state elections. \(^{165}\) According to the Court, a state’s interest in voting is limited to the power to fix voter qualifications that “promote the intelligent use of the ballot.” \(^{166}\) Because “wealth, like race, creed or color, is not germane to one’s ability to participate intelligently in the electoral process,” states that introduce “wealth or payment of a fee as a measure of a voter’s qualification . . . introduce a capricious and irrelevant factor.” \(^{167}\) Therefore, the Court concluded that distinguishing among citizens based on their ability to pay by conditioning voting on affluence or the payment of money violates the Equal Protection Clause. \(^{168}\) In addition, the Court emphasized that voter qualifications based on affluence violate the Constitution no matter how large or small the degree of resulting discrimination. \(^{169}\) To the Court, any discrimination based on financial condition is unacceptable. \(^{170}\)

---

160. *Id.* at 666.

161. The Twenty-Fourth Amendment prohibits the use of a state poll tax or any other tax to “deny or abridge” the right of citizens to vote in federal primary and general elections. U.S. CONST. amend. XXIV. In *Harman v. Forssenius*, the U.S. Supreme Court held that the Twenty-Fourth Amendment also prohibits requirements that are the functional equivalent of a poll tax. 380 U.S. 528, 540–41 (1965).


165. *Id.* at 668. *Harper* was decided two years after the passage of the Twenty-Fourth Amendment in 1964, which prohibited the collection of poll taxes in federal elections. U.S. CONST. amend. XXIV. Poll tax requirements for state elections were valid until the Court’s decision in *Harper*. 383 U.S. at 666.


167. *Id.* at 668.

168. *Id.*

169. *Id.* at 665.

170. *Id.*
B. States Cannot Distinguish Between Groups of Offenders Based on Their Ability to Pay Where Fundamental Rights Are at Stake

Although the U.S. Supreme Court has permitted distinctions between offenders and other citizens for voting rights purposes, it has been less tolerant of distinctions between groups of offenders, especially where the distinctions fall along economic lines. The Court has held that the length of offenders’ incarceration may not be conditioned on their ability to pay legal-financial obligations.\(^{171}\) Furthermore, the Court has held that although states are not constitutionally required to provide appellate courts, if they do provide appellate review, it cannot be based on offenders’ ability to pay money.\(^{172}\)

1. States Cannot Prevent Offenders from Regaining Their Liberty Because They Are Unable to Pay Legal-Financial Obligations

States cannot restrict offenders’ liberty longer than the maximum time permitted by statute because they are unable to pay the legal-financial obligations imposed at sentencing.\(^{173}\) In *Williams v. Illinois*,\(^{174}\) the Court invalidated an Illinois law requiring indigent offenders who were unable to pay their legal-financial obligations to remain in prison until they had worked off their debts.\(^{175}\) According to the Court, the Illinois requirement violated the Equal Protection Clause because it determined which offenders would be released and which ones would not based on an ability to pay.\(^{176}\) Although the Court recognized that “the sentence was not imposed upon appellant because of his indigency but because he had committed a crime,” the Court held the law was nonetheless discriminatory because it presented some offenders with the opportunity to regain their freedom while effectively denying the same opportunity to other offenders because of their economic circumstances.\(^{177}\)

The *Williams* Court acknowledged that judges have considerable discretion when sentencing offenders and may take into account a wide


\(^{172}\) Griffin v. Illinois, 351 U.S. 12, 18 (1956).

\(^{173}\) Williams, 399 U.S. at 243.


\(^{175}\) Id. at 243–44.

\(^{176}\) Id. at 244.

\(^{177}\) Id. at 242.
range of factors. According to the Court, however, the Illinois law went beyond sentencing discretion to punish some offenders longer—beyond even the statutory maximum—solely based on their financial condition. The Court noted that Illinois had imposed “different consequences on two categories of persons” and held that “the Fourteenth Amendment requires that the statutory ceiling placed on imprisonment for any substantive offence be the same for all defendants irrespective of their economic status.” In later cases, the Court has construed its holding in Williams as prohibiting states from imposing a fine and then converting it to jail time because the offender does not have the financial resources to pay the fine.

At the conclusion of it’s constitutional analysis, the Court recognized that preventing states from incarcerating offenders for failure to pay legal-financial obligations would make collection somewhat more difficult, but the Court noted that it would not eliminates a state’s ability to collect the debts altogether. Dismissing the added burden to the state, the Court stated “the constitutional imperatives of the Equal Protection Clause must have priority . . . .” over the state’s inconvenience. Further, the Court reasoned that “administrative inconvenience would be minimal” because states have a wide variety of other debt collection options available that do not restrict offenders’ liberty through additional incarceration. Consequently, in Williams, the Court held that the denial of liberty through continued incarceration was unnecessary to advance the state’s debt collection interest.

178. Id. at 243.
179. Id.
180. Id. at 244.
181. Tate v. Short, 401 U.S. 395, 398 (1971). Further, in Bearden v. Georgia the Court held that due process prohibits states from incarcerating offenders for failure to pay legal-financial obligations without inquiring into the willfulness of the nonpayment and without considering alternative forms of punishment that do not restrict liberty interests. 461 U.S. 660, 672–73 (1983). The Court stated that automatic incarceration violates the Fourteenth Amendment because it is fundamentally unfair to incarcerate offenders solely because they lack the financial resources to pay their legal-financial obligations. Id.
182. Williams, 399 U.S. at 245.
183. Id.
184. Id.
185. Id. at 244–45.
2. States Cannot Restrict Offenders’ Access to Appellate Review Based on Their Inability to Pay Court Costs

The fundamental right of offenders to access the courts may not be constrained by an inability to pay court costs.186 In Griffin v. Illinois,187 the U.S. Supreme Court invalidated an Illinois law that required defendants challenging a criminal conviction to purchase a copy of their trial court transcript.188 According to the Court, Illinois’ law violated the Fourteenth Amendment because “there can be no equal justice where the kind of trial a man gets depends on the amount of money he has.”189

As in the voting rights cases,190 the Court held in Griffin that, although the right to access state courts is not mandated in the text of the U.S. Constitution, an offender’s right to access appellate review is constitutionally required where the state provides an appellate court system.191 The Court emphasized that states cannot provide for appellate review “in a way that discriminates against some convicted defendants on account of their poverty” because the ability to pay costs “bear[s] no rational relationship to a defendant’s guilt or innocence.”192 As a result of the Court’s decision in Griffin, distinctions between convicted defendants based on financial condition, which have the effect of restricting poor defendants’ access to the court system, violate the Equal Protection Clause.193

188. Id.
189. Id. Seven years later, the Court extended the principle in Griffin by holding that states must also provide indigent defendants counsel for their first appeal of right. Douglas v. California, 372 U.S. 353, 357–58 (1963). However, in Ross v. Moffitt, the Court declined to extend this right to discretionary appeals. 417 U.S. 600, 617–18 (1974).
192. Id. at 17–18.
193. Id.
C. States Cannot Condition Fundamental Rights on the Payment of Pre-existing Debts Unless the Restriction is Necessary to Advance a Compelling State Interest

Courts apply strict scrutiny review to laws that restrict fundamental rights by imposing payment requirements. To survive strict scrutiny, a restriction must be necessary to advance a compelling state interest and be narrowly tailored to effectuate only that interest. The U.S. Supreme Court set forth these principles in *Zablocki v. Redhail*, a case in which plaintiffs argued that a Wisconsin statute violated the Equal Protection Clause because it restricted the fundamental right to marry. Wisconsin’s law required marriage license applicants to be current in child support payments and to demonstrate that their children would not become public charges. The Court agreed with the plaintiffs, concluding that the statute was an impermissible restraint on the fundamental right to marry. The Court held that Wisconsin’s law failed strict scrutiny because it was not necessary to advance the state’s interest collecting child support, nor was it narrowly tailored to advance only that interest.

Holding that the statute violated the Equal Protection Clause, the *Zablocki* Court recognized that the denial of a marriage license was a broad restriction that, at best, provided the state with an indirect method of child support collection. The Court noted that the state’s act of denying marriage licenses to those who were delinquent in child support payments may have pressured some people to pay, but the denial did not in itself increase the amount of money Wisconsin collected. The only direct effect of Wisconsin’s law was the restriction of a fundamental right. According to the Court, the indirect pressures were unnecessary

195. *Id.* at 387.
197. *Id.* at 379.
198. *Id.*
199. *Id.* at 388. *See also* Boddie v. Connecticut, 401 U.S. 371 (1971) (holding that state laws denying divorce to married couples based on an inability to pay court costs unconstitutionally restricts the fundamental right to marry).
201. *Id.*
202. *Id.*
203. *Id.*
in part because Wisconsin already had a number of other, “at least as effective” options for encouraging payment of child support obligations, including garnishment of wages, civil contempt proceedings, and criminal penalties.204

As illustrated above, in a variety of contexts, the U.S. Supreme Court has held that fundamental rights may not be conditioned on the ability to pay money to the state. The Court’s insistence that the right to vote, the right to liberty, the right of access to the courts, and the right to marry cannot be conditioned on one’s financial resources suggests that offenders’ voting rights also cannot be conditioned on their ability to pay money to the state. The following section draws on these analogies to argue that Washington’s re-enfranchisement law, which requires offenders to pay their legal-financial obligations before regaining their voting rights, violates the Equal Protection Clause.205

V. WASHINGTON’S FELON RE-ENFRANCHISEMENT LAW VIOLATES THE EQUAL PROTECTION CLAUSE BECAUSE IT CONDITIONS A FUNDAMENTAL RIGHT ON THE PAYMENT OF MONEY TO THE STATE

Washington State’s requirement that offenders pay their legal-financial obligations before re-enfranchisement violates the Equal Protection Clause because it conditions the right to vote on offenders’ ability to pay. The U.S. Supreme Court’s decision in Ramirez permits Washington to distinguish between offenders and other citizens through felon disenfranchisement laws,206 but it does not address re-enfranchisement. Although the constitutional requirements of re-enfranchisement laws are undetermined, the Court has suggested that ordinary Equal Protection principles apply where voting rights laws discriminate on a basis other than criminal conviction.207 Washington’s re-enfranchisement law impermissibly distinguishes among offenders based on their economic circumstances, rather than permissibly distinguishing between offenders and other citizens based on their status

204. Id. at 389–90.
205. See infra Section V.B and V.C.
206. Ramirez v. Richardson, 418 U.S. 24, 54 (1974); see also supra notes 130–34 and accompanying text.
207. See Hunter v. Underwood, 471 U.S. 222, 233 (1985); see also supra notes 142–45 and accompanying text.
Felon Re-Enfranchisement

as an offender. Because Washington’s re-enfranchisement law implicates the fundamental right to vote, it must withstand strict scrutiny. Washington’s re-enfranchisement law fails strict scrutiny because it is not narrowly tailored to achieve Washington’s interests in debt collection, responsible voting, or punishment. Consequently, requiring offenders to pay their legal-financial obligations before re-enfranchisement violates the Equal Protection Clause.

A. Notwithstanding Ramirez, Traditional Equal Protection Principles Apply to Washington’s Felon Re-enfranchisement Law

Under Ramirez, Washington is permitted to disenfranchise felons upon conviction because states may constitutionally distinguish for voting rights purposes between offenders and all other citizens. However, Washington’s felon re-enfranchisement law involves distinctions among offenders not distinctions between offenders and other citizens. The U.S. Supreme Court’s opinion in Ramirez does not address whether distinctions among offenders are constitutional. Consequently, Ramirez’s holding does not apply to felon re-enfranchisement, and its sanction of disenfranchisement should not be translated into a sanction of discriminatory re-enfranchisement laws.

Although Ramirez is inapplicable, the U.S. Supreme Court’s holding in Hunter, which also deals with disenfranchisement, should be applied to felon re-enfranchisement laws. Unlike Ramirez, the issue in Hunter was not the state’s decision to disenfranchise persons convicted of crimes while granting voting rights to other citizens. Rather, the issue in Hunter was Alabama’s decision to distinguish between black and white offenders when selecting crimes that would require disenfranchisement. By applying a traditional Equal Protection analysis to Alabama’s disenfranchisement law, the Court demonstrated

208. See infra notes 176–181 and accompanying text (establishing that states cannot distinguish between offenders based on financial circumstances) and notes 128–134 and accompanying text (establishing that states can distinguish among citizens based on offender status).
210. See infra Part V.C.
211. Ramirez, 418 U.S. at 54.
212. Id.
214. Id.
that ordinary Equal Protection principles apply when states decide felon voting rights on a basis other than offender status. Therefore, the Hunter analysis applies to any felon voting rights law—including a re-enfranchisement law—that distinguishes among offenders on a basis other than their criminal conviction. Washington’s felon re-enfranchisement law distinguishes among offenders based on ability to pay. Thus, the Hunter analysis applies and Washington’s re-enfranchisement restriction must satisfy traditional Equal Protection principles.

B. Washington’s Re-Enfranchisement Process Must Pass Strict Scrutiny

Because normal Equal Protection principles apply to felon re-enfranchisement processes, courts must also determine the level of judicial review appropriate for re-enfranchisement laws. Generally courts make this determination by identifying the basis for the disparate treatment and the right at stake. Washington’s felon re-enfranchisement law separates offenders into two groups based on whether or not they have paid their legal-financial obligations. In addition, Washington’s felon re-enfranchisement laws involve a fundamental right: voting. Laws that infringe a fundamental right based on an individual’s financial circumstances require strict scrutiny. Consequently, Washington’s re-enfranchisement law must survive strict scrutiny under the Equal Protection Clause.

215. Id. at 233; see also supra notes 142–45 and accompanying text.
216. For example, in Hunter, if Alabama had purposefully made it more difficult for blacks to be re-enfranchised than whites, it is hard to imagine that the result in the case would have been any different. See supra notes 135–45 and accompanying text.
217. WASH. CONST. art. VI, § 3 and WASH. REV. CODE § 9.94A.637 (2002) (re-enfranchising only those offenders who have paid their legal financial obligations and received a sentence discharge).
1. **Washington’s Re-Enfranchisement Law Treats Offenders Differently Based Solely on Their Ability to Pay Legal-Financial Obligations.**

By requiring that offenders pay their legal-financial obligations before re-enfranchisement, Washington separates offenders into two groups—those who have sufficient financial resources to pay their legal-financial obligations and those who do not—and grants only one group the right to vote. Washington’s requirement thus results in the same unconstitutional separation of offenders based on ability to pay as the requirement struck down by the U.S. Supreme Court in *Williams*. In *Williams*, the U.S. Supreme Court held that states cannot deny indigent offenders their liberty by requiring them to remain in prison beyond the statutory maximum in order to work off their debts. The Court’s reasoning in *Williams* suggests that states cannot create opportunities for non-indigent offenders to gain access to fundamental rights like liberty when the same opportunities are necessarily out of reach for indigent offenders. Contrary to *Williams*, Washington law provides only those offenders with sufficient financial resources to pay their legal-financial obligations access to voting rights, and as a result the fundamental right to vote is necessarily out of indigent offenders’ reach.

As the *Williams* Court held, the U.S. Constitution forbids disparate treatment based on individuals’ financial resources, even if the state imposes the legal-financial obligations as part of a criminal sentence. According to the Court, a law based on payment of legal-financial obligations is discriminatory when it presents some offenders with the opportunity to regain their fundamental rights but effectively denies the same opportunity to other offenders because of their economic circumstances. In identifying the discrimination, the Court suggested that the focus should be on the effect of the state requirement rather than its source. Therefore, courts should focus on the discriminatory effect of Washington’s legal-financial obligations requirement (keeping poor...
offenders, but not rich ones, from the franchise) rather than the source of the legal financial obligation (a criminal conviction).

2. Given the Opportunity to Vote Through Washington’s Re-Enfranchisement System, an Offender’s Right to Vote Becomes Fundamental

Because wealth is not a suspect classification within the U.S. Supreme Court’s interpretation of the Equal Protection Clause, laws that visit different consequences on offenders based on economic circumstances must also restrict a fundamental right in order to receive strict scrutiny. Washington’s re-enfranchisement law satisfies this second requirement because it implicates the fundamental right to vote. Though states do not have an affirmative, constitutional duty to provide citizens with voting rights, the Court has established that once states provide citizens with the opportunity to vote, that opportunity becomes a fundamental right protected by the U.S. Constitution.

Like its grant of voting rights to all citizens, Washington does not have an affirmative, constitutional duty to give offenders the right to vote. In fact, the U.S. Supreme Court has held that Section 2 of the Fourteenth Amendment allows Washington to deny offenders the right to vote while granting it to all other citizens. Therefore, Washington may permanently disenfranchise felons—all felons—without any constitutional problems. However, once Washington grants some offenders the right to vote, it has stepped outside the protection of Section 2 because it is no longer distinguishing between citizens based on whether they have been convicted of a felony, but is distinguishing among offenders based on some other criteria. Hence, Washington’s decision to re-enfranchise offenders has the same effect as its decision to grant all citizens voting rights: an offender’s right to vote again becomes fundamental and it cannot be restricted in ways that violate the Equal Protection Clause.

232. Id.
233. Id.
Felon Re-Enfranchisement

In Johnson, the federal district court mistakenly refused to recognize an offender’s fundamental right to vote.\textsuperscript{234} According to the court, an offender’s right to vote, within the context of re-enfranchisement, is not fundamental because that right has already been, and could continue to be, constitutionally withheld by the state.\textsuperscript{235} In other words, the court suggested that felons lose all stake in the right to vote, and consequently offenders do not receive the Constitution’s voting right protections during re-enfranchisement.\textsuperscript{236} The Johnson court’s reasoning is problematic for two reasons. First, it ignores the U.S. Supreme Court’s analysis in Hunter that demonstrates that voting rights laws must survive a traditional Equal Protection analysis when they are discriminatory on a basis other than offender status.\textsuperscript{237} Second, and more importantly, the court’s holding is inappropriate because it ignores the fact that voting’s status as a fundamental rights is directly linked to—and arises from—the state’s grant of the right to vote. Therefore, while states can constitutionally take away all felons’ voting rights and never restore them, states that choose to re-grant the right to vote to offenders thereby also invoke voting’s status as a fundamental right. Thus, Washington may not establish re-enfranchisement requirements that grant the right to vote to some offenders while effectively denying it to others without satisfying the Equal Protection Clause’s strict scrutiny standard. To survive this standard, the restriction must be necessary to achieve a compelling governmental interest.\textsuperscript{238}

C. Washington’s Restriction of Offenders’ Voting Rights Based on the Failure to Pay Legal-Financial Obligations Does Not Survive Strict Scrutiny

Washington’s re-enfranchisement requirement based on the payment of legal-financial obligations commits the same harm as the statute struck

\textsuperscript{234} 214 F. Supp. 2d. 1333, 1342–43 (S.D. Fla. 2002).
\textsuperscript{235} Id. at 1343.
\textsuperscript{236} Id.
\textsuperscript{237} See supra notes 215–17 and accompanying text. For example, despite Hunter’s implication to the contrary, the Johnson court’s reasoning suggests a state could decide to re-enfranchise only offenders with last names beginning with the letters A-M without running into any constitutional problems. Faced with this hypothetical and Hunter’s holding, the Johnson court’s assertion that re-enfranchisement requirements are completely outside the scope of the Constitution’s voting rights protections seems unlikely.
\textsuperscript{238} Dunn v. Blumstein, 405 U.S. 330, 337 (1972).
down by the U.S. Supreme Court in Zablocki: it conditions a fundamental right on the payment of legal-financial obligations without advancing a compelling state interest.239 In Zablocki, the Court accepted for the purposes of argument that Wisconsin’s interest in debt collection was compelling.240 However, the compelling interest alone did not save Wisconsin’s statute from constitutional defect.241 The Court held that despite the presence of a compelling interest, the statute violated the Equal Protection Clause because the restriction was unnecessary to the advancement of that interest.242 Compelling interests may also motivate Washington’s restriction, but nevertheless the restriction fails strict scrutiny because it is not a necessary means of achieving those interests, and thus violates the Equal Protection Clause.

Although Washington’s legislative record does not reveal the state’s purpose behind conditioning the restoration of voting rights on the payment of legal-financial obligations,243 court opinions and legal commentators have suggested three possible state interests.244 First, states have an interest in collecting payment on outstanding legal-financial obligations.245 Second, states have an interest in promoting responsible use of the ballot.246 Third, states have an interest in punishing offenders for their crimes and for refusing to pay their legal-financial obligations.247 While there is surely debate about the extent to which these state interests are “compelling,” like the Zablocki Court, this Comment assumes arguendo that interests in debt collection, responsible voting, and punishment are compelling governmental interests and further that they are actual interests of Washington state. This Comment

240. Id.
241. Id.
242. Id.
243. See supra note 49 and accompanying text.
244. See supra note 50 and accompanying text.
247. In the last 100 years, states and courts have generally not advanced this interest when promoting disenfranchisement laws. Keyssar, supra note 50, at 162–63. However, in a motion for summary judgment in the case of Johnson v. Bush, the state of Florida did advance its punishment interests when explaining its re-enfranchisement restrictions. Defendant’s Memorandum in Support of Motion for Summary Judgment, Johnson v. Bush, (Jan. 4, 2002) (00-3542-CIV-King). The trial court, rejecting the plaintiff’s claim that the restitution requirement constituted an unconstitutional poll tax, did not address Florida’s interest in punishment. 214 F. Supp. 2d 1333, 1342–43 (S.D. Fla. 2002).
accepts these interests as compelling to emphasize the fact that the
central flaw in Washington’s re-enfranchisement requirement is not the
governmental interest put forward to support it. Rather, the fatal flaw in
Washington’s requirement is that the restriction is an unnecessary
infringement on the right to vote and is not narrowly tailored to achieve
compelling governmental interests. In requiring that restrictions of
fundamental rights must achieve a compelling governmental interest, the
U.S. Supreme Court has set the constitutional standard high. The Court
requires states to demonstrate that their laws not only support a
compelling interest, but also that they are essential and narrowly tailored
to the interest’s achievement.\footnote{Washington’s re-enfranchisement
restriction, analyzed in the context of its three interests, does not meet
this standard.}

First, Washington’s interest in collecting payment on outstanding
legal-financial obligations does not justify the broad infringement on
offenders’ voting rights because the restriction is not necessary to the
advancement of the state’s debt collection interests. Like the law struck
down in \textit{Zablocki}, Washington’s denial of voting rights until payment of
legal-financial obligations is, at best, an indirect method designed to
influence offenders’ decision to pay their legal-financial obligations.\footnote{For example, offenders who are denied re-enfranchisement may or may
not pay their outstanding legal-financial obligations because nothing in
Washington’s re-enfranchisement restriction compels them to pay. Yet,
as the U.S. Supreme Court indicated in \textit{Zablocki} and \textit{Williams}, state
restrictions on fundamental rights that only indirectly advance the
“collection device rationale” are unconstitutional, especially where the
state has a number of more direct collection options available.\footnote{In fact,
Washington has a number of more direct, and almost certainly more
effective, methods for collecting payment on offenders’ legal-financial
obligations. For example, Washington may enforce collection through

\textit{See} NOWAK & ROTUNDA, \textit{supra} note 44, at 639.

\textit{Zablocki}, 434 U.S. at 389–90. Discouraging regulation of voting rights as a way to achieve
unrelated state interests, the Court held that “the use of the franchise to compel compliance with
other, independent state objectives is questionable in any context.” Hill v. Stone, 421 U.S. 289, 299
(1975).

\textit{Zablocki}, 434 U.S. at 389–90; \textit{Williams}, 399 U.S. at 245.

\textit{Williams}, 399 U.S. at 245. Because the act of denying offenders’ voting rights until payment does not in itself collect
additional money, it is impossible to quantify the effect of Washington’s re-enfranchisement
restriction on the collection of legal-financial obligations.
garnishment of wages and sale of real property. And, parties other than the state can collect the legal-financial obligation of restitution through a civil action. These direct measures provide a sharp contrast to Washington’s re-enfranchisement restrictions, in which the only direct effect is the denial of the right to vote. Following the Court’s decision in Zablocki, Washington’s denial of offenders’ voting rights is not a narrowly tailored means of advancing its debt collection interests because the state’s other collection devices do not infringe on offenders’ fundamental rights.

Second, Washington’s interest in promoting the responsible use of the ballot is not advanced by its restriction of offenders’ voting rights based on non-payment of legal-financial obligations. Courts have accepted a state’s interest in responsible voting for some voting rights restrictions such as literacy tests and felon disenfranchisement. In Johnson, the federal district court also cited Florida’s responsible voting interests as justification for its re-enfranchisement restrictions, stating that states have an interest in determining felons’ “readiness to return to the electorate.” For Washington’s payment requirement to be necessary for the advancement of a responsible voting interest, the state must demonstrate that offenders who have paid their legal-financial obligations are more responsible voters, and therefore more ready to return to the electorate, than offenders who have not paid them. A successful demonstration of this sort, however, is severely limited by the U.S. Supreme Court’s opinion in Harper, which made clear that possessing the financial resources to pay a fee is not relevant to a citizen’s ability to participate intelligently in the electoral process.

253. Legal-financial obligations are enforceable in the same manner as a civil judgment, which may be enforced through the sale of real property. See id. §§ 9.94A.760(4); 4.56.190.
254. Id. § 9.94A.760(4).
257. See supra note 50 and accompanying text.
in promoting the intelligent use of the ballot. Consequently re-enfranchising only those offenders who have paid their legal-financial obligations is not a necessary and narrowly tailored means of advancing responsible voting interests.

Third, Washington’s re-enfranchisement restriction is not a narrowly tailored means of advancing the state’s punishment interest. In citing a punishment interest, Washington is likely targeting one of two types of offender misbehavior: willful failure to pay legal-financial obligations or criminal activity. As shown below, in both cases of misbehavior, Washington’s punishment interest is not advanced by its re-enfranchisement restriction.

Washington cannot use continued disenfranchisement as a way of punishing delinquent debtors because the voting rights restriction is not narrowly tailored to those offenders who are willfully refusing to pay their legal-financial obligations. Under Washington’s re-enfranchisement law, voting rights are withheld not only from offenders who are willfully refusing to pay their legal-financial obligations but also from indigent offenders who are unable to pay. However, indigency does not equal misbehavior. Poor offenders will necessarily take longer to repay their legal-financial obligations because their limited financial resources will require lower monthly payments. As a result, withholding re-enfranchisement until payment of legal-financial obligations does not punish only misbehaving offenders, but it punishes all offenders who have limited financial resources. The re-enfranchisement restriction is not a narrowly tailored means of punishing misbehaving offenders who refuse to pay their debts.

Washington’s felon re-enfranchisement requirement is also not narrowly tailored to advance its interest in punishing criminal behavior because the extent of offenders’ punishment through disenfranchisement

260. Id.

261. See generally Bearden v. Georgia, 461 U.S. 660 (1983) (holding that it is fundamentally unfair to incarcerate offenders for non-payment of legal-financial obligations without considering the willfulness of the non-payment).

262. In fact, Washington law requires that the sentencing court establish offenders’ legal-financial obligations payment schedule based on an offender’s financial resources. WASH. REV. CODE §§ 9.94A.760(1); 9.94A.760(10) (2002). Because the statute requires sentencing judges to base the payment schedule on an offender’s financial resources, poor offenders will be ordered to pay lower monthly payments than non-indigent offenders. The lower monthly payments guarantee that poor offenders will be denied voting rights for a longer period than non-indigent offenders with the same legal-financial obligation requirement. See id.
is conditioned on the irrelevant factor of their ability to pay, rather than on their criminal activity. 263 Offenders are sentenced to a number of distinct forms of punishment, of which legal-financial obligations and disenfranchisement are two mandatory and independent types. 264 However, in conditioning the restoration of voting rights on payment of legal-financial obligations, Washington blends these two distinct forms of punishment, causing the extent of offenders’ disenfranchisement to be conditioned on their ability to pay legal-financial obligations. 265 As a result, certain offenders receive the additional punishment of continued disenfranchisement based not on the severity of their crime, or any other legitimate sentencing consideration, but on their ability to pay the legal-financial obligations.

The U.S. Supreme Court’s opinion in Williams is a useful guide for understanding the constitutional flaw in this punishment arrangement. The Williams Court declared that states cannot incarcerate offenders longer than the statutory maximum just because they are poor, 266 essentially holding that states cannot impose a fine and then convert it to extended jail time if the offender does not have the means to pay the fine. 267 Williams’ principles apply to Washington’s re-enfranchisement requirement because Washington imposes legal-financial obligations, and subsequently conditions the duration of disenfranchisement on offenders’ ability to pay, in the same way that Illinois conditioned the duration offenders’ jail time on their ability to pay imposed fines. Accordingly, Washington’s re-enfranchisement restriction, which reserves extra punishment for poor offenders, is not narrowly tailored to advance Washington’s interest in punishing all criminal behavior.

In sum, requiring payment of legal-financial obligations before re-enfranchisement is a broad infringement on the fundamental right to vote. The re-enfranchisement restriction is not justified by Washington’s three compelling interests because, like Zablocki, the infringement of the fundamental right is not necessary or narrowly tailored to the

263. See Harper, 383 U.S. at 668 (holding that ability to pay is an irrelevant factor in voting).
264. The sentencing court must assess at least $500 in legal-financial obligations and it must revoke the offender’s right to vote. See supra Part I.A.; Part I.B.
265. Cf. Williams v. Illinois, 399 U.S. 235, 243 (1970) (holding that Illinois could not blend distinct forms of punishment—imprisonment and fines—by requiring poor offenders to remain in prison until the fine was paid.).
266. Williams, 399 U.S. at 243.
267. In Tate v. Short, the Court endorsed this reading of Williams’ holding. 401 U.S. 395, 398 (1971).
governmental objectives. Therefore, Washington’s re-enfranchisement law fails strict scrutiny. To comply with Equal Protection requirements, Washington should eliminate the requirement that offenders pay their legal-financial obligations before regaining voting rights.

VI. CONCLUSION

Washington’s re-enfranchisement law treats offenders differently—granting some, but not others, the right to vote—based solely on their different abilities to pay legal-financial obligations. What is more, voting is the only fundamental right Washington restricts in this way. The disparate treatment inherent in Washington’s re-enfranchisement requirement is particularly troubling given the U.S. Supreme Court’s directive that affluence or the payment of money cannot be a condition of voting. Washington’s re-enfranchisement provisions must be held to the same Equal Protection standard as laws enfranchising all other citizens: restrictions are only permissible if narrowly tailored to achieve a compelling governmental interest. Washington’s re-enfranchisement law does not meet this strict scrutiny standard because the infringement of offenders’ voting rights is unnecessary to advance Washington’s compelling governmental interests. Consequently, it violates the Equal Protection Clause, and Washington should eliminate its requirement that offenders pay their legal-financial obligations before regaining voting rights.
WATER, PROPERTY, AND THE CLEAN WATER ACT

Janis Snoey

Abstract: In PUD No. 1 of Pend Oreille County v. Department of Ecology, the Supreme Court of Washington held that Washington State has authority under the Clean Water Act to impose a minimum stream flow requirement on a hydroelectric project seeking to amend its federal license, regardless of whether the flow requirement affects an existing water right. A water right is property protected by the U.S. Constitution’s prohibition on taking without just compensation. If a state’s imposition of a minimum flow requirement under the Clean Water Act restrains a project from diverting the full quantity of an existing water right, the water right holder could challenge the state’s action as an unconstitutional taking. This Comment argues that courts should not analyze the constitutionality of the state action under a physical invasion or exaction analysis. Instead, courts should employ a regulatory taking analysis because the state is acting in a regulatory capacity and the regulatory takings test allows for consideration of both the complex interests inherent in a water right and the high degree of control a state exercises over the right.

Water quantity affects water quality.1 The United States Supreme Court acknowledged this relationship in PUD No. 1 of Jefferson County v. Department of Ecology [hereinafter Elkhorn II] in 1994.2 In Elkhorn II, the Court held that a state has authority under the Clean Water Act to impose a minimum stream flow requirement on a hydroelectric project seeking a federal license.3 Eight years later, the Supreme Court of Washington examined the meaning of Elkhorn II in Department of Ecology v. PUD No. 1 of Pend Oreille County4 [hereinafter Sullivan Creek]. Although Elkhorn II and Sullivan Creek involve rivers on opposite sides of Washington State, the cases share similar facts. In both cases, the public utility district proposed a hydroelectric power project intended to divert water and discharge it far downstream.5 Both projects imperiled fish as a result of inadequate volumes of water in the bypass reach,6 the section of river that would be bypassed by the diversion. In both cases, Washington State, in reviewing the project under the Clean

1. PUD of Jefferson County v. Dep’t of Ecology, 511 U.S. 700, 719 (1994) [hereinafter Elkhorn II]. In water law, cases concerning hydroelectric projects are traditionally named for either the body of water or the dam at issue.
2. Id.
3. Id. at 723.
4. 146 Wash. 2d 778, 51 P.3d 744 (2002) [hereinafter Sullivan Creek].
5. Elkhorn II, 511 U.S. at 708–09; Sullivan Creek, 146 Wash. 2d at 786–87, 51 P.3d at 748–49.
6. Elkhorn II, 511 U.S. at 709; Sullivan Creek, 146 Wash. 2d at 787, 51 P.3d at 749.
Water Act (CWA)\(^7\) agreed to the project on the condition that the operator limit its proposed diversion to allow enough water in the bypass reach to support fish.\(^8\) What is the difference? The public utilities district in *Elkhorn II* had not secured a right to divert water\(^9\) while the district in *Sullivan Creek* had an existing water right.\(^10\) The *Sullivan Creek* court held that the CWA does not prohibit the state from requiring the operator to maintain minimum stream flow, regardless of the effect on the applicant’s existing water right.\(^11\)

If the public utility district in *Sullivan Creek* builds the power project as currently designed and meets the minimum flow requirements insisted upon by the state, the district cannot withdraw the entire year-round quantity allowed under its water right.\(^12\) Some would see this as a government attack on a valuable property right.\(^13\) Others would view it as a reasonable imposition on a user of a public resource.\(^14\) Both views raise the issue of whether Washington State can order this without invoking the Fifth Amendment requirement of government compensation for the taking of property.

This Comment argues that the determination of whether Washington State’s action is constitutional must be the result of a fact-intensive regulatory taking analysis in which the holder’s loss is balanced against the state’s interest in promoting the general welfare. Part I describes the authority of a state to impose minimum flow requirements under the CWA and thereby affect an existing water right. Part II summarizes the relevant U.S. Supreme Court takings jurisprudence and identifies particular problems in evaluating diminution of a water right as a Fifth Amendment taking. Part III argues that if a water right holder challenges

---

8. *Elkhorn II*, 511 U.S. at 709; *Sullivan Creek*, 146 Wash. 2d at 787, 51 P.3d at 749.
11. *Id*.
12. See Petitioner’s Opening Brief at 5–17, *Sullivan Creek*, 146 Wash. 2d 778, 51 P.3d 744 (2002) (No. 70272-8). The briefing for *Sullivan Creek* does not describe resulting harm to the proposed project; the author presumes the diminished diversion would reduce the hydraulic head or power generation capacity. The public utility district did not plead an unconstitutional taking at the trial court level.
this state action as an unconstitutional taking, the court should analyze it as a regulatory taking.\textsuperscript{15}

I. STATES HAVE AUTHORITY UNDER THE CLEAN WATER ACT TO IMPOSE MINIMUM FLOW REQUIREMENTS

Congress enacted the CWA to control water pollution throughout the nation, but structured the legislation to encourage state participation.\textsuperscript{16} A party invokes the CWA by seeking a federal permit for an activity that results in discharge to the nation’s waters.\textsuperscript{17} The state becomes involved because the project proponent must ask the state to certify that the proposed project meets state water quality standards before a federal agency can issue the permit.\textsuperscript{18} In \textit{Elkhorn II}, the U.S. Supreme Court held that a state has authority under the CWA to require minimum stream flow as a condition to certification.\textsuperscript{19} In \textit{Sullivan Creek}, the Supreme Court of Washington held that the authority to require minimum stream flow exists even when an existing water right may be affected.\textsuperscript{20}

A. The Clean Water Act Applied to Activities in States

The CWA is a water pollution control law intended to restore and maintain the quality of national water resources.\textsuperscript{21} The CWA’s jurisdiction encompasses “the waters of the United States, including territorial seas,” otherwise known as “navigable waters.”\textsuperscript{22} Courts interpret this to include non-navigable waters, such as wetlands, when they are associated with the navigable watercourses because the water quality of navigable waters and non-navigable water are interdependent.\textsuperscript{23} The CWA prohibits discharge of toxic wastes,

\begin{itemize}
  \item[15.] This Comment addresses the right to divert water from surface watercourses, but does not address groundwater diversion.
  \item[16.] See infra Part I.A.
  \item[18.] Id.
  \item[19.] Elkhorn II, 511 U.S. 700, 723 (1994).
  \item[20.] Sullivan Creek, 146 Wash. 2d 778, 784, 51 P.3d 744, 747 (Wash. 2002).
  \item[22.] Id. § 1362(7).
  \item[23.] See 33 C.F.R. § 328.3(a)(3) (2002).
  \item[24.] The U.S. Supreme Court considered the CWA’s jurisdiction in \textit{Solid Waste Agency of Northern Cook County v. United States Army Corps of Engineers}, 531 U.S. 159 (2001). The Court confirmed that CWA jurisdiction over non-navigable waters associated with navigable waters is valid. See id. at 167. However, the Court held the Corps’ Migratory Bird Rule (which viewed any
promotes research and planning, and provides for federal financial assistance.\textsuperscript{25}

The CWA requires the Environmental Protection Agency to establish minimum national water quality standards.\textsuperscript{26} States may establish their own standards, but only if such standards are the same or stricter than federal standards.\textsuperscript{27} In addition, the CWA requires “[a]ny applicant for a federal license or permit to conduct any activity . . . which may result in any discharge into [] navigable waters” to obtain certification from the state that the activity will meet the state’s water quality standards.\textsuperscript{28} This is known as § 401 certification.\textsuperscript{29} No federal agency may issue a license without state certification or state waiver of certification authority.\textsuperscript{30}

Federal permits requiring § 401 certification cover a wide variety of projects. They include permits for point source discharges,\textsuperscript{31} discharges of dredged and fill material,\textsuperscript{32} activities affecting navigation under the Rivers and Harbors Act,\textsuperscript{33} and licenses for hydroelectric projects issued by the Federal Energy Regulatory Commission (FERC).\textsuperscript{34} A § 401 requirement to maintain minimum flow would most likely affect facilities using a water diversion such as a hydroelectric power plant or a cooling tower for an industrial facility.

A public utility district seeking to build a dam provides an example of a state’s authority in the federal permitting process. First, the district must seek a license from FERC.\textsuperscript{35} FERC requires the district to obtain § 401 certification from the state because the water expelled from the watercourse used by migratory birds to be under CWA jurisdiction) impermissibly extended jurisdiction to unassociated non-navigable waters. \textit{Id.} at 174.

\textsuperscript{25} 33 U.S.C. § 1251. This Comment focuses on the CWA’s function in the issuance of federal permits.

\textsuperscript{26} \textit{Id.} § 1313.

\textsuperscript{27} \textit{Id.}

\textsuperscript{28} \textit{Id.} § 1341(a)(1).


\textsuperscript{30} 33 U.S.C. § 1341(a)(1).

\textsuperscript{31} Debra L. Donahue, \textit{The Untapped Power of Clean Water Act Section 401}, 23 ECOLOGY L.Q. 201, 219 (citing the OFFICE OF WATER, EPA, WETLANDS AND 401 CERTIFICATION—OPPORTUNITIES AND GUIDELINES FOR STATES AND ELIGIBLE INDIAN TRIBES 10 (Apr. 1989)).

\textsuperscript{32} \textit{Id.}

\textsuperscript{33} \textit{Id.}

\textsuperscript{34} \textit{Id.}

Water, Property, and the Clean Water Act

dam is a discharge to the nation’s waters. Second, if dam construction results in discharge of fill into the water, the district also must seek a permit from the U.S. Army Corps of Engineers (Corps). Like FERC, the Corps will require the district to obtain § 401 certification from the state. If the state finds the project violates state water quality standards under either the FERC license or the Corps permit, then the state has the option of refusing to certify the project—in which case the federal agency would deny the permit—or certifying the project if the applicant fulfills certain conditions (such as maintaining minimum stream flow) to assure the activity meets state water quality standards.

B. Congressional Intent Behind the Clean Water Act

In 1994, the U.S. Supreme Court examined the authority of a state to require minimum stream flow under the CWA in Elkhorn II. In Elkhorn II, the Public Utility District of Jefferson County (PUD) sought to construct the Elkhorn Hydroelectric Project on the Dosewallips River in Washington State. As required by its application for a FERC license, the PUD sought state § 401 certification that the project met state water quality standards. As planned, the Elkhorn project would have diverted water from a 1.2 mile reach of the Dosewallips River and reduced water flow in the bypass reach to less than the minimum necessary to protect salmon and steelhead populations. The State Department of Ecology (Ecology) issued certification with the condition that the PUD maintain minimum stream flow in the bypass reach. When the Supreme Court of Washington upheld Ecology’s authority to apply such conditions, the PUD appealed to the U.S. Supreme Court.

38. Id. at § 1341(a)(1).
39. Id. §§ 1341(a), 1341(d).
41. Id. at 709.
42. Id.
43. Id.
44. PUD No. 1 of Jefferson County v. Dep’t of Ecology, 121 Wash. 2d 179, 189, 849 P.2d 646, 651 (1993) [hereinafter Elkhorn I].
45. Elkhorn II, 511 U.S. at 710.
In *Elkhorn II*, the Court recognized that Congress intended states to have considerable authority under § 401 of the CWA and thus held that a state has power to require an applicant for a federal permit to maintain minimum water flow in compliance with state law as a condition of § 401 certification. The Court agreed that a minimum flow requirement is a valid water quality standard. Further, the Court held that it was reasonable for the state to find that reducing stream flow below the level necessary for sustaining fish populations would violate the state’s water quality policy.

The *Elkhorn II* Court rejected the notion that the CWA exempted water quantity from water quality standards. Writing for the majority, Justice O’Connor observed that the CWA “preserve[s] the authority of each State to allocate water quantity as between users; [but does] not limit the scope of water pollution control that may be imposed on users who have obtained, pursuant to state law, a water allocation.” The Court arrived at this conclusion by examining the legislative history behind the Wallop Amendment to the CWA. This provision, adopted in 1977, affirms that each state, not the federal government, has authority over water allocation. The Court determined that although Congress intended to prevent subversion of state allocation systems, it also acknowledged that effects on individual rights are permissible if such effects arise from valid concerns for water quality. Further, the Court clarified that § 401 certification does not determine the holder’s proprietary interest in the water right, it “merely determines the nature of the use to which that proprietary right may be put under the [CWA].”

---

46. *Id.* at 723.
47. *Id.* at 719.
49. *See id.* at 720.
50. Justice Stevens concurred in the result. *Id.* at 723. Justice Thomas wrote a dissent in which Justice Scalia joined. *Id.* at 724.
51. *Id.* at 720.
52. The amendment is named for Senator Wallop of Wyoming who, along with Senator Hart of Colorado, introduced it during the 1997 Senate negotiations to amend the CWA. *See Gregory J. Hobbs, Jr. & Bennett W. Raley, Water Rights Protection in Water Quality Law, 60 U. COLO. L. REV. 841, 852–56 (1989).*
55. *See Elkhorn II*, 511 U.S. at 721.
56. *Id.*
In *Elkhorn II*, the Court recognized Congress’ intent that states exercise substantial authority under the CWA and thus held that a state can apply conditions to a project to protect water quality even though water allocation may be affected. After *Elkhorn II*, commentators questioned whether a state could condition a project using an existing water right because the PUD in *Elkhorn II* did not have an existing right to divert water. Further, the Court was ambiguous as to whether the fact that a new right was at issue was central to the holding. While the Court confirmed that a requirement for minimum stream flow does not correspond to a proprietary water right, the Court also stated that the PUD had yet to obtain a water right.

The Supreme Court of Washington resolved the debate for Washington State in *Sullivan Creek*. The case concerned a hydroelectric project that had been constructed in 1907, but was decommissioned in 1956. Since that time, the Public Utility District No. 1 of Pend Oreille County (the District) had operated the project under a FERC nongenerating license. This allowed the District to release water from the project’s reservoir so that it could flow in the natural streambed to benefit other hydroelectric projects far downstream. In 1992, the District proposed to reestablish power generation by diverting water under the District’s existing water right and sending it through a pipeline to a powerhouse approximately three miles downstream from the diversion. In the process of amending its FERC license to allow power generation, the District sought § 401 water quality certification from the state as required by the CWA. Ecology found that the project’s proposed reduction of flow in the bypass reach violated the state’s water quality standards because it endangered fish habitat and diminished a

57. *Id.*
60. *Id.* at 720–21.
61. *See id.* at 721.
63. *Id.* at 784–87, 51 P.3d at 747–49.
64. *Id.* at 778, 51 P.3d at 747.
65. *Id.*
67. *Id.*
significant recreation and aesthetic resource. Ecology issued the certification with the condition that the District maintain minimum stream flow in the bypass reach. Consequently, the District could not withdraw the same year-round quantity of water allowed under its existing water right. The Supreme Court of Washington granted a motion for direct review upon appeal from an administrative hearing affirming Ecology’s authority to impose the minimum flow requirement.

The Sullivan Creek court determined that federal law did not preclude Ecology from requiring the hydroelectric power project to maintain minimum stream flow, regardless of the effect on the applicant’s existing water right. Justice Madsen, writing for an eight to one majority, first declared that under Elkhorn II a state can impose a minimum stream flow requirement. The court rejected the District’s argument that Elkhorn II did not apply to the facts in Sullivan Creek because, unlike the project in Sullivan Creek, Elkhorn II concerned a project without an existing water right. The District argued that the Wallop Amendment precluded a state from imposing a condition that might affect an applicant’s existing water right. The court disproved this claim by examining the legislative history of the Wallop Amendment. The Supreme Court of Washington concluded that the Wallop amendment was a legislative compromise to limit the CWA’s jurisdictional reach, pointing to Senator Wallop’s remarks that the purpose of the amendment was to assure that the regulatory authorities used the CWA to protect water quality and not to

68. See id.
69. Id. at 787, 51 P.3d at 748–49.
70. See Petitioner’s Opening Brief at 13, Sullivan Creek, 146 Wash. 2d 778, 51 P.3d 744 (2002) (No. 70272-8).
71. Sullivan Creek, 146 Wash. 2d at 789, 51 P.3d at 749–50. In Sullivan Creek, the Supreme Court of Washington also ruled on change in point of diversion, water right relinquishment, and water right abandonment law. Id. at 784, 51 P.3d at 747. This Comment does not address issues other than a state’s authority under the CWA.
72. Id. at 784, 51 P.3d at 747. The court also disagreed that state law prohibited Ecology from imposing minimum stream flow requirements. Id. at 818, 51 P.3d at 764 .
73. Justice Sanders dissented from the majority opinion. Id. at 822, 51 P.3d at 776 (Sanders, J., dissenting).
74. Id. at 811, 51 P.3d at 761.
75. Id. at 812, 51 P.3d at 761.
76. Id.
77. Id. at 813–15, 51 P.3d at 762–63.
78. Id. at 814, 51 P.3d at 762 (citing Hobbs, Jr. & Raley, supra note 52).
Water, Property, and the Clean Water Act

assert federal control over water allocation. Based on this history, the Supreme Court of Washington reasoned that a condition imposed by the state affecting water quantity but intended to protect water quality was a legitimate state action under the CWA. Further, the court acknowledged federal court decisions reasoning that the Wallop amendment does not immunize a water right holder against legitimate water quality regulation. Thus, under Sullivan Creek, Ecology can impose minimum instream flow requirements in Washington State regardless of whether it affects an existing water right.

When Congress enacted the CWA, it intended that states be involved in the effort to control water pollution. The CWA assures state involvement by requiring that an applicant for a federal permit seek certification from the state that the proposed project meets state water quality standards. In Elkhorn II, the U.S. Supreme Court held that a state’s certification authority includes the power to require minimum stream flow. The Supreme Court of Washington held in Sullivan Creek that the state has this power regardless of whether it affects an existing water right. However, neither the U.S. Supreme Court nor the Supreme Court of Washington decided whether this is an unconstitutional taking of a property right.

II. FEDERAL TAKINGS JURISPRUDENCE AND WATER RIGHTS

When a state acting under the CWA requires an existing water right holder to maintain minimum stream flow, the state is potentially preventing the holder from withdrawing the full quantity of water allowed by the right. This gives rise to the question of whether the holder is due compensation. Most states recognize a water right as a

79. Id. (citing 123 CONG. REC. 39, 211–12 (1977) (statement of Senator Wallop)).
80. See id. at 814–16, 51 P.3d at 762–63.
81. Id. at 816–17, 51 P.3d at 763–64. The federal cases were Riverside Irrigation District v. Andrews, 758 F.2d 508 (10th Cir. 1985) and United States v. Akers, 785 F.2d 814 (9th Cir. 1986).
82. See supra Part I.A.
85. Sullivan Creek, 146 Wash. 2d at 784, 51 P.3d at 747.
86. See id. at 787, 51 P.3d at 749.
87. See infra Part II.A.
constitutionally protected form of property. Federal takings jurisprudence provides guidance on how courts should analyze a water right takings challenge in the context of state action under the CWA. However, this guidance has limited utility because a water right is a form of property significantly different from the property considered in federal takings cases. Nonetheless, at least one federal court has applied federal takings law to a claim involving minimum stream flow and an existing water right.

A. Federal Takings Jurisprudence

Most states recognize a water right as a form of property deserving constitutional protection against governmental taking without just compensation. Sovereign governments, such as the United States and each individual state, possess the power of eminent domain, allowing each to take privately owned property upon payment of compensation. The Fifth Amendment of the U.S. Constitution guarantees just compensation when the government takes private property for public use. Many state constitutions have similar or more protective provisions. Courts construe the Fifth Amendment’s prohibition on taking to be more than just physical seizure of property. For example, a taking also can occur when a government’s regulatory power impairs the use or value of private property.

The U.S. Supreme Court recognizes three basic situations in which a governmental taking can occur. The first, known as “physical invasion”,

---

89. See infra Part II.A.
90. See infra Part II.B.
91. See infra Part II.C.
92. See supra note 88.
94. Id.
95. U.S. CONST. amend. V.
97. NOWAK & ROTUNDA, supra note 93, § 11.12 at 472.
98. Id.
occurs when the government physically occupies private property. The second, known as “exaction,” occurs when the government demands private property for public use in exchange for the grant of a development permit. Finally, the Court has recognized “regulatory takings” in which the government neither occupies private property nor exacts it for public uses, but instead regulates the property in a way that diminishes its value.

A physical invasion results in a compensable taking to the extent of the government’s occupation of the property, regardless of whether the occupation serves the public interest or causes only minor economic harm to the property owner. Thus, a determination of physical invasion results in a per se taking, affording a very high level of protection to a property interest. The U.S. Supreme Court enunciated the rule of physical invasion in *Loretto v. Teleprompter Manhattan CATV Corp.*, where the government compelled property owners to allow cable television equipment to be installed on their building. Although the government was not physically present on the private property, the Court characterized the government action as occupation and held that the owners had suffered a taking of property for which compensation was due.

In contrast, when a government requires a property owner to dedicate private property for public use in exchange for development approvals—an exaction—the fact that the government action can be characterized as occupation does not result in an unconstitutional per se taking. Instead, the determination of whether the government has caused a compensable taking depends on the outcome of two tests. The first test asks whether there is a “nexus” or relationship between the interests a government is seeking to advance and the condition exacted by the government. If

---

103. 458 U.S. 419 (1982).
104. *Id.* at 438.
106. *Id.*
108. *Id.*
109. *Id.* at 386 (citing *Nollan v. Cal. Coastal Comm’n*, 483 U.S. 825, 837 (1987)).
there is no nexus, then the government has unconstitutionally taken property. If there is a nexus, the second, “rough proportionality”, test evaluates whether the nature and extent of the exaction is proportionate to the impact created by the proposed development. If the dedication of private property appears to be roughly proportionate to the impact created by the proposed development, then the government’s action is not an unconstitutional taking.

The exaction doctrine recognizes that the government has a legitimate interest in regulating land use through its development approval process. Nonetheless, the doctrine also recognizes that a landowner should not be forced to give up property without compensation when a proposed development does not cause the impact the government is seeking to mitigate. In *Dolan v. City of Tigard*, the U.S. Supreme Court considered whether an unconstitutional taking occurred when a city exacted a portion of a business owner’s property for a bicycle trail when the owner applied for a permit to expand her business. The Court determined that a nexus existed because there was a relationship between the city’s interest in reducing traffic congestion and the proposed bicycle trail. However, the city failed to satisfy the rough proportionality test because it did not show how enlarging the business would generate additional traffic necessitating the trail. Accordingly, the Court reversed the Supreme Court of Oregon’s decision that the city’s act was valid and remanded the case for further proceedings. Subsequently, the U.S. Supreme Court suggested that the exaction doctrine’s rough proportionality test might be confined to circumstances where land use approvals are conditioned on the dedication of private land to public use.

110. *Id.*
111. *Id.* at 391.
112. *Id.* at 395. The Court required no “precise mathematical calculation,” only some level of quantification. *Id.*
113. *Id.* at 384–85.
114. *Id.* at 385.
116. *Id.* at 377.
117. *Id.* at 395.
118. *Id.*
119. *Id.* at 396.
120. City of Monterey v. Del Monte Dunes, Ltd., 526 U.S. 687, 702–03 (1999) (refusing to extend the rough proportionality test to a circumstance where a city had repeatedly denied a permit).
The third circumstance in which a property owner may claim a compensable taking is when the taking is alleged to result from government regulation. In this circumstance, compensation may be due depending on whether the regulation deprives the owner of all economic value, and if not, on a balancing of factors such as the economic hardship and the character of the government action. The regulatory takings doctrine recognizes government authority to enforce regulations advancing legitimate state interests such as the protection of health, safety, and general welfare, and the enhancement of quality of life. A regulation may substantially diminish the value of a property without resulting in a compensable taking. However, if a regulation deprives the property owner of all economic use of the property, it is a per se taking requiring compensation.

In *Lucas v. South Carolina Coastal Council*, the Court considered whether a state caused a compensable taking when it imposed regulations limiting development on two seashore properties. The Court reasoned that property owners expect new regulations to restrict property uses, but eliminating all economically valuable use is inconsistent with the Fifth Amendment. The Court reasoned that a total loss of economic value is constitutional only when background principles of state law present when the owner took title to the property would have allowed it. The Court accepted the trial court’s finding that newly imposed regulations rendered the seashore properties valueless and remanded the case to determine if background principles of nuisance and property law would have reached the same result.

125. *See id.* at 131 (citing *Euclid v. Ambler Realty Co.*, 272 U.S. 365 (1926) (75% diminution in value) and *Hadacheck v. Sebastian*, 239 U.S. 394 (1915) (87.5% reduction in value)).
127. *Id.* at 1003.
128. *Id.* at 1007.
130. *See id.* at 1029 (providing the example of a lakebed owner who would be prohibited from filling if the effect is to flood the land of others).
131. *Id.* at 1009.
132. *See id.* at 1031–32.
Subsequently, in *Palazzolo v. Rhode Island*, the Court made clear that showing that a regulation existed at the time an owner acquires property does not bar the owner from bringing a takings challenge; a court must still examine the reasonableness, and thus the constitutionality, of the regulation. In *Palazzolo*, the Court held that a state prohibition on filling wetlands was not a per se taking because Palazzolo, a residential landowner, retained some economic value in an entire parcel. However, the Court remanded the case for further consideration, noting that even where a regulation does not result in a complete taking, a court must analyze government action to determine if an unconstitutional taking has nonetheless occurred.

The Court recently considered a regulatory takings challenge in *Tahoe-Sierra Preservation Council, Inc. v. Tahoe Regional Planning Agency*. In *Tahoe-Sierra*, the Court addressed whether a development moratorium constituted a per se taking. Before deciding that it did not, the Court summarized several aspects of its takings jurisprudence. Justice Stevens, writing for the majority, characterized a physical taking as a “rare, easily identified” circumstance in which the government seizes property for its own use. In contrast, a regulatory takings challenge results from a government program intended to benefit the “common good.” The Court emphasized that the property as a whole must be considered and discouraged isolating discrete rights in property for purposes of a takings analysis; the “destruction of one strand” in a “bundle of property rights” does not constitute a taking. In addition, the Court confirmed its preference for “ad hoc, factual

---

134. *Id.* at 629–30.
135. *Id.* at 630.
136. *Id.* at 632.
137. 122 S. Ct. 1465 (2002).
138. *Id.* at 1470.
139. *Id.* at 1486.
140. Six justices joined in the majority. *Id.* at 1470. Chief Justice Rehnquist wrote a dissent in which Justices Thomas and Scalia joined. *Id.* Justice Thomas also filed a dissenting opinion in which Justice Scalia joined. *Id.*
141. *Id.* at 1479–80.
142. *Id.* at 1480.
143. *Id.* at 1481.
144. *Id.* (citing Andrus v. Allard, 444 U.S. 51, 65–66 (1979) (internal quotations omitted)).
inquiries” as opposed to the adoption of per se rules where partial regulatory takings are at issue.\footnote{145. \textit{Id.} (citing Lucas v. S.C. Coastal Council, 505 U.S. 1003, 1015 (1992)).}

In both \textit{Tahoe-Sierra}\footnote{146. \textit{See id.} at 1481, 1483–84.} and \textit{Palazzolo},\footnote{147. \textit{Palazzolo v. Rhode Island}, 533 U.S. 606, 617–18 (2001).} the Court indicated that when a regulation does not result in a complete taking, courts should analyze a takings challenge by balancing a number of factors first described in \textit{Penn Central Transportation Co. v. New York City}.\footnote{148. 438 U.S. 104 (1978).} In \textit{Penn Central}, New York City’s landmark preservation ordinance prevented developers from constructing an office tower above a train station.\footnote{149. \textit{Id.} at 115–17.} The developers claimed that the regulation deprived them of their property interest in the air space above their building.\footnote{150. \textit{Id.} at 130.} The Court held that New York had not effected an unconstitutional taking.\footnote{151. \textit{Id.} at 138.} The Court arrived at this conclusion by balancing (1) the economic impact on the property owner; (2) the extent of the regulation’s interference with “distinct investment-backed expectation;” and (3) “the character of the governmental action.”\footnote{152. \textit{Id.} at 124.} To clarify the last factor, Justice Brennan, writing for the majority,\footnote{153. Justice Rehnquist filed a dissenting opinion in which Chief Justice Burger and Justice Stevens joined. \textit{Id.} at 106.} noted that an unconstitutional taking “may more readily be found when the [government’s] interference with property can be characterized as a physical invasion . . . than when interference arises from some public program adjusting the benefits and burdens of economic life to promote the common good.”\footnote{154. \textit{Id.} at 124–25.} The Court’s analysis stressed that the government has the authority to employ “laws or programs that adversely affect recognized economic values” in order to promote the government’s legitimate interest in the public’s welfare.\footnote{155. \textit{Id.} at 147 (Rehnquist, J., dissenting).} Although the \textit{Penn Central} landowner claimed a multimillion dollar loss,\footnote{156. \textit{Id.} at 124.} the Court reasoned that the regulation was valid because it promoted general welfare and gave the landowner reasonable use of the
property.\footnote{157} Thus, the Court held that the preservation ordinance did not effect a regulatory taking.

\textbf{B. The Difficulty of Applying Federal Takings Jurisprudence to a Water Right}

The primary difficulty in applying federal takings jurisprudence to a water right is that the right does not fit neatly among the other forms of property considered by courts in the taking cases.\footnote{158} This creates a challenge because courts find more or less protection depending on the type of property,\footnote{159} the degree of control traditionally exerted over the property,\footnote{160} and whether the government is occupying or regulating the property.\footnote{161} The last factor is particularly important because courts afford the greatest level of protection where the government permanently and physically occupies property.\footnote{162} Physical invasion, no matter how little space is occupied, is per se taking requiring compensation.\footnote{163} In contrast, if the takings challenge arises from government regulation, courts may not find an unconstitutional taking even though the government action significantly reduces the value of the property.\footnote{164} Commentators considering takings jurisprudence in the context of water rights usually assume that courts will apply a regulatory taking analysis; but those who have suggested a physical invasion analysis admit doubt about its

\footnotetext[157]{Id. at 138.}


\footnotetext[159]{See JAN G. LAITOS, LAW OF PROPERTY RIGHTS PROTECTION, LIMITATIONS ON GOVERNMENTAL POWERS § 9.03[B], 9-15 (2002 Supp.) (contending that land generally receives the highest level of protection). But see Armstrong v. United States, 364 U.S. 40, 48–49 (1960) (holding total destruction of materialmen’s liens to be a taking).}

\footnotetext[160]{Lucas v. S.C. Coastal Council, 505 U.S. 1003, 1027–28 (1992).}


\footnotetext[163]{Lucas, 505 U.S. at 1015 (describing how the physical invasion in Loretto occupied only 1.5 cubic feet of property). Exactions are an exception to the per se rule when the government occupies the property. See supra notes 108–19 and accompanying text.}

\footnotetext[164]{See supra note 125.}
Water, Property, and the Clean Water Act

applicability. The correct analysis is unclear given the unique characteristics of a water right.

First, a water right is a usufructuary right, meaning that the holder does not own the water but possesses the right to use it. The water belongs to the public or to the state. The public’s ownership of the water may be subject to appropriation or the state may have a paramount right to use waters for a public purpose. However, an appropriation is still subject to the public interest in the water. For example, once water is withdrawn from the watercourse and flows in a private irrigation ditch, it becomes personal property in some states. Nonetheless, the right holder can use the water only once and then must return unconsumed portions to the public and other appropriators.

Second, a water right is unique because of the extent its entitlements and duties vary from state to state. The United States has two primary legal regimes for allocating water rights. The first is riparianism, in which the right to use water attaches to the land next to a watercourse.


166. 1 WELLS A. HUTCHINS, WATER RIGHTS LAWS IN NINETEEN WESTERN STATES 151 (1971).


168. Anderson, supra note 167, at 12-48–12-55. As used here, appropriation is the term for the claim an individual may legitimately make to withdraw water in order to apply it to a beneficial use.


171. See 1 GEORGE VRAKESH, COLORADO WATER LAW § 3.1, at 121 (1987).

172. Id.

173. Beck, supra note 170, at 4-1. Many states have adopted permit systems.

174. 78 AM. JUR. 2D Waters § 49 (2002). “Riparian rights” includes such activities as building wharves and swimming in water. Joseph W. Dellapenna, Introduction to Riparian Rights, in 1 WATER AND WATER RIGHTS § 6.01(a), at 6-3–6-4 (Robert E. Beck ed., 2001). This Comment’s reference to riparian rights is limited to consumptive use.
The second is prior appropriation, which allows for water to be severed from the land so it can be used at a great distance from the watercourse.175 The “prior” in prior appropriation comes from the fact that earlier established rights are “senior” and later established rights are “junior.”176 In times of drought, junior rights give way to senior rights.177 Twenty-nine states have riparian systems and all of them are in the eastern United States.178 Nine states have prior appropriation systems179 and the remaining twelve states use hybrid systems.180

A third distinctive characteristic of a water right is that it is not a precise amount of water, but a bundle of entitlements and duties.181 Under riparianism, a riparian landowner may divert water for some, but not all, consumptive uses.182 The riparian landowner’s use must be “reasonable;”183 meaning that it does not unreasonably interfere with another riparian landowner’s water use.184 In contrast, under prior appropriation, a water right holder’s essential entitlements are the priority of the right relative to others, and the right to apply water to a particular “beneficial use.”185 Beneficial use, rather than water quantity, is “the basis, the measure and the limit of the right to the use of water.”186 If an appropriator does not use the entire amount of the water right to achieve the beneficial use for which the water was appropriated, then the appropriator must make the excess water available for appropriation by

177. Id.
179. Id. at 7. (Ala., Ariz., Colo., Id., Mont., Nev., N.M., Utah, and Wyo.).
180. In hybrid states, state law recognizes both, or parts of both, the prior appropriation and riparian systems. Id. at 8 (Cal., Kan., Miss., Neb., N.D., Okla., Or., S.D., Tex., Wash.). Louisiana and Hawaii each have unique systems based on laws established before each state’s entry into the United States. Id.
181. See Dellapenna, supra note 174, at §§ 6-1–6-2; Hutchins, supra note 166, at 8–19.
182. See Dellapenna, supra note 174, at § 6.01(a)(4), at 6–49–6-50.
185. Laitos, supra note 158, at 906–07; Vranesh, supra note 171, § 3.1, at 121; Hutchins, supra note 166, at 9.
186. N. M. Const. art. XVI, § 3; Hutchins, supra note 166, at 9; Bell & Johnson, supra note 88, at 5.
The concept of beneficial use ensures that appropriators use the public resource to its maximum potential. Thus, no one can “waste” the water.

Finally, all states exercise a high degree of control over water rights without implicating a compensable taking. For example, all states control the transfer of a water right to some extent. Some riparian states allow riparians to convey rights away from riparian lands while others do not. In contrast, prior appropriation states allow a water right to be sold separately from land, unless the transfer causes harm to another right. Further, some states expressly prohibit the transfer of a water right if it harms the public interest. Riparian rights are correlative; consequently, the amount used by one riparian may shrink or grow relative to the amounts used by other riparians. In some circumstances, a prior appropriation state may reduce the quantity of a right if it finds that a right holder has not been using the water or has “wasted” the water.

Thus, a water right is a form of property that is distinctively different from land or personal property usually considered in federal takings jurisprudence. The holder does not possess the water, only the right to use it. A water right is not a precise measure of water but a bundle of

187. See Laitos, supra note 158, at 907–08; Vranesh, supra note 171, § 3.2 at 154; Bell & Johnson, supra note 88, at 5 (referring to the use-it-or-lose-it principle).

188. See supra note 187.

189. Lock, supra note 158, at 86 (citing A-B Cattle Co. v. United States, 589 P.2d 57, 61 (Colo. 1978)).


191. Dellapenna, supra note 184, §§ 7.04 at 7-90 & 7.04(a)(3) at 7-97–7-98.

192. Lock, supra note 158, at 82 (citing Samuel C. Wiel, 1 Water Rights in Western States § 277, at 303–04 (3d ed. 1911)).


195. Dellapenna, supra note 182, at § 6.01(a)(4)–6-50.

196. An example of such a circumstance is an adjudication, “a special form of quiet title action to determine all existing rights to the use of water from a specific body of water.” Dep’t of Ecology v. Grimes, 121 Wash. 2d 459, 466, 852 P.2d 1044, 1048 (1993).

197. See id. at 471–72.

198. See Thompson, supra note 158, at 48; Laitos, supra note 158, at 922; Sax, supra note 14, at 260–61; Lock, supra note 158, at 79 (2000).

199. See Hutchins, supra note 166.
entitlements and duties, which varies greatly from state to state. In many circumstances, the state can affect a water right without violating the Fifth Amendment prohibition on taking without compensation. The characteristics of the property interest in a water right make application of federal takings jurisprudence difficult in water rights cases.

C. Judicial Application of Takings Jurisprudence in the Context of Minimum Stream Flow and Water Rights

At least one court has found a physical taking in a matter concerning minimum stream flow and water rights. In Tulare Lake Basin Water Storage District v. United States, the United States Court of Federal Claims found a physical invasion when the U.S. Bureau of Reclamation (BOR) withheld a portion of water that it had contracted to supply to water service districts. The BOR had not delivered the entire amount because low instream water volumes threatened endangered fish species. Although the court did not question the State of California’s authority to alter the water rights at issue, it concluded that the federal government could not justify its refusal to deliver the entire amount because the state had not reduced the quantity allowed under the water rights. Reasoning that the BOR had displaced the contract holder, the court characterized the government’s action as a physical intrusion.

The Court of Federal Claims based its finding that a physical invasion analysis is appropriate to apply to a water right on three cases. First, the court compared the case to United States v. Causby where the U.S. Supreme Court found a compensable taking had occurred when low flying military planes occupied the air space above a chicken farm, depriving the private property owners of the use and enjoyment of their

---

200. See supra notes 182–89 and accompanying text.
201. See supra notes 173–80 and accompanying text.
202. See supra notes 191–97 and accompanying text.
204. Id. at 318–19.
205. Id. at 316.
206. Id. at 315.
207. Id. at 324.
208. Id. The U.S. Bureau of Reclamation (BOR) was acting in response to the Endangered Species Act. Id. at 315.
209. Id. at 319.
210. Id.
211. 328 U.S. 256 (1946).
The court in *Tulare Lake Basin* found that the BOR’s act of withholding water mirrored *Causby’s* circumstances because the BOR had eliminated the water districts’ use and enjoyment of the amount of water withheld in order to preserve endangered fish.213

Further, the Court of Federal Claims analogized the facts of *Tulare Lake Basin* to *International Paper Co. v. United States*,214 a case where the government requisitioned all the power a New York power company could generate during World War I.215 To increase the power company’s capacity, the government directed the power company to use all the water from its power canal, thus cutting off water the power company leased to International Paper.216 The U.S. Supreme Court held that this was a governmental taking of the amount of leased water.217 The Court recognized that the property taken was International Paper’s “right . . . to the use of water.”218 The *Tulare Lake Basin* court viewed this as a confirmation of its characterization of a water right as property subject to a physical invasion analysis.219

The third case, *Dugan v. Rank*,220 was a complex injunction suit filed in 1947 to prevent construction of the Friant Dam.221 Prior to building the dam, the property owners refused to sell their rights to the government.222 Once built, the dam impounded water that would have been delivered to the landowners but went instead to Central Valley irrigation districts.223 The district court held that Congress had not authorized the federal government to seize water rights, but to acquire such rights exclusively through judicial proceedings.224 The U.S. Supreme Court disagreed, holding that the government had the power to “to acquire the water rights of respondents by physical seizure.”225 However, the Court emphasized

---

212. *Id.* at 266.
215. *Id.* at 405.
216. *Id.* at 405–06.
217. *Id.* at 408.
218. *Id.* at 407.
221. *Id.* at 610.
222. *Id.* at 613–14.
223. *Id.* at 612–13.
224. See *id.* at 616.
225. *Id.* at 619.
that the government intended to purchase the water rights and ultimately held that the property owners were due damages resulting from the otherwise valid seizure of water rights.

The Tulare Lake Basin court found three factors in Dugan to support an application of a physical invasion takings analysis to water rights. First, in Dugan, the U.S. Supreme Court referred to the method the government used to acquire water rights as “physical seizure.” Second, when the Court clarified that a physical invasion of land was not necessary to claim a taking, the Court observed that a seizure of a water right can be “analogized to interference or partial taking of air space over land.” Finally, the Court held that the government effected a compensable taking when it seized water rights rather than compensating the deprived landowners. Thus, the Tulare Lake Basin court held that a physical invasion analysis is the appropriate test in a water rights takings challenge.

A question of unconstitutional takings arises when state action restrains a water right holder from withdrawing the entire quantity allowed under a right. Federal takings tests for physical intrusion, exaction or regulatory taking can inform a court considering a takings challenge under the CWA. Determining which test is appropriate in a water rights case is complicated by the fact that a water right is a form of property distinctly different from the exclusively owned private property usually considered in takings jurisprudence. However, one federal court has applied physical intrusion analysis to a claim resulting from the effect of a minimum stream flow requirement on existing water rights.

---

226. Id. at 623.
227. See id. at 624.
229. Id. (citing Dugan v. Rank, 372 U.S. 609, 625 (1963)).
230. Id.
231. See supra note 88 and accompanying text.
232. See supra Part II.A.
233. See supra Part II.B.
234. See supra Part II.C.
III. COURTS SHOULD ANALYZE A TAKINGS CHALLENGE TO
STATE ACTION IMPOSING MINIMUM FLOW
REQUIREMENTS UNDER THE CLEAN WATER ACT AS A
REGULATORY TAKING

Courts recognize that a water right is property protected against
unconstitutional governmental taking.\textsuperscript{235} A state action under the CWA
that prevents a water right holder from withdrawing the full measure of a
water right gives rise to a takings challenge.\textsuperscript{236} Courts should not analyze
this state action under the physical intrusion analysis because the state is
not seizing the water right and the complex nature of a water right is not
suited to an inflexible, per se analysis.\textsuperscript{237} Further, a court should not
analyze the takings challenge as an exaction because the analysis is
inconsistent with the nature of the government action and, more
importantly, it could yield an unfair result.\textsuperscript{238} Instead, a court should use
the regulatory takings analysis because it is appropriate to the nature of a
water right, consistent with the type of government action, and most
likely to reach a fair conclusion.\textsuperscript{239}

A. The Physical Invasion Analysis is Inappropriate to a Water Rights
Takings Challenge Under the Clean Water Act

A court should not analyze a water right taking challenge under the
CWA as a physical invasion. Despite the finding of the Court of Federal
Claims in \textit{Tulare Lake Basin},\textsuperscript{240} a water right is not a suitable subject for
an analysis in which the slightest diminution results in a per se
compensable taking. The physical invasion analysis is inappropriate
because it is inconsistent with both the legal nature of a water right and
the character of the state action.\textsuperscript{241} Further, the cases on which the \textit{Tulare
Lake Basin} court bases its conclusion do not apply when a state imposes
minimum flow requirements under the CWA.\textsuperscript{242}

\textsuperscript{235} See supra note 88 and accompanying text.
\textsuperscript{236} See supra Part II.A.
\textsuperscript{237} See infra Part III.A.
\textsuperscript{238} See infra Part III.B.
\textsuperscript{239} See infra Part III.C.
\textsuperscript{240} 49 Fed. Cl. 313 (2001).
\textsuperscript{241} See infra Part III.A.1.
\textsuperscript{242} See infra Part III.A.2.
1. A Physical Invasion Analysis is Contrary to the Nature of a Water Right and the Character of the State Action

The Tulare Lake Basin court’s finding that a physical invasion analysis is appropriate to a water right\textsuperscript{243} contravenes the legal nature of a water right in three ways. First, the water right holder does not have exclusive interest in the water.\textsuperscript{244} Water in a natural watercourse is the property of the public or the state.\textsuperscript{245} Even after water is diverted, it is subject to the interests of others in both riparian\textsuperscript{246} and prior appropriation\textsuperscript{247} systems. In any system, a water right holder does not own the water; but possesses only the right to use it.\textsuperscript{248} The quantity diverted by a riparian landowner may be reduced if the diversion unreasonably prevents fellow riparian landowners from asserting their rights.\textsuperscript{249} In prior appropriation states, a water right holder is obligated to use water efficiently,\textsuperscript{250} apply it only once to the beneficial use for which it was granted, and then return the unused portions to the public.\textsuperscript{251} The shared interests inherent in a water right make it a substantially different type of property from the land in Lucas,\textsuperscript{252} the specified amount of water agreed to by private parties in International Paper,\textsuperscript{253} and the contracted amount of water in Tulare Lake Basin.\textsuperscript{254}

Second, unlike ownership in land, the entitlements attaching to a water right vary significantly from state to state.\textsuperscript{255} A takings rule adopted in one state may not be correct under another state’s law. This supports the

\begin{itemize}
  \item \textsuperscript{243} Tulare Lake Basin, 49 Fed. Cl. at 319.
  \item \textsuperscript{244} See supra Part II.B.
  \item \textsuperscript{245} See HUTCHINS, supra note 166, at 5–6; Anderson, supra note 167, at 12-48–12-55.
  \item \textsuperscript{246} See DELLAPPENNA, supra note 174, at 6–49–6-50; Harris v. Brooks, 283 S.W.2d 129, 133–34 (Ark. 1955); 78 AM. JUR. 2D Waters § 49 (2002); DELLAPPENNA, supra note 184, at 7-47.
  \item \textsuperscript{247} See VRANESH, supra note 171 § 3.1, at 121; GETCHES, supra note 176, at 101; Laitos, supra note 158, at 907–08; VRANESH, supra note 171 § 3.2, at 154; BELL & JOHNSON, supra note 88, at 5; Dep’t of Ecology v. Grimes, 121 Wash. 2d 459, 466, 852 P.2d 1044, 1048 (1993).
  \item \textsuperscript{248} See HUTCHINS, supra note 166, at 151.
  \item \textsuperscript{249} Harris v. Brooks, 283 S.W.2d 129, 133–34 (Ark. 1955); 78 AM. JUR. 2D Waters § 49 (2002); DELLAPPENNA, supra note 184, at 7-47.
  \item \textsuperscript{250} LOCK, supra note 158, at 86; Dep’t of Ecology v. Grimes, 121 Wash. 2d 459, 471–72, 852 P.2d 1044, 1054–55 (1993).
  \item \textsuperscript{251} VRANESH, supra note 171, at 121.
  \item \textsuperscript{252} Lucas v. S.C. Coastal Council, 505 U.S. 1003, 1006 (1992).
  \item \textsuperscript{253} INT’L Paper Co. v. United States, 282 U.S. 399, 405 (1931).
  \item \textsuperscript{254} Tulare Lake Basin Water Storage Dist. v. United States, 49 Fed. Cl. 313, 314 (2001).
  \item \textsuperscript{255} See Beck, supra note 170, at 4-1; 78 AM. JUR. 2D Waters § 49 (2002); BELL & JOHNSON, supra note 88, at 6; GETCHES, supra note 176, at 5–8.
\end{itemize}
conclusion that an inflexible, per se taking rule is inappropriate to a water right. Further, avoiding a per se rule is consistent with the U.S. Supreme Court’s opinion in Tahoe-Sierra which expressed a preference for “ad hoc, factual inquiries” rather than per se rules. 256

Third, a water right is not a right to a certain quantity of water alone; it is a collection of entitlements and duties. 257 In riparian states, water rights are correlative; the quantity diverted by a riparian landowner can be reduced if the use unreasonably infringes on the rights of other riparian landowners to divert water. 258 In prior appropriation states, the fundamental entitlements of a water right are (1) the right to priority of use relative to others and (2) the right to apply water to a beneficial use, which is limited by the maximum quantity of water that may be diverted under the right. 259 If a right holder can accomplish the same beneficial use with less water than the maximum amount allowed by a right, then the excess water must be made available for other uses. 260 In other words, in either system, a water right is not altered in its entirety by a lesser quantity of water. As the Court reiterated in Tahoe-Sierra, affecting one stick in a bundle of rights does not constitute a per se taking. 261 Thus, the government does not cause a per se compensable taking when a water quality regulation prevents diversion of the full amount of a water right. 262

2. The Tulare Lake Basin Court Incorrectly Applied Precedent

Further, the three cases on which Tulare Lake Basin is based are distinguishable from circumstances where a government, acting in its regulatory capacity, imposes minimum flow requirements on the use of

257. See Dellapenna, supra note 174, at 6–1–6–2; Hutchins, supra note 166, at 8–19.
258. Harris v. Brooks, 283 S.W.2d 129, 133–34 (Ark. 1955); 78 A.M. JUR. 2D Waters § 49 (2002); Dellapenna, supra note 184, at 7–47.
259. Laitos, supra note 158, at 906–07.
260. See id. at 907–08; Vranesh, supra note 171, § 3.2 at 154; Bell & Johnson, supra note 88.
262. A state legislature could redefine a water right as a purely quantified amount without the elements of beneficial or reasonable use, but would risk increasing a state’s potential liability in managing water. States could incur substantial costs if every holder could claim a per se taking whenever water is inadvertently withheld or unavailable for reasons other than those expressly stated in the law.
an existing water right. In the first case, *United States v. Causby*,263 the Court found a compensable taking when military planes occupied a portion of private property—the air space above the land—and by doing so deprived the owners of the use and enjoyment of their land.264 However, government action affecting air space is not a per se physical intrusion; courts can analyze the effect of government action on air space as a regulatory taking.265

In contrast to *Causby*, the government is not occupying a portion of a water holder’s property when it imposes minimum flow requirements; it is merely regulating how the holder may use the property.266 The full measure of the right remains available to the holder to use in a manner that does not create the condition that the government is seeking to mitigate by imposing the minimum flow requirement.267 For example, a holder could redesign a project to reduce the length of the bypass reach and thereby eliminate water quality concerns. Also, unlike the airplanes in *Causby* that affected a portion of the property beyond that actually occupied by the airplanes,268 minimum flow requirements does not deprive a right holder of the use and enjoyment of the entirety of a water right. The quantity allowed under a water right is only one stick in the bundle of entitlements inhering in the right.269 Therefore, a physical intrusion analysis would be inconsistent with *Tahoe-Sierra*, where the Court emphasized that the government does not create a per se taking when it affects one stick in a bundle of rights.270

Further, the *Tulare Lake Basin* court’s reliance on *International Paper* is misplaced.271 *International Paper* concerned not water rights, but a contract between private parties that created a right to a certain amount of leased water.272 State water law, not a contract, creates a holder’s property interest in a water right.273 Under state law, the maximum

---

263. 328 U.S. 256 (1946).
264. *Id.* at 266.
265. *See id.*
267. *See id.*
268. *See United States v. Causby, 328 U.S. 256, 266 (1946).*
269. *See Dellapenna, supra* note 174, at 6–1–6–2; *Hutchins, supra* note 166, at 8–19.
272. *Int’l Paper Co. v. United States, 282 U.S. 399, 405 (1931).*
amount of water that a holder may divert under the right is but one element in that bundle of rights and duties. Further, state law allows for considerable control over a water right without resulting in a compensable taking.

Finally, the Tulare Lake Basin court misunderstood the U.S. Supreme Court’s language in Dugan. The Tulare Lake Basin court correctly observed that the Dugan Court referred to the taking of water rights as a physical seizure, comparing it to interference with air space. However, the Court used the term “physical seizure” to describe taking a right without consent as compared to acquiring the right through an eminent domain proceeding. The Court used the air space analogy to explain that a physical invasion of land is not necessary for a taking to have occurred. It did not use the air space analogy to hold that government action affecting a water right is a per se taking. Further, the key issue in Dugan was that the Central Valley project took water from water rights claimants and gave equivalent rights to others. In contrast, when a state exercises its § 401 certification authority to impose minimum flow requirements, it is not taking one right and giving it to another. The state is simply conditioning the way in which a holder may use the water right. As described in Tahoe-Sierra, this is not a physical taking, but an effect on a property right arising from a program intended to benefit the common good.

The court in Tulare Lake Basin incorrectly concluded that a water right takings challenge should be analyzed under the physical invasion analysis. An analysis in which the slightest diminution results in a per se compensable taking contravenes the nature of a water right and the government action. Further, the cases on which the Tulare Lake Basin court relied are distinguishable from a situation where a state imposes minimum flow requirements under the CWA.

274. See Dellapenna, supra note 174, at 6–1–6–2; Hutchins, supra note 166, at 8–19.
278. Id. at 625.
279. See id. at 612–13.
281. See supra Part III.A.1.
282. See supra Part III.A.2.
B. An Exaction Analysis is Inappropriate to a Water Rights Takings Challenge

Imposing minimum flow requirements under § 401 resembles an exaction because it permits a water rights holder to build a project, but only if the holder agrees to divert less than the maximum amount of water allowable under an existing water right. However, an exaction analysis is inappropriate to a water rights takings challenge because it is not consistent with the regulatory character of the government action and is potentially unfair to the water right holder.

A water rights takings challenge under the CWA is not the same as a land takings challenge in *Dolan*,284 where the city required a business owner who wanted to build a bigger store to hand over privately-held land for a public bicycle trail.285 In *Dolan*, the government would have physically occupied the property.286 In contrast, a water right holder retains the property that is purportedly the subject of the exaction. Further, the holder may use the full measure of the right provided that the use does not result in poor water quality that the government intended the minimum flow requirement to avoid.287 Therefore, the government is not occupying, but merely regulating the property.

Second, the exaction analysis is potentially unfair to a water right holder because the doctrine does not allow balancing the interests of the state with the reasonable expectations of the right holder.288 If the exaction analysis is applicable to a water right challenge, then the state is likely to justify imposing minimum flow requirements under its § 401 authority in nearly all circumstances. There is a clear nexus between (1) the exaction (minimum water flow) and (2) the harm created by the proposed project (too little water). If the amount of water not diverted equals the amount of water necessary to meet water quality standards, then the exaction is roughly proportional to the harm. But, the court does not consider the extent of the holder’s loss unless the holder loses all economic value in use of the right.289 This is potentially unfair because

285. *Id.* at 378.
286. *Id.* at 380.
288. *See* *Dolan*, 512 U.S. at 386–96 (defining the exaction analysis tests without reference to the economic expectations of the landowner).
289. If a minimum flow requirement extinguishes all economically viable use of the right, it is a per se taking. *See* *Lucas v. S.C. Coastal Council*, 505 U.S. 1003, 1015 (1992).
the holder may have reasonably expected to use the full measure of the water right granted under state law and may have invested funds in seeking to use the right. The exactions analysis is not suited to a water rights taking challenge because it considers neither the extent of the harm to the holder, the reasonableness of the holder’s expectation or the legitimacy of the state action.290

C. A Water Right Takings Challenge is Best Analyzed as a Regulatory Taking

One disadvantage of the regulatory takings analysis is that, compared to exaction analysis, the outcome of judicial scrutiny is less predictable.291 Under an exaction analysis, a state’s action will be valid as long as the state can demonstrate its minimum flow requirement is tailored to meet the water quality standard.292 Under the regulatory takings analysis, however, the state’s action may solve a water quality problem but still be an unconstitutional taking if it has a substantial economic impact on a water right holder.293 At the same time, the court could also find that a taking has not occurred if the protection afforded to the public by the requirement outweighs the impact suffered by the holder.294 Despite the lesser degree of predictability, a regulatory takings analysis is more likely than other takings tests to reach a fair result in a water right takings challenge.

Under a regulatory takings analysis, if a government’s regulatory action does not deprive a water right holder of all economic value,295 the court balances factors such as the economic hardship and the character of the government action to determine if an unconstitutional taking has occurred.296 Regulatory takings analysis is the most appropriate test for a court to use when considering a challenge to state imposed minimum flow requirements under CWA for three reasons. First, the regulatory

290. Additionally, the U.S. Supreme Court in City of Monterey v. Del Monte Dunes, Ltd., has, without explanation, intimated that the exaction doctrine’s rough proportionality test is appropriate only to land use decisions. 526 U.S. 687, 702 (1999).
291. The following assumes the requirement does not destroy all economically viable use of the right. If it does, it is a per se taking. See Lucas, 505 U.S. at 1015.
292. See supra Part III.B.
294. Id.
295. If a minimum flow requirement extinguishes all economically viable use of the right, it is a per se taking. See Lucas, 505 U.S. at 1015.
takings analysis accounts for and balances the many interests that permeate a water right. Second, a regulatory takings analysis is in harmony with the legal nature of a water right. Finally, the test acknowledges that the state is acting in its regulatory capacity under state law and the CWA.

A regulatory takings analysis is more likely to be fair because it requires the court to consider all of the interests inherent in the right. The test protects the right holder by requiring the court to examine the economic impact on the water right holder and the extent to which the minimum flow regulations interferes with the holder’s reasonable investment-backed expectations. The test protects the public by recognizing the government’s authority to enforce laws and programs promoting the public welfare, even though the economic value of private property may be adversely affected. This balancing of interests makes the regulatory takings analysis more appropriate to a water right takings challenge than other takings analyses employed by the courts.

Second, a regulatory takings analysis accords with the legal nature of a water right. The factors that cause a water right to be unsuitable for a physical invasion analysis render it appropriate for a regulatory takings analysis. A water right is a collection of entitlements and duties, not a right to a precise quantity of water. The entitlements and duties vary greatly from state to state. There are circumstances in which a state can adjust a water right without leading to a compensable taking. The unique legal character of a water right, and the degree of control that states exercise over it, make it appropriate to a regulatory takings analysis because the test compels the court to balance several factors on a factual, case-by-case basis. This is consistent with the U.S. Supreme Court’s acknowledgement in Lucas that the degree of control that states traditionally exercise over the property affects the outcome of a takings

297. Id. at 124.
301. Id.
302. See supra Part III.A.
303. See Dellapenna, supra note 174, at 6-1–6-2; Hutchins, supra note 166, at 8–19.
304. See Beck, supra note 170, at 4-1; Getches, supra note 176, at 5–8.
analysis. It also reflects the Court’s emphasis in *Tahoe-Sierra* that a takings analysis must consider the property as a whole.  

Finally, a regulatory analysis is appropriate to apply to a water rights taking challenge because the government is acting in its regulatory capacity. When a state imposes minimum flow requirements, it does not affect the water right holder’s proprietary interest in the water. Instead, it is regulating how the holder may use the water. The holder may use the full allowable quantity of the right if the use does not cause the condition that the government intended the minimum flow requirement to mitigate.

When a state imposes a minimum flow requirement under its § 401 certification authority, a water right holder may challenge the state’s action as an unconstitutional taking of property. A government has effected a per se taking if the water right holder retains no economically viable use of the property. If the water right remains useful, the court should not analyze the challenge as either a physical intrusion or as an exaction, but as a regulatory taking because this analysis will be more likely to accurately consider the interests intrinsic to the right. Further, the analysis respects the legal nature of a water right, and reflects the fact that the state is acting in its regulatory capacity.

### IV. CONCLUSION

States have the power to impose minimum flow requirements under the CWA and thereby affect an existing water right. If a water right holder cannot withdraw the full quantity of his or her right because of the minimum flow requirements and is not compensated by the state, then the holder can challenge the state action as an unconstitutional taking. A court should not examine this challenge under the physical intrusion

---

306. *Id.* at 1027–28.
309. *Id.*
310. *See id.*
311. *See supra* note 88.
313. *See supra* Part III.A.
314. *See supra* Part III.B.
315. *See supra* Part III.C.
316. *See supra* Part III.C.
analysis because the state is not physically intruding on the property and an inflexible, per se test is not appropriate to the legal nature of a water right. Similarly, a court should not analyze the takings challenge as an exaction because the state is not occupying the water right but merely regulating how it can be used, and an exaction analysis could conclude in an unfair result. Thus, a court should analyze this challenge against the state as a regulatory taking because the state is acting in its regulatory capacity, the analysis respects the legal nature of a water right, and the result will accurately consider the diverse interests inherent in a water right.
CLASSES, PERSONS, EQUAL PROTECTION, AND
VILLAGE OF WILLOWBROOK V. OLECH

Robert C. Farrell*

Abstract: In most contexts, the Equal Protection Clause serves as a limitation on
government classifications, but it has also been used as a protector of individual rights. These
competing versions of equal protection are contradictory, but courts have for the most part
ignored this problem. In Village of Willowbrook v. Olech, the United States Supreme Court
determined that an individual homeowner had stated a valid equal protection claim when she
alleged that she alone, without regard to her membership in any class, had been treated
differently from other similarly situated homeowners. The Court’s decision in Olech has
created a powerful precedent for other individual persons complaining of wrongful treatment
by government officials. It also suggests a method of resolving the conflict between the two
competing views of equal protection.

In most contexts, the basic role of the Equal Protection Clause is to act
as a limit on government classifications. Tussman and tenBroek formally
articulated this role in their influential article, The Equal Protection of
the Laws.1 The United States Supreme Court has provided substantial
support for this view on numerous occasions. Some have gone so far as
to suggest that this limitation on government classification is the only
role of the Equal Protection Clause.² However, there has always been a
less well-known, less influential version of the Equal Protection Clause
that emphasizes, not classifications, but the protection of individual
persons without regard to their membership in any particular class. This
alternate view of equal protection occupies an uneasy relation with the
more predominant class-based version. The two are apparently
incompatible and might easily come into direct conflict, with one view
emerging as the winner. But this has not happened. Rather, like trains
riding on parallel tracks that never meet, the two apparently incompatible
views of equal protection do not come into direct conflict, but simply
ignore each other. Recently, in Village of Willowbrook v. Olech,³ the

(1949).
2. The author of this Article was one of those who once made such an assertion. See Robert C.
Farrell, Equal Protection: Overinclusive Classifications and Individual Rights, 41 ARK. L. REV. 1
(1988). That position I now recant with this Article.
3. 528 U.S. 562 (2000). For previous commentary on this case, see generally Hortensia S.
Carreira, Protecting the “Class of One,” 36 REAL PROP. PROB. TR. J. 331 (2001); Erwin
Chemerinsky, Suing The Government for Arbitrary Actions, 36 TRIAL 89 (May 2000); J. Michael

* B.A., Trinity College; J.D., Harvard University; Professor, Quinnipiac University School of Law.
Court suggested, although inadvertently, how the conflicting versions of equal protection might be reconciled. Although the Court’s opinion in Olech is extremely brief and makes no mention of the problem, the case does provide a roadmap for assigning each of the views of equal protection to its proper sphere. The conceptual result is a greater level of clarity in the making of equal protection arguments. The practical result will probably be the proliferation in the federal district courts of cases where an individual person claims that governmental officials have treated him or her unequally.

This Article will set forth the two variant views of equal protection and attempt to identify the appropriate context for each. Part I of this Article sets forth the traditional version of equal protection as a limit on government classification. Part II introduces the personal, individual-rights view of equal protection. Part III closely examines Olech, which reinvigorated the individual rights version under the rubric, “class of one.” Part IV examines how Olech has changed litigation in the federal courts. Finally, Part V takes a second look at the class-based and individual-rights versions of equal protection in light of Olech.

I. EQUAL PROTECTION AS A LIMIT ON GOVERNMENT CLASSIFICATION

A. Tussman and tenBroek’s View of Equal Protection

The notion of equality could not possibly require that all persons be treated the same. For example, a doctor who prescribed an aspirin to all his patients would not be treating them equally. And thus, at least since

---


4. The term “inadvertently” is used here because not only did the Court not explicitly claim to be reconciling a conflict between two different views of the Equal Protection Clause, it did not even advert to the fact that there was any such conflict. Nor did the Court cite, much less distinguish, any of the cases that spoke of equal protection as a limit on classifications rather than an individual right. The Supreme Court’s opinion in Olech is discussed at greater length in Part III.B.

5. See infra Part IV.C.

6. See RONALD DWORKIN, *TAKING RIGHTS SERIOUSLY* 227 (1977) (discussing “the right to treatment as an equal, which is the right, not to receive the same distribution of some burden or benefit, but to be treated with the same respect and concern as anyone else. If I have two children,
Aristotle, it has been understood that equality requires, not that everyone be treated the same, but that those similarly situated be treated similarly. Tussman and tenBroek applied this paradigm to the notion of equality as it applies to legislation. In opposition to the notion that equality requires all persons to be treated the same, Tussman and tenBroek explained that it is not possible to demand that all laws be of a general and universal character that make no distinction between persons. “It is clear that the demand for equal protection cannot be a demand that laws apply universally to all persons.” Rather, “[f]rom the very necessities of society, legislation of a special character . . . must often be had.” The legislature, if it is to act at all, must impose special burdens upon or grant special benefits to special groups or classes of individuals.

But once it is admitted that the legislature is free to enact laws with less than universal impact, thus treating different groups differently, then “[w]e . . . [have] arrive[d] at the point at which the demand for equality confronts the right to classify.” This is what Tussman and tenBroek identified as the basic paradox, that “[t]he equal protection of the laws is a ‘pledge of the protection of equal laws.’” But laws may classify. And ‘the very idea of classification is that of inequality.’ The way out of this bind turns out to be the doctrine of “reasonable classifications” where “the reasonableness of a classification is the degree of its success in treating similarly those similarly situated.”

Tussman and tenBroek explained this doctrine as follows. First, the process of classification involves the “designat[ion of] a . . . trait . . . the possession of which, by an individual, determines his membership in or inclusion within a class.” Thus, for example, the legislature could create a class by reference to a trait such as gender, age, or citizenship. Once a class is established, Tussman and tenBroek explained that it must be

---

8. Tussman & tenBroek, supra note 1, at 343 (“A state . . . is not compelled to ‘run all its laws in the channels of general legislation.’” (citing Bachtel v. Wilson, 204 U.S. 36, 41 (1907))).
9. Id.
10. Id.
11. Id.
12. Id. at 343–44.
13. Id. at 344.
14. Id.
15. Id.
tested for reasonableness, that is, its success in treating similarly those similarly situated.\textsuperscript{16} But how is the legislature to determine who is similarly situated? Clearly, this standard requires more than that every member of the class possess the classifying trait.\textsuperscript{17} By such a circular standard, all classifications would be reasonable. Rather, according to Tussman and tenBroek, in order to test for similarity of situation, “we must look beyond the classification to the purpose of the law. A reasonable classification is one which includes all persons who are similarly situated with respect to the purpose of the law.”\textsuperscript{18}

In applying this test, Tussman and tenBroek explained that it is helpful to think of the class by reference to the trait that makes one a member of it, and to think of the purpose of the law by reference to the evil or mischief that the law seeks to eliminate.\textsuperscript{19} With this terminology, then, the Tussman and tenBroek test requires us to compare the class that bears the trait with the class that is tainted by the mischief.\textsuperscript{20} If a classification is perfect, the two classes completely coincide. In the real world, however, classifications are far more likely to be overinclusive, underinclusive, or both, but courts are quite deferential to legislative judgments and ordinarily will tolerate a great deal of imprecision. Of the differing kinds of imprecision, Tussman and tenBroek pointed out that the more objectionable is the overinclusive classification, the one that “imposes a burden upon a wider range of individuals than are included in the class of those tainted with the mischief at which the law aims.”\textsuperscript{21} The overinclusive classification “reach[es] out to the innocent bystander, the hapless victim of circumstance or association,”\textsuperscript{22} and thus is the most inconsistent with “our traditional antipathy to assertions of mass guilt and guilt by association.”\textsuperscript{23}

It should be obvious that the Tussman and tenBroek view of equal protection as a limitation on classification provides no protection for individuals who are harmed by a classification, so long as the classification satisfies the requirement of reasonableness. For them, at

\textsuperscript{16} Id.
\textsuperscript{17} Id. at 345.
\textsuperscript{18} Id. at 346.
\textsuperscript{19} Id.
\textsuperscript{20} Id. at 346–47.
\textsuperscript{21} Id. at 351.
\textsuperscript{22} Id.
\textsuperscript{23} Id. at 352.
least as applied to legislation, the equal protection clause does not create individual, personal rights.

B. The View of Equal Protection as a Limitation on Classifications in the Courts

1. U.S. Supreme Court Precedent

The U.S. Supreme Court has repeatedly spoken of the Equal Protection Clause as a limit on government classifications. In *Nordlinger v. Hahn*, for example, it explained that “[o]f course, most laws differentiate in some fashion between classes of persons. The Equal Protection Clause does not forbid classifications. It simply keeps governmental decisionmakers from treating differently persons who are in all relevant respects alike.” This basic explanation led the Court to specify the traditional deferential rule for evaluating classifications is that, “the Equal Protection Clause requires only that the classification rationally further a legitimate state interest.” The *Nordlinger* Court then concluded that the different treatment accorded to older and newer homeowners was rationally related to governmental interests in neighborhood stability and the protection of reliance interests. It was of no concern to the Court, applying this version of the Equal Protection Clause, that the plaintiff in the case was paying property taxes five times higher than her neighbors were paying in similar homes. So long as the classification was reasonable, individual harm or unfairness to a particular person was not part of the equal protection calculus.

The notion that the Equal Protection Clause’s principal work is to limit government classifications is repeated frequently by the Court. Another example is *Romer v. Evans* where the Court stated that:

The Fourteenth Amendment’s promise that no person shall be denied the equal protection of the laws must coexist with the practical necessity that most legislation classifies for one purpose or another, with resulting disadvantage to various groups or

---

25. *Id.* at 10.
26. *Id.*
27. *Id.* at 12.
28. *Id.* at 7.
persons. We have attempted to reconcile the principle with the reality by stating that, if a law neither burdens a fundamental right nor targets a suspect class, we will uphold a legislative classification so long as it bears a rational relation to some legitimate end.\textsuperscript{30}

The Court in \textit{Romer} then ruled that the classification at issue in that case, one that explicitly singled out gay persons for disadvantageous treatment, was not one that satisfied the reasonable classification requirement.\textsuperscript{31} The problem was both that the provision at issue “sing[ed] out a certain \textit{class} of citizens for disfavored legal status or general hardships”\textsuperscript{32} and that it “raise[d] the inevitable inference that the disadvantage imposed is born of animosity toward the \textit{class} of persons affected.”\textsuperscript{33} “The Court’s opinion invalidating the provision made no reference to individual rights.

The Supreme Court’s opinion in \textit{Massachusetts Board of Retirement v. Murgia}\textsuperscript{34} is a particularly clear expression of the view that equal protection is a limit on classification rather than a protector of individual rights. In that case, the Court reviewed a Massachusetts statute requiring all state police officers to retire at age fifty.\textsuperscript{35} The purpose of the law was to ensure the fitness of officers in order to promote public safety “by assuring physical preparedness of its uniformed police.”\textsuperscript{36} Robert Murgia, the plaintiff in the case, was more than fifty years of age, but extremely fit.\textsuperscript{37} He challenged the law as a denial of equal protection.\textsuperscript{38} In effect, Murgia was arguing that he was similarly situated to those under fifty who were allowed to stay on their jobs and thus he should receive the same treatment.\textsuperscript{39} According to Tussman and tenBroek’s analysis, the

\begin{itemize}
  \item \textsuperscript{30} \textit{Id.} at 631 (citations omitted).
  \item \textsuperscript{31} \textit{Id.} at 632.
  \item \textsuperscript{32} \textit{Id.} at 633 (emphasis added).
  \item \textsuperscript{33} \textit{Id.} at 634 (emphasis added).
  \item \textsuperscript{34} \textit{427 U.S. 307} (1976).
  \item \textsuperscript{35} \textit{Id.} at 307.
  \item \textsuperscript{36} \textit{Id.} at 308.
  \item \textsuperscript{37} \textit{Id.} at 314.
  \item \textsuperscript{38} \textit{Id.} at 311.
  \item \textsuperscript{39} On the one hand, the mandatory retirement at age 50 was designed “to remove from police service those whose fitness for uniformed work presumptively has diminished with age.” \textit{Id.} at 315. On the other hand, Murgia had proved through individualized testing that “his excellent physical and mental condition rendered him capable of performing the duties of a uniformed officer.” \textit{Id.} at 311. Thus, it was Murgia’s claim that in relation to fitness, he was in fact similar to the presumptively fit group of those under 50 rather than similar to the presumptively unfit group of those over 50.
\end{itemize}
issue is whether the class of state troopers over fifty was similarly situated with the class of state troopers under fifty in relation to the purpose of ensuring fitness to promote safety. Once the Court determined that fitness tends to decline with age, then the classes were not similarly situated because it is probably true that a greater percentage of those over fifty were unfit for the rigors of police work than the percentage of those under fifty who were unfit. Murgia was thus part of a class that was not similarly situated with the class of those under fifty in relation to the purpose of the law. Therefore, the Equal Protection Clause did not require that the members of Murgia’s class be treated the same as those under fifty.

But what about Robert Murgia himself? Murgia had demonstrated through individual testing that he himself was extremely fit and quite capable of handling the rigors of police work. Wouldn’t that entitle him to the same treatment as other fit individuals? No, it would not. Under the view of equal protection as a limitation on government classification, all that was required was that the classification meet the required standard, without regard to the classification’s effect on individual persons. Once it was determined that the age classification was reasonably correlated with fitness for police work, and thus that those who were members of the over fifty group were not similarly situated with those in the under fifty group, it did not matter what adverse effect the classification had on an individual member of the class. What was at work was a classification that made use of a generalization about age that was reasonably accurate. It did not matter that the generalization was not true as to a particular member of the class. Thus, the ordinary equal protection doctrine, at least as applied to legislative or administrative rules, appears to provide no protection for individuals from being harmed by classifications that embody generalizations that are not true as to the individual. In this context, it is very hard to make sense of the claim that equal protection is a personal, individual right.

Kimel v. Florida Board of Regents provides further support for the view that equal protection serves only as a limitation on classifications.

40. Tussman & tenBroek, supra note 1, at 346 (indicating that “A reasonable classification is one which includes all persons who are similarly situated with respect to the purpose of the law.”).
41. Id. at 315.
42. Id. at 311 (“Appellee Murgia had passed . . . [a rigorous physical] examination four months before he was retired, and there is no dispute that, when he retired, his excellent physical and mental health still rendered him capable of performing the duties of a uniformed officer.”).
In that case, the Supreme Court was concerned with the limits of Congressional power to enforce the Fourteenth Amendment through appropriate legislation.\textsuperscript{44} The particular issue was whether Congress had the power under Section 5 of the Fourteenth Amendment to extend the prohibitions of the Age Discrimination in Employment Act of 1967 (ADEA)\textsuperscript{45} to bind state governments. In the course of answering this question in the negative,\textsuperscript{46} the Court used a test that required that there be “a congruence and proportionality between the injury to be prevented or remedied and the means adopted to that end.”\textsuperscript{47} In order to apply this test, the Court had to compare the kind of age discrimination that would be prohibited by the Equal Protection Clause with the kind of discrimination that was in fact prohibited by the statute.\textsuperscript{48} In making this comparison, the Court made it clear that, unlike the ADEA, the Equal Protection Clause does not protect individuals. Specifically, the Court stated:

Under the Fourteenth Amendment, a State may rely on age as a proxy for other qualities, abilities, or characteristics that are relevant to the State’s legitimate interests. The Constitution does not preclude reliance on such generalizations. That age proves to be an inaccurate proxy in any individual case is irrelevant.\textsuperscript{49}

The \textit{Kimel} Court went on to explain that the constitutionality of age classifications was not to be determined “on a person by person basis.”\textsuperscript{50} On the other hand, the statutory language of the ADEA specifically made unlawful “discriminat[ion] against any individual . . . because of such individual’s age.”\textsuperscript{51} This statutory language thus created “a presumption in favor of requiring . . . [an] individualized determination.”\textsuperscript{52} Once the Court had identified this clear distinction between the Equal Protection Clause (which allows for generalizations that may not be true in individual cases) and the ADEA (which protects each individual person

---

\textsuperscript{44} U.S. CONST. amend. XIV, § 5.
\textsuperscript{46} \textit{Kimel}, 528 U.S. at 67 (“We conclude that the ADEA does contain a clear statement of Congress’ intent to abrogate the States’ immunity, but that the abrogation exceeded Congress’ authority under [Section] 5 of the Fourteenth Amendment.”).
\textsuperscript{47} Id. at 81 (quoting City of Boerne v. Flores, 521 U.S. 507, 520 (1997)).
\textsuperscript{48} Id. at 82–86 (comparing the conduct prohibited by the statute with the conduct prohibited by the Equal Protection Clause).
\textsuperscript{49} Id. at 84 (emphasis added).
\textsuperscript{50} Id. at 85–86.
\textsuperscript{52} \textit{Kimel}, 528 U.S. at 87.
from discrimination), the Court determined that Congress did not have authority under Section 5 to adopt the ADEA because the conduct prohibited by the statute was much broader than, and thus not congruent and proportional to, the conduct prohibited by the Equal Protection Clause. The Court’s opinion in Kimel is thus very strong evidence that the Equal Protection Clause does not protect individual rights.

In fact, on at least one occasion, the Supreme Court has suggested that limiting classifications is the only role of the Equal Protection Clause. In Oyler v. Boles, the Court reviewed a claim alleging selective enforcement of a habitual criminal statute. In rejecting that claim, the Court explained that “the conscious exercise of some selectivity in enforcement is not itself a federal constitutional violation.” What was missing from the allegations in the Oyler case was a claim that “the selection was deliberately based upon an unjustifiable standard such as race, religion, or other arbitrary classification.” Without such an “arbitrary classification,” the Court concluded, “grounds supporting a finding of a denial of equal protection were not alleged.” The Court’s Oyler opinion is perhaps its strongest statement that equal protection does no more than limit government classifications.

2. Lower Federal Court Precedent

Some lower federal courts have been even more explicit in the view that equal protection is a limit on government classifications and nothing more. One of the strongest versions of this argument was made by the U.S. Court of Appeals for the Sixth Circuit in Futernick v. Sumpter Township. In that case, the plaintiff alleged an equal protection violation when town officials selectively enforced state environmental regulations against him. In rejecting that claim, the Sixth Circuit, citing Oyler, found that a claim of selective prosecution must be based on an

53. Id. at 86 (“The Act, through its broad restriction on the use of age as a discriminating factor, prohibits substantially more state employment decisions and practices than would likely be held unconstitutional under the applicable equal protection, rational basis standard.”).


55. Id. at 456.

56. Id.

57. Id.

58. 78 F.3d 1051 (6th Cir. 1996).

59. Id. at 1056 (“The heart of Futernick’s case is the claim that state officials . . . violated his constitutional right to equal protection of the law by selectively enforcing Michigan state environmental regulations.”).
unjustifiable standard on the grounds of “race, religion, or other arbitrary classification.” 60. The court explained that an “arbitrary classification” 61 implied that “group distinctions could be a basis for liability” 62 but that not every act of arbitrariness directed at an individual would be constitutionally impermissible. 63. The court determined that Futernick’s equal protection claim failed because he did “not claim to be a member of any group” 64 but had simply alleged town officials had acted in bad faith toward him individually. 65. At least with regard to equal protection claims of selective enforcement, the Sixth Circuit in Futernick read Supreme Court precedent as requiring “arbitrary classifications” 66 as a basis of liability. According to the Sixth Circuit, “choosing to enforce the law against a particular individual is [not] a ‘classification.’” 67.

Futernick further explained that limiting equal protection claims to allegations of arbitrary classifications, that is, decisions aimed at a particular group, was essential in order to limit the work of the federal courts and to give proper deference to the judgment of local officials. 68. For example, the Sixth Circuit explained, a court could require that equal protection claims of bad faith toward an individual be accompanied by the requirement of also showing that “others who are similarly situated in ‘all relevant aspects’ have not been regulated.” 69. The focus on “similarly situated” 70 persons would act as a screening device, in that not every act of bad faith directed at an individual would be an equal protection violation. Rather, equal protection would only prohibit those acts of bad faith where one individual had been singled out while other similarly situated individuals were left unregulated. Such a rule would protect individuals without requiring them to identify any class-based discrimination.

60. Id. at 1057 (citing Oyler, 368 U.S. at 456).
61. Id.
62. Id.
63. Id. at 1057 n.8.
64. Id. at 1057.
65. Id.
66. Id.
67. Id. at 1058.
68. Id. (“Furthermore, we see compelling reasons that the sundry motivations of local regulators should not be policed by the Equal Protection Clause of the United States Constitution, absent the intent to harm a protected group or punish the exercise of a fundamental right. The sheer number of possible cases is discouraging.”) (emphasis in original).
69. Id. (citing Rubinowitz v. Rogato, 60 F.3d 906, 910 (1st Cir. 1995)).
70. Id.
But, according to the *Futernick* court, such a rule would also make it too easy for litigious plaintiffs to get past the motion to dismiss stage and advance to summary judgments because:

Determining ‘all relevant aspects’ of similar situations usually depends on too many facts (and too much discovery) to allow dismissal on a Rule 12(b)(6) motion. If we require defendants to wait until summary judgment, we burden local and state officials with the regular prospect of ‘fishing expeditions’ and meritless suits. In the meantime we federalize and constitutionalize what are essentially issues of local law and policy.71

The Sixth Circuit was concerned about the possibility of expanding equal protection to simple claims of arbitrariness or animosity not related to group identity.72 The state needed to retain the discretion to determine against whom to enforce its laws, especially because the state has never been in a position to enforce all laws at all times against all infractions. It is both appropriate and necessary “for a state somehow to choose to prosecute or regulate only part of the group of people who may be violating the law, and to do so without subjecting the selection decision to micro-management by courts.”73 The requirement that allegations of arbitrariness or animosity by governmental officials must include a reference to group identity would exclude most of these claims.74 Thus, according to *Futernick*, the Equal Protection Clause is only relevant when a plaintiff has alleged an arbitrary or invidious classification. A mere claim of arbitrary or invidious conduct toward an individual person will not suffice.

*Powell v. Montgomery*75 is another recent case where a federal district court made an extremely strong statement supporting the class-based version of equal protection. In that case, the plaintiff had been fired from his job with the fire department for failing to obey an order to have himself weighed to determine his compliance with the department’s

---

71. *Id.* at 1058–59 (internal citations omitted).
72. *Id.* at 1059.
73. *Id.* at 1060.
74. *Id.* at 1058–59 (indicating that, without a requirement that an equal protection claimant include an allegation of an intent to harm a protected group or punish the exercise of a fundamental right, “[t]he sheer number of possible cases is discouraging,” and suggesting that the requirement of alleging an invidious classification is part of a view that “counsels against expanding a federal right to protection from non-group animosity on the part of local officials”).
75. 56 F. Supp. 2d 1328 (M.D. Ala. 1999).
weight management regulations. In response to his equal protection claim that the regulations had been applied unequally to him, the court insisted that his equal protection claim had to include an allegation of purposeful discrimination. In explaining how one proved purposeful discrimination, a federal district court for the Middle District of Alabama cited the U.S. Supreme Court’s decision in Personnel Administrator of Massachusetts v. Feeney. There the Court stated that discriminatory purpose “implies that the decisionmaker . . . selected or reaffirmed a particular course of action at least in part ‘because of,’ not merely ‘in spite of,’ its adverse effects upon an identifiable group.” The federal district court in Powell interpreted this language to mean that the plaintiff’s equal protection claim needed to include an allegation that the defendant’s conduct had been “deliberately based on an unjustifiable, group-based standard.” Because the plaintiff in Powell had only alleged unequal treatment toward himself as an individual, his claim was rejected.

Judge J. Skelly Wright, while he was Chief Judge of the U.S. Court of Appeals for the District of Columbia Circuit, also shared the view of equal protection as being no more than a limit on government classification. In the course of an argument against the notion that equal protection violations required proof of purposeful discrimination, he included the following:

The key here is that the [E]qual [P]rotection [C]lause is primarily concerned with classes or groups, not individuals. As I am confident Mr. Justice Frankfurter must have written somewhere, a case invoking the [E]qual [P]rotection [C]lause, if it is to succeed, must allege something more than a tort, personal to the plaintiff.

All of the cases cited in this section constitute strong support for the view that equal protection operates to limit government classifications, but does no more than that. However, as the next section will

76. Id. at 1331.
77. Id. at 1332.
79. Id. at 279.
80. Powell, 56 F. Supp. 2d at 1333 (quoting E & T Realty v. Strickland, 830 F.2d 1107, 1114 (11th Cir. 1987)) (emphasis in Powell).
81. Id.
demonstrate, there is an alternate view of equal protection that will necessarily call that wisdom into question.

II. EQUAL PROTECTION AS A PROTECTOR OF INDIVIDUAL RIGHTS

The alternate view of equal protection focuses, not on limiting governmental classifications, but on protecting individual rights. This view of equal protection has an obvious contextual basis; the Fourteenth Amendment itself provides that no state shall deny to any person the equal protection of the laws. Shelley v. Kramer was one of the earliest cases expressing this version of equal protection. In that case, the defendants had argued that the state courts were not treating black residents unequally in enforcing a racially restrictive covenant because the courts would be equally willing to enforce such a covenant against white persons. The U.S. Supreme Court rejected this “equal discrimination” argument as amounting to no more than the “indiscriminate imposition of inequalities.” The Court explained that, even if courts were willing to penalize all races equally, that would not satisfy the mandate of the Equal Protection Clause because “[t]he rights created by . . . the Fourteenth Amendment are, by its terms, guaranteed to the individual. The rights established are personal rights.” Thus, an individual black person who is unable to buy a home because of his race has been treated unequally, without regard to a court’s willingness to prevent a white person from buying a home because of his race.

Although the Court in Shelley rejected the “equal application” defense on the basis of a personal rights view of equal protection, it is quite clear that there was no need to resort to that explanation in order to reach that result. In Loving v. Virginia, the Court considered and rejected, without any reference to personal rights, a similar argument that had been made in defense of a statute that prohibited interracial marriage. In Loving, the statute was defended on the ground that the state was willing to

83. U.S. Const. amend. XIV.
84. 334 U.S. 1 (1948).
85. Id. at 21.
86. Id. at 22.
87. Id.
88. 388 U.S. 1 (1967).
89. Id. at 8.
prohibit interracial marriage equally by both black and white persons.\textsuperscript{90} The Court rejected that argument on the grounds that the statutes contained “racial classifications,” and “the fact of equal application does not immunize the statutes from the very heavy burden of justification which the Fourteenth Amendment has traditionally required of state statutes drawn according to race.”\textsuperscript{91} Thus, the problem the Court identified in \textit{Loving}, and which it ought to have identified in \textit{Shelley}, was an indefensible classification, not an invasion of a personal right.

However, notwithstanding \textit{Loving}, the view of equal protection as an individual right has had several more recent expressions in a series of affirmative action cases, where it has generally been used by those opposed to affirmative action, or at least by those opposed to a lesser standard of review for affirmative action. In \textit{Regents of the University of California v. Bakke},\textsuperscript{92} Justice Powell argued that all racial classifications, whether invidious or benign, should be subjected to strict scrutiny.\textsuperscript{93} In support of this conclusion, Justice Powell cited the personal nature of the equal protection guarantee.\textsuperscript{94} He began by citing the above-quoted language in \textit{Shelley} that the rights established by the equal protection clause are personal rights.\textsuperscript{95} He then insisted that “[t]he guarantee of equal protection cannot mean one thing when applied to one individual and something else when applied to a person of another color. If both are not accorded the same protection, then it is not equal.”\textsuperscript{96} Justice Powell criticized the “artificial line of a ‘two-class theory’ of the Fourteenth Amendment,”\textsuperscript{97} and insisted that “[n]othing in the Constitution supports the notion that individuals may be asked to suffer otherwise impermissible burdens in order to enhance the societal standing of their ethnic groups.”\textsuperscript{98} Justice Powell’s concluding words on the subject were a strong endorsement of the individual rights position. He said, “[i]f it is the individual who is entitled to judicial protection against classifications based upon his racial or ethnic background because such distinctions

\textsuperscript{90} Id.
\textsuperscript{91} Id. at 9.
\textsuperscript{92} 438 U.S. 265 (1978).
\textsuperscript{93} Id. at 291.
\textsuperscript{94} Id at 289–90.
\textsuperscript{95} Id. at 289 (quoting Shelley v. Kramer, 334 U.S. 22 (1948)).
\textsuperscript{96} Id.
\textsuperscript{97} Id. at 295. The “two-class theory” to which Powell here refers is the view that the Equal Protection Clause is directed solely at discrimination “based on differences between ‘white’ and Negro.” Id. at 295 (citing Hernandez v. Texas, 347 U.S. 475, 478 (1954)).
\textsuperscript{98} Id. at 298.
impinge upon personal rights, rather than the individual only because of his membership in a particular group then constitutional standards may be applied consistently . . . . The Constitution guarantees that right to every person regardless of his background."

Subsequently, in Adarand Constructors Inc. v. Pena, the Supreme Court articulated an extremely strong version of the individual rights view of equal protection. In Adarand, the Court, for the first time in a majority opinion, held that strict scrutiny was the proper standard of review for all racial classifications, without regard to the level of government that was making the classification. In so holding, the Court cited Justice Powell’s Bakke opinion at some length as strong support for the individual rights view of equal protection. Going beyond that, the Court in Adarand argued that its previous cases had established three general propositions with regard to racial classifications: skepticism, consistency, and congruence. According to the Court, these three propositions:

All derive from the basic principle that the Fifth and Fourteenth Amendments to the Constitution protect persons not groups. It follows from that principle that all governmental action based on race—a group classification long recognized as ‘in most circumstances irrelevant and therefore prohibited,’—should be subjected to detailed judicial inquiry to ensure that the personal right to equal protection of the laws has not been infringed.

The Court explained further that “whenever the government treats any person unequally because of his or her race, that person has suffered an injury that falls squarely within the language and spirit of the Constitution’s guarantee of equal protection.”

The majority criticized Justice Stevens’ dissenting opinion, which had argued that it was appropriate to review benign racial classifications more generously than

---

99. Id. at 299.
101. Id. at 224 (“Taken together, these three propositions lead to the conclusion that any person, of whatever race, has the right to demand that any governmental actor subject to the Constitution justify any racial classification subjecting that person to unequal treatment under the strictest judicial scrutiny.”).
102. Id. at 224–25.
103. Id. at 223–24.
104. Id. at 227 (quoting Hirabayashi v. United States, 320 U.S. 81, 100 (1943)) (emphasis in original).
105. Id. at 229–30 (emphasis added).
invidious racial classifications,\textsuperscript{106} as inconsistent “with the long line of cases understanding equal protection as a personal right.”\textsuperscript{107}

Outside of the affirmative action context, the Court has also used personal rights rhetoric to explain why gender classifications should be subjected to a heightened level of scrutiny. In \textit{Frontiero v. Richardson},\textsuperscript{108} the plurality opinion explained that sex, “like race and national origin, is an immutable characteristic determined solely by accident of birth”\textsuperscript{109} and thus to impose disabilities on the basis of sex “would seem to violate ‘the basic concept of our system that legal burdens should bear some relationship to individual responsibility.’”\textsuperscript{110} Further, because “the sex characteristic frequently bears no relation to ability to perform or contribute to society,”\textsuperscript{111} classifications based on sex “often have the effect of invidiously relegating the entire class of females to inferior legal status without regard to the actual capabilities of its individual members.”\textsuperscript{112} These statements seem to suggest that what is objectionable about gender classifications is that they fail to take into account individual differences within the sexes, that is, they involve broad generalizations that are not universally true, and thus, wrongfully ignore individual merit. Thus, equal protection would appear to protect individual rights.

However, the Supreme Court’s assertions of the essentially personal nature of the right to equal protection are striking in their inconsistency with the host of Supreme Court cases identifying equal protection as a limit on governmental classifications\textsuperscript{113} and in their willingness to ignore those cases. For example, consider Justice Powell’s personal rights language in \textit{Bakke} as it might apply to the facts of \textit{Murgia}.\textsuperscript{114} In \textit{Murgia}, the issue was precisely the Court’s unwillingness to look at Robert Murgia’s personal situation, that is, his own proven qualifications for police work.\textsuperscript{115} The Court apparently was uninterested in the argument that “it is the individual who is entitled to equal protection against

\textsuperscript{106} See id. at 243 (Stevens, J., dissenting).
\textsuperscript{107} Id. at 230.
\textsuperscript{108} 411 U. S. 677 (1973) (plurality opinion).
\textsuperscript{109} Id. at 686.
\textsuperscript{110} Id. (citing Weber v. Aetna Cas. & Sur. Co., 406 U.S. 164, 175 (1972)) (emphasis added).
\textsuperscript{111} Id.
\textsuperscript{112} Id. at 687 (emphasis added).
\textsuperscript{113} See supra Part I.B.
\textsuperscript{114} See supra notes 34–42 and accompanying text.
\textsuperscript{115} See supra note 42 and accompanying text.
Classes or Persons?

classifications based on [age] because such distinctions impinge on personal rights.” Rather, what was important for the Court in Murgia was the reasonableness of the age classification which had the effect of terminating Murgia’s employment. The Court apparently viewed Murgia as someone who was to be considered “only because of his membership in a particular group.” Or, if Adarand is to be used as precedent, why was it not true of Murgia that “whenever the government treats any person unequally because of his or her [age], that person has suffered an injury [under the Equal Protection Clause]”? While it is true that race classifications are more suspicious than age classifications and thus should be more strictly scrutinized, the Court has given no reasonable explanation for treating equal protection as a personal right in the affirmative action cases, but not in most other areas of equal protection. In fact, it appears that the two views of equal protection are quite inconsistent and thus can survive together only so long as they continue to ignore each other. However, in Olech, the Supreme Court, apparently inadvertently and implicitly, identified a way in which the two versions of equal protection can be reconciled.

III. VILLAGE OF WILLOWBROOK V. OLECH

A. Village of Willowbrook v. Olech in the Lower Courts

Village of Willowbrook v. Olech did not begin as a case likely to lead to a significant precedent in the U.S. Supreme Court. Rather, it began rather trivially, when Grace Olech’s well broke and could not be repaired. Fortunately for Mrs. Olech, the public water supply was close at hand and the Village of Willowbrook was willing to connect her to it. Unfortunately, the Village demanded from Mrs. Olech, as part of

116. This was the language Justice Powell used in his Bakke opinion, although with regard to race classifications rather than age classifications. See Regents of the Univ. of Cal. v. Bakke, 438 U.S. 265, 299 (1978).
117. Id.
118. This was the language the Court used in Adarand Constructors Inc. v. Payna, 515 U.S. 200, 229–30 (1995), although, once again, with regard to race rather than age.
120. Id. at *1.
121. Id.
the connection process, a 33-foot easement “while only requiring a 15-foot easement from other Village residents.” Mrs. Olech complained, but the Village initially would not reduce the size of the required easement. This dispute between Mrs. Olech and the Village was the classic “garden variety” land use dispute that rarely gets to federal court and, if it does, is likely to be dismissed very quickly at the early stages. Mrs. Olech’s case did get to federal district court, but was dismissed for failure to state a claim.

According to Mrs. Olech’s allegations before a district court in the Northern District of Illinois, the Village had initially demanded more of her than of others because village officials still harbored “substantial ill will” against her because she had successfully sued the Village in another matter six years earlier. During the delay caused by this dispute, Mrs. Olech was without running water for three months and it was “this three-month delay that Olech claim[ed] deprived her of her rights under the Equal Protection Clause.” A federal district court for the Northern District of Illinois had no difficulty in disposing of Mrs. Olech’s action. The court cited Esmail v. Macrane, an earlier case from the Seventh Circuit that the Olech district court read as requiring an “orchestrated campaign[] of official harassment directed against [a plaintiff] out of sheer malice” in order for a plaintiff to prevail. The court concluded that Mrs. Olech’s complaint did not describe the “malignant animosity” or the “orchestrated campaign of official harassment” required by Esmail. Her assertions of vindictiveness and retaliation were mere “conclusory assertions” that did not include factual underpinnings sufficient to support them.

122. Id. at *2.
123. Id.
124. Id. at *3–4.
125. See id. at *2.
126. Id.
127. 53 F.3d 176 (7th Cir. 1995).
129. Id.
130. Id.
131. Id.
Classes or Persons?

2. The Seventh Circuit Reverses on the Basis of Judge Posner’s Vindictive Action Theory of Equal Protection

Mrs. Olech appealed to the U.S. Court of Appeals for the Seventh Circuit and Chief Judge Posner wrote the opinion for that court.\(^{132}\) It turns out that, at least from Mrs. Olech’s perspective, Judge Posner was exactly the right person to consider her claim.

a. Earlier “Class of One” Claims in the Seventh Circuit

For several years before Olech, Judge Posner had been arguing in favor of a particular view of equal protection that was not shared by all the judges of the Seventh Circuit.\(^ {133}\) The particular issue was the nature of the “class of one” claim by an individual against the government. While it had been clear since Ciechon v. City of Chicago\(^ {134}\) that “class of one” equal protection claims were recognized in the Seventh Circuit, the conceptual underpinning of these claims was not at all clear. In Ciechon, the Seventh Circuit upheld the equal protection claim of a paramedic who had been discharged for the allegedly inadequate treatment of a patient, while the paramedic who had been working with her and had been equally responsible for the patient was not punished.\(^ {135}\) The court explained that “[e]qual protection demands at a minimum that a municipality must apply its laws in a rational and nonarbitrary way.”\(^ {136}\) The effect of this rule was to “protect[] against intentional invidious discrimination by the state against persons similarly situated.”\(^ {137}\)

The Ciechon decision appeared to require a successful claimant to prove, as Ciechon herself had done, that she had been treated differently from a similarly situated person. In addition, a claimant would have to prove that this different treatment was “intentional” as opposed to being the result of “error or mistake in the application of the law.”\(^ {138}\)

\(^{132}\) Olech v. Vill. of Willowbrook, 160 F.3d 386 (7th Cir. 1998).

\(^{133}\) This view of equal protection that differed from Judge Posner’s is set forth infra notes 140–54 and accompanying text.

\(^{134}\) 686 F.2d 511 (7th Cir. 1982). See also Levenstein v. Salafsky, 164 F.3d 345, 353 (7th Cir. 1998) (“So called ‘class of one’ equal protection claims, cases ‘in which the governmental body treated individuals differently who were identically situated in all respects rationally related to the government’s mission’ have been allowed in this circuit since at least Ciechon.”) (citation omitted).

\(^{135}\) Ciechon, 686 F.2d at 516–17.

\(^ {136}\) Id. at 522.

\(^{137}\) Id. at 522–23.

\(^{138}\) See id.
this, the *Ciechon* court made no mention of any requirement of vindictiveness or ill will, unless one could read such a requirement into a phrase the court did use, “intentional invidious discrimination.” But several post-*Ciechon* cases from the Seventh Circuit specifically held, not only that proof of animosity was not necessary to make out an equal protection violation, but also that it was not sufficient. For example, in *Falls v. Town of Dyer*, Judge Easterbrook upheld the equal protection claim of a convenience store owner who had alleged that he was “the only person against whom the town enforce[d] [its] portable sign ordinance.” According to Judge Easterbrook, if the plaintiff could show a pattern of selectivity under which everyone else was allowed to use portable signs but the plaintiff was not, that would make out a valid claim. On the other hand, according to Judge Easterbrook, even if the plaintiff could prove “that someone in local government had a vendetta against [plaintiff] but that the law is being enforced rationally against others,” he would not survive a motion for summary judgment. Thus, for Judge Easterbrook, the essence of a “class of one” claim was the different treatment of a single individual compared with everyone else, without regard to the motivation behind that treatment.

A subsequent opinion by the Seventh Circuit made it even clearer that a vindictive motivation by a government official did not make out an equal protection claim. In *Wroblewski v. City of Washburn*, the Seventh Circuit rejected the plaintiff’s “class of one” claim arising out of allegations that the city had singled out the plaintiff arbitrarily and made his employment at a city marina virtually impossible. According to the court, the plaintiff’s allegation that he was singled out by the local officials out of animosity did not make out a valid equal protection claim. For purposes of a motion to dismiss, even if the allegation that

139. *Id.*
140. 875 F.2d 146 (7th Cir. 1989).
141. *Id.* at 147 (emphasis in original).
142. See *id.* at 149.
143. *Id.*
144. 965 F.2d 452 (7th Cir. 1992).
145. *Id.* at 453.
146. See *id.* at 459 (“He simply alleges that he, and he alone, was singled out by the City Parents out of animosity and for no good reason. His equal protection claim would fail under the principle stated in *New Burnham Prairie Homes.*” (New Burnham Prairie Homes Inc. v. Vill. of Burnham, 910 F.2d 1474 7th Cir. (1990))). The court in *Wroblewski* had earlier cited *New Burnham* for the proposition that “an equal protection claim requires discrimination because of membership in a
the city acted out of animosity were assumed to be true, it would still be “insufficient to defeat the City policy’s presumed rationality.”147 “[A]nimosity is not necessarily inconsistent with a rational basis.”148 It might be rational for the city to conclude that the animosity it felt toward the plaintiff, even if not the fault of the plaintiff, would be “likely to doom any future association to failure.”149 Thus, for the Seventh Circuit in Wroblewski, the essence of an equal protection claim had nothing to do with subjective ill will; rather, it required an allegation of different treatment of similarly situated persons.150 Since the plaintiff in Wroblewski had failed to identify anyone similarly situated to himself who had been treated differently, he did not make out a successful equal protection claim.

Another Seventh Circuit panel also rejected a vindictive action equal protection claim. In Herro v. City of Milwaukee,151 the plaintiff had alleged that the city had denied him a tavern license because of the city’s animosity toward him.152 The court found the claim wanting, at least in part because “claims of state action motivated by personal vendettas ‘are hardly the type of allegations necessary to sustain an equal protection claim.’”153 The Seventh Circuit in Herro cited with approval a case that had dismissed an equal protection claim when it had been based only “on allegations that [a] state official engaged in vendetta to destroy plaintiff.”154

b. Judge Posner’s Vindictive Action Theory

These four cases served as background for Judge Posner when he entered the arena in 1995 to write the opinion for the Seventh Circuit in Esmael v. Macrane.155 Judge Posner upheld an equal protection claim by a liquor store owner whose license was not renewed because of a “deep-

---

147. Id. at 965 F. 2d. at 458–59 (citing New Burnham, 910 F.2d at 1481–82) (emphasis in Wroblewski only).
148. Id. at 460.
149. Id.
150. Id. at 459.
151. 44 F.3d 550 (7th Cir. 1995).
152. Id. at 552.
154. Id.
155. 53 F.3d 176 (7th Cir. 1995).
seated animosity"\textsuperscript{156} toward him by city officials. Judge Posner conceded that “[t]his is an unusual kind of equal protection case”\textsuperscript{157} because it involved neither a charge of singling out members of a vulnerable group nor a challenge to a law alleged to make irrational distinctions.\textsuperscript{158} Rather, the claim involved only a charge “that a powerful public official picked on a person out of sheer vindictiveness.”\textsuperscript{159} Although the case bore similarities to those of “selective prosecution,”\textsuperscript{160} Judge Posner pointed out that selective enforcement of the law is not an equal protection problem unless “the decision to prosecute is made either in retaliation for the exercise of a constitutional right . . . or because of membership in a vulnerable group.”\textsuperscript{161} Esmail’s claim did not meet any of these criteria, but Judge Posner found that he had made out a claim nonetheless.

According to Judge Posner, “equal protection does not just mean treating identically situated persons identically.”\textsuperscript{162} Judge Posner cited \textit{Ciechon} to support the rule that conduct motivated by vindictiveness toward a particular individual violates the equal protection clause.\textsuperscript{163} This was the correct view, according to Judge Posner, because “[i]f the power of the government is brought to bear on a harmless individual merely because a powerful state or local official harbors a malignant animosity toward him, the individual ought to have a remedy in federal court.”\textsuperscript{164} Although the unfair treatment was directed at an individual rather than a group, this fact was not determinative because “neither in terms nor in interpretation is the [equal protection] clause limited to protecting members of identifiable groups.”\textsuperscript{165} Thus, Esmail had satisfied his burden by alleging that “the action taken by the state . . . was a spiteful effort to ‘get’ him for reasons wholly unrelated to any legitimate state objective.”\textsuperscript{166}

\textsuperscript{156} Id. at 178.
\textsuperscript{157} Id.
\textsuperscript{158} Id.
\textsuperscript{159} Id.
\textsuperscript{160} Id. at 178–79 (explaining the two meanings of the term “selective prosecution” but concluding that the plaintiff’s claim in \textit{Esmail} was “not pleaded as a case of selective prosecution in any of the above senses”).
\textsuperscript{161} Id. at 179.
\textsuperscript{162} Id.
\textsuperscript{163} Id.
\textsuperscript{164} Id.
\textsuperscript{165} Id. at 180.
\textsuperscript{166} Id.
c. Reconciling Vindictive Action Claims with Comparative Right Claims

However, even if one agrees with Judge Posner that vindictively-motivated conduct directed at an individual by a government official “ought to have a remedy in federal court,”167 it is not immediately obvious why the source of this remedy ought to be the Equal Protection Clause. Although Judge Posner had conceded that “[t]his is an unusual kind of equal protection case,”168 he understated how unusual it was, because there is ordinarily widespread agreement that equal protection arguments, whether they involve classifications or not, are inherently comparative.169 A claimant insists that he or she has a right to a certain treatment because someone else, allegedly similarly situated, is getting that same treatment. Judge Posner’s vindictive action claims involve no comparison, but rather simply a claim that the government has treated one person badly because of subjective ill will.170

However, there are two ways in which the vindictive action claim might be reconciled with the traditional notion of equal protection as a comparative right. The first is to argue that a vindictive action claim assumes an implied comparison with a class or person, that is, there is an implied comparison with the group or person whom the government has treated rationally, without any ill will.171 By means of this implied comparison to this implied group, one could argue that a vindictive action claim is in fact a comparative claim. However, the rather large amount of implication that this defense requires makes it not particularly compelling.

167. Id. at 179.
168. Id. at 178.
169. See, e.g., Buckles v. Columbus Mun. Airport Auth., 2002 WL 193853, at *13 (S.D. Ohio Jan. 14, 2002) (“Nevertheless, even a class of one must show that he or she was treated differently than similarly situated individuals. An equal protection claim simply cannot exist absent an allegation that, compared to others, the plaintiff was treated less favorably.”) (emphasis in original).
170. E.g., Esmail, 53 F.3d at 180 (indicating that, to establish a violation of the Equal Protection Clause, a claimant must prove that “the action taken by the state . . . was a spiteful effort to ‘get’ him for reasons wholly unrelated to any legitimate state objective”).
171. See, e.g., Kiser v. Naperville Cnty. Unit, 227 F. Supp. 2d 954, 972 (N.D. Ill. 2002) (“[T]he District argues that [the plaintiff, a lawyer] cannot state an equal protection claim because he did not allege that other lawyers were employed by the District. The potential relevance of other lawyers is obvious from the complaint, so the District can reasonably foresee a subsequent allegation that such lawyers existed. For this reason, the omission of a specific allegation of fact concerning other lawyers in the complaint does not mandate dismissal.”).
An alternative justification has some support from Tussman and tenBroek and the U.S. Supreme Court. As Tussman and tenBroek pointed out, one cannot tell who is similarly situated to whom without reference to the purpose of a law, and any useful consideration of purposes must close off certain purposes as impermissible, so that the constitutional standard requires a classification to be rationally related to a permissible purpose. It is not a large step from here to conclude that government action designed to achieve an impermissible purpose violates the Equal Protection Clause. The Supreme Court has taken this very step in two cases where it analyzed a neutral rule that was improperly applied because of an inappropriate motivation.

In the first of these cases, *Yick Wo v. Hopkins*, decided more than one hundred years before *Olech*, the Court found a violation of the Equal Protection Clause in the manner in which the San Francisco Board of Supervisors applied an ordinance, neutral on its face, that required a person operating a laundry to obtain the consent of the Board unless the laundry was located in a building made of brick or stone. The Board had rejected the applications of all Chinese applicants but had consented to all others. As the Court explained, the problem did not involve a general rule “for the regulation of the use of property for laundry purposes, to which all similarly situated may conform.” No one would have a constitutional problem with a general rule about wooden or brick buildings for laundries. But the problem in *Yick Wo* was not the rule itself, but rather its application. Although the law itself was impartial on its face, it was applied “with a mind so unequal and oppressive as to amount to a practical denial by the state of . . . equal protection.” The law was “applied and administered by public authority with an evil eye and an unequal hand, so as practically to make unjust and illegal discriminations between persons in similar circumstances.”

Although the oppressive application of the laundry rule in *Yick Wo* was directed at a class of Chinese persons rather than one individual

---

172. See supra notes 15–18, 25–26 and accompanying text.
173. 118 U.S. 356 (1886).
174. Id. at 357.
175. Id. at 359.
176. Id. at 368.
177. Id. at 373.
178. Id. at 373–74.
person, the case made clear that the Equal Protection Clause was broad enough to prohibit not only facially invalid classifications appearing in statutes, but also the oppressive application of impartial rules to individual persons. The determining factor was the presence of the evil eye, the unequal hand, or the oppressive mind. These factors seem to cover much of the same territory as Judge Posner’s “subjective ill will” theory.

Additionally, there is a much more recent precedential case on point. In 1985, in City of Cleburne v. Cleburne Living Center, Inc., the Court found an equal protection violation in the city’s refusal to grant a special use permit for a group home for the mentally retarded. The Court found it unnecessary to decide whether the regulation requiring a special use permit for “hospitals for the insane or feebleminded” was facially invalid since the Court found it easier to conclude that the regulation was unconstitutional as applied. The Court sifted through the city’s purported justifications for denying the permit and found them all either impermissible or not credible. It then concluded that the denial of the special use permit to this particular home was unconstitutional because it could be explained only by “an irrational prejudice against the mentally retarded, including those who would occupy the [group home].” This “as applied” version of equal protection that looks to the motivation behind the administrative application of a local regulation also appears to

---

179. Id. at 359 (indicating that 150 natives of China had been arrested for operating a laundry without the required consent while 80 persons not subjects of China were allowed to operate their laundries without the required consent, and that these numbers demonstrate “a system of oppression of one kind of men and favoritism to all others”).

180. Id. at 373–74 (referring to “a mind so unequal and oppressive” and “an evil eye and an unequal hand”).


182. Id. at 435.

183. Id. at 436.

184. Id. at 447 (“We inquire first whether requiring a special use permit for the Featherstone home in the circumstances here deprives respondents of the equal protection of the laws. If it does, there will be no occasion to decide whether the special use permit provision is facially invalid where the mentally retarded are involved, or to put it another way, whether the city may never insist on a special use permit for a home for the mentally retarded in an R-3 zone. This is the preferred course of adjudication since it enables courts to avoid making unnecessarily broad constitutional judgments.”). Having determined that it was preferable to decide the constitutionality of the ordinance on an as applied basis, the Court then found it unconstitutional on that basis. Id. at 450.

185. Id. at 448–50.

186. Id. at 450.
be very similar to Judge Posner’s “subjective ill will.”\textsuperscript{187} It is also quite clearly a precedent for the conclusion that the protection of the Equal Protection Clause extends to individuals harmed by administrative decisions, without the necessity of showing that the harm results from an improper classification. In any case, Judge Posner’s view of the equal protection clause as prohibiting government action that is motivated by ill will, while not in the mainstream of equal protection arguments, does have some support from the U.S. Supreme Court.

It was with this background that Judge Posner wrote the opinion for the Seventh Circuit in \textit{Olech}. For Judge Posner, his previous opinion in \textit{Esmail} was dispositive. Judge Posner in \textit{Esmail} had found a constitutional violation when there was proof of “a spiteful effort to ‘get’ [a person] for reasons wholly unrelated to any legitimate state objective.”\textsuperscript{188} Judge Posner found that Olech’s complaint satisfied this standard in that it alleged that her previous lawsuit against the Village had generated “substantial ill will” and that, as a result, she had been treated differently from other property owners in the village in terms of the size of the easement demanded.\textsuperscript{189} Judge Posner noted that the District Court had read too much into \textit{Esmail}’s reference to an “orchestrated campaign of official harassment,”\textsuperscript{190} since, according to Judge Posner, a showing of subjective ill will does not require evidence of a “general” orchestration.\textsuperscript{191} Judge Posner concluded that he was troubled “by the prospect of turning every squabble over municipal services, of which there must be tens or even hundreds of thousands every year, into a federal constitutional case,”\textsuperscript{192} but he was confident that the requirement of proving “totally illegitimate animus” toward the plaintiff would be a substantial check on the likely success of any such claim.\textsuperscript{193}

\textsuperscript{187} Cf. \textit{Esmail v. Macrane}, 53 F.3d 176, 180 (7th Cir. 1995) (indicating that an Equal Protection Clause claimant must prove that “the action taken by the state . . . was a spiteful effort to ‘get’ him for reasons wholly unrelated to any legitimate state objective”).

\textsuperscript{188} \textit{Olech v. Vill. of Willowbrook}, 160 F.3d 386, 387 (7th Cir. 1998) (citing \textit{Esmail v. Macrane}, 53 F. 3d 176, 180 (7th Cir. 1995)).

\textsuperscript{189} \textit{Id.} at 387–88.

\textsuperscript{190} \textit{Id.} at 388 (citing \textit{Esmail}, 53 F.3d at 179).

\textsuperscript{191} \textit{Id.} at 388.

\textsuperscript{192} \textit{Id.}

\textsuperscript{193} \textit{Id.}
B. The U.S. Supreme Court Endorses Mrs. Olech’s Individual Equal Protection Claim

It may have been expected that the decision of the Seventh Circuit would end this “garden variety” land-use dispute, but the Village sought certiorari in the U.S. Supreme Court, and, surprisingly, the Court decided to hear the case.\textsuperscript{194} Also surprising had to be the Supreme Court opinion that followed,\textsuperscript{195} particularly its brevity and the casualness with which the Court treated the problem before it. The entire per curiam opinion consisted of five paragraphs.\textsuperscript{196} The first three paragraphs summarized the facts and procedural history below, and then concluded with the Court’s identification of the issue as “whether the Equal Protection Clause gives rise to a cause of action on behalf of a ‘class of one’ where the plaintiff did not allege membership in a class or group.”\textsuperscript{197} The Court’s entire substantive response to this question consisted of exactly two short paragraphs.

In the first of these paragraphs, the Court stated that “[o]ur cases have recognized successful equal protection claims brought by a ‘class of one,’ where the plaintiff alleges that she has been intentionally treated differently from others similarly situated and there is no rational basis for the difference in treatment.”\textsuperscript{198} This statement of the “class of one” problem was obviously different from the manner in which Judge Posner had posed it below. Judge Posner had been concerned with the allegation that the Village had vindictively singled out Mrs. Olech in retaliation for a previous lawsuit.\textsuperscript{199} Judge Posner’s theory requires a plaintiff to produce evidence of the defendant officials’ subjective motivation, but not necessarily evidence of similarly situated persons who were treated differently. The Supreme Court’s explanation, on the other hand, requires a plaintiff to produce evidence that similarly situated persons were treated differently, but not any evidence of the defendant’s subjective motivation for the conduct being challenged.\textsuperscript{200} The Supreme Court

\begin{footnotesize}
\begin{itemize}
  \item \textsuperscript{194} Vill. of Willowbrook v. Olech, 527 U.S. 1067 (1999) (mem.).
  \item \textsuperscript{195} Vill. of Willowbrook v. Olech, 528 U.S. 562 (2000).
  \item \textsuperscript{196} Id. at 563–65.
  \item \textsuperscript{197} Id. at 564.
  \item \textsuperscript{198} Id.
  \item \textsuperscript{199} Olech v. Vill. of Willowbrook, 160 F.3d 386, 387 (7th Cir. 1998).
  \item \textsuperscript{200} See Olech, 528 U.S. at 564 (recognizing an equal protection claim where a plaintiff alleges different treatment from others similarly situated); id. at 565 (indicating that the Court’s decision was not based on the theory of “subjective ill will”).
\end{itemize}
\end{footnotesize}
explicitly distanced itself from the “subjective ill will” theory that had been relied on by Judge Posner below.\footnote{Id. at 565.} 

1. The U.S. Supreme Court’s Surprising Citations in Support of its Holding

The U.S. Supreme Court cited two cases in support of its “class of one” holding: \textit{Sioux City Bridge Co. v. Dakota County},\footnote{260 U.S. 441 (1923).} a largely ignored case from 1923, and \textit{Allegheny Pittsburgh Coal Co. v. County Commission of Webster County},\footnote{488 U.S. 336 (1989).} a more recent case that the Court had previously attempted to limit to its facts. These two supporting cases were as far removed from the pantheon of influential equal protection cases as one could imagine.\footnote{In the fifty years before \textit{Olech}, the Supreme Court had cited \textit{Sioux City Bridge} in the text of a majority opinion only one time. That single citation was in \textit{Allegheny Pittsburgh Coal}, the case that the Court had later limited to its facts. In \textit{Nordlinger v. Hahn}, 505 U.S. 1 (1992), the petitioners argued that the result in \textit{Allegheny Pittsburgh Coal} required invalidation of a very similar property tax assessment scheme in their case. The \textit{Nordlinger} Court rejected the argument, purporting to find “an obvious and critical factual difference,” \textit{id.} at 14, and concluding that \textit{Allegheny Pittsburgh Coal} “was the rare case where the facts precluded any plausible inference that the reason for the unequal assessment practice was to achieve the benefits of an acquisition-value tax scheme.” \textit{id.} at 16. See also Robert C. Farrell, \textit{Successful Rational Basis Claims in the Supreme Court from the 1971 Term Through Romer v. Evans}, 32 IND. L. REV. 357, 404–05 (1999) (arguing that “[t]he effect of \textit{Nordlinger} . . . is to limit \textit{Allegheny Pittsburgh Coal} to its very narrow set of facts”).} However, both of these supporting cases involved successful attempts by individual plaintiffs who challenged excessive property tax appraisals on real property when other similar properties had been appraised at comparatively lower rates.

In \textit{Sioux City Bridge}, the Court ruled that a bridge company had proven a violation of the Equal Protection Clause when it demonstrated that its property was assessed at one hundred percent of its valuation while other properties were assessed at fifty-five percent of their valuations.\footnote{Sioux City Bridge, 260 U.S. at 445.} The Court then explained that “the purpose of the [E]qual [P]rotection [C]lause . . . is to secure every person within the State’s jurisdiction against intentional and arbitrary discrimination, whether occasioned by the express terms of a statute or by its improper execution through duly constituted agents.”\footnote{Id. (citing Sunday Lake Iron Co. v. Wakefield Tpk., 247 U.S. 350, 352 (1918)).} The Court emphasized that it was
“intentional systematic undervaluation” that would violate the Equal Protection Clause, not “mere errors of judgment.” The assessment in Sioux City Bridge was viewed, not as an example of arbitrary discrimination by the legislature in the enactment of a statute, but rather of intentional discrimination by a local board, the “duly constituted agents,” in setting the tax value for one particular piece of property.

Allegheny Pittsburgh Coal involved a coal company that complained about an acquisition value assessment scheme that had the effect of appraising its property, based on its recent purchase price, at a level that was thirty-five times the level applied to owners of comparable property who had purchased their property at earlier times. In concluding a violation of the Equal Protection Clause, the Court once again emphasized “intentional systematic undervaluation” of comparable property and concluded that “[t]he equal protection clause . . . protects the individual from state action which selects him out for discriminatory treatment by subjecting him to taxes not imposed on others of the same class.”

Although no property owner had an independent right to have his or her individual property assessed at less than its fair market value, the Equal Protection Clause conferred on each property owner a comparative right to have his or property assessed at the same standard as other owners, even if that standard resulted in a valuation substantially below market value.

Although Sioux City Bridge and Allegheny Pittsburgh Coal, broadly interpreted, provide some support for the Court’s holding in Olech, they do not do so with much force. They were, of course, tax cases, not land use cases as Olech was. Although both cases involved a single, individual plaintiff, a “class of one” whose equal protection claim succeeded, neither of the decisions explicitly alluded to that fact. Nor had the Supreme Court, before Olech, ever cited them as “class of one” cases. And it certainly strains credulity to claim, as the Court did, that the

207. Id.
208. Id. at 447.
209. Id. at 445.
211. Id. at 345–46 (citing Hillsborough v. Cromwell, 326 U.S. 620, 623 (1946)).
212. Id. at 346 (indicating that “the fairness of one’s allocable share of the total property tax burden can only be meaningfully evaluated by comparison with the share of others similarly situated relative to their property holdings” and concluding that “relative undervaluation” of comparable property . . . over time therefore denies petitioners the equal protection of the law”) (emphases added).
“class of one” equal protection claim had been “recognized” since these two cases had been decided.

But in any case, having determined that these two cases provided adequate support for its holding, the Olech Court found that Mrs. Olech’s complaint stated a valid claim. Her complaint was sufficient because she had alleged “that the Village intentionally demanded a 33-foot easement as a condition of connecting her property to the municipal water supply where the Village required only a 15-foot easement from other similarly situated property owners,”213 and that this demand was “irrational and wholly arbitrary.”214 In support of its conclusion that Mrs. Olech’s complaint could “fairly be construed”215 as making adequate allegations, the Court cited Conley v. Gibson,216 a case that had adopted an extremely forgiving standard in evaluating plaintiffs’ complaints.217 The Court concluded its opinion by making clear that its decision was not based on the Seventh Circuit’s “subjective ill theory” of equal protection, but had been decided on the “similarly situated” standard, “quite apart” from any reliance on the village’s subjective motivation as a basis for its decision.218

Justice Breyer wrote a brief concurring opinion in Olech in which he expressed that, without a limiting principle, the majority’s opinion could “transform many ordinary violations of city or state law into violations of the constitution.”219 Breyer argued that an appropriate limiting principle could be found, and that principle was precisely the “subjective ill will” theory adopted by the court of appeals, but rejected by the majority.220 Breyer concurred only because the presence of that “added factor” (ill will) would “minimize any concern about transforming run-of-the-mill zoning cases into cases of constitutional right.”221

214. Id.
215. Id. at 565.
217. Olech, 528 U.S. at 565 (citing Conley, 355 U.S. at 41, 45–46, which had approved the “accepted rule that a complaint should not be dismissed for failure to state a claim unless it appears beyond doubt that the plaintiff can prove no set of facts in support of his claim that would entitle him to relief”).
218. Id.
219. Id. (Breyer, J., concurring).
220. Id. at 565.
221. Id. at 566.
2. Ignoring Precedential Cases and a Possible Explanation

The Olech Court’s choice of precedents to ignore is as revealing as its choice of precedents to cite. Once the Court had cited two cases in support of its substantive position, it did not identify or distinguish any cases that might have suggested a different result. However, if the Court had been looking for a third property tax assessment case beyond the two it cited, it might have mentioned Nordlinger v. Hahn. That case would have made it harder for the Court to explain its result in Olech. In Nordlinger, the individual plaintiff was badly mistreated in relation to her neighbors in her property tax assessment and paid property taxes almost five times as high as some of her neighbors. Yet the Court upheld that different treatment because it was the result of a reasonable system of classification. In Nordlinger, the Court demonstrated no concern for the individual person unfairly treated. Likewise, the Court in Massachusetts Board of Retirement v. Murgia showed no identifiable concern for the individual person adversely affected by governmental decisionmaking. Surely Robert Murgia would take issue with the assertion that, under Supreme Court precedents, the Equal Protection Clause protects individuals from arbitrary treatment.

However, a close reading of the Court’s citation to Sioux City Bridge suggests an explanation for the apparently inconsistent results. There, the Court explained that the Equal Protection Clause prohibits intentional and arbitrary discrimination “whether occasioned by express terms of a statute or by its improper execution through duly constituted agents.” This statement explicitly identifies two different forms of state action subject to the Equal Protection Clause: legislation, on the one hand, and “execution,” on the other. It also implicitly suggests a different equal protection standard to be applied to each form of state action that is consistent with both views of equal protection. The different standards would result because of the essential difference between government legislating, on the one hand, and government acting on the other, through

---

224. Id. at 7.
225. Id. at 11–13.
226. See supra notes 34–42 and accompanying text.
227. See supra notes 205–09 and accompanying text.
228. Id.
individual decisions of government officials that assign government benefits or impose government burdens on individual persons.

Most of the cases in which the Court has spoken of equal protection as a limitation on classification have in fact involved legislative classification.229 When legislatures enact statutes, they ordinarily make use of broad generalizations about large numbers of people.230 They assume that people who have a certain trait, like age or gender, are similarly situated in relation to a particular purpose.231 In short, legislatures when legislating usually classify.232 But if it is generally accurate to assert that legislatures ordinarily classify, it follows that the Equal Protection Clause imposes limits on that process of classification.233 However, in this context, it does no more than that. Individual persons who are harmed by legislative generalizations that are not true as applied to them have no remedy, unless the classification itself is unreasonable.234 In these situations, equal protection is not an individual right.

It is quite a different story, on the other hand, when government officials make any of their millions of individual determinations daily. These include some of the most basic decisions involved in running a government, such as who gets hired for a government job, who gets fired from a government job, who gets a zoning waiver, who gets a government contract, who gets a subdivision approval, who gets a building permit, who gets a government grant, or who gets arrested. These governmental decisions are not legislative and do not amount to


230. E.g., Murgia, 427 U.S. 307 (reviewing statute that accepts generalization that fitness declines with age).

231. E.g., (statute assumes that those over 50 are similarly situated in respect to fitness); Frontiero v. Richardson, 411 U.S. 677 (1973) (plurality opinion) (statute assumes that military wives are financially dependent on their spouses but that military husbands are not financially dependent on their spouses).

232. See supra notes 8–23 and accompanying text. It is not unheard of for a legislature to enact a law directed at one particular person or entity, rather than at a class, but on the few occasions that legislatures do this, there are provisions of the Constitution other than the Equal Protection Clause that limit legislative action. See Consol. Edison Co. of N.Y. Inc. v. Pataki, 292 F.3d 338 (2d Cir. 2002) (holding New York statute explicitly prohibiting Consolidated Edison Co. from recovering from ratepayers costs associated with outage at Indian Point 2 Nuclear Facility violated Bill of Attainder Clause of the U.S. Constitution).

233. See supra Part I.

234. See supra notes 24–26 and accompanying text.
broad generalizations about large numbers of persons. These are
dividual decisions. And here, according to Olech, the Equal Protection
Clause creates a personal right.  

In fact, upon careful observation, it must be admitted that one need not
be engaged in the act of classifying at all in order to call the Equal
Protection Clause into play. Because equality arguments are inherently
comparative, there must be at least two persons. However, the basic
mandate of equality that requires similarly situated persons to be treated
similarly does not require classifications. If a parent has only two
children and treats them differently, then that can amount to inequality
without reference to classifications. If a local government treats two
similarly situated neighbors differently in the size of its demand for an
easement, that can constitute inequality. In neither of these cases need a
plaintiff identify a class in order to make an equality claim. A plaintiff
need only prove that he or she was treated differently from one similarly
situated person. In such a case, equal protection is a personal right and
there should be no necessity of referring to a “class of one” because there
is no necessity of identifying a classification.

The Court’s opinion in Olech clearly supports such a conclusion, but
the Court did not make clear the conceptual underpinning of that
conclusion. While the result in the case suggests that an individual
plaintiff harmed by an unequal administrative decision has an equal
protection claim without proving membership in a class, the Court’s
opinion suggests a slightly different explanation. The language chosen by
the Court in Olech suggests that it is necessary to identify a class, but
that the class can consist of only one member. Although these two
alternate formulations of the equal protection right recognized in Olech
probably do not create practical differences, it is ironic that the Olech
opinion, in the act of providing the strongest support for an individual
rights view of equal protection, does so by using the traditional language
of equal protection as a limit on classifications. Although the Supreme
Court in Olech purported to recognize an existing right and thus not
change existing law, the opinion was far more transformative than the
Court intimated. Olech in fact has had a dramatic effect on subsequent
litigation in federal courts. The next section examines these effects.

236. Id. at 564 (“Our cases have recognized successful equal protection claims brought by a ‘class
    of one.’”).
237. Id.
IV. THE AFTERMATH OF OLECH

After the U.S. Supreme Court’s decision in Olech, it was clear that the Equal Protection Clause does protect individual rights, at least in an appropriately limited context.238 But what was the exact nature of that right? According to the Court, an individual has a valid claim under the Equal Protection Clause when “she has been intentionally treated differently from others similarly situated and there is no rational basis for the difference in treatment.”239 This is the “similarly situated” equal protection claim. In addition, according to the Seventh Circuit and several other courts of appeals, an individual also has a valid claim under the Equal Protection Clause when a government acts vindictively toward that person, with the motivation to “get” him.240 This is the “vindictive action” equal protection claim. Although the Supreme Court did not affirm the “vindictive action” portion of the Seventh Circuit’s Olech opinion, it did not overrule it either.241 The following section examines how lower federal courts treated the Supreme Court’s Olech decision and the “vindictive action” equal protection claim.

A. Did Olech Change Existing Law?

In Olech, the U.S. Supreme Court carefully explained that its decision validating the “class of one” equal protection claim was nothing new. According to the Court, “[o]ur cases have recognized successful equal protection claims brought by a ‘class of one.’”242 The Court purported to identify two cases that supported this assertion, one of them decided in 1923.243 Surely then, the Court was of the view that it was simply restating existing law. Although ultimately that claim is defensible, it is not without its problems. First, the Court made no effort to distinguish those cases where it spoke of equal protection as a limitation on government classifications, with no mention of any protection of individual rights. Nor did the Court explain why in those cases the harm

238. Although the Equal Protection Clause does not protect individuals from adverse affects of reasonable legislative classifications, it does protect an individual from an arbitrary individual administrative decision by a government official. See supra Part III.B.
240. See supra Part III.A.2.
241. Olech, 528 U.S. at 565 (“[W]e therefore affirm the judgment of the Court of Appeals, but do not reach the alternative theory of ‘subjective ill will’ relied on by that court.”).
242. Id. at 564 (emphasis added).
243. Id. (citing Sioux City Bridge Co. v. Dakota County, 260 U.S. 441 (1923)).
to individuals was considered unimportant. Second, the Court did not explicitly state that its “class of one” holding was limited to the execution of law by government officials but did not extend to the enactment of laws by legislatures. However, if the holding of the case is appropriately limited, then Olech can be viewed as consistent with previous understandings of the Equal Protection Clause.

The issue of the consistency of Olech with pre-existing precedent arises explicitly in lower federal courts when defendants raise the defense of qualified immunity. This occurs when plaintiffs seek money damages from individual government defendants for violations of constitutional rights. In these cases, the Supreme Court has announced a rule that “government officials performing discretionary functions generally are shielded from liability for civil damages insofar as their conduct does not violate clearly established statutory or constitutional rights of which a reasonable person would have known.” When a plaintiff seeks damages for a “class of one” claim that arose before February 23, 2000, the date of the Supreme Court opinion in Olech, it then becomes necessary to determine whether the “class of one” claim was “clearly established” before that date.

Of the few lower federal court cases that have explicitly considered this question, there is support for the conclusion that Olech did not change existing law. When the Olech litigation itself was remanded to a federal district court in the Northern District of Illinois, that court addressed the issue of qualified immunity for the defendants and whether or not the “class of one” equal protection claim was “clearly established” before 2000. The lower court answered in the affirmative, citing both Seventh Circuit precedent and Sioux City Bridge and Allegheny Pittsburgh Coal. In McWaters v. Rick, a federal district court for the Eastern District of Virginia found that Olech “confirmed that government officials constitutionally may not apply a law arbitrarily to . . . a person who is similarly situated to others to whom the law has been applied.” Thus, the holding in Olech “was not based on novel legal theory . . . [but]
was confirming and clarifying, rather than revealing for the first time, that such a ‘class of one’ claim is constitutionally cognizable.”

But there is some dissent on this point. In *Taylor v. Russell*, a federal district court for the Eastern District of Texas insisted that, “[p]rior to Village of Willowbrook, it was generally understood in this [Fifth] Circuit that an equal protection claim could only succeed if the disparate treatment at issue was premised upon a person’s membership in a protected class or invocation of a constitutional right.” According to *Taylor*, it was not at all clear before *Olech* that a claim of personal vindictiveness “would be enough to support an equal protection claim without some other class-based discrimination.” Further, in *Anderson v. Anderson*, a federal district court for the Northern District of Ohio found that “the *Olech* opinion invalidates the Futernick rule that cognizable equal protection claims must be based on class-based or group-based treatment.” The federal district courts in these cases found that *Olech* had forced change in the law of their circuits. At the very least, then, it must be said that, even if the Supreme Court’s *Olech* opinion did not change the law of the Supreme Court itself, it certainly forced several lower federal courts to change their views of the Equal Protection Clause. Surprisingly, however, not every circuit appears to have received the message. As late as September 2002, more than two years after *Olech*, a panel in the Tenth Circuit stated in an unpublished opinion that, to establish an equal protection violation, “plaintiffs must show that they are members of a protected class and that the defendants purposefully discriminated against them because of their membership in that class.”

**B. Limiting the Effect of Olech**

Both Judge Posner and Justice Breyer expressed the fear that the Court’s *Olech* decision, unless properly limited, could open the

251. *Id.* at 672.
252. *Id.*
254. *Id.* at *4* (citing Summers v. City of Raymond, 105 F. Supp. 2d 549, 551–52 (S.D. Miss. 2000)).
255. Brown v. Millard County, 47 F. Appx. 882, 890 (10th Cir. 2002) (citing Jones v. Union County, 296 F.3d 417, 426 (6th Cir. 2002)). Judge Hartz, in a concurring opinion, disagreed on this point. *Id.* at 890–91 (Hartz, J., concurring).
floodgates in federal courts to a host of insignificant yet time-consuming lawsuits.\textsuperscript{256} In response to this concern, a number of federal courts have attempted to interpret \textit{Olech} in a way that would limit its effect. One such strategy was to read \textit{Olech} as requiring proof of vindictive motivation as well as different treatment of similarly situated persons.\textsuperscript{257} A second strategy to limit \textit{Olech} was to interpret the term "intentionally different treatment" very strictly.\textsuperscript{258} A third strategy was to interpret the phrase "similarly situated" very narrowly.\textsuperscript{259} The following section examines all three of these attempts to limit the effect of the Supreme Court’s \textit{Olech} opinion.

1. \textbf{Limiting Olech by Requiring Both Unequal Treatment and Subjective Ill Will}

   The U.S. Supreme Court’s opinion in \textit{Olech} was short and apparently rather simple, but some of the federal courts of appeal treated it like a complex puzzle, to be mined for hidden meaning. Almost immediately after it was reported, both the Seventh and Second Circuits engaged in what they must have viewed as damage control, that is, an attempt to limit \textit{Olech} so that it would not overrun the federal courts with garden variety disputes involving claims against local government.

   In \textit{Hilton v. City of Wheeling},\textsuperscript{260} the first appellate case to cite \textit{Olech}, Judge Posner once again wrote for the Seventh Circuit. In his Seventh Circuit \textit{Olech} opinion, Judge Posner had found the constitutional problem to be the subjective ill will that the town officials allegedly held toward Mrs. Olech. On the other hand, the Supreme Court decided the case on the ground that the town had intentionally treated Mrs. Olech different from similarly situated property owners without adequate justification.\textsuperscript{261} It has been noted that these two explanations of the result

\textsuperscript{256} Vill. of Willowbrook v. Olech, 528 U.S. 562, 565 (2000) (Breyer, J., concurring) ("The Solicitor General and the Village of Willowbrook have expressed concern lest we interpret the Equal Protection Clause in this case in a way that would transform many ordinary violations of city or state law into violations of the Constitution."); Olech v. Vill. of Willowbrook, 160 F.3d 386, 388 (7th Cir. 1998) ("Of course, we are troubled, as was the district judge, by the prospect of turning every squabble over municipal services, of which there must be tens or even hundreds of thousands every year, into a federal constitutional case.").

\textsuperscript{257} See infra Part IV.B.1.

\textsuperscript{258} See infra Part IV.B.2.

\textsuperscript{259} See infra Part IV.B.3.

\textsuperscript{260} 209 F.3d 1005 (7th Cir. 2000).

\textsuperscript{261} Vill. of Willowbrook v. Olech, 528 U.S. 562, 564 (2000).
in Olech are quite different, and the Supreme Court seemed to have adopted only one.

But Judge Posner was not willing to concede this point. In Hilton, Judge Posner addressed a claim of unequal provision of police protection as an equal protection violation. In rejecting that claim, Judge Posner explained that, while the police may have been inept or may have been deceived by the plaintiff’s neighbors, it did not matter. What mattered, said Judge Posner, “is the absence of evidence of an improper motive.” How could Judge Posner have read the Olech opinion to require evidence of improper motive? Judge Posner pointed to the language of the Supreme Court opinion indicating that the allegation of dissimilar treatment must include the claim that there is “no rational basis” for the difference in treatment and that the different treatment was “irrational and wholly arbitrary.” According to Judge Posner, this “no rational basis” language must necessarily include a gloss that “to make out a prima facie case the plaintiff must present evidence that the defendant deliberately sought to deprive him of the equal protection of the laws for reasons of a personal nature unrelated to the duties of the defendant’s position.” Judge Posner then cited his own Seventh Circuit Olech opinion, rather than the Supreme Court’s opinion, to show that what was required was “proof that the cause of the differential treatment of which the plaintiff complains was a totally illegitimate animus toward the plaintiff by the defendant.”

Judge Posner’s opinion in Hilton is surprising. First, it ignores the per curiam opinion of the Supreme Court, which explicitly stated that it did “not reach the alternative theory of ‘subjective ill will’ relied on by the [Seventh Circuit].” Second, it exalts Justice Breyer’s concurring opinion, which did require “ill will” as a component of a “class of one” claim, as well as Judge Posner’s previous Seventh Circuit opinion, to the status of Supreme Court majority opinion. Judge Posner’s reading of Olech is inconsistent with the per curiam opinion in that case. But whether grounded in precedent or not, Judge Posner’s view, both because

---

262. See supra notes 198–201 and accompanying text.
263. Hilton, 209 F.3d at 1006–07.
264. Id. at 1008.
265. Id.
266. Id.
268. Id. (citing Olech v. Vill. of Willowbrook, 160 F.3d 386, 388 (7th Cir.1998)).
269. Olech, 528 U.S. at 565.
of his scholarly reputation and his position as Chief Judge of the Seventh Circuit and because Hilton was one of the first cases to interpret Olech, gave that position a great deal of credibility and influence. This created problems for subsequent courts interpreting Olech.

It certainly created difficulties for the Seventh Circuit. Shortly after Hilton, a different panel of Seventh Circuit judges adopted apparently contradictory interpretations of Olech. The court initially held that a “class of one” equal protection claim could be made either by showing different treatment of similarly situated persons or by showing a “spiteful effort to ‘get’ [a person].” Later in the same opinion, the court also held that in order to make a “class of one” claim, a plaintiff must show different treatment of similarly situated persons and “totally illegitimate animus.” Two subsequent Seventh Circuit opinions approvingly cited Hilton as requiring proof of illegitimate animus in order to make out a claim under Olech. These three cases suggest that the Seventh Circuit is free to come up with its own interpretations of Supreme Court precedents.

However, more recent opinions have retreated from the more extreme view expressed in Hilton. For example, in Nevel v. Village of Schaumburg, the Seventh Circuit explicitly stated the two “class of one” tests disjunctively. In Nevel, the court concluded that, to succeed under Olech, the plaintiff must prove either intentionally different treatment of similarly situated persons or illegitimate animus toward the plaintiff. A subsequent district court opinion from the Seventh Circuit reviewed the confusion on the meaning of Olech and concluded that Nevel:

[S]tates the proper standards governing class of one equal protection claims. Nevel is more faithful to the Supreme Court’s opinion in Olech . . . . The Seventh Circuit’s subsequent attempts in Hilton and Purze to narrow the range of options available to class of one equal protection plaintiffs simply cannot be squared with Olech.

---

270. Albiero v. City of Kankakee, 246 F.3d 927, 932 (7th Cir. 2001).
271. Id.
272. Cruz v. Town of Cicero, 275 F.3d 579, 587 (7th Cir. 2001); Purze v. Vill. of Winthrop Harbor, 286 F.3d 452, 455 (7th Cir. 2002).
273. 297 F.3d 673 (7th Cir. 2002).
274. Id. at 681.
The Second Circuit has had its own problems in interpreting the meaning of the Supreme Court’s Olech opinion. Even before Olech, the Second Circuit had already developed a two-part test in cases of selective treatment that anticipated the alternative interpretations of Olech. In Leclair v. Saunders,276 the Second Circuit considered an allegation of an equal protection violation in the selective enforcement of a dairy farm regulation.277 The court explained that liability would depend on proof that “(1) the person, compared with others similarly situated, was selectively treated; and (2) that such selective treatment was based on impermissible considerations such as race, religion, intent to inhibit or punish the exercise of constitutional rights, or malicious or bad faith intent to injure a person.”278 This two-part test seems to be virtually the same conjunctive test used by Judge Posner in his interpretation of Olech. But how much of Judge Posner’s test has survived the Supreme Court’s Olech opinion?

The Second Circuit has had a difficult time in determining exactly how much is left of the LeClair standard after Olech. In three separate attempts at explaining the effect of Olech, the Second Circuit did not resolve the issue of whether a plaintiff must prove both elements of the LeClair standard or only one of them.279 In each of the cases, the court found that it did not need to decide that issue because in each case, the plaintiff was unable to prove either element.280 However, the most recent cases from the district courts within the Second Circuit now seem to be of the view that in order to make a “class of one” claim, a plaintiff must prove either intentionally different treatment of similarly situated persons or subjective ill will.281

276. 627 F.2d 606 (2d Cir. 1980).
277. Id. at 607–08.
278. Id. at 609–10.
279. Giordano v. City of New York, 274 F.3d 740 (2d Cir. 2001); Harlen Assocs. v. Vill. of Mineola, 273 F.3d 494 (2d Cir. 2001); Gelb v. Bd. of Elections, 224 F.3d 149 (2d Cir. 2000).
280. Giordano, 274 F.3d at 751 (“[W]e decline to resolve this issue because its resolution would not affect the outcome of this appeal.”); Harlen, 273 F.3d at 500 (“We need not decide which reading is the correct one in order to resolve this case, as Harlen’s claim fails even if no showing of animus is required.”); Gelb, 224 F.3d at 157 (“Although we do not foreclose the possibility of summary judgment in favor of the City Board, we note that summary judgment is generally inappropriate where questions of intent and state of mind are implicated.”); but see Jackson v. Burke, 256 F.3d 93, 97 (2d Cir. 2001) (assuming that “proof of subjective ill will is not an essential element of a ‘class of one’ equal protection claim,” but dismissing case because plaintiff had not shown any evidence that he was being treated from others similarly situated).
281. Payne v. Huntington Union Free Sch. Dist., 2002 WL 31039460, at *5 (E.D.N.Y. July 26, 2002) (indicating that to survive motion for summary judgment, plaintiff must prove that the
Classes or Persons?

The most recent cases from the Seventh and Second Circuits thus seem to make clear that previous attempts to limit Olech by requiring subjective ill will does not work. The work of limiting Olech, if it is to succeed, must adopt a different strategy. The next section examines one such alternate method.

2. Limiting Olech By Imposing a Strict Intent Standard

The U.S. Supreme Court in Olech held that a “class of one” claim was made out when the plaintiff alleged “that she has been intentionally treated differently from others similarly situated and that there is no rational basis for the difference in treatment.”282 An alternative method to limit Olech is to interpret the term “intentionally” in a very restrictive way. This was the strategy adopted by the Second Circuit in Giordano v. City of New York.283

In Giordano, a police officer was terminated from his position because of his use of the blood thinner, Coumadin.284 The plaintiff’s equal protection claim was that there was another New York City police officer, also using Coumadin and thus similarly situated, who had not been terminated.285 The court rejected this claim on the grounds that, because there was no evidence that the police officials who had made the decision to terminate Giordano were also aware of this other officer, they could not have intended to treat Giordano differently from other officers.286 This interpretation of the word “intentionally,” is at first glance counterintuitive because it seems to reward ignorance by local officials. As long as an official makes sure he does not know the status of other similarly situated persons, he cannot be found to have intentionally

motivation for the disparate treatment was ill will or for wholly arbitrary reasons lacking any rational basis); Padilla v. Harris, 2002 WL 750856, at *2 (D. Conn. Apr. 24, 2002) (“The Court cannot ignore the plain language of the Supreme Court in Olech that allegations that a plaintiff ‘has been intentionally treated differently from others similarly situated and that there is no rational basis for the difference in treatment’ were sufficient to state a claim for denial of equal protection.”); Barstow v. Shea, 196 F. Supp. 2d 141, 148 (D. Conn. 2002) (“To prevail on a ‘class of one’ equal protection claim, plaintiff must ‘show, not only “irrational and wholly arbitrary” acts, but also intentional disparate treatment.’”); Tuchman v. Bechem Transp., Inc., 185 F. Supp. 2d 169, 173 (D. Conn. 2002) (granting defendant’s motion to dismiss because plaintiffs had alleged neither dissimilar treatment nor animus).

283. 274 F.3d. 740 (2d Cir. 2001).
284. Id. at 742.
285. Id.
286. Id. at 751–52.
violated a plaintiff’s equal protection rights. A more obvious interpretation of the term “intentionally” would simply have meant that police officials intended to terminate Giordano because of his use of Coumadin, without regard to their conscious awareness of other officers in the same situation who were not being terminated. It is of little comfort to the officer laid off because of his use of Coumadin that, even though other officers in his situation received better treatment, police officials were not aware of those other officers.

However, there is ample support for the restrictive use of the term “intentionally” that was adopted by the court in *Giordano*. In 1944, long before *Olech*, the Supreme Court considered what was necessary to show a violation of the Equal Protection Clause in *Snowden v. Hughes*. In *Snowden*, the plaintiff alleged that certain officials had violated state law in not certifying him for the Republican nomination to the state assembly and, in so doing, had violated his equal protection rights. The Court rejected the claim, explaining that:

> [N]ot every denial of a right conferred by state law involves a denial of the equal protection of the laws, even though the denial of the right to one person may operate to confer it on another. [A]n erroneous or mistaken performance of [a] statutory duty, although a violation of the statute, is not without more a denial of the equal protection of the laws.

The “something more” that would turn such an act into an equal protection violation, explained the Court, is “an element of intentional or purposeful discrimination.” For example, according to the Court, if state officials did not assess property uniformly for the purpose of imposing property taxes, “[i]t is not enough to establish a denial of equal protection that some are assessed at a higher valuation than others. The difference must be due to a purposeful discrimination.” Thus, in *Snowden*, the Court found that the plaintiff had not adequately alleged that the Canvassing Board purposefully discriminated against him in favor of another. The Second Circuit’s opinion in *Giordano* is consistent with this restrictive interpretation.

---

288. *Id.* at 5.
289. *Id.*
290. *Id.*
291. *Id.* at 9.
292. *Id.* at 10.
Other post-\textit{Olech} courts have also interpreted the term “intentionally” in this same strict manner with the effect of limiting the reach of \textit{Olech}. In \textit{Payne v. Huntington Union Free School District},\textsuperscript{293} the court rejected the equal protection claim of a part-time teacher who had been terminated because of the absence of evidence of intentionally different treatment.\textsuperscript{294} The court explained that in order to meet this standard, the plaintiff would have to prove “that the Board knew it was treating her differently than it [was treating] other similarly situated individuals.”\textsuperscript{295} Thus, even if the plaintiff could identify individuals similar to herself whom the school board treated differently, that would not establish an equal protection claim in the absence of evidence that this different treatment was intentional. This was a standard too difficult for the plaintiff in \textit{Payne} to satisfy.

In \textit{Pariseau v. City of Brockton},\textsuperscript{296} the court rejected an equal protection claim arising out of a failure by police to dispatch a cruiser in response to a 911 call reporting a robbery.\textsuperscript{297} The court found that the claim did not satisfy the intent standard set forth in \textit{Olech}, since “[t]he arbitrariness of a law enforcement decision is not, without more, sufficient to state an equal protection claim.”\textsuperscript{298} The court, citing Justice Breyer’s concurring opinion in \textit{Olech}, explained that when differential treatment results from ineptness rather than design, there is no violation.\textsuperscript{299} Thus, the court in \textit{Pariseau} found that, “[e]ven if the decision not to dispatch a cruiser immediately lacked a rational basis, the equal protection claim cannot succeed unless Plaintiffs can make a threshold showing of the requisite discriminatory intent.”\textsuperscript{300} The plaintiffs in \textit{Pariseau} were unable to show either a “purposeful scheme not to protect white complainants” or “a custom or policy . . . to provide less protection to victims of a particular kind of crime.”\textsuperscript{301}

Even in those cases where plaintiffs succeed in making “class of one” claims, they must overcome the hurdle of a strict “intent” standard. When the \textit{Olech} case itself was remanded to the district court, that court

\textsuperscript{293}. 2002 WL 31039460 (E.D.N.Y. July 26, 2002).
\textsuperscript{294}. \textit{Id.} at *8.
\textsuperscript{295}. \textit{Id.}
\textsuperscript{297}. \textit{Id.} at 260.
\textsuperscript{298}. \textit{Id.} at 263.
\textsuperscript{299}. \textit{Id.} (citing Vill. of Willowbrook v. Olech, 528 U.S. 562, 565 (Breyer, J., concurring)).
\textsuperscript{300}. \textit{Id.} at 264.
\textsuperscript{301}. \textit{Id.}
refused to grant the defendant’s motion for summary judgment, at least in part because of the question of intent.\textsuperscript{302} While it was clear that the Village had “intended” to demand the 33-foot easement from Mrs. Olech, it was not clear whether the Village “intended” to treat her differently from other similarly situated individuals.\textsuperscript{303} The court found that there was a genuine issue of material fact as to whether the Village in fact knew of other similar landowners.\textsuperscript{304} Only if they did know of that other treatment at the time they demanded something different from Mrs. Olech could it be proven that they “intended” to treat Mrs. Olech differently.\textsuperscript{305}

In \textit{McWaters v. Rick}, another post-\textit{Olech} case, a federal district court in the Eastern District of Virginia imposed a strict intent standard. The plaintiff’s equal protection claim was that a local board, of which she was a member, had investigated her travel expenses but not those of a similarly situated colleague, and also that the board had refused to reimburse her legal expenses incurred in the investigation while paying those of her colleague.\textsuperscript{306} Although the court upheld her claim at the pleading stage, it did insist that she satisfy a demanding intent standard. The court explained, “to prove that a statute has been administered [in a discriminatory manner], more must be shown than the fact that a benefit was denied to one person while conferred on another. A violation is established only if the plaintiff can prove that the state intended to discriminate.”\textsuperscript{307} The court then showed that the required intent means “more than intent as volition or awareness. It implies that the decisionmaker . . . selected or reaffirmed a particular course of action at least in part ‘because of,’ not merely ‘in spite of’ its adverse effects upon an identifiable group [or person].”\textsuperscript{308}

This demanding standard would require McWaters to prove, not only that the board had in fact treated her differently from another board member, but that the board had done so because it wanted to discriminate against her. Such a standard would have the effect of eliminating equal protection claims where the conduct, although arbitrary and not defensible, results from mere inattention or accident. In \textit{McWaters},

\begin{itemize}
  \item \textsuperscript{302} Olech v. Vill. of Willowbrook, 2002 WL 31317415, at *13 (N.D. Ill. Oct. 10, 2002).
  \item \textsuperscript{303} Id.
  \item \textsuperscript{304} Id.
  \item \textsuperscript{305} Id.
  \item \textsuperscript{306} McWaters v. Rick, 195 F. Supp. 2d 781, 787 (E.D. Va. 2002).
  \item \textsuperscript{307} Id. at 792.
  \item \textsuperscript{308} Id. at 781 (citing Pers. Adm’r of Mass. v. Feeney, 442 U.S. 256, 279 (1979)).
\end{itemize}
Classes or Persons?

however, the federal district court found that the plaintiff had in fact alleged such intentional discrimination and thus refused to dismiss her claim on the pleadings.\footnote{309}{Id. at 793 (concluding that plaintiff McWaters had alleged an “arbitrary, irrational, and intentionally discriminatory act to investigate [her] alone” and that thus her complaint had alleged “a cognizable equal protection claim”).}

It seems clear that the requirement that plaintiffs prove “intentional” discrimination as construed in the preceding cases is likely to limit the success of “class of one” claims and is thus also likely to limit the number of such claims.

3. Limiting Olech By A Restrictive Interpretation of “Similarly Situated”

Another technique to limit the effect of Olech is to apply a very restrictive interpretation of the term “similarly situated.” As argued earlier, the oldest version of equality is the idea that similarly situated persons must be treated similarly.\footnote{310}{See supra note 7 and accompanying text.} But it has always been a vexing problem to determine what it means to be similarly situated. As Tussman and tenBroek demonstrated, the idea of “similarly situated” is incoherent in the abstract.\footnote{311}{Tussman & tenBroek, supra note 1, at 345 (“First, ‘similarly situated’ cannot mean simply ‘similar in the possession of the classifying trait.’ All members of any class are similarly situated in this respect and consequently, any classification whatsoever would be reasonable by this test.”).} Since each human is like all other humans in an infinite variety of ways (and thus we are all similar) and, at the same time, each human is different from every other human in an infinite number of ways (and thus we are all different). Tussman and tenBroek found a way out of this incoherence by insisting that we relate the classification to the purpose on account of which it was made in order to determine who is similarly situated to whom.\footnote{312}{See supra notes 17–18 and accompanying text.} But their analysis also suggests that the concept of who is similarly situated to whom is a manipulable device. On the one hand, the concept can be stretched by identifying a large number of persons who are similar to the plaintiff, thus making it easy for a plaintiff to insist on similar treatment. On the other hand, the concept can be shrunk by identifying only a small number of persons, or no persons, who are similar to the plaintiff, thus making it difficult for a plaintiff to claim a right to similar treatment. This latter strategy has been used by some courts to limit the effect of Olech.

\footnote{309}{Id. at 793 (concluding that plaintiff McWaters had alleged an “arbitrary, irrational, and intentionally discriminatory act to investigate [her] alone” and that thus her complaint had alleged “a cognizable equal protection claim”).}

\footnote{310}{See supra note 7 and accompanying text.}

\footnote{311}{Tussman & tenBroek, supra note 1, at 345 (“First, ‘similarly situated’ cannot mean simply ‘similar in the possession of the classifying trait.’ All members of any class are similarly situated in this respect and consequently, any classification whatsoever would be reasonable by this test.”).}

\footnote{312}{See supra notes 17–18 and accompanying text.}
The problem arises because the Supreme Court’s opinion in Olech, with its reference to “intentionally treated differently from others similarly situated,” appears to leave the impression that “similarly situated” is a self-defining term, without reference to any exterior criterion. This leaves the lower courts a great deal of discretion in deciding to whom the plaintiff ought to be compared and thus who is similar to him or her. Of course at the very least, it seems that a plaintiff needs to identify other similarly situated individuals who were treated differently. Since equality claims necessarily involved a comparison, it is not enough to allege that you alone have been treated unfairly. But the “similarly situated” requirement can be used to demand quite a bit more.

For example, in Payne, a federal district court for the Eastern District of New York considered the termination of a part-time teacher whose husband happened to be the superintendent of the school district. Plaintiff’s equal protection claim was that, while she had been fired because of her relationship to the superintendent, there were over one hundred other, similarly situated individuals, that is, employees of the school district who were related to other employees, who had not been fired. In the alternative, the plaintiff could have accepted a narrower identification of the similarly situated class, that is, those “related individuals who are in supervisory-subordinate relationships.” The court rejected both comparison groups as not sufficiently similar. The court’s interpretation of the term required that the comparison [of] individuals be “very similar indeed,” and “similarly situated in all material respects.” Because no employee of the school district (other than plaintiff’s husband who was the superintendent) had the power to supervise all other employees, the result was that there was no other employee similarly situated to the superintendent and thus no other

314. See, e.g., Presnick v. Town of Orange, 152 F. Supp. 2d 215, 224–25 (D. Conn. 2001) (citing Nassau County v. County of Nassau, 106 F. Supp. 2d 433, 440 (E.D.N.Y. 2000) for the proposition that “class of one” plaintiffs are not relieved from the burden of showing that other similarly situated people were treated differently”).
315. But see supra Part III.2.B for Judge Posner’s somewhat different “vindictive action” version of equal protection, which apparently does not require an explicit comparison.
317. Id. at *5.
318. Id.
319. Id at *6.
320. Id.
relative similarly situated to the plaintiff. Thus the plaintiff could not make a valid equal protection claim.

The federal district court in *Payne* might have treated the plaintiff’s claim more generously by applying the “similarly situated” requirement in a more relaxed fashion. For example, the comparison class could have been viewed as all those employees who are directly supervised by a relative. There were such employees in the school system and not all of these relationships were prohibited. But by insisting that any prospective members of the comparison class not differ in any way from the plaintiff, the court virtually foreordained the result that the plaintiff could not find any similarly situated persons and thus her equal protection claim would fail.

In *Campagna v. Commonwealth of Massachusetts Department of Environmental Protection*, the plaintiff was a state employee who had been disciplined and who alleged unequal treatment in violation of the Equal Protection Clause. A federal district court for the District of Massachusetts rejected the claim, finding that “the applicability of the ‘class of one’ theory to an employment-based equal protection claim seems dubious.” The court was concerned that:

> [A]ny public employee convinced that someone similarly situated is being treated more favorably could sue his or her employer under the Fourteenth Amendment for a violation of equal protection. Since practically every employee, public or private, is bound to be convinced at some point that he or she is getting the short end of the stick, it is not hard to imagine the bee hive of constitutional litigation that would be generated by this variant of the “class of one” doctrine.

But the court in *Compagna* found an easy way out of this dilemma by applying a very strict measure of what it meant to be “similarly situated.” The plaintiff was an environmental engineer. He was

---

321. *Id.* at *7.
322. *Id.* at *12 (granting Defendants’ motion for summary judgment).
323. *Id.* at *5.
325. *Id.* at 121.
326. *Id.* at 126.
327. *Id.* at 127.
328. *Id.*
329. *Id.* at 121.
disciplined, *inter alia*, for failing to complete a required form.\textsuperscript{330} His equal protection argument was that two other inspectors had also failed to file the required forms, but had not been disciplined.\textsuperscript{331} The court rejected the claim on the ground that the two other inspectors were not similarly situated to the plaintiff. Specifically, they were different first, because “rightly or wrongly” the department considered that they did their work competently while the plaintiff did not, and second, because the plaintiff had performed his inspection as part of his private after-hours business, while the two others had performed their inspections as part of the work as state employees.\textsuperscript{332} Thus, according to the court, these differences “warranted stricter treatment for [the] plaintiff.”\textsuperscript{333}

However, the plaintiff in *Compagna* would have identified the proper comparison differently. In the plaintiff’s view, all those who are required to complete an inspection form should be treated the same.\textsuperscript{334} On the other hand, from the perspective of the state defendant and for the federal district court, the requirement that a form be completed did not apply in the same way to all form-filers. Those whom the department considered to do competent work or who filled out forms as department employees were not held to the same standard as others.\textsuperscript{335} Or to put it in other terms, all form filers are equal, but some are more equal than others. While the federal district court may well have been correct in concluding that the plaintiff in *Compagna* had no valid equal protection claim, the result in the case also suggests that a court can always find differences if it is so inclined.

A federal district court for the District of Massachusetts in *Lakeside Builders Inc. v. Planning Board of the Town of Franklin*\textsuperscript{336} also applied a strict interpretation of “similarly situated” in rejecting the plaintiff’s claim. In that case, the plaintiff, a builder, had requested a waiver of a subdivision requirement about the length of dead-end roads, but the request was denied.\textsuperscript{337} Plaintiff’s equal protection claim arose from the

\begin{itemize}
  \item \textsuperscript{330} *Id.* at 122–23.
  \item \textsuperscript{331} *Id.* at 123.
  \item \textsuperscript{332} *Id.* at 127.
  \item \textsuperscript{333} *Id.*
  \item \textsuperscript{334} *Id.* (“Plaintiff contends that he was similarly situated to [two other employees] who were not fined despite the fact that they also failed properly to complete the form in conjunction with their inspections of the Westfield property.”).
  \item \textsuperscript{335} *Id.*
  \item \textsuperscript{336} 2002 WL 31655250 (D. Mass. Mar. 21, 2002).
  \item \textsuperscript{337} *Id.* at *1.
\end{itemize}
fact that the planning board had approved waivers of that requirement in twenty-one subdivision developments, and, according to plaintiff, these other situations were similar to its own and thus required the planning board to treat it similarly. But the court found the plaintiff’s complaint to be inadequate because it “[did] not propose any standard by which to judge whether one applicant for a waiver was ‘similarly situated’ to another. Rather, it simply asserts in conclusory fashion that the plaintiffs were treated differently from other similarly situated applicants.” The court was unwilling to conclude “that all applicants should be considered ‘similarly situated’ simply because they had all made requests for waivers of the dead-end street length regulation.”

It was not enough that all of the comparison group had applied for a waiver. The court indicated that it would “want to know a good deal more about the merits of individual applicants before deciding who was similarly situated to whom.” As the court noted, it would not make sense to assume that all applicants to a particular college were similarly situated, and would thus have to be treated similarly, simply because they had all applied to that particular college. Likewise, the court found that in the case before it, the plaintiff had not alleged sufficient facts for the court to determine who, of all those requesting waivers, was similar to whom. As the Lakeside Builders court explained, “Exact correlation is neither likely nor necessary, but the cases must be fair congeners. In other words, apples must be compared to apples.” Thus, once again a relatively strict interpretation of the term “similarly situated” lead to the dismissal of a plaintiff’s “class of one” claim.

C. How Olech Made It Easier For Plaintiffs in Federal Courts

1. Creating a Powerful New Precedent

...
federal courts to limit its effect, *Olech* is a powerful new precedent that has changed litigation in the federal courts. The Supreme Court’s decision explicitly validated the individual equal protection claim that alleges intentionally different treatment of similarly situated persons.\textsuperscript{346} In addition, while not endorsing the concept, the *Olech* opinion also called attention to the “vindictive action” version of that claim that had been widespread in the Seventh Circuit.\textsuperscript{347} Since *Olech*, plaintiffs citing the case have made successful “class of one” arguments in a surprisingly high percentage of cases.\textsuperscript{348}

Before *Olech*, this success would not have been expected. It has long been understood that rational basis equal protection claims, that is, those that do not involve heightened scrutiny because of a suspect class or fundamental right, have very little chance of success, because courts usually adopt an extremely deferential attitude.\textsuperscript{349} During a recent twenty-five year period, the Supreme Court decided one hundred ten rational basis cases and the plaintiff prevailed in only ten of these, for a success rate of only nine percent.\textsuperscript{350} One would imagine then, that “class of one” equal protection claims, which rarely involve suspect classes or fundamental rights, would rarely be successful. After *Olech*, that is no longer true. In the first eighty-six federal district court opinions after *Olech* that cite *Olech* and the term “class of one,” plaintiffs prevailed in thirty,\textsuperscript{351} for a success rate of thirty-five percent. This is an unexpectedly

\textsuperscript{346} Vill. of Willowbrook v. Olech, 528 U.S. 562, 564 (2000)
\textsuperscript{347} See supra notes 199–200 and accompanying text.
\textsuperscript{348} See infra note 351.
\textsuperscript{349} See Farrell, supra note 204, at 358–59.
\textsuperscript{350} Id. at 370.
high percentage which suggests both that *Olech* is having a significant impact on the federal courts, and that Justice Breyer’s and Judge Posner’s concerns about the explosion of federal cases were well-justified.

A number of these successful claims appear to involve rather trivial matters that traditionally would not have made their way into federal court. For example, a state employee prevailed where the state required her to complete a medical incident report before leaving work because of illness.\(^{352}\) An air ambulance service prevailed where local government selected its rival as a provider.\(^{353}\) A builder prevailed where a town issued a stop work order on construction work already begun.\(^{354}\) The owner of a mobile home park prevailed where he disagreed with the town’s decision to install a particular kind of water and sewer meters at his park.\(^{355}\) And a developer prevailed where a town refused to approve his subdivision plan.\(^{356}\) Although other cases citing *Olech* were of a more substantial nature, the successes just cited are evidence of an equal protection jurisprudence that must be quite far removed from the intent of the framers.

This next section examines some of these successful “class of one” equal protection claims in the federal courts. Most of these successes have come at the pre-trial stages of a lawsuit, that is, at the time of a motion to dismiss or at the time of a motion for summary judgment. The cases examined will demonstrate how difficult it can be, after *Olech*, for a government defendant to get a suit dismissed at the early stages of a lawsuit.

2. Rejecting Defendants’ Motions To Dismiss

Probably the most significant effect of the U.S. Supreme Court’s Olech opinion is that it is now far more difficult for government defendants to have a case dismissed on the pleadings. Although it has never been possible for local government officials to prevent equal protection law suits from being filed against them, the damage from those suits to local government can be minimized if the suits can be dismissed at a very early stage, before discovery has taken place, before trial preparations have been made, and before the trial itself. Under traditional equal protection doctrine, rational basis claims are very commonly dismissed on the pleadings as a result of the extremely deferential standard that courts have traditionally applied in rational basis cases. The Supreme Court articulated a sweeping version of this deferential attitude in *Federal Communications Commission v. Beach Communications Inc.*, 357 where it stated that “[i]n areas of social and economic policy, a statutory classification . . . must be upheld against equal protection challenge if there is any reasonably conceivable state of facts that could provide a rational basis for the classification.” 358 Under this standard, “those attacking the rationality of the legislative classification have the burden ‘to negative every conceivable basis which might support it,’” 359 and “legislative choice is not subject to courtroom fact-finding and may be based on rational speculation unsupported by evidence or empirical data.” 360

When courts apply the rationality standard deferentially, a complaint cannot be drafted that would survive a motion to dismiss. A plaintiff would have to “hypothesize all conceivable justifications for a statutory classification and then prove that no legislative body could ‘rationally have believed’ that the classification served [any of] the hypothesized purpose[s].” 361 But federal district courts seem to have read Olech as overturning much of this accepted wisdom, at least in the context of equal protection challenges to administrative decisions by local officials.

358. Id. at 313.
359. Id. at 315.
360. Id.
Classes or Persons?

In contrast to the insurmountable burden for plaintiffs that previous courts suggested, some post-Olech cases suggest a far more relaxed pleading standard. In *Russo v. City of Hartford*, the plaintiff challenged his suspension from the police force on equal protection grounds. A federal district court for the District of Connecticut rejected the city’s motion to dismiss because “Russo has alleged that similarly situated individuals were treated differently and that the defendants have not expressed any legitimate basis for the differential treatment.” This statement suggests that the court assumes that the defendant has the burden of showing the legitimate basis for the different treatment. But that assumption is surely inconsistent with the strong presumption of validity accorded government action in rational basis cases and with the “any conceivable basis” language of *Beach Communications*. The court in *Russo* was in fact quite aware of these presumptions, acknowledging that “defendants are under no obligation to provide a rational basis for the alleged violation; the court can dismiss the count on the pleadings if it can conceive of any rational basis for the classification.” Notwithstanding that acknowledgment, the court went on to say that it “refuses, however, to speculate as to conceivable rational bases for the defendants’ actions.” As a result of the court’s refusal to speculate, the case was not dismissed on the pleadings and thus the city of Hartford was forced to continue to litigate the case.

*Stone v. Hope, Indiana* is another post-Olech case where a federal district court in the Southern District of Indiana appeared to bend over backward to help the plaintiff survive a motion to dismiss. In that case, the plaintiff complained that the town had arbitrarily refused his request for individual water meters at his mobile home park while treating other trailer park owners differently. Even though the plaintiff had not made his allegations with particularity, the court refused to dismiss the case. “All reasonable inferences are to be drawn in favor of the Plaintiff.

363. Id. at 190.
364. Id. at 195 (emphasis added).
366. Id. at 195–96 (citing Connolly v. McCall, 254 F.3d 36, 42 (2d Cir. 2001)).
367. Id. at 196.
368. Id. at 197–98.
370. Id. at *4.
371. Id.
Under this standard, the Plaintiff has provided Defendants with fair notice of his claims and the grounds upon which they are based. This “fair notice” standard announced by the court in Stone is a very far distance from the “any conceivable basis” the Supreme Court used in Beach Communications. Beach Communications suggests that the burden is entirely on the plaintiff to identify and disprove all of the defendant’s possible justifications. The standard in Stone is almost exactly the opposite. It suggests that the plaintiff need only give notice in a general way of his equal protection claim and that will be sufficient to get him to the next stage of the litigation.

Although some courts have interpreted the “similarly situated” requirement quite strictly in order to eliminate claims, that standard has sometimes been treated very generously in order to allow plaintiffs to survive a motion to dismiss. In Kiser v. Naperville Community Unit, the plaintiff had been terminated from his position as executive administrator and attorney for a school district. In moving to dismiss the claim, the school district argued that the plaintiff had not alleged that any other lawyers were employed by the district, and that, without such a group against which to compare the plaintiff, he had not stated an equal protection claim. A federal district court in the Northern District of Illinois rejected that argument on the grounds that “the potential relevance of other lawyers is obvious from the complaint” and thus that the plaintiff need not make the comparison explicitly. Thus, on the basis of a rather vague complaint that had made no attempt to identify specific individuals who were similar but had been treated differently, the court found that a valid equal protection claim had been made.

In addition to the problems that government defendants have had in getting “similarly situated” “class of one” claims dismissed, courts have also made it more difficult for a governmental defendant to have “vindictive action” claims dismissed. In Singleton v. Chicago School Reform Board, a federal district court for the Northern District of Illinois refused to dismiss, “since this Court cannot examine the

372. Id.
374. See supra notes Part IV. B.3.
376. Id at *2.
377. Id. at *14.
378. Id.
pleadings and magically determine the Defendants’ state of mind at the
time of their alleged unlawful actions, we must accept Plaintiff’s
allegations [of sheer vindictive purposes] to be true." In Northwestern
University v. City of Evanston, a federal district court for the Northern
District of Illinois explained that, at the motion to dismiss stage, the
university did not need to prove its claims in order to avoid dismissal. It
was enough “simply [to] allege that it was treated differently and that
the City’s actions were irrational, arbitrary and even vindictive.” The
Evanston court found this standard satisfied in the university’s
allegations that the city’s decision to include some of the university’s
property in a historic district was “motivated by the . . . illegitimate
desire to disregard the University’s . . . right to be exempt from property
taxes,” and “by vindictiveness against the University for its refusal to
accede to the City’s demand for revenue payments in lieu of property
taxes.” Finally, in Hyatt v. Town of Lake Lure, a federal district court
for the Western District of North Carolina, although conceding that
claims “must be alleged with sufficient specificity to avoid being
conclusory,” insisted that “there is no heightened pleading requirement
imposed on the plaintiff.” The Hyatt court found that the plaintiff had
satisfied these standards with a nonspecific claim that “she ha[d] been
treated differently from others similarly situated,” and that “Defendants
acted with ‘personal malice’ towards her and that the Defendants acted
arbitrarily and capriciously.”

The relaxed standard adopted by the courts in these cases appears to
turn on its head the previous presumption in rational basis cases that
plaintiffs must counter all conceivable justifications for government
action. Instead, these cases make it substantially easier for a plaintiff to
survive a motion to dismiss and thus force the government defendant to
continue to litigate the case, through discovery, summary judgment, and
possibly trial.

380. Id. at *10.
382. Id. at *3.
383. Id.
384. Id.
386. Id. at *11 (citing McWaters v. Rick, 195 F. Supp. 2d 781, 791 (E.D. Va. 2002)).
3. Rejecting Defendants’ Motions for Summary Judgment

Traditionally, courts have also been very deferential to defendants at the summary judgment stage when deciding rational basis equal protection claims. A motion for summary judgment is not to be granted unless “there is no genuine issue as to any material fact and . . . the moving party is entitled to judgment as a matter of law.” But where a court has determined that the actual motivation of a government actor is not relevant, then the parties’ dispute over that motivation does not constitute a material fact. Thus, government defendants have traditionally been able to have cases thrown out at the summary judgment stage. But, after Olech, that traditional wisdom has been drawn into question in “class of one” cases, where the courts now appear to be far more deferential to plaintiffs.

In Evanston, the court, contrary to the traditional wisdom, considered evidence of actual motive as part of a motion for summary judgment. In that case, the university complained that it had been singled out for inclusion in an historic preservation district in retaliation for its refusal to adjust its tax exempt status. With regard to the claim of illegitimate animus and vindictive motivation as an equal protection violation, the court explained that “[a] vindictive action equal protection claim, unlike its traditional counterpart, requires scrutiny of the legislature’s actual subjective motivation.” Because the city and the university disagreed about what had motivated the city’s actions, the court found that there was a genuine issue of material fact and that it would not grant summary judgment. Noticeably missing from the Evanston court’s opinion was any suggestion that the court or the city might justify the city’s action by reference to “any conceivable basis” that might be hypothesized.

In Barstow v. Shea, the plaintiff, a state employee, alleged that she had been required to complete a medical incident report before leaving.

388. See Farrell, supra note 361, at 41, n.233.
389. FED. R. CIV. P. 56(c).
390. See Farrell, supra note 361, at 41.
391. Id. at 41, n.233.
393. Id.
394. Id.
395. Id.
396. 196 F. Supp. 2d 141 (D. Conn. 2002).
work for illness while other employees had not been so required.\textsuperscript{397} A federal district court for the District of Connecticut denied the defendant’s motion for summary judgment because the plaintiff presented the depositions of two fellow employees who testified that they had left work due to illness but had not been required to complete the form.\textsuperscript{398} Therefore, “the jury could conclude that plaintiff was treated differently from other similarly situated employees”\textsuperscript{399} and “the jury could [also] discredit the reason given by the defendant for the differential treatment,”\textsuperscript{400} and thus summary judgment was inappropriate.

In \textit{Oneto v. Town of Hamden},\textsuperscript{401} the plaintiff was a police officer who had been passed over for promotion and then claimed a denial of equal protection.\textsuperscript{402} The town moved for summary judgment on the grounds that the plaintiff had not introduced any evidence of similarly situated persons who were treated differently.\textsuperscript{403} “The town argued that since Oneto was promoted “only after resort to the courts, and well outside the ordinary course of civil service procedure, ” there was no one else similarly situated to him.”\textsuperscript{404} But a federal district court for the District of Connecticut did not identify the comparison class so narrowly. Instead, the court identified three other persons who were promoted outside the civil service rules (and were thus similarly situated) but had not been subjected to a special investigation like Oneto (and were thus treated differently).\textsuperscript{405} Although the court conceded that Oneto’s claim ultimately would be difficult to establish, “there [was] sufficient evidence in the record, when taken together and with all inferences drawn in Oneto’s favor, from which a jury could conclude that [the town official’s] actions were motivated by reasons unrelated to a legitimate investigative objective.”\textsuperscript{406} Thus summary judgment for the defendant was not appropriate.
4. **Final Judgments in Favor of Plaintiffs**

Most successful “class of one” arguments have been won at the pretrial stage, that is, plaintiffs have successfully opposed defendants’ motions to dismiss or motions for summary judgment. There are very few reported cases of plaintiffs’ ultimate victory on the merits of “class of one” claims. This shortage of reported victories could mean that “class of one” claims can now survive longer through the litigation, but ultimately are not successful. An alternate, and perhaps more probable, explanation is that both settlements of cases and final judgments, particularly when they are the result of jury verdicts, do not lead to reported opinions and thus are not readily retrievable through traditional legal databases. Thus the shortage of reported final judgments where plaintiffs succeeded does not necessarily mean that they do not exist. In any case, there are two reported post-**Olech** cases of final victory in “class of one” claims.

The first of these cases, **Cruz v. Town of Cicero**, was a “vindictive action” claim that involved a particularly egregious case of ill will and malice by a public official toward a private citizen. **Cruz** was one more in that long line of Seventh Circuit “class of one” cases. The court affirmed a jury verdict of $402,000 for a violation of the Equal Protection Clause. The plaintiffs were in the business of converting apartment buildings to condominiums. The jury’s verdict was based in part on testimony that the town President, Betty Loren-Maltese, had, through her friends in city government, seen to it that the plaintiffs would not get the certificates because they had been unwilling to make contributions to support her political career. Thus, the certificates were denied in order to punish the plaintiffs. On appeal, Seventh Circuit held that the evidence supported the verdict in that it demonstrated a “totally illegitimate

---

407. 275 F.3d 579 (7th Cir. 2001).
408. *Id.* at 589 (“[A] reasonable jury could have concluded (as this one did) that the trouble the Gonzalez parties had obtaining certificates of compliance . . . had nothing to do with [the merits]. Instead, these troubles stemmed from [the town president’s] desire to punish Gonzalez for not repaying her ‘help’ with significant financial contribution of some kind.”).
409. *Id.* at 582–83.
410. *Id.* at 583.
411. *Id.* at 583–85.
412. *Id.* at 588–89.
413. *Id.* at 589.
animus,” not related to the duties of government and not rationally related to a legitimate government interest.414

Caudell v. City of Toccoa415 was a rare example when a legislature, rather than a local government official, singled out an individual for invidious treatment and thereby created a “class of one” claim. As indicated above, for the most part legislatures act by passing laws that make use of broad generalizations.416 In these situations, the Equal Protection Clause serves only as a limit on classification.417 However, in Caudell, the Georgia legislature approved an act directed at one person only.418 The act provided that no one could serve on a city commission while also serving as a member of any hospital authority.419 The plaintiff was the only person in the entire state of Georgia affected by the act.420 A federal district court for the Northern District of Georgia found that he had been “singled out for a special burden to which others have not been subjected,”421 and that the state had “offered no legitimate state purpose whatsoever to justify this legislative classification,”422 and thus it violated the Equal Protection Clause. The court granted final judgment for plaintiff declaring the act unconstitutional and enjoined city officials from taking any actions to enforce it.423

V. RE-EVALUATING THE INDIVIDUAL RIGHTS OR CLASS-BASED NATURE OF EQUAL PROTECTION AFTER OLECH

Although inadvertently and only implicitly, but nonetheless definitively, the U.S. Supreme Court in Olech has resolved the problem addressed in this article. That problem is that there are two views of the equal protection clause that are conceptually in conflict but in practice have ignored each other. With the benefit of the Supreme Court’s opinion in Olech, we can identify an appropriate resolution of the conflict by reference to the following two rules.

414. Id. at 589.
416. See supra Part I.A.
417. See supra Part I.B.
419. Id. at 1375.
420. Id. at 1378.
421. Id.
422. Id.
423. Id. at 1381.
**Rule 1.**

When legislatures enact broad rules based on generalizations about persons, the Equal Protection Clause operates as a limitation on legislative classifications, but does no more. It does not protect individual rights, and thus, the individual person who is treated unfairly because a reasonable legislative generalization is not true as to him has no equal protection claim. Administrative rulemaking is subject to these same limits.

**Rule 2.**

When government officials make individual decisions to grant or withhold a benefit to a particular person, or to impose or eliminate a burden on a particular person, the Equal Protection Clause does protect individual rights. The Supreme Court in *Olech* requires a plaintiff in these situations to prove that “she has been intentionally treated differently from others similarly situated and that there is no rational basis for the difference in treatment.” The plaintiff in such a case need not identify a classification, but must simply show at least one other person who was similarly situated but received different treatment. In the courts of appeal, particularly the Seventh Circuit, there is an alternative way to make out an individual rights claim under the Equal Protection Clause. A plaintiff does so by proving that a state official acted vindictively with an intention to “get” the plaintiff.

With these two rules as guides, it is now possible to see what is wrong with some of the earlier individual rights equal protection arguments. The individual rights argument has been made most frequently to demonstrate what is wrong with affirmative action. As that argument goes, affirmative action is based on group rights rather than individual rights and assumes that all persons of a particular race are effectively fungible, that is, they all share the same views and thus will all contribute in the same way to a “diverse” environment, or that they have all suffered from the same wrongs of racial discrimination and thus are all equally entitled to a race-based remedy. This, according to the critics, ignores the fact that each person of any race is an individual and therefore does not necessarily hold any particular view attributable to his race, nor will any particular individual necessarily have suffered discrimination based on his race. Further, race-based affirmative action classifications necessarily assume that white persons all hold the same viewpoints on race, and thus could not contribute to a diverse

environment, and also implicitly assume that all white persons somehow are responsible for past racial discrimination and should thus be willing to take a back seat today. It follows from this line of argument that race-based affirmative action is wrong because it confers benefits on all members of a racial minority group even though the individual members of that group do not individually deserve those benefits, and imposes burdens on all members of the white majority, even though there are many individual, innocent white persons who bear no responsibility for the status of minorities in today’s society.

Whatever the ultimate resolution of the affirmative action issue under the Equal Protection Clause, it should not be resolved by resort to this individual rights reasoning. Since virtually all the forms of affirmative action have been subject to lawsuits have been broad rule-based programs rather than individual decisions directed at one particular person, they are thus subject only to the limit that the classifications they create satisfy the relevant equal protection standard. When the classification is based on race, the standard is strict scrutiny, so that the racial classification must be necessary to achieve a compelling interest in order to be upheld. But so long as the classification survives that test, the fact that some individuals seem to be treated unfairly does not give rise to a constitutional issue. The Court should be no more concerned for Allan Bakke than it was for Robert Murgia. Any decision that overturns an affirmative action program because it treats individuals as members of a group rather than as individual persons is simply not in accord with the U.S. Supreme Court precedent. If the Supreme Court ultimately determines that affirmative action programs violate the Constitution, that determination should follow from the conclusion that a racial classification has been used improperly, not because the classification harmed an individual person.

However, the discussion of individual rights claims in affirmative action cases does strike a proper chord in a much more limited context. In determining what classifications should be given heightened scrutiny, the Court has traditionally looked at a number of factors, including whether or not the trait is immutable, whether or not the trait “frequently bears no relation to ability to perform or contribute to society,”\(^\text{425}\) and whether the use of that trait is consistent with our commitment that “legal

---

burdens should bear some relationship to individual responsibility.”

Thus, in deciding what traits should receive heightened scrutiny on the ground that their use by government is suspicious because usually invidious, it might well be appropriate to consider the harm to individual persons that result from unthinking generalizations about persons that turn out, as they usually do, not to be universally true. But, once the Court has identified those traits that will receive heightened scrutiny, when it actually applies that scrutiny, the review should be only of the classification, not of the individual person unfairly harmed.

*Olech* teaches us that the equal protection clause does protect individual rights in a certain context, but the case clearly does not support any wholesale changes in the way courts review legislative classifications. *Olech* does support claims by individual persons who have been treated unequally by an executive branch official. That limited right itself will probably create substantial extra work for the federal courts. But *Olech* does not create any such individual right against legislative or administrative classification.

---

TRADITIONAL EQUITY AND CONTEMPORARY PROCEDURE

Thomas O. Main*

Abstract: This Article offers extensive background on the development and eventual merger of the regimes of law and equity, and suggests that the procedural infrastructure of a unified system must be sufficiently elastic to accommodate the traditional jurisdiction of equity. As the Federal Rules of Civil Procedure become increasingly more elaborate and technical, strict application of those procedural rules can generate mischievous results and hardship. This Article suggests that equity remains a source of authority for district judges to avoid the application of a procedural rule when technical compliance would produce an inequitable result. A separate system of equity provided a forum for hardship created by the procedures of the common law system. Because the jurisdiction of equity was preserved by the procedural merger of law and equity, mischief and hardship created by the contemporary procedures of a unified system of law and equity need not be tolerated and may be corrected in a manner consistent with traditional principles of equity.

“Let judgment run down as waters, and righteousness as a mighty stream.”

Much of the grand history of Anglo-American law could be characterized as an epic struggle between the regimes of law and equity. The roots of this conflict run deep and straight—to Aristotle, who recognized that universal laws could promote injustice as well as justice,

---

* Assistant Professor of Law, University of the Pacific, McGeorge School of Law. I wish to thank Professors Greg Pin gee, Kevin Stack, and Peter Nicolas for their insights and comments on an early draft. And I wish to thank my student-colleagues Grant Wahlquist, Mat Larsen, and William Diedrich for their able research assistance. This work was supported by the law school’s generous summer research grant, for which I am grateful.

1. Amos 5:24. See also WILLIAM SHAKESPEARE, THE MERCHANT OF VENICE, act 4, sc. 1, lines 184–87 (“The quality of mercy is not strain’d / It droppeth as the gentle rain from heaven / Upon the place beneath. It is twice blest: / It blesseth him that gives and him that takes.”); 2 THE WORKS OF JOHN MILTON 307 (F. Patterson ed., 1931) (“temper . . . Justice with Mercie”).

2. See generally ROSCOE POUND, AN INTRODUCTION TO THE PHILOSOPHY OF LAW 54 (rev. ed. 1954) (“Almost all of the problems of jurisprudence come down to a fundamental one of rule and discretion, of administration of justice by law and administration of justice by the more or less trained intuition of experienced magistrates.”); KENNETH C. DAVIS, DISCRETIONARY JUSTICE: A PRELIMINARY INQUIRY 17 (1969) (“Every governmental and legal system in world history has involved both rules and discretion. No government has ever been a government of laws and not of men in the sense of eliminating all discretionary power. Every government has always been a government of laws and of men.”); RICHARD A. POSNER, THE PROBLEMS OF JURISPRUDENCE 24–25 (1990) (“[F]or more than two millennia, the field of jurisprudence has been fought over by two distinct though variegated groups. One contends that law is more than politics and in the hands of skillful judges yields . . . correct answers to even the most difficult legal questions. The other contends that law is politics through and through and that judges exercise broad discretionary authority.”). See also BARBARA J. SHAPIRO, PROBABILITY AND CERTAINTY IN SEVENTEENTH CENTURY ENGLAND: A STUDY OF THE RELATIONSHIPS BETWEEN NATURAL SCIENCE, RELIGION, HISTORY, LAW, AND LITERATURE 163–93 (1983).
and thus fashioned a notion of juridical equity to temper the strict application of laws. Equity moderates the rigid and uniform application of law by incorporating standards of fairness and morality into the judicial process. Equity assures just results in each application of strict law and eliminates the need for elaborate legislative drafting to contemplate all conceivable applications. Naturally, there exists some tension between the two regimes: law ensures strict uniformity and predictability, while equity tempers law to offer relief from hardship. Yet although there is tension between the two regimes, they are also


4. Equity presupposes that certain applications of law can frustrate the laws of nature, the administration of “justice,” or the common good. See Anton-Hermann Chroust, The “Common Good” and the Problem of “Equity” in the Philosophy of Law of St. Thomas Aquinas, 18 NOTRE DAME L. REV. 114, 117 (1942–1943) (“Equity does not intend to set aside what is right and just, nor does it try to pass judgment on a ‘strict Common Law rule’ by claiming that the latter was not well made. It merely states that, in the interest of a truly effective and fair Administration of Justice, the ‘strict Common Law’ is not to be observed in some particular instance.”); Colin P. Campbell, The Court of Equity—A Theory of its Jurisdiction, 15 GREEN BAG 108, 111 (1903) (Equity can “recognize and enforce principles which actually govern society in general, whether embodied in the so-called rules of law or not.”). See also infra notes 87–98 and accompanying text.

5. See generally Campbell, supra note 4, at 111–12 (“No set of prohibitive or declaratory words which the ingenuity of legislatures or courts can devise will do justice in all cases or will provide for all situations. Hence, both the statutes and the opinions must some time fall short in future cases of that which the peculiar demand of the occasion requires.”). See also ROSCOE POUND, “TOWARD A NEW JURISCIENTUM,” IDEOLOGICAL DIFFERENCES AND WORLD ORDER 1, 9 (1949) (“Men with very different conceptions of the social order, groups of men with one ideal or picture of what ought to be and other groups with wholly divergent pictures, must live together and work together in a complex social organization.”).

6. See infra notes 86–90 and accompanying text.
complementary, and for centuries separate systems of law and equity combined to administer the laws with both certainty and discretion.\footnote{See Frederic William Maitland, Equity and The Forms of Action 17 (Chaytor ed. 1909) (“[F]or two centuries before the year 1875 the two systems had been working together harmoniously.”). See also infra notes 80–85 and accompanying text.}

The image of separate systems of law and equity is, however, an increasingly fading memory. In the middle of the nineteenth century, procedural codes merged law and equity into a single unified system in most American state courts.\footnote{See infra notes 213–34 and accompanying text.} The Judicature Acts of 1873 and 1875 accomplished much the same for law and equity courts in England.\footnote{See 1873, 36 & 37 Vict., c. 66 (Eng.); 1875, 38 & 39 Vict., c. 77 (Eng.). See also infra notes 284–88 and accompanying text.} Since 1938 the federal district courts of the United States have recognized one merged form of action under the Federal Rules of Civil Procedure.\footnote{See Fed. R. Civ. P. 2 (“There shall be one form of action to be known as ‘civil action.’”). See also infra notes 235–83 and accompanying text.} Under these unified systems, memories of a divided bench have receded into the past.\footnote{See generally Mertens v. Hewitt Assocs., 508 U.S. 248, 256–57 (1993) (construing a statutory reference to “equitable remedies” and stating “memories of the divided bench, and familiarity with its technical refinements recede into the past”). See also Great-West Life & Annuity Ins. Co. v. Knudson, 534 U.S. 204, 212–17 (2002) (discussing legal and equitable origins of restitution).} Law schools have eliminated the separate course in Equity,\footnote{See generally Walter Wheeler Cook, The Place of Equity in Our Legal System, 3 AM. LAW SCH. REV. 173 (1912) (urging the elimination of equity as a separate law school course); Ralph E. Kharas, A Century of Law-Equity Merger in New York, 1 SYRACUSE L. REV. 186, 186 (1949) (discussing Harvard Law School’s decision to abolish the course in Equity and to teach those principles instead in Contracts, Torts, Procedure and other courses); Jerome Frank, Civil Law Influences on the Common Law—Some Reflections on “Comparative” and “Contrastive” Law, 104 U. PA. L. REV. 887, 895 (1956) (“In several of our leading university law schools, there is now no course on ‘equity.’ I teach a course on the subject at Yale Law School, but until June 1956, I have been required to call it ‘Procedure III.’ One of my esteemed colleagues, Judge Charles Clark, has been a leader in this sort of eradication of the word ‘equity’ from the law school curriculum, rejecting it almost as if it were an obscene term.”); Lester B. Orfield, The Place of Equity in the Law School Curriculum, 2 J. LEG. EDUC. 26, 27 (1949–50) (“The movements to abolish Equity as a separate course have appeared at twenty-year intervals, first in 1909, then in 1930, and now following World War II when so many law schools are revising their curricula.”); Douglas Laycock, The Triumph of Equity, 56 LAW & CONTEMP. PROBS. 53, 53 (1993) (“Most lawyers I meet are incredulous that anyone my age ever taught a course in equity.”); Jack B. Weinstein & Eileen B. Hershenov, The Effect of Equity on Mass Tort Law, 1991 U. ILL. L. REV. 269, 272 (1991) (“equity was taught as a separate course until the 1950s”); Mary Brigid McManamon, The History of the Civil Procedure Course: A Study in Evolving Pedagogy, 30 ARIZ. ST. L.J. 397 (1998); Stephen B. Burbank, The Bitter with the Sweet: Tradition, History and Limitations on Federal Judicial Power—A Case Study, 75 NOTRE DAME L. REV. 1291, 1292 (2000) (civil procedure teachers instead have long filled their courses with a heavy diet of constitutional topics, such as choice of law and personal}
equity are a few vestigial distinctions and labels, such as joining claims for legal and equitable relief in a single action,\textsuperscript{13} or determining the availability \textit{vel non} of a jury trial.\textsuperscript{14}

This Article argues that in merging the regimes of law and equity, reformers may have swept away part of the wisdom that had guided the development and operation of dual systems. One virtue of an autonomous system of equity was its authority to act in opposition to the strict law when the unique circumstances of a particular case demanded intervention.\textsuperscript{15} The architects of the merger took great pains to sustain this virtue by preserving the substantive principles of both law and equity; only the procedure was modified, they insisted.\textsuperscript{16} But even assuming that the antagonistic substantive regimes of law and equity can co-exist and be applied contemporaneously within a single unified procedural system, a fundamental flaw inheres in the procedural infrastructure of a merged system. In denying equity any structural autonomy, there remains no relief from the procedures of the merged system itself when the modes of proceedings in that system are inadequate.\textsuperscript{17} Indeed, the jurisdiction of equity is impaired if equity cannot operate as a check upon the “strict law” that is codified in the procedures of the merged system.\textsuperscript{18}

Parts I, II and III of this Article are but a history lesson. “People need not so much to be told as to be reminded,”\textsuperscript{19} and the reminder is essential context for my thesis. Part I discusses the development of law and equity as separate systems from the Middle Ages through the eighteenth jurisdiction. For a statement in support of a separate course in equity, see Robert S. Stevens, \textit{A Brief on Behalf of a Course in Equity}, 8 J. LEGAL EDUC. 422 (1956).

\textsuperscript{13} See generally Fed. R. Civ. P. 18(a) (“A party asserting a claim to relief . . . may join . . . as many claims, legal [or] equitable . . . as the party has against an opposing party.”). See also Fed. R. Civ. P. 1–2.


\textsuperscript{15} See infra notes 88–98 and accompanying text.

\textsuperscript{16} See infra notes 221, 278–81, 288 and accompanying text.

\textsuperscript{17} See infra notes 306–84 and accompanying text.

\textsuperscript{18} See infra notes 385–400 and accompanying text.

\textsuperscript{19} Melvin M. Johnson, Jr., \textit{The Spirit of Equity}, 16 B.U. L. REV. 345, 345 (1936) (quoting “a former president of the United States” without further attribution). See also infra notes 289–90 and accompanying text.
century. I emphasize the utility and uniqueness of having complementary and rival systems effect the “wise adjustment of law to experience.”

Part II focuses primarily on the publication of Sir William Blackstone’s *Commentaries on the Laws of England*. Blackstone recognized substantive rights apart from procedural remedies and, in doing so, marginalized the law/equity distinction by suggesting that the difference between the two systems was merely procedural. A substance/procedure distinction assumed primary significance, with substance as the actual rules that we apply; and procedure no more than a practical and ancillary means to an end. The perceived elasticity of procedure made change in the light of practical details inevitable, if not noble. Broad acceptance of this paradigm ultimately paved the way for the so-called merger of law and equity, beginning in the mid-nineteenth century.

In Part III, I present the history of that merger, which purported to fuse only the procedure of law and of equity, while leaving the substance of each regime otherwise intact. The reformers envisioned a unified procedural apparatus that would permit judges to jointly administer the jurisprudence of both law and equity.

In Part IV, I argue that an important ingredient of the jurisprudence of equity was displaced by the procedural merger of law and equity in the federal courts. A merged system offers no recourse from insufficiency created by the procedural apparatus of the merged system. The argument for institutional autonomy is compelling even though unified Federal Rules of Civil Procedure incorporated much of the philosophy

22. See infra notes 183–88 and accompanying text.
24. See infra notes 195–212 and accompanying text.
25. See infra notes 211–88 and accompanying text.
and practices of equity procedure. First, the Federal Rules have not been immune to the complication, trivialization and ossification pathogens that have plagued earlier procedural systems. Second, procedural rules featuring discretion and flexibility within a unified system cannot replicate the administration of justice pursuant to a separate system of equity. The Federal Rules, like any codification, are unavoidably a product of experience developed by reason, and reason tested by experience. Yet generalizations in laws cannot always be completely general, and human calculations are imperfect. Indeed, the unimaginable is inevitable and, when the unexpected occurs, some authority must stand in the breach and supply that which prevents the general rules from meeting the immediate necessity. A separate system of equity viewed each lawsuit from the standpoint of the “wrong” presented. The powers of that court were as vast, and its processes and procedure as elastic, as all the changing emergencies of increasingly complex relations could demand. By contrast, the common law system.

27. For a description of the equitable origins of the Federal Rules, see infra notes 262–77 and accompanying text.

28. See infra notes 307–48 and accompanying text. For a contrary view, expressed by the current Reporter to the Advisory Committee on Civil Rules, see Edward H. Cooper, Simplified Rules of Federal Procedure?, 100 Mich. L. Rev. 1794, 1795 (2002), stating that “It may be inevitable that a continuing revision process lengthens the rules and adds complexity to them. Doubts grow up around old solutions, and new problems appear. The Civil Rules have not escaped this effect. Yet time and again, the Rules adhere to a pervading characteristic. The effort is less to provide detailed controls and more to establish general policies that guide discretionary application on a case-specific basis. Many a district judge may view one provision or another as an unwarranted intrusion on the proper sovereignty of a trial court, but vast discretion remains at virtually every turn. It does not seem fair to charge the revision process with a descent into the naggling detail and sterile ossification that have overtaken earlier procedural systems.”

29. See generally POUND, supra note 5, at 2.

30. Aristotle was quick to point out that such imperfection is not the fault of the law or of the legislature. Rather “the law is no less correct on this account; for the source of the error is . . . the nature of the object itself, since that is what the subject-matter of actions is bound to be like.” Nichomachean Ethics, supra note 3, at 17–19. See generally Eric G. Zahnd, Note, The Application of Universal Laws to Particular Cases: A Defense of Equity in Aristotelianism and Anglo-American Law, 59 Law & Contemp. Probs. 263 (1996). See also supra note 5.

31. See generally Campbell, supra note 4, at 111–12.

32. See Charles D. Frierson, A Certain Fundamental Difference in Viewpoint Between Law and Equity as Illustrated by Two Maxims, 22 Case & Comment 403, 410 (1915). See also supra notes 173–82, 385–400 and accompanying text.

viewed every lawsuit as a test of the existence and scope of the alleged “right.”\(^{34}\) Whether by design or by unfortunate evolution, the procedural infrastructure of the unified system today increasingly resembles the latter vision, compromising fair and just results at the behest of formalism.\(^{35}\)

I offer contemporary procedure in mass tort cases as an illustration of the problem. The impact of mass torts on the justice system has been labeled “overwhelming,”\(^{36}\) “elephantine,”\(^{37}\) “unprecedented,”\(^{38}\) “bizarre,”\(^{39}\) “pathological,”\(^{40}\) and an “emergency.”\(^{41}\) Mass tort cases may be especially likely to present new and unforeseeable challenges for trial courts trying to process these cases fairly and efficiently.\(^{42}\) Yet the administration of these cases is governed largely by the procedural templates that also govern all other types of civil cases.\(^{43}\) As applied in

\(^{34}\) See Frierson, supra note 32, at 410. See also infra notes 156–72 and accompanying text.

\(^{35}\) See infra notes 307–48 and accompanying text.


\(^{38}\) Weinstein & Hershenov, supra note 12, at 270.


\(^{42}\) See, e.g., Edward H. Cooper, Aggregation and Settlement of Mass Torts, 148 U. PA. L. REV. 1943, 1944 (2000) (“It is often observed that each new mass tort presents different problems, requiring different procedural solutions than any of its predecessors.”); Martha Minow, Judge for the Situation: Judge Jack Weinstein, Creator of Temporary Administrative Agencies, 97 COLUM. L. REV. 2010, 2019 n.51 (1997) (discussing mass torts and the traditional model of adjudication); Carrie Menkel-Meadow, Ethics and the Settlement of Mass Torts: When the Rules Meet the Road, 80 CORNELL L. REV. 1159, 1160 (1995) (suggesting that resolution of mass torts presents novel issues). See also infra notes 349–84 and accompanying text. But see Linda S. Mullenix, Mass Tort as Public Law Litigation: Paradigm Misplaced, 88 NW. U. L. REV. 579, 581 (1994) (“[T]he essential mass tort case is nothing more or less than an injured plaintiff, represented by a personal-injury, contingency-fee lawyer, suing the product’s manufacturer.”); Siliciano, supra note 41, at 991 (noting there are “few interesting or novel questions of doctrine” in mass tort cases and the problems attributed to the size of such cases are not unique).


435
certain circumstances, then, existing procedural rules may be inefficient, too complicated, or otherwise deficient for the unique circumstances presented in some mass tort cases.\footnote{See Cooper, supra note 42, at 1944 (referring to “statutes and rules framed for the last war”).}

In the current system, the procedural rules typically are applied as drafted, and the response to the procedural mischief is to amend the rules specifically to address that problem. But amendment, if any,\footnote{See generally id.} may come too late for the problem that occasioned the amendment.\footnote{See also infra notes 385–400 and accompanying text.} Furthermore, the amended language itself may become the “strict law” that, in turn, creates the insufficiency for the next generation of mass tort.\footnote{See generally Arthur Miller, An Overview of Federal Class Actions: Past, Present and Future 58–59 (1977) (reviewing criticisms and proposed reforms); Responses to the Rule 23 Questionnaire of the Advisory Committee on Civil Rules, reprinted in 5 Class Action Rep. 3, 10, 17 (1978) (reporting results of survey suggesting that a majority of practitioners and judges favored prompt revision of Rule 23); Stephen Berry, Ending Substance’s Indenture to Procedure: The Imperative for Comprehensive Revision of the Class Damage Action, 80 Colum. L. Rev. 299 (1980) (surveying flaws and suggesting reforms to Federal Rule 23); Report and Recommendations of the Special Committee on Class Action Improvements, 110 F.R.D. 195, 199 (1986); Elizabeth Barker Brandt, Fairness to the Absent Members of a Defendant Class: A Proposed Revision of Rule 23, 1990 B.Y.U. L. Rev. 909 (1990); Jean Wegman Burns, Decorative Figureheads: Eliminating Class Representatives in Class Actions, 42 Hastings L.J. 165 (1990); Mark W. Friedman, Constrained Individualism in Group Litigation: Requiring Class Members to Make a Good Cause Showing Before Opting Out of a Federal Class Action, 100 Yale L.J. 745 (1990); Jonathan R. Macey & Geoffrey P. Miller, The Plaintiffs’ Attorney’s Role in Class Action and Derivative Litigation: Economic Analysis and Recommendations for Reform, 58 U. Chi. L. Rev. 1, 27–33 (1991); Report of Advisory Committee on Civil Rules, 167 F.R.D. 535, 537 (1996); George F. Sanderson, Congressional Involvement in Class Action Reform: A Survey of Legislative Proposals Past and Present, 2 N.Y.U. J. Legis. & Pub. Pol’y 315 (1999); Thomas Merton Woods, Wielding the Sledgehammer: Legislative Solutions for Class Action Jurisdictional Reform, 75 N.Y.U. L. Rev. 507 (2000). See also infra notes 385–400 and accompanying text.} Because equity traditionally complemented the administration of justice by...
offering, as an independent ground for equitable relief, relief from such procedural insufficiencies. 48 I argue here that the tradition of equity is impaired in a merged system if the trial judge cannot escape the rigors of that infrastructure in the exercise of her magisterial good sense and offer relief from such hardship.

Finally, in Part V, I offer a proposal to resurrect the curative purpose of equity in circumstances when the procedural apparatus of our merged system fails to provide plain, adequate and complete relief. I urge judges to use and credit equity as a source of authority to avoid applications of the Federal Rules that, although achieving technical compliance, result in inequitable outcomes in violation of the spirit of the Rules. Also, I urge procedural rulemakers to draft amendments to the Federal Rules that better accommodate the enduring jurisprudence of equity.

I. THE DEVELOPMENT OF COMPLEMENTARY SYSTEMS OF LAW AND EQUITY

Much has been written about the origins of law and equity. 49 We know that both of these legal systems are derived from the royal prerogative of English kings to interfere with the ordinary legal processes of the communal courts. 50 The ultimate and supreme power of kings to do

48. See infra notes 385–400 and accompanying text.


50. Until the latter part of the twelfth century, ordinary law and justice in England was governed by custom and was administered rather informally (if not crudely) by the shire courts and the courts of the hundred motes (in the time of Saxons and Danes, dating back to the seventh century) and by the county, borough and manor courts (in the early Norman period beginning with the Norman Conquest in 1066). The forms of trial were, in large part, appeals to the supernatural. See generally 1 POLLOCK & MAITLAND, supra note 49, at 14–22; 1 HOLDSWORTH, ENGLISH LAW, supra note 49, at 40; GEORGE L. CLARK, PRINCIPLES OF EQUITY 3 (1948); Adams, supra note 49, at 91 & n.10.
justice in any case between their subjects extended, of course, to any matter brought to their attention. To exercise that authority, kings issued brevia or writs which had the effect of removing the cause directly to the king’s court or council. The repeated issuance of writs based upon similar circumstances led an astute King Henry II to realize that certain standardized writs could be issued. He established a Curia Regis to administer a national law based on these writs. The Chancellor was then the king’s secretary, and Chancery was the secretariat of the state. Writs, like all other state papers, were prepared there, and the clerks in Chancery issued the standardized writs whenever a complainant

(discussing the king’s “prerogative machinery”); Frederick Pollock, English Law Before the Norman Conquest 14 L.Q. REV. 291, 297 (1898).

51. The operative principle was that the king was the fountainhead of all justice, and in him, resided the final power to do whatever was just and righteous. See MILLAR, supra note 23, at 12–13; WILLIAM F. WALSH, OUTLINES OF THE HISTORY OF ENGLISH AND AMERICAN LAW 69–70 (1923). The king, of course, was the source of all power and authority—whether legislative, executive or judicial—and the sovereign lord of all land; he was the state. See Kittle, supra note 49, at 23 (“It was the firm policy of the Norman kings to concentrate all power within themselves.”); 1 POLLOCK & MAITLAND, supra note 49, at 85–87; KERLY, supra note 49, at 13–14. The administration of justice in England was originally confided to the Aula Regis, or great Court or Council of the King. See 1 STORY, supra note 49, § 39, at 30.

52. See Adams, supra note 49, at 89 (discussing the new “judicial machinery” brought into England at the Norman Conquest); Garrard Glenn & Kenneth Redden, Equity: A Visit to the Founding Fathers, 31 VA. L. REV. 753, 760 (1945) (“Justice did open the door, of course, and it was the royal hand that was on the knob.”); W. R. Vance, Law in Action in Medieval England, 17 VA. L. REV. 1, 9–10 (1930) (discussing writ to gain seisin of land held by the King in reversion); MAITLAND, supra note 7, at 4–5.


54. Petitioners generally had to purchase writs from the king’s secretary, the Chancellor. The writs constituted the king’s law, derived personally from his power and authority as the fountain of all justice. Accordingly these were not natural rights. See Adams, supra note 49, at 89 (noting the new procedure and the new machinery were “the king’s private property”); William Searle Holdsworth, The Relation of the Equity Administered by the Common Law Judges to the Equity Administered by the Chancellor, 26 YALE L.J. 1, 3 (1916) (discussing writs de cursu “which any litigant could purchase”).

55. The judges of the Curia Regis (court of the king) were the king’s councilors and advisors, a group of secular and ecclesiastical dignitaries. See Kittle, supra note 49, at 25 (the curia regis “may be regarded as the parent stem from which the courts of law and equity have sprung”).

presented facts contemplated by a writ.\textsuperscript{57} Forms of action developed around the writs.\textsuperscript{58}

The writs largely displaced the customary laws of the different parts of the country and became the foundation of our common law actions.\textsuperscript{59} In the earliest stage of this new common law, writs were construed liberally to apply to new cases where justice seemed to require that an action be allowed.\textsuperscript{60} Many forms of action and relief, including those that we would now label “equitable,” were administered by the king’s judges as part of the common law during this period.\textsuperscript{61} However, this dynamic vision of justice threatened the power of entrenched English barons, and later the Parliament, who perceived the power to issue writs as a power

\begin{footnotes}
\item[57.] See Walsh, supra note 56, at 100–01; Maitland, supra note 7, at 3; Robert Severns, Nineteenth Century Equity: A Study in Law Reform, 12 CHI-KENT L. REV. 81, 92 (1934) (“[b]y the end of the thirteenth century the number of petitions had become very large and the work of reading them was onerous.”); see also infra notes 157–61 and accompanying text. For a sampling of early writs, see John H. Baker, An Introduction to English Legal History 437–47 (2d ed. 1979).
\item[58.] See Sherman Steele, The Origin and Nature of Equity Jurisprudence, 6 AM. L. SCH. REV. 10, 10–11 (1926). See also infra notes 162–72 and accompanying text.
\item[59.] See Baker, supra note 57, at 49; Walsh, supra note 51, at 86–88; Pollock & Maitland, supra note 49, at 129–130. See also Joseph H. Koffler & Alison Reppy, Common Law Pleading 18 (1969) (“Substantive law grew out of procedure. Courts were organized to handle a series of specific cases, the division of which gradually developed theories of rights and liabilities. Our rights and liabilities as defined by substantive law, then, had their origin in and developed out of procedural law.”). Of course, it bears emphasis that the rights that were recognized were almost exclusively property rights; there were no personal rights, political rights, civil rights as we understand them. See deFuniak, supra note 56, at 56.
\item[60.] Henry de Bracton described a class of writs, breve magistralia, that were very freely issued. See Millar, supra note 23, at 18 (1952). See also 2 Holdsworth, English Law, supra note 49, at 245 (use of these writs was the immediate and effective cause of the rapid development of the law during this period).
\item[61.] See Willard Barbour, Some Aspects of Fifteenth-Century Chancery, 31 HARV. L. REV. 834, 834 (1918) (“It is now more than thirty years since Justice Holmes in a brilliant and daring essay set on foot an inquiry which has revealed the remote beginnings of English equity. Equity and common law originated in one and the same procedure and existed for a long time, not only side by side, but quite undifferentiated from each other . . . . There was no equity as a separate body of law; for the king’s justices felt themselves able to dispense such equity as justice required.”); Adams, supra note 49, at 91–92 (recognizing common law and equity as an “undifferentiated system in the effort of the king to carry out his duty of furnishing security and justice”); Holdsworth, supra note 54, at 1 (accumulating evidence that common law judges in the twelfth through fourteenth centuries “administered both law and equity”); Aaron Friedberg, The Merger of Law and Equity, 12 St. John’s L. REV. 317, 318 n.2 (1938) (“during the reign of Henry II, both equity and common law were administered under the same system of procedure and were quite indistinguishable from each other”); Severns, supra note 57, at 91 (“It is obvious that in the thirteenth and fourteenth centuries, no distinction can be drawn between common law and equity.”).
\end{footnotes}
to make new law. To limit the king’s authority, the common law system became “a hard and fast system with certain clearly defined things which it could do and with equally clearly defined things which it could not do.” The universe of writs was fixed and their construction by law judges narrowly circumscribed; precise and technical rules of pleading, procedure and proof cabined judicial discretion within the form of action. The only remedy that the law courts could award in personal actions was monetary damages.

The common law courts gradually became an institution that was separate from the king, but the royal prerogative endured. Litigants

---

62. See Holdsworth, supra note 54, at 3. See also Adams, supra note 49, at 96, n.27, n.28 (discussing various attempts to regulate and limit the issue of writs); Plucknett, supra note 23, at 26.

63. Adams, supra note 49, at 96. See also deFuniak, supra note 56, at 57 (“A growing worship of formalism and technicality also began to obsess the courts of law.”); George Palmer Garrett, The Heel of Achilles, 11 VA. L. REV. 30, 30 (1924–1925) (“The common law made a fetish of procedure.”).

64. For example, a provision in Magna Charta (1215) significantly diminished the scope of the royal writ in respect to titles to land. Also, the Provisions of Oxford (1258) expressly forbade the Chancellor to issue any new writs “without the commandment of the King and his council who shall be present.” The Provisions were annulled five years later, but the common law courts nevertheless were transformed during the 13th century into a rigid system of formal actions. See 1 Holdsworth, English Law, supra note 49, at 196; 2 Holdsworth, English Law, supra note 49, at 291; Millar, supra note 23, at 18 (citing Frederic William Maitland, The Forms of Action 41 (1936); 1 Holdsworth, English Law, supra note 49, at 58–59).

65. See Steele, supra note 58, at 10–11 (“In accordance with its technical mode of procedure, every species of legal wrong was supposed to fit into some one of a limited number of classes; for each class an appropriate remedy was provided, obtainable only by the use of some one of a limited number of ‘forms of action.’ An action was begun by the issuance of a writ appropriate to the form of action; in time these writs became standardized, and, where the facts of a case were without precedent, no writ to cover them was found, and hence no action could be brought.”); Garrett, supra note 63, at 31 (discussing “form-mad common lawyers”); James Fosdick Baldwin, The King’s Council in England During the Middle Ages 61–62 (1913) (referring to the common law’s “formulaic procedure”); Holdsworth, supra note 54, at 22 (discussing the “complicated machinery” of the law courts). See also infra notes 156–72 and accompanying text.

66. See Elias Merwin, The Principles of Equity and Equity Pleading 17 (1895) (discussing inability of common law courts to compel the performance of duties); Kittle, supra note 49, at 28 (“[T]he remedies which the law courts gave were often wholly inadequate. They were as bad as no remedy at all.”).

67. See Adams, supra note 49, at 96, n.27, n.28 (discussing the multiplication and classification of writs in the thirteenth and fourteenth centuries as an indication of this separation); Leonard J. Emmerglick, A Century of the New Equity, 23 TEX. L. REV. 244, 246 (1945) (noting independence of common law courts as of the fourteenth century); Holdsworth, Equity, supra note 49, at 294 (“In the latter half of the 14th and in the 15th centuries the common law tended to become a fixed and rigid system. It tended to be less closely connected with the king, and therefore less connected with, and sometimes even opposed to, the exercise of . . . royal discretion.”).
confronting the power of a great lord who could unduly influence the regular common law officials, juries and judges continued to petition the king for relief.\textsuperscript{69} The Crown remained the only source of relief for circumstances that did not fit within the narrow range of available writs.\textsuperscript{70} As England transitioned from an agricultural to a commercial nation, the more frequent became situations involving rights not previously contemplated and for which no writ and, thus, no remedy, was available.\textsuperscript{71}

Appeals to the king, instead of to his courts, became numerous, and about the time of Edward I,\textsuperscript{72} it became usual to refer such petitions for consideration and disposition to the Lord Chancellor.\textsuperscript{73} As "the keeper of the king's conscience," the Lord Chancellor was a churchman who was familiar with both the ecclesiastical and the civil or Roman law.\textsuperscript{74} The

\textsuperscript{68} See HENRY L. MCCINTOCK, HANDBOOK OF EQUITY § 2 at 3 (1936) ("As the common-law courts came to be recognized more clearly as separate legal institutions . . . [t]he justices in eyre, who were the king's representatives, as well as legal judges, continued to exercise the[ir] power."); Friedberg, supra note 61, at 318 n.2 ("[T]he provisions of Oxford in 1258, by forbidding the Chancellor to frame new writs without the consent of the king and Council, drew a definite line of demarcation between the two systems of law.").

\textsuperscript{69} Common law writs were expensive luxuries for individuals too poor to avail themselves of the remedy. See supra note 54. Further, the adversary could be so rich and powerful that it would be hopeless to proceed in the law courts. See Barbour, supra note 61, at 856 (reprinting sample petitions filed in chancery). See also Hohfeld, supra note 49, at 547, n.9 (noting until nearly the end of the fifteenth century, most petitions to the king were founded on some suggestion of inequality between the parties); Oliver Wendell Holmes, Early English Equity, 1 L. QUART. REV. 162–63 (1885) (discussing different substantive doctrines developed in chancery); Glenn & Redden, supra note 52, at 763–69 (reprinting and translating sample bills filed in Chancery). There was no charge for obtaining a bill in equity. See Severns, supra note 57, at 88.

\textsuperscript{70} See Glenn & Redden, supra note 52, at 760 (discussing the limited range of the law courts); MAITLAND, supra note 7, at 3 ("Though these great courts of law have been established (King's Bench, Common Pleas, etc.) there is still a reserve of justice in the king.").

\textsuperscript{71} See deFuniak, supra note 56, at 56.

\textsuperscript{72} Edward I reigned from 1272–1307. For relevant background on King Edward I, see SMITH, supra note 53, at 162–187.

\textsuperscript{73} See Glenn & Redden, supra note 52, at 760–61 (describing that the Chancellor took over the task of reviewing the petitions and ultimate had them addressed to him directly); Walsh, supra note 56, at 106 (same).

\textsuperscript{74} See Glenn & Redden, supra note 52, at 760–61 (1945). According to Professor Glenn, much pomp accompanied the early chancellors when they marched in state. A graphic description appears in GEORGE CAVENDISH, THE LIFE OF THOMAS WOLSEY (1893), which was written by a gentleman usher of a chancellor. See also Walter E. Sparks, The Origin, Growth, and Present Scope of Equity Jurisprudence in England and the United States, 16 W. JURIST 473, 475 (1882) ("From the time of the reign of Henry VI [chancery] constantly grew in importance, and in the reign of Henry VII it expanded into a broad and almost boundless jurisdiction under the fostering care and ambitious wisdom and the love of power of Cardinal Wolsey.").
Chancellor unrolled a vast body of legal principle that we know as “equity” to offer relief in those cases where, because of the technicality of procedure, defective methods of proof, and other shortcomings in the common law, there was no “plain, adequate and complete” remedy otherwise available. Equity channeled extraordinary powers to afford relief when there were procedural or substantive deficiencies of the law courts. Intervention was premised on the notion that justice incorporated the moral sense of the community, existing as a function not only of a community’s technical rules, but also of “magisterial good sense, unhampered by rule.” The Chancellor had “the right and the powers, in fact, to do as he likes, whatever hard law and still harder practice may dictate.”

In the fourteenth and fifteenth centuries, the Court of Chancery developed into a distinct court. There were no writs and no forms of

75. See MILLAR, supra note 23, at 24 (1952); 1 STORY, supra note 49, § 33, at 22-26. See also infra notes 385–400 and accompanying text.

76. Severns, supra note 57, at 84.

77. See Pound, Justice, supra note 49, at 701–02 (“[O]ne function of the administration of justice is to adjust the relations of individuals to each other so as to accord with the general moral sense. Rules in many of these matters are needed to guide the weak judge and to save us from his lack of will and lack of judgment. But these same rules may serve only to hamper the strong judge and to prevent application of the full measure of his good sense and sound judgment to the case in hand.”). See also infra notes 173–82 and accompanying text.

78. Severns, supra note 57, at 89.

79. See Steele, supra note 58, at 11 (the “practice of referring to the Chancellor all of these special appeals to the kind led to the establishment of a tribunal which by the time of Edward III (1327–1377) had become recognized as a distinct and permanent court, with its separate jurisdiction and mode of procedure and its seat at Westminster”); Holdsworth, supra note 54, at 6 (describing that all cases which called for equity were “handed over to a tribunal which, in time, came to be perfectly distinct from any of the common law courts”); Walsh, supra note 56, at 107 (suggesting that Chancery as a court of equity was taking form “around the 14th century”); Adams, supra note 49, at 97 (dating origins of a separate system of equity to the fourteenth century); George Burton Adams, The Continuity of English Equity, 26 YALE L.J. 550, 556 n.17 (1917) (“The chancellor’s court had become distinct from the Council before the end of the 15th century.”); 1 HOLDSWORTH, ENGLISH LAW, supra note 49, at 404 (suggesting that the Chancellor first made a decree on his own authority in 1474); Severns, supra note 57, at 96 (“It was not until the end of the fifteenth century that purely equity matters go to the chancellor alone.”); Sparks, supra note 74, at 474 (quoting the proclamation of 22 Edward III addressed to the sheriffs of London “commanding them that, whatsoever business relating as well to the common law of our kingdom, as our special grace, cognizable before us, from henceforth to be prosecuted as followeth; viz., The common law business before the Archbishop of Canterbury, elect, our chancellor, by him to be dispatched, and the other matters grantable by our special grace be prosecuted before our special chancellor, or our well beloved clerk, the keeper of the privy seal, so that they, or one of them, transmit to us such petitions of business which, without consulting us, they cannot determine, together with their advice thereupon, without any further prosecution to be had before use for the same.”). In an effort to date the commencement of a court of
action. Equity as administered by this court served as both an “appendix” and a competitor for the common law. On one hand, as already described above, Chancery was doing some convenient and useful works that could not be done, or could not easily be done by the law courts. By requiring the specific performance of contracts, developing a law for vendors and purchasers of land, enjoining some of the more common torts such as waste, trespass, and nuisance before they were committed in the first instance, and by reforming or rescinding contracts that were tainted by fraud and mistake, equity supplemented the common law. Chancery could “adjust their decrees so as to meet most, if not all, of these exigencies; and they [would] vary, qualify, restrain, and model the remedy, so as to suit it to mutual and adverse claims, controlling equities, and the real and substantial rights of all the parties.” In these instances, the “appendix” characterization is especially appropriate, because equity also served as a catalyst for significant reforms to the common law.

Yet in every case in which the result reached in equity differed materially from the judgment a court of law would give, there was a rival chancery, it bears mention that the earliest writers of the common law, such as Bracton, Glanville, Britton and Fleta make no reference to an equitable jurisdiction of a court of chancery. See also 10 Selden Society, Select Cases in Chancery A.D. 1364 to 1471 (William Paley Bailldon ed., London, Bernard Quaritch 1896); id. at xix (“It seems clear that the Chancellor had and exercised judicial functions of his own as early as the reign of Richard II if not Edward III.”). See generally Joseph Parkes, A History of the Court of Chancery (1828).

80 See Severns, supra note 57, at 88 (“No form was necessary and no strict procedure had to be followed.”); Walsh, supra note 56, at 106 (“Relief was given without a write. The bill [in equity] was generally in simply form, without formality, and free from the technical rules which applied to writs.”); Barbour, supra note 61, at 854 (“Less exactness of pleading was required than by the law, and even if a bill were ‘misconceived’ the complaint was not out of court.”); see also infra notes 173–82 and accompanying text.

81 See Maitland, supra note 7, at 19 (“I do not think that any one has expounded or ever will expound equity as a single, consistent system, an articulate body of law. It is a collection of appendixes between which there is no very close connection.”).

82 See supra notes 69–71 and accompanying text.

83 See Walsh, supra note 53, at 28 (describing purpose of uses was to avoid the rigors of the common law which forbade testamentary gifts of land as well as inter vivos transfers except by livery of seisin); Maitland, supra note 7, at 4–7; Sidney Post Simpson, Fifty Years of American Equity, 50 Harv. L. Rev. 171 (1936).

84 1 Story, supra note 49, § 28, at 19.

85 For example, although law courts initially would not enforce instruments that had been lost or destroyed, they ultimately adopted the equity practice of admitting secondary evidence of contents. See generally Glenn & Redden, supra note 52; Maitland, supra note 7, at 6–7; William F. Walsh, Is Equity Decadent?, 22 Minn. L. Rev. 479, 483–86 (1938) (discussing “the reforming influence of equity”). See also infra note 297.
system. Broad categories of cases fared differently in the two systems. Equity also incorporated standards of morality into the calculus of resolving any dispute. Indeed, equity corrected the law by applying, in circumstances where the ordinary rules would lead to unwarranted hardship, considerations of what was fair and just. “When the social needs demand one settlement rather than another, there are times when we must bend symmetry, ignore history and sacrifice custom in the pursuit of larger ends.”

The regimes of law and equity thus approached a given set of facts from opposite angles—invoking distinctive traditions, applying different reasoning, and pursuing separate aims.

The principles of equity are, of course, merely a part of the larger concept of fairness and justice upon which all law must be based. The law’s dilemma long has been to develop a jurisprudence that recognizes when unique circumstances justify a departure from rigid rules. On the one hand, there is no more fundamental social interest than that law should be uniform and impartial. Commenting upon Lord Mansfield’s statement that “we must act alike in all cases of like nature,” Judge Henry J. Friendly termed this “the most basic principle of jurisprudence.” Indeed, “the normal and necessary marks, in a civilized community, of justice administered according to law, are generality, equality, certainty.” At the same time, however, hardly any two cases

86. See Holdsworth, supra note 54, at 15 (referring to law and equity as rival systems); Smith, supra note 53, at 211 (crediting Professor A.B. White’s characterization of equity as “the upstart jurisdiction”).
87. See Barbour, supra note 61, at 834 (“Equity is outside the common law, even antagonistic to it.”). See also infra note 191.
88. See Holdsworth, Equity, supra note 49, at 293 (“the root . . . of equity [is] the idea that the law should be fairly administered and that hard cases should as far as possible be avoided”).
89. BENJAMIN N. CARDOZO, THE NATURE OF THE JUDICIAL PROCESS 65 (1921).
90. See Frierson, supra note 32, at 411.
91. See Campbell, supra note 4, at 110 (noting the intimacy of the relations among the basic principles of “natural justice, equity, honesty, generosity and good conscience”).
93. CARDOZO, supra note 89, at 112.
95. NEWMAN, supra note 92, at 19–20 (quoting FREDERICK POLLOCK, JURISPRUDENCE 37 (5th ed. 1923)). See also GIROLAMO DEL VECCHIO, JUSTICE 173 n.13 (Edinburgh ed., 1952) (“the worst misfortune of a civilized people is doubt about the impartiality of justice”) (internal citation and quotation omitted); GEOFFREY DE Q. WALKER, THE RULE OF LAW: FOUNDATION OF CONSTITUTIONAL DEMOCRACY 25–27 (1988); infra note 436.
are the same: every case presents a moral problem, and almost all moral problems are unique.

The expanding role of equity in the broader administration of justice was controversial, yet constant. The early chancellors decided cases with little or no regard for precedent, basing their decisions largely upon their idiosyncratic ideas of “conscience.” The applicable rule depended upon the notions of right and wrong possessed by each chancellor, leading to Selden’s well-known aphorism: “Equity is a roguish thing. For law we have a measure... equity is according to the conscience of him that is Chancellor, and as that is larger or narrower, so is equity. Tis all one as if they should make the standard for the measure a Chancellor’s foot.”

Echoing the institutional resistance centuries

96. Roscoe Pound, Law and Morals 65 (1924) (“Cases are seldom exactly alike.”).
98. Newman, supra note 92, at 20. See Johnson v. United States, 186 F.2d 588, 590 (2d Cir. 1951) (Hand, L., J.) (“Nor is it possible to make use of general principles, for almost every moral situation is unique; and no one could be sure how far the distinguishing features of each case would be morally relevant to one person and not to another.”).
99. See Severns, supra note 57, at 82 (“no court has been so vigorously hated as the system called Equity and the tribunal known as the High Court of Chancery”); Sparks, supra note 74, at 473 (as equity “slowly but surely [was] enlarging and extending its mighty arm... the encroachments it was making... seemed almost sacrilegious, so detestable, owing to the fact that its principles were largely derived from the Roman or civil law, and its chancellors were generally ecclesiastics, and the people generally did not desire to have the church gain so strong a hold upon their courts of justice or the affairs of State, and it was evident the Court of Chancery was gaining, for fear that the church would eventually assume control, dictate to the people, usurp their rights, and virtually subject to them the Church of Rome”).
100. See Severns, supra note 57, at 99 (at least through the Fifteenth century, “the Chancellor did not consider himself bound by any sort of fixed principles... [T]here was no tendency, as in the common law courts, to feel bound by precedent.”). See generally W.H.D. Winder, Precedent in Equity, 57 L.Q. Rev. 245 (1941).
101. See 1 John Norton Pomeroy, A Treatise on Equity Jurisprudence § 385, at 524 (2d ed. 1892) (“It is undeniable that courts of equity do not recognize and protect the equitable rights of litigant parties, unless such rights are, in pursuance of the settled juridical notions of morality, based upon conscience and good faith.”); Pound, Justice, supra note 49, at 698, n.9; Vidal v. Girard’s Exrs., 42 U.S. (2 How.) 127, 193 (1846) (Justice Story describing the Chancellor’s reports as “shadowy, obscure and flickering”); see also 3 John Reeves, History of English Law 384–85 (2d ed. 1880) (“[i]t jurisdiction did not comprehend a great extent and the exercise of it was feeble and imperfect”). Cf. Barbour, supra note 61, at 840 (suggesting that such a contemptuous view was erroneous); Glenn & Redden, supra note 52, at 758, n.15 (chancellors were guided by a system of law); Walsh, supra note 85, at 481. n.4 (discussing the jurisdiction of equity); Grupo Mexicano de Desarrollo, S.A. v. Alliance Bond Fund, Inc., 527 U.S. 308, 322 (1999) (Scalia, J.) (suggesting that equity is flexible but not omnipotent).
102. John Selden, The Table-Talk 64 (The Legal Classics Library 1989). See also 6 Bulstrode White洛克, Commons Journals 373 (1650) (“The proceedings in Chancery are
earlier to the establishment of the common law courts, Parliament again was threatened and—with the support of the common law judges— resisted the kings’ and chancery’s discretionary exercise of power under the guise of equity and natural law.

Jealousy and conflict persisted until the relative authority of the two rival systems was decided in the early seventeenth century, in a contest between two of the great lawyers of all time, Coke and Bacon, in a drama that could carry an opera. Throughout his career, Coke had defended his venerable law—what he called “the perfection of reason”—against the encroaching jurisdiction of equity. The contest began with the entry of a judgment in an action before Coke, apparently because the defendant’s material witnesses were somehow detained at an inn by agents of the plaintiffs. Defendant thereafter sought relief in equity, and Coke induced the plaintiffs to secure indictments against their opponents for attacking a judgment of the King’s Court. The case was referred to

_ secundum arbitrium boni viri, and this arbitrium differeth as much in several men as their countenances differ. That which is right in one man’s eyes is wrong in another’s._); Severns, supra note 57, at 82 (mocking the jurisprudence of equity as “some sort of Philosopher’s Stone by which injustice is whisked into justice by the simple method of preparing a form of petition lately called a ‘bill’”).

103. See supra notes 57–66 and accompanying text.

104. See Sparks, supra note 74, at 477 (noting that equity was “strenuously opposed by the courts of common law and its hosts of disciples, who, with watchful and jealous eyes, were the first to see the encroachments [equity] was making”).

105. Shortly after Edward III (1327–1377) had issued an ordinance directing that matters of grace be referred to the Chancellor, Parliament denounced with severe penalties those “who so sue in any other court to defeat or impeach the judgment given in the King’s [common law] Court.” 27 Edw. III, c. 1. (1353). See Severns, supra note 57, at 88 (“[C]riticism came from the ruling class . . . [In the fourteenth century,] the bitterest criticism comes from the very defendants before the chancellor.”); EDWARD COKE, THE FOURTH PART OF THE INSTITUTES OF THE LAWS OF ENGLAND 82–84 (1797). The House of Commons petitioned against Chancery ten times between the reigns of Richard II (1377–1399) and Henry VI (1422–1461). In 1653, the House of Commons voted that the High Court of Chancery of England should be eliminated. See generally ALFRED HENRY MARSH, HISTORY OF THE COURT OF CHANCERY (1890); Munger, supra note 49, at 52 (linking the resistance to equity with a greater respect for English constitutional law and lex scripta); Pound, Justice, supra note 49, at 711, n.48 (suggesting that rigid laws discourage corruption); CHARLES E. PHELPS, JURIDICAL EQUITY § 10 (1894); Holdsworth, Equity, supra note 49, at 297 (suggesting that progress of equity was hindered by Parliament).


108. See id. at 46.
the law officers who, under the leadership of Bacon in 1616, sustained Chancellor Ellesmere.109 Recognizing that justice required not only the certainty of law but also the discretion of equity, King James I110 established both the legitimacy and the primacy of equity within a dual system.111

Shortly thereafter, in 1618, Bacon became the Lord Chancellor, and a fundamental transformation of chancery was underway.112 For many centuries the sweeping jurisdiction of equity had been untrammeled by any definite rule.113 But chancery could not remain a “fountain of unlimited dispensations.”114 To reform the “heterogeneous medley of empirical remedies,”115 Bacon issued one hundred rules of equity that were “wisely conceived, and expressed with the greatest precision and perspicuity.”116 Continuing thereafter, particularly under the

109. King James I had little inclination to act otherwise since upholding the authority of the chancellor was consistent with the royal prerogative. This finally settled the power of Chancery to make good its decree, though directly opposed to the results in the same controversy at law, by the exercise of its power in personam over the parties to the litigation. See Munger, supra note 49, at 45–46. See also Glenn & Redden, supra note 52, at 777 (contrasting Coke with Sir Thomas More who, nearly a century earlier, averted such conflict by “inviting the judges to dinner, so that the matter could be pleasantly discussed”).

110. James I reigned from 1603–1625. For relevant background on King James I, see SMITH, supra note 53, at 303–18.

111. See Munger, supra note 49, at 45–46.


113. See supra notes 72–90 and accompanying text. See also 1 JOHN FONBLANQUE, A TREATISE OF EQUITY § 3 (London, A. Strahan & W. Woodfall, 1st ed. 1793) (“So there will be a necessity of having recourse to natural principles, that what is wanting to the finite may be supplied out of that which is infinite. And this is properly what is called equity, in opposition to strict law . . . . And thus in chancery every particular case stands upon its own particular circumstances; and, although the common law will not decree against the general rule of law, ye chancery doth, so as the example introduce not a general mischief. Every matter, therefore, that happens inconsistent with the design of the legislator, or is contrary to natural justice, may find relief here.”).

114. Frederick Pollock, The Transformation of Equity, in FREDERICK POLLOCK, ESSAYS IN JURISPRUDENCE AND ETHICS 293 (1882) (Chancery became “as regular a court of jurisdiction as any other”); MAITLAND, supra note 7, at 9 (“In the second half of the sixteenth century the jurisprudence of the court is becoming settled.”).

115. Smith, supra note 20, at 315.

116. 2 JOHN LORD, LIVES OF THE LORD CHANCELLORS AND KEEPERS OF THE GREAT SEAL OF ENGLAND 134 (5th ed. 1868) (“They are the foundation of the practice of the Court of Chancery, and are still cited as authority.”).
Chancellorships of Lord Nottingham and Lord Hardwicke, the exercise of equity became more circumscribed, if not predictable. Chancery no longer “decide[d] every individual case according to the result of a sort of ransacking search for the particular set of conscientious principles applicable to the case.” Indeed, as Nottingham and Hardwicke “deliberately set out to reduce equity to a system of rules established by precedent,” the jurisdiction of equity “crystallized.”

But one commentator’s crystallization is another’s ossification. As the jurisdiction of equity lost its youthful exuberance, so also its freedom, elasticity and luminance. The administration of equity, much like the administration of law became bound and confined by the channels of its own precedents and the technicalities of its own procedures. Chancery,

117. Lord Nottingham served as Lord Chancellor from 1673 to 1682. See generally 4 CAMPBELL, supra note 116, at 236–79. Lord Hardwicke served from 1736 to 1756. See generally 6 CAMPBELL, supra note 116, at 158–304.

118. See Sparks, supra note 74, at 477 (“as time passed on . . . opposition gradually diminished”). See, e.g., Bond v. Hopkins, 1 Sch. & Lef. 413, 428 (1802) (“The cases which occur are various, but they are decided on fixed principles. Courts of equity have in this respect no more discretionary power than courts of law. They decide new cases, as they arise, by the principles on which former cases have been decided, and may then illustrate or enlarge the operation of those principles; but the principles are as fixed and certain as the principles on which the courts of common law proceed.”).

119. See Hanbury, supra note 119, at 205 (Nottingham “stiffened and rationalized old ideas and turned them to permanent and practical use.”); id. at 196 (detailing “the transformation from a heterogenous medley of isolated, empirical beliefs into a stable and increasingly rigid system of rules.”). See also James O’Connor, Thoughts About the Common Law, 3 CAMBRIDGE L.J. 161, 164 (1928) (referring to the “crystallized conscience” of equity). See generally MAITLAND, supra note 7, at 9 (noting that during the sixteenth century, “[t]he day for ecclesiastical Chancellors is passing away”); Paul Vinogradoff, Reason and Conscience in Sixteenth Century Jurisprudence, 24 LAW Q. REV. 373 (1908); SHAPIRO, supra note 2.

120. See Hanbury, supra note 119, at 205 (Nottingham “stiffened and rationalized old ideas and turned them to permanent and practical use.”); id. at 196 (detailing “the transformation from a heterogenous medley of isolated, empirical beliefs into a stable and increasingly rigid system of rules.”). See also Charles Synge Christopher & Baron Bowen, Progress in the Administration of Justice During the Victorian Period, in 1 SELECT ESSAYS IN ANGLO-AMERICAN LEGAL HISTORY 516, 529 (1907) (“‘No man, as things now stand,’ says in 1839 Mr. George Spence, the author of the well-known work on the equitable jurisdiction of the Court of Chancery, ‘can enter into a Chancery suit with any reasonable hope of being alive at its termination, if he has a determined adversary.’”). A vivid picture of the technicalities, delays, and expense involved in a suit in chancery is to be found in the case of Jarndyce v. Jarndyce, as related in Charles Dickens’ BLEAK HOUSE (Houghton Mifflin Co. 1956) (1853). Some have suggested that Dickens’ negative depiction is exaggerated. See

121. See Johnson, supra note 19, at 345 (“Equity became handcuffed by a rigorous body of rules and concepts.”); see also id. at 351 (“The times were not suitable for reasoned discretion. The public demanded certainty.”).

122. See Charles Synge Christopher & Baron Bowen, Progress in the Administration of Justice During the Victorian Period, in 1 SELECT ESSAYS IN ANGLO-AMERICAN LEGAL HISTORY 516, 529 (1907) (“‘No man, as things now stand,’ says in 1839 Mr. George Spence, the author of the well-known work on the equitable jurisdiction of the Court of Chancery, ‘can enter into a Chancery suit with any reasonable hope of being alive at its termination, if he has a determined adversary.’”). A vivid picture of the technicalities, delays, and expense involved in a suit in chancery is to be found in the case of Jarndyce v. Jarndyce, as related in Charles Dickens’ BLEAK HOUSE (Houghton Mifflin Co. 1956) (1853). Some have suggested that Dickens’ negative depiction is exaggerated. See
too, became a *jus strictum* differing little from the common law except in point of identity of the judicial decisions it had made its own. Indeed, by the first quarter of the nineteenth century, equity had become "so fixed, so certain, that lawyers could say, ‘There is nothing new in equity.’"

Meanwhile, the early American courts were modeling the English method of complementary systems of law and equity. Even prior to the American Revolution, "courts of chancery had existed in some shape or other in every one of the thirteen colonies." Pursuant to Article III,
Section 2 of the United States Constitution, the jurisdiction of the federal courts could extend to certain cases “in Law and Equity.” \[128\] Although Congress did not create a separate court of equity in the Judiciary Act of 1789, it contemplated that the federal court system would administer law and equity on different “sides” of the court and by different procedures.\[129\] The federal courts tried cases at law and suits in equity within a “temple of justice . . . [metaphorically] constructed with a partition extending from the foundation to the roof.” \[130\] Federal judges thus alternately played the role of common law judge or of chancellor.\[131\] Suits in equity were decided in accordance with the principles and practice of equity jurisdiction as established in the High Court of Chancery in England.\[132\]

---

128. U.S. CONST. art. III, § 2. See generally Kansas v. Colorado, 206 U.S. 46, 64 (1907) (federal judicial power extends to cases in law and in equity). See also U.S. CONST. amend. XI (declaring that judicial power shall not be construed to extend to any suit in law and equity prosecuted against the States).  
129. See Schurmeier v. Conn. Mut. Life Ins. Co., 171 F. 1, 16 (1909) (Sanborn, J., dissenting) (“The union of legal and equitable causes of action in one suit is prohibited by § 913, Revised Statutes (United States Comp. St., 1901, at 683), and in removal cases, when such a union is permitted in the state courts from which they come, the causes of action must be separated into distinct actions at law and suits in equity in the national courts.”). See generally Ingersoll, supra note 49, at 63–65 (recognizing two sides to the court “and between them there is no possible connection”); Glenn & Redden, supra note 52, at 757 (“the same judge would do equity work one day and sit as a common law judge on another”); U.S. COMP. STAT. § 913 (1901) (the forms and mode of procedure in equity shall be according to the rules and usages of courts of equity); id. § 917 (authorizing Supreme Court to prescribe rules of practice in courts of equity); JAMES LOVE HOPKINS, THE NEW FEDERAL EQUITY RULES 2–4 (7th ed. 1930).  
Equity and Procedure

The test of equitable jurisdiction long has been whether the law courts could provide “plain, adequate and complete” relief. Whenever a court of law was competent to take cognizance of a right and had the power to proceed to a judgment that afforded plain, adequate and complete relief, the plaintiff had to proceed at law because, inter alia, the defendant had a right to a trial by jury, which was available only in the law courts. Courts of equity thus steadily refused to entertain jurisdiction of actions where the law courts both recognized a right and offered a remedy. In order to deny the jurisdiction of equity the remedy at law had to be as “plain,” “certain,” “prompt,” “adequate,” “full,” “practical,” subject neither to limitation nor restraint by state legislation, and is uniform throughout the United States).

133. See Kittle, supra note 49, at 29 (“It was not . . . until the law courts began to administer justice in a more fixed and certain manner that the equity courts adopted the rule that they would not take jurisdiction where there is a complete, adequate and plain remedy at law.”); I STORY, supra note 49, § 33, at 22–26; GEORGE COOPER, A TREATISE ON PLEADING ON THE EQUITY SIDE OF THE HIGH COURT OF CHANCERY 128–29 (1813); MERWIN, supra note 66, at 17 (“equity will not take jurisdiction whenever there is a plain, adequate, and complete remedy at common law.”). See, e.g., Earl of Oxford’s Case, 1 Ch. Rep. 1 (1615). See also U.S. COMP. STAT. § 723 (1901); Jones v. Mut. Fid. Co., 123 F. 506, 517 (1903); Thompson v. Cent. Ohio R.R., 73 U.S. (6 Wall.) 134, 137 (1867); Farwell v. Colonial Trust Co., 147 F. 480, 482–83 (1906); Williams v. Neely, 134 F. 1 (1904); Brown v. Arnold, 131 F. 723, 727 (1904); Wiener v. Louisville Water Co., 130 F. 246, 250 (1903); Monmouth Invest. Co. v. Means, 151 F. 160, 165 (1906); Miller v. Steele, 153 F. 714 (1907); Wolf v. Lovering, 159 F. 91 (1908); Root v. Lake Shore & M.S.R. Co., 105 U.S. 207, 215–16 (1881); Hartford Fire Ins. Co. v. Bonner Mercantile Co., 44 F. 151, 155 (1890); McMullen Lumber Co. v. Strother, 136 F. 302 (1905); Payne v. Kan. & Ark. Valley R.R., 46 F. 546, 552 (1891); Lewis v. Cocks, 90 U.S. (23 Wall.) 466, 470 (1874); Pennsylvania v. Wheeling & B. Bridge Co., 59 U.S. (18 How.) 460, 462 (1855); Grether v. Wright, 75 F. 742, 749 (1896).


135. See, e.g., Jenkins v. Haunan, 26 F. 657, 663–64 (1885) (no equity jurisdiction where plain and adequate remedy at law by an action of ejectment for the recovery of the possession of the lands the mesne profits); Lacassagne v. Chapuis, 144 U.S. 119, 124–25 (1892).


137. Brun v. Mann, 151 F. 145, 154 (1906); Brewster, 140 F. at 816.
“just,” “final,” “complete,” and “efficient” as the remedy in equity. Naturally, this language left much to the discretion of the chancellor, and consistent with the general principle of equity to address new or unforeseen circumstances, the equities in each case controlled the court’s exercise of that broad discretion.


139. Payne v. Hook, 74 U.S. (7 Wall.) 425, 430 (1868); Am. Nat'l Bank, 89 F. at 839; W. Assurance Co., 75 F. at 342; Mann, 151 F. at 154; Brewer, 140 F. at 816; Walla Walla, 172 U.S. at 12; Sunderland, 130 U.S. at 505–15; Springfield Mill Co., 81 F. at 266.

140. Brewer, 140 F. at 816.

141. Tyler, 143 U.S. at 95; Payne, 74 U.S. at 430; Am. Nat'l Bank, 89 F. at 839; W. Assurance Co., 75 F. at 342; Walla Walla, 172 U.S. at 12; Springfield Mill Co., 81 F. at 266.


143. W. Assurance Co., 75 F. at 342; Springfield Mill Co., 81 F. at 266.

144. Tyler, 143 U.S. at 95; Payne, 74 U.S. at 430; Am. Nat'l Bank, 89 F. at 839; W. Assurance Co., 75 F. at 342; Brun, 151 F. at 154; Brewer, 140 F. at 816; Walla Walla, 172 U.S. at 12; Kilbourn, 130 U.S. at 505–15; Springfield Mill Co., 81 F. at 266.

145. Barber v. Barber, 62 U.S. (21 How.) 582, 591 (1858); Tyler, 143 U.S. at 95; Payne, 74 U.S. (7 Wall.) at 430; Am. Nat'l Bank, 89 F. at 839; W. Assurance Co., 75 F. at 342; Brun, 151 F. at 154; Brewer, 140 F. at 816; Walla Walla, 172 U.S. at 12; Kilbourn, 130 U.S. at 505–15; Springfield Mill Co., 81 F. at 266.


147. See, e.g., Watson v. Sutherland, 72 U.S. (5 Wall.) 74, 79–80 (1866); Boyce v. Grundy, 28 U.S. (3 Pet.) 210, 218–20 (1830); Sullivan v. Portland & K. R. Co., 94 U.S. 806, 811–12 (1876); Mut. Life Ins. Co. v. Pearson, 114 F. 395, 396 (1902). See also Toledo, A.A. & N.M. Ry. v. Penn. Co., 54 F. 746, 751 (C.C.N.D. Ohio 1893) (“[T]he powers of a court of equity are as vast, and its processes and procedure as elastic, as all the changing emergencies of increasingly complex . . . relations and the protection of rights can demand.”) (quoting Chicago, R.I. & P. Ry. v. Union Pac. Ry., 47 F. 15 (C.C. Neb. 1891)), cited in Wasserman, supra note 33, at 623. This same principle is reflected centuries later in the legislation enabling the Supreme Court to promulgate rules of equity procedure. See U.S. COMP. STAT. § 913 (1901) (“The forms and modes of proceeding in suits of equity . . . shall be according to the principles, rules and usages which belong to courts of equity . . . except when it is otherwise provided by statute or by rules of court made in pursuance thereof.”).

148. See supra notes 84 and 101.
II. THE DISAGGREGATION OF SUBSTANCE AND PROCEDURE

In 1753, with the process of transforming equity into a regular system of settled principles and rules well underway, Sir William Blackstone, in opening the first course on English law ever offered in an English university said that “law is to be considered not only as a matter of practice but also as a rational science.” Blackstone undertook to discern these scientific principles in four immensely influential volumes of *Commentaries on the Laws of England* published from 1765 to 1769. We might have expected the *Commentaries* to provide as valuable a picture of the condition of equity as Blackstone has given us of the condition of other parts of English law at the same period. Yet that expectation is not realized. In fact, Blackstone largely ignored equity, finding the law/equity distinction to be superficial.

149. See supra notes 113–25 and accompanying text.


152. See Holdsworth, supra note 150, at 1.

153. See Munger, supra note 49, at 49 (“Writing his treatise in 1765 in volumes comprising more than a thousand pages, [Blackstone] finds room only for a scant eight pages for a discussion of equity.”).

154. See infra notes 186–88 and accompanying text.
The substance/procedure dichotomy as the fundamental elements of law worthy of attention.\textsuperscript{155} The new paradigm had a significant effect on both the law courts and the equity courts, where substance and procedure long had been inextricably intertwined.

\textbf{A. The Integration of Substance and Procedure Prior to Blackstone}

For centuries prior to Blackstone the substance of the English common law had been buried in the cumbersome procedure of the law courts—and particularly in its pleading rules.\textsuperscript{156} The two defining characteristics of common law pleading were its processes of issue formation and its system of forms of action.\textsuperscript{157} Regarding the former, the parties by successive pleadings conceding or contesting the various contentions would reduce the dispute to a single issue then to be tried—often by a jury.\textsuperscript{158} Single-issue pleading precluded the joinder of multiple claims or defenses, the joinder of multiple parties, and pleading in the alternative.\textsuperscript{159} The system required an intricate network of highly

\begin{itemize}
  \item \textsuperscript{155} See infra notes 183–93 and accompanying text; see also Subrin, supra note 23, at 929–30 (“Blackstone atomized the study of law by separating not only rights from wrongs, but also the methods of enforcement from both. He treated English law as a rational, objective science, congruent with natural law. Blackstone, thus, disassociated the learning of rights, wrongs, and methods of enforcement from the socioeconomic-political environment.”).
  \item \textsuperscript{156} For general background on the role of pleading in procedure, see Henry John Stephen, Principles of Pleading in Civil Actions 7 (2d ed. 1901) (“The subjects of Pleadings, Practice and Evidence comprise what is commonly called the law of procedure.”); Henry R. Gibson, The Philosophy of Pleading, 2 Yale L.J. 181 (1893).
  \item \textsuperscript{157} For the historical significance and uniqueness of single-issue pleading and the forms of action, see Stephen, supra note 156, at § 132 (“As the object of all pleading or judicial allegation is to ascertain the subject for decision, so the main object of that system of pleading established in the common law of England is to ascertain it by the production of an issue. And this appears to be peculiar to that system. To the best of the author’s information, at least, it is unknown in the present practice of any other plan of judicature. In all courts, indeed, the particular subject for decision must, of course, be in some manner developed before the decision can take place; but the methods generally adopted for this purpose differ widely from that which belongs to the English law.”); and Koffler & Reppy, supra note 59, at 32 (“The Common-Law Forms of Action had their Origin in the Action and Inter-action which took place between the Chancellor and the Three Royal Courts, King’s Bench, Exchequer and Common Pleas, whereby individual litigants applied to the chancery for Original Writs authorizing one of the three Courts to try a Specific Action.”).
  \item \textsuperscript{158} See generally Stephen, supra note 156; Benjamin Shipman, Handbook of Common-Law Pleading (1894); R. Ross Perry, Common-Law Pleading (1897); Koffler & Reppy, supra note 59, at 532 (“The reduction of the controversy to Issues is the great Object of Pleading.”).
  \item \textsuperscript{159} See supra note 62; see also S.F.C. Milsom, Historical Foundations of the Common Law 70–81 (2d ed. 1981); Perry, supra note 158, at 109 (“No action could be grounded on two original writs, nor could one writ be in two forms. Consequently only such counts could be joined as could properly be grouped under one and the same original writ.”).
\end{itemize}
technical rules designed to aid or force the parties’ dispute to converge upon a single issue of law or fact. These rules earned common law pleading the dubious distinction as “the most exact, if not the most occult, of the sciences.”

The other distinguishing feature of common law pleading was the form of action. Each writ incorporated a distinct method of procedure adapted to that particular form of action. The writ governed the whole course of litigation from beginning to end, and the plaintiff selected the most appropriate writ at his peril. Comparing litigation to battle and

160. There are three fundamental rules to single-issue pleading. First, after a declaration, the parties must at each stage (i) demur; (ii) plead by way of traverse; or (iii) plead by way of confession and avoidance. Second, upon a traverse issue must be tendered. And third, the issue when well tendered must be accepted. Either by virtue of the first rule a demurrer takes place which is a tender of an issue in law, or, by the joint operation of the first two rules, the tender of an issue in fact. And then, by virtue of the second and third rules, the issue so tendered, whether in fact or in law, is accepted and becomes finally complete. It is by these rules that the production of an issue is effected. See generally Stephen, supra note 156, at § 136. Encyclopedic volumes of supplemental rules and principles ensure the production of an issue that is truly but one issue, see, e.g., id. §§ 137–49, 264–339, that is material, see, e.g., id. §§ 170–74, 340–45, and is unified, see, e.g., id. §§ 175–90, 346–71, and is certain, see, e.g., id. §§ 191–228, 372–430, and is neither obscure nor confusing, see, e.g., id. §§ 229–43, 431–52, and will lead to neither prolixity nor delay in pleading, see, e.g., id. §§ 244–49, 453–65. See also id. §§ 250–59, 466–81 (“Certain Miscellaneous Rules”); Perry, supra note 158, at 231–81 (explaining the rules and mechanics of issue pleading).

161. 2 Pollock & Maitland, supra note 49, at 612. See also 1 Pollock & Maitland, supra note 49, at 559 (explaining that, within this system “the whole fate of a lawsuit depends upon the exact words that the parties utter when they are before the tribunal”); Milsom, supra note 159, at 335 (“it is only by confusing the issues that legal development becomes possible”); supra notes 63–66. For a more positive depiction of the common law writ system, see Robert Bone, Mapping the Boundaries of a Dispute: Conceptions of Ideal Lawsuit Structure from the Field Code to the Federal Rules, 89 COLUM. L. REV. 1, 20–21 n.41 (1989).

162. See Maitland, supra note 7, at 3 (“a form of action’ has implied a particular original process, a particular mesne process, a particular final process, a particular mode of pleading, of trial, of judgment”); 2 Samuel Warren, A Popular and Practical Introduction to Law Studies, 759 (3d ed. London 1863) (a form of action is the “technical mode of framing the writ, and pleadings appropriate to the particular injury.”); Koffler, supra note 59, at 66 (“[A] form of Action’ may be defined as a Procedural Device whereby the primitive mind gave concrete expression to a theory of liability; it is a mechanism through which the doctrine or principle of Law applicable to the Statement of a Plaintiff’s Cause of Action may be enforced.”).

163. See Baker, supra note 57, at 52; Koffler, supra note 59, at 39 (“When the plaintiff petitioned the Chancellor for an Original Writ, he was under great pressure to select the right Writ for the facts of his case . . . . If he selected a Form of Writ which did not fit his case . . . . he could not succeed.”); Charles Hepburn, The Historical Development of Code Pleading in America and England 46 (1897) (“If a wrong action was adopted, the error was fatal to the whole proceeding, however clearly the facts of the controversy might have been brought before the proper court . . . . It was not enough that he stood within the temple of justice, he must have entered through a particular door.”).
the forms of action to an armory, Pollock and Maitland offer this vivid imagery:

[The system of common law forms of action] contains every species of medieval weapon from a two handed sword to the poniard. The man who has a quarrel with his neighbor comes hither to choose his weapon. The choice is large; but he must remember that he will not be able to change weapons in the middle of the combat and also that every weapon has its proper use and may be put to none other. If he selects a sword, he must observe the rules of sword-play; he must not try to use his cross-bow as a mace.\(^{164}\)

The procedural apparatus was fundamental to the common law, not simply because lawyers were more punctilious about forms than they now are, but also because the procedural institutions preceded the substantive law as it is now understood.\(^{165}\) Indeed, this glorification of form led Sir Henry Maine to suggest “that substantive law has at first the look of being gradually secreted in the interstices of procedure.”\(^{166}\) Prior to Blackstone, these forms of action were the objects of both legal reform and legal study.\(^{167}\) The principles of the common law had not been mapped out in the abstract, but instead grew around the forms by which justice was centralized and administered by the law courts.\(^{168}\) “There was no substantive law to which pleading was adjective. These were the terms in which the law existed and in which lawyers thought.”\(^{169}\)

\(^{164}\) 2 Pollock & Maitland, supra note 49, at 559. See also Hepburn, supra note 163, at 47–48 (using the same metaphor, “[a]ll the weapons of juridical warfare are here”).

\(^{165}\) See Baker, supra note 57, at 49. See also supra note 59 and accompanying text.

\(^{166}\) Henry Sumner Maine, Dissertations on Early Law and Custom 389 (Arno Press 1975) (1886). For a broader discussion of the procreative significance of procedure vis-à-vis substance, see Plunknett, supra note 23, at 379–81; Millar, supra note 23, at 3 (“Procedure belongs to the institutions of earliest development. . . . At a time when substantive legal conceptions are visible only in the faintest of outline, procedure meets us as a figure already perfected and exact.”); Maitland, supra note 7, at 3.

\(^{167}\) See Steele, supra note 58, at 10–11; 2 Pollock & Maitland, supra note 49, at 559 (“Our forms of action are not mere rubrics nor dead categories; they are not the outcome of a classificatory process that has been applied to pre-existing materials; they are institutes of the law; they are, we say it without scruple, living things.”); Perry, supra note 158, at 3 (“It may be thought these are extravagant expressions of men who were educated to see excellence in anything that was technical and abstruse. When Littleton says that the law is proved by the pleading, and when Coke adds, approvingly, ‘as if pleading were the living voice of the law itself,’ they are not using mere figures of rhetoric.”); see also infra note 170.

\(^{168}\) See Plunknett, supra note 23, at 380–81 (discussing efforts of Glanville, Bracton and Littleton). See also Baker, supra note 57, at 49–52.

\(^{169}\) Milsom, supra note 159, at 59.
Accordingly, a substantive law of, say, torts, could only be explained through the actions of trespass, case and trover. “[O]ne could say next to nothing about actions in general, while one could discourse at great length about the mode in which an action of this or that sort was to be pursued and defended.” The common law “became so interested in forms that they allowed the substance to escape.”

Meanwhile in equity courts, for centuries prior to Blackstone, procedure had been consumed by a broad substantive mandate. Whereas the common law over-emphasized form, chancery historically had eschewed it. There were no forms of action nor emphasis upon the formation of a single issue:

In the equity procedure one encounters no bewildering rules as to the name or classification of the particular suit, or according to the nomenclature at law, “forms of action.” When from an investigation of the law and facts, counsel has determined that the client has a good cause for equitable relief, he is saved the problem of wasting brain-sweat in deciding whether he shall sue in debt, assumpsit, or covenant, in trover or replevin, in trespass vi et armis or trespass on the case. He simply decides to file a “bill in equity.”

The bill in equity was to perform only two functions: to state the facts upon which the claims were based and to outline the discovery sought

---

170. See KOFFLER, supra note 59, at 65 (“The Law was required to express itself through the Limited System of Writs and Forms of Action sanctioned by precedent.”).


173. See supra note 80. See also Garrett, supra note 63, at 33, 36 (with its “crazy-quilt of allegation and counterallegation . . . procedure is, as Chancery, the vulnerable spot in an otherwise almost perfect legal system”).

174. In its earliest stages, a bill could be filed informally and there were no technical rules of pleading. See, e.g., the bills contained in SELDEN SOCIETY, supra note 79. See generally Holdsworth, supra note 54, at 12–14 (discussing procedure on bills in chancery). Over time, equity procedure became rather technical, drawing upon the principles of the English ecclesiastical courts. See JOSEPH STORY, COMMENTARIES ON EQUITY PLEADING § 13 (1857). Because the pleadings were sworn statements providing the facts upon which the case was decided, they tended to be quite detailed. See CHARLES A. KEIGWIN, CASES IN EQUITY PLEADING (2d ed. 1933); MAITLAND, supra note 7, at 6; Kittle, supra note 49; CHRISTOPHER COLUMBUS LANGDELL, SUMMARY OF EQUITY PLEADING 9–11 (2d ed. 1883).

175. EDWIN B. MEADE, LILE’S EQUITY PLEADING AND PRACTICE § 95, at 59 (3d ed. 1952) (emphasis added).
from the defendant.\textsuperscript{176} After the filing of the answer by the defendant, the only other pleading was a formal replication by the plaintiff.\textsuperscript{177} Indeed, animated by the juristic principles of discretion, natural justice, fairness and good conscience,\textsuperscript{178} the essence of a jurisprudence of equity is somewhat inconsistent with the establishment of formal rules.\textsuperscript{179} Efforts to define the jurisdiction of equity surrendered to circularity—e.g., “that body of rules which is administered only by those courts which are known as Courts of Equity.”\textsuperscript{180} Hence the characterization of equity as

---

\textsuperscript{176} See Langdell, supra note 174, at 9–11.

\textsuperscript{177} See id.; see also Meade, supra note 175, § 93, at 58 (explaining that the replication is “[t]he very brief pleading by which the plaintiff takes issue on the facts set up in defendant’s plea or answer. It is a bare denial of such facts, and its purpose is simply to put the defendant on notice that his defensive allegations of fact are not admitted, but must be established by evidence. On the filing of the replication, the parties are supposed to be at issue, and the cause matured and ready for the taking of testimony.”); Millar, supra note 23, at 25–26 (explaining procedure); W. S. Simkins, A Federal Equity Suit 462–64 (2d ed. 1911) (same). See generally Richard Marcus, Completing Equity’s Conquest? Reflections on the Future of Trial Under the Federal Rules of Civil Procedure, 50 U. Pitt. L. Rev. 725, 726 (contrasting trial practices of law and equity courts).

\textsuperscript{178} See Roscoe Pound, The Decadence of Equity, 5 Colum. L. Rev. 20, 20 (1905); see also supra notes 86–92 and accompanying text.

\textsuperscript{179} See generally Maitland, supra note 7, at 12–22; John Salmond, The First Principles of Jurisprudence I (1893) (suggesting there is no body of rules for equity). See Baldwin, supra note 65, at 64 (equity a court “of indefinite powers and unrestricted procedure”). This same principle is reflected centuries later in the legislation enabling the Supreme Court to promulgate rules of equity procedure. See U.S. Comp. Stat. § 913 (1901) (“The forms and modes of proceeding in suits of equity . . . shall be according to the principles, rules and usages which belong to courts of equity . . . except when it is otherwise provided by statute or by rules of court made in pursuance thereof.”).

\textsuperscript{180} Maitland, supra note 7, at 1. See also 1 Pomeroy, supra note 101, § 67, at 70–71 (defining equity as “those doctrines and rules, primary and remedial rights and remedies, which the common law, by reason of its fixed methods and remedial system, was either unable or inadequate, in the regular course of its development, to establish, enforce, and confer, and which it therefore either tacitly omitted or openly rejected”); Edmund H.T. Snell, The Principles of Equity 2 (18th ed. 1920) (“Equity . . . may be defined as that portion of natural justice which, though of such a nature as properly to admit of being judicially enforced, was, from circumstances hereafter to be noticed, omitted to be enforced by the Common Law Courts—an omission which was supplied by the Court of Chancery.”); Melville M. Bigelow, Elements of Equity 9 (1879) (“The jurisdiction of courts of chancery now extends to all civil cases proper in good conscience and honesty for relief or aid as to which the procedure of the common-law courts is unsuited to give an adequate remedy, or as to which the common-law courts, when able to extend their aid, have refused to do so.”); Phelps, supra note 105, at 192 (“By juridical equity is meant a systematic appeal for relief from a cramped administration of defective laws to the disciplined conscience of a competent magistrate, applying to the special circumstances of defined and limited classes of civil cases the principles of natural justice, controlled in a measure as well by considerations of public policy as by established precedent and by positive provisions of law.”); 1 Story, supra note 49, § 25, at 18 (“[E]quity jurisprudence may . . . properly be said to be that portion of remedial justice, which is exclusively administered by a court of equity, as contradistinguished from that portion of remedial
“loose and liberal, large and vague.” The broad substantive mandate dominated the jurisprudence of equity much as the procedural forms captured the jurisprudence applied in the law courts.

B. Blackstone Introduces the Substance/Procedure Paradigm

Against this backdrop, Sir William Blackstone introduced a new paradigm for understanding English law. Blackstone restated the entire corpus of English law in the form of substantive rules that he derived from fundamental principles through “solid, scientific method.” The Commentaries purport to expose and then resolve a discontinuity between fundamental moral principles, on one hand, and certain technicalities of the legal system, on the other. Blackstone constructed “a general map of the law,” connecting its “primary rules” with “the law of nature” and “the civil transactions of the kingdom.” Setting to one side the maze of legal precedents and modern procedure, Blackstone focused on the fundamental moral rights that, he argued, represented the wisdom of the ages and underlay all of English law.

According to Blackstone, these moral rights transcended the boundaries of the traditional court systems of law and of equity:

Equity then, in its true and genuine meaning, is the soul and spirit of all law: positive law is construed, and rational law is made, by it. In this, equity is synonymous to justice; in that, to the true sense and sound interpretation of the rule. But the very terms of a court of
equity, and a court of law, as contrasted to each other, are apt to confound and mislead us: as if the one judged without equity, and the other was not bound by any law. Whereas every definition or illustration to be met with, which now draws a line between the two jurisdictions, by setting law and equity in opposition to each other, will be found either totally erroneous, or erroneous to a certain degree.\footnote{186}

Indeed, he attempted to demonstrate that the rules administered by the courts of law and of equity were substantially the same.\footnote{187} The difference between the two systems, he argued, was not in their primary rules but rather in their judicial machinery: “Such then being the parity of law and reason which governs both species of courts, wherein (it may be asked) does their essential difference consist? It principally consists in the different modes of administering justice in each; in the mode of proof, the mode of trial, and the mode of relief.”\footnote{188}

Blackstone may have underrated the effect of these procedural differences upon the substantive rules of the two systems.\footnote{189} The procedural differences had “given rise to many substantial differences, which tended to grow more fundamental, as the variant effects of the procedures were worked out in detail.”\footnote{190} By taking sides on the heady jurisprudential question about the role of equity vis-à-vis law,\footnote{191}

\footnote{186. 3 BLACKSTONE, supra note 183, *429, at 269 (emphasis added).}
\footnote{187. See Holdsworth, supra note 150, at 5–6.}
\footnote{188. 3 BLACKSTONE, supra note 183, *436, at 272.}
\footnote{189. See generally James Barr Ames, Law and Morals, in Lectures on Legal History 435, 443–44 (1913) (“Blackstone has asserted that the common-law judges, by a liberal interpretation of the Statute of Westminster, by means of the action on the case, might have done the work of a court of equity. Such an opinion betrays a singular failure to appreciate the fundamental difference between law and equity, namely, that the law acts in rem, while equity acts in personam. The difference between the judgment at law and the decree in equity goes to the root of the whole matter.”).}
\footnote{190. Holdsworth, supra note 150, at 14.}
\footnote{191. For support of the Blackstone vision, see, e.g., HENRY HOME, LORD KAMES, PRINCIPLES OF EQUITY 42 (2d ed. 1767) (“equity commences at the limits of the common law, and in certain circumstances neglected by common law . . . . And thus a court of equity, accompanying the law of nature in its general refinements, enforces every natural duty that is not provided for by common law”); MAITLAND, supra note 7, at 17 (equity came “not to destroy the law but to fulfil it”); E. C. CLARK, PRACTICAL JURISPRUDENCE 1 (1883) (equity “supplements” existing rules of law by reference to current standards of morality); Christopher Columbus Langdell, Brief Survey of Equity Jurisdiction, 1 HARV. L. REV. 55, 58 (1887) (“Equity cannot therefore, create personal rights which are unknown to the law . . . nor can it impose upon a person or a thing an obligation which by law does not exist. . . . To say that equity can do any of these things would be to say that equity is a separate and independent system of law, or that it is superior to law.”); JOHN ADAMS, JR., THE
Blackstone offered little guidance on vexing if practical questions presented by separate systems of law and equity. Even if, at some level of abstraction, equity and law were the same or, at least, consistent, at a practical level, the discretionary nature of equity lacked the formal characteristics that made the common law rationally separable into parts substantive and procedural. Indeed, as a practical matter, the jurisdiction of equity could not even be defined, much less parsed. In Blackstone’s defense, it may be important to note that his Commentaries were intended for teaching purposes, and were not necessarily intended as a blueprint for reform.

C. The Substance/Procedure Dichotomy Finds Traction

Blackstone’s systematic exposition resonated with the scientific rationalism of the eighteenth century. Moreover, by downplaying the significance of any meaningful conflict between law and equity,
Blackstone’s approach had the additional virtue of imparting coherence and simplicity, familiar touchstones of successful reform efforts. One may fairly question whether the substance/procedure dichotomy is primarily theoretic, but the theory was well-received, and the consequences of its broad acceptance were very real. Blackstone demoted “procedure” to secondary status within a new hierarchy of legal precepts. Procedure was a conceptually separate apparatus, and one that was neither sacred nor fundamental. Procedure was but a set of practical and ancillary functional rules designed to remedy the wrongs that transgressed substantive rights. Commentators often credit Blackstone for liberating the substance of law from the antiquated procedural machinery that stunted the growth and progress of substantive law. True though that may be, it is procedure that was then destined to be transformed. The perceived elasticity for procedure made change in the light of practical details inevitable, if not noble. In this view procedure could be based on the experience of the ages, but unlike

---


196. See generally Stephen M. Feldman, From Premodern to Modern American Jurisprudence: The Onset of Positivism, 50 VAND. L. REV. 1387, 1411–12 (1997) (arguing that substance and procedure fundamentally are inseparable); Thurman Arnold, The Role of Substantive Law and Procedure in the Legal Process, 45 HARV. L. REV. 617, 643 (1932) (“The difference between procedure and substantive law is a movable dividing line which may be placed wherever an objective examination of our judicial institutions indicates is necessary.”); Duncan Kennedy, Form and Substance in Private Law Adjudication, 89 HARV. L. REV. 1685, 1701 (1976) (arguing from a critical perspective that form and substance are linked in subtle ways); Alan Watson, Comment, The Structure of Blackstone’s Commentaries, 97 YALE L.J. 795, 804–05 (1988) (discussing Blackstone’s largely unsuccessful effort to disentangle substantive and adjective law).


198. See PLUCKNETT, supra note 23, at 381 (discussing the lowly status of procedure); Pound, The End of Law, supra note 49, at 204–08.

199. See generally Lemmings, supra note 151.

200. See generally id.; see also Watson, supra note 196, at 811, n.62.

201. See, e.g., MILLAR supra note 23, at 4; PLUCKNETT, supra note 23, at 381.
substance, aged procedure is more likely to be viewed as senility, rather than wisdom.\textsuperscript{202}

As noted above, early American federal jurisprudence modeled the English experience. Consistent therewith, the early American court structure recognized substantive and procedural categories throughout the new systems of law and equity.\textsuperscript{203} The power was expressly conferred upon the Supreme Court to promulgate rules of procedure for equity cases,\textsuperscript{204} and Chief Justice Marshall upheld the delegation of powers to the courts to make rules in regulation of their practice.\textsuperscript{205} The Court exercised this procedural rule-making power, and the equity rules were subject to several revisions.\textsuperscript{206} By the Act of 1792, commonly known as the “Process Act,” Congress confirmed the modes of common law proceeding then used in the federal courts.\textsuperscript{207} Subsequent process and conformity acts repeated the pattern of requiring federal trial courts to apply the procedure of the state in which the federal court sat.\textsuperscript{208} Formal substantive and procedural differences were recognized throughout the regimes of law and equity, and the substance-procedure dichotomy became a fundamental characterization issue for many legal doctrines.\textsuperscript{209}

\begin{itemize}
\item \textsuperscript{202} See Thurman Arnold, \textit{The Role of Substantive Law and Procedure in the Legal Process}, 45 Harv. L. Rev. 617, 643 (1932).
\item \textsuperscript{203} See Friedman, supra note 197, at 56–57. See also Thomson v. Wooster, 114 U.S. 104, 119–20 (1885) (discussing English authorities for applying the adopted English practice). For a general discussion of the reception of the Commentaries in America, see Nolan, supra note 151; Julian S. Waterman, \textit{Thomas Jefferson and Blackstone’s Commentaries}, 27 U. Ill. L. Rev. 629 (1932).
\item \textsuperscript{204} See 28 U.S.C. § 723 (“The forms and modes of proceeding in suits of equity . . . shall be according to the principles, rules and usages which belong to courts of equity . . . except when it is otherwise provided by statute or by rules of court made in pursuance thereof.”); id. § 730. See also Act of September 30, 1789, c. 21, 1 Stat. L. 93; Act of May 8, 1792, c. 36, 1 Stat. L. 276.
\item \textsuperscript{205} See Wayman v. Southard, 23 U.S. (10 Wheat.) 1, 22–23 (1825).
\item \textsuperscript{206} See generally Hopkins, supra note 129 (reprinting and discussing the equity rules of 1822, 1842, 1866 and 1912); von Moschzisker, supra note 127, at 294 (explaining that equity rules always reflected the original assumption of Chief Justice John Jay, in 1792, that such procedure would be guided by the great tradition of equity jurisprudence as developed in the English chancery courts).
\item \textsuperscript{207} See Rev. Comp. Stat. § 913.
\item \textsuperscript{208} See 1 Julius Goebel, Jr., History of the Supreme Court of the United States: Antecedents and Beginnings to 1801 509–51 (1971) (discussing the process acts of the 1780s and 1790s); Henry Hart & Herbert Wechsler, The Federal Courts and the Federal System 17–18, 581–86 (1963) (discussing the repeated pattern of state law governing unless superseded by federal law in the process and conformity acts of the 1700s and 1800s) cited in Subrin, supra note 23, at 930 n.114.
\item \textsuperscript{209} For early cases recognizing the substantive, formal and procedural differences of law and equity, see, e.g., \textit{Green v. Mills}, 69 F. 852, 857 (1895) (jurisprudence of the United States has always recognized substantive and procedural distinctions between common law and equity); \textit{Owens v. Heidbreder}, 78 F. 837, 839 (1897) (distinction between actions at law and suits in equity is one of
\end{itemize}
Reinforcing Blackstone’s vision, scholars have referred to the disaggregation of substance and procedure as a sign of the “maturity” of a legal system.210

III. THE PROCEDURAL MERGER OF LAW AND EQUITY

The perception that parallel court systems were applying substantially similar substantive rules of law under different procedural schemata led inevitably to the notion of merger. There was no tolerance for the delays, the expense, and the technical complications that resulted from the separation of the courts of law and equity.211 Widespread and escalating contempt for procedure suggested that any distinctions were impractical and unnecessary. Procedure could better fulfill its functional and secondary role if a single set of procedural rules facilitated the joint administration of the substantive principles of both law and equity.212 This Part briefly details the experience of the merger of law and equity in American state and federal courts, as well as in the English courts. My effort here is a modest one, with a limited focus on establishing the procedural nature of these mergers. I demonstrate here that in each of these instances merging law and equity, the merger purportedly left the substantive principles of both law and equity in each jurisdiction intact.

Beginning in the middle of the nineteenth century, a reform effort to simplify legal procedure originated in the State of New York.213 The


210. See, e.g., PLUCKNETT, supra note 23, at 381 (“the power to think of law apart from its procedure . . . can only develop when civilisation has reached a mature stage”). See Pound, The End of Law, supra note 49, at 204–08.

211. See Holksworth, supra note 150, at 7.

212. See infra notes 214–18 and accompanying text.

213. Texans may disagree. See Emmerglick, supra note 67, at 244–45 (upon being admitted into the Union in 1845, Texas provided in its constitution that jurisdiction in equity would be exercised by its law courts); SMKINS, supra note 177, at 7 (referring to Texas as “a pioneer in this blended
reformers were frustrated with the practical and theoretical complexities of parallel systems of law and equity. Enticed by the rhetoric of uniformity, these reformers sought to unify law and equity into a single system of codes. Such codes offered a simple set of uniform rules better suited for the practical task of procedure to efficiently process the more important issues of substantive law. One commentator described the technicalities of common law pleading as “needless distinctions, scholastic subtleties and dead forms which have disfigured and encumbered our jurisprudence.” The reform effort was successful, as Section 62 of the new New York Code of Civil Procedure declared for New York state courts:

The distinction between actions at law and suits in equity, and the forms of all such actions and suits heretofore existing, are abolished; and there shall be in this state, hereafter, but one form of action, for the enforcement or protection of private rights and the redress or prevention of private wrongs, which shall be denominated a civil action.

Nevertheless, credit for the reform generally is accorded David Dudley Field, a New York City commercial lawyer and legal reformer. See generally Stephen N. Subrin, David Dudley Field and the Field Code: An Historical Analysis of an Earlier Procedural Vision, 6 L. & Hist. Rev. 311 (1988) (discussing the roles of Field and another New Yorker, Apthaxed Loomis, who also played a prominent role in drafting the original code); Subrin, supra note 23, at 928; Mildred Coe & Lewis Morse, Chronology of the Development of the David Dudley Field Code, 27 Cornell L.Q. 238 (1942); Kharas, supra note 12, at 186; Charles E. Clark, Handbook of the Law of Code Pleading 15 (1928) (“In 1848 came the first of the American reforms of pleading with the adoption of the New York Code.”). For an account of the nearly simultaneous English reform effort, see infra notes 284–88 and accompanying text.

214. See generally Subrin, supra note 213; Alison Reppy, David Dudley Field: Centenary Essays (1949).
215. For a discussion of the rhetoric of procedural uniformity, see supra note 195 and accompanying text.
217. For a discussion of the subservience of procedure to substance, see supra notes 198–200.
The Field Code abolished the common law forms and merged law and equity in a greatly simplified procedure.\textsuperscript{220} Code reformers took great pains to emphasize that the new codes reorganized only the procedure of law and equity.\textsuperscript{221} Accepting Blackstone’s view that substance and procedure were conceptually distinct,\textsuperscript{222} the Field Code took the additional step of recognizing the divisibility in fact of substance and procedure: “The legislative mandate of the Commissioners was reform in procedure—not alteration of the substantive rules of equity or the common law.”\textsuperscript{223}

The merged procedure of the codes borrowed heavily from equity practice.\textsuperscript{224} Much like the old bills in equity, the Field Code provided that the pleadings should state the facts;\textsuperscript{225} thus the codes, like equity, de-emphasized the importance of framing an issue.\textsuperscript{226} The Code adopted for all actions numerous equity practices and processes, including latitude in the joinder of claims and parties.\textsuperscript{227} Further, echoing King James I’s

\textsuperscript{220} Id. See N.Y. Const. art. VI, § 27 (1846) (commission appointed to “revise, reform, simplify, and abridge the rules of practice, pleading, forms, and proceedings of the courts of record of this state, and to report thereon to the legislature.”). See also Bone, supra note 161.

\textsuperscript{221} See Kharas, supra note 12, at 187. See also Philemon Bliss, A Treatise Upon the Law of Pleading 15 (3d ed. 1894) (codes “affect modes of procedure”); Coe & Morse, supra note 213, at 240–43; Subrin, supra note 213, at 329–30.

\textsuperscript{222} See supra notes 183–93 and accompanying text.

\textsuperscript{223} First Report of the Commission on Practice and Pleadings 74 (1878). See also Kharas, supra note 12, at 187; Walsh, supra note 85, at 488 (“Law and equity are simply brought together in code merger, without changing equity and without changing law, except that in cases of former conflict the equity rule necessarily displaces the legal rule.”); Gould v. Cayuga County Nat’l Bank, 86 N.Y. 75, 83 (1881) (“The distinctions between legal and equitable actions are as fundamental as that between actions ex contractu and ex delicto, and no legislative fiat can wipe it out.”).

\textsuperscript{224} Walsh, supra note 85, at 497 (emphasis added) (“Code merger makes equity far more important than before. Instead of eliminating equity or converting it into law, code merger has brought it into the modern legal system freed of the old restraints, with all its principles and practices unchanged and unimpaired, and operating directly in all cases.”).

\textsuperscript{225} The Field Code complaint was to provide “[a] statement of the facts constituting the cause of action, in ordinary and concise language, without repetition, and in such a manner as to enable a person of common understanding to know what is intended.” N.Y. Laws., c. 379 § 120(2) at 521 (71st Sess., Apr. 12, 1848) (reorganized in 1849 as N.Y. Laws, c. 438 § 142 (1849)). See also N.Y. Laws, c. 479 § 1 (1851) (“A plain and concise statement of the facts showing a cause of action without unnecessary repetition.”).

\textsuperscript{226} Clark, supra note 213, at 23 (“Instead of the issue pleading of the common law there was to be fact pleading.”); Subrin, supra note 213, at 327–38.

\textsuperscript{227} Under the Field Code, plaintiffs could be joined if they had “an interest in the subject of the action, and in obtaining the relief demanded,” and defendants if they had “an interest in the controversy, adverse to the plaintiff.” N.Y. Laws, c. 379 §§ 97–98 (1848). Several identified causes
resolution of the dispute between Bacon and Coke three centuries prior, any conflict between the substantive doctrines of law and equity was to be resolved in favor of equity.

The innovative codes proved popular elsewhere and were adopted in most states. The system inaugurated by the New York Code of 1848 was adopted promptly by Missouri and Massachusetts in 1849 and 1850, respectively. In 1851, California adopted a version of the Field Code, and prior to the outbreak of the Civil War, Iowa, Minnesota, Indiana, Ohio, the Washington Territory, Nebraska, Wisconsin and Kansas likewise enacted similar procedural codes. Within twenty-five years, procedural codes had been adopted in a majority of the states and territories. Additionally, the Field Code had at least some influence in all states, as all states departed somewhat from the common law system of pleading in response to the proliferation of the codes. For example, some of the states that did not model the codes nevertheless modified their pleading rules by statutes, allowing the assertion of equitable defenses in actions at law.

Nevertheless, the reform effort that was remarkably successful in the state courts initially drew only skepticism from the federal courts. Although law and equity were administered on different “sides” of the

---

228. See supra notes 106–11 and accompanying text.

229. See Ingersoll, supra note 49, at 70 (“whenever there is a conflict between Law and Equity, the doctrines and maxims of the latter are dominant in all civil controversies in all their courts”). See also supra note 223. See generally MAITLAND, supra note 7, at 17 (in matters of conflict or variance, “the rules of equity shall prevail”).

230. See Timothy Walker, Law Reform in Missouri, 6 W. L. JOURNAL 431 (1849); see also SWISHER, supra note 218, at 349; Report of Commissioners Appointed to Revise and Reform the Procedure in the Courts of Justice of this Commonwealth, in 2 BENJAMIN R. CURTIS, A MEMOIR OF BENJAMIN ROBINS CURTIS 149 (1879) (discussing a partial merger of law and equity under procedural codes). But see CLARK, supra note 213, § 8 (classifying Massachusetts as a “quasi-code” state, and admiring a system that “may serve as a model for code pleaders”).

231. See FRIEDMAN, supra note 197, at 394; HEPBURN, supra note 163; MILLAR, supra note 23. In 1861 Nevada adopted the Field Code, and by the end of the century so had the Dakotas, Idaho, Arizona, Montana, the Carolinas, Wyoming, Utah, Colorado, Oklahoma and New Mexico. See FRIEDMAN, supra note 197, at 394; CLARK, supra note 213, § 8.

232. See CLARK, supra note 213, at 19–20 (detailing the spread and contemporary extent of code pleading).

233. See id. at 20–22; see also Subrin, supra note 23, at 938–39.

same federal courts, a commitment to the formal separation of law and
equity was venerated and, arguably, constitutionally grounded. Justice
Grier emphasized the significance of the separation in an 1858 opinion of
the Court:

This [dual] system, matured by the wisdom of ages, founded upon
principles of truth and sound reason, has been ruthlessly abolished
in many of our States, who have rashly substituted in its place the
suggestions of sociologists, who invest new codes and systems of
pleading to order. But this attempt to abolish all species, and
establish a single genus, is found to be beyond the power of
legislative omnipotence. They cannot compel the human mind not
to distinguish between things that differ. The distinction between
the different forms of actions for different wrongs, requiring
different remedies, lies in the nature of things; it is absolutely
inseparable from the correct administration of justice in common
law courts.

Bolstered by constitutional references to systems of law and of equity,
commentators long sustained the argument that “the Federal courts
cannot adopt the blended system, nor can Congress change the present
Federal system, because it is fixed by the Constitution of the United
States.”

However, the resolve for separate systems weakened as popular
confusion and dissent mushroomed. A primary source of the confusion
and dissent was federal procedure, which, both prior and subsequent to
state adoption of the procedural codes, followed state procedure in law

235. See supra notes 126–32 and accompanying text.
237. See supra note 128; see also Bennett v. Butterworth, 52 U.S. 669 (1850). See THE
FEDERALIST NO. 83, at 569 (Alexander Hamilton) (Jacob E. Cooke ed., 1961) (advocating separate
equity courts, he wrote “to unify the [equity] jurisdiction . . . with the ordinary jurisdiction, must
have a tendency to unsettle the general rules, and to subject every case that arises to a special
determination; while a separation of the one from the other has the contrary effect of rendering one a
sentinel over the other” (emphasis added).
238. SIMKINS, supra note 177, at 3. See also Alexander Holtzoff, Equitable and Legal Rights and
Remedies Under the New Federal Procedure, 31 CAL. L. REV. 127, 130 (1942–1943) (“While the
explanation for the separation is to be found in what may be called accidents of history rather than
any a priori reasoning, this circumstance does not detract from the conclusion that the distinction
between substantive rules of law and equity is so much a part of the warp and woof of our
jurisprudence and is so deeply imbedded in it, that the classification cannot be discarded without
completely demolishing some of the foundation stones of our legal system. It is clear that no such
result was intended by the draftsmen of the rules.”).
cases and a uniform federal procedure in equity cases.\textsuperscript{239} Thus, there was a uniform simplified procedure in equity for the federal courts throughout the country. Yet in law cases the various federal courts were applying the procedure of the corresponding state court.

Federal equity practice was a model of simplicity and uniformity. Somewhat paradoxically, federal procedure in equity cases was actually a product of a certain hostility toward equity among the early colonists.\textsuperscript{240} Conformity to state practice seems to have been demanded, but it became necessary to follow the English equity procedure because a number of the states adopted no equity procedure to which conformity could be had.\textsuperscript{241} The first set of Federal Equity Rules, promulgated by the Supreme Court in 1822, contained thirty-three very concise rules of practice and procedure.\textsuperscript{242} A few of the rules were mandatory,\textsuperscript{243} but most generously accorded federal judges with broad discretionary authority.\textsuperscript{244} Moreover, after the extension of the doctrine of Swift v. Tyson,\textsuperscript{245} to equity cases in 1851, the federal courts enunciated their own views of the principles of equity jurisprudence, without restriction by the decisions of state courts.\textsuperscript{246} The Federal Equity Rules proved quite durable and were

\textsuperscript{239} Prior to the proliferation of the codes, federal procedure in law cases was governed by Congressional enactments that adopted the corresponding state court procedure as a given date; this static conformity usually was keyed to the date of the state’s admission to the union. To address some of the confusion caused by the codes, Congress passed the Conformity Acts, which required federal procedure in law cases to conform dynamically with the corresponding state procedures. \textit{See generally} Subrin, supra note 23, at 957–58, n.284; Stephen Burbank, \textit{The Rules Enabling Act of 1934}, 130 U. PA. L. REV. 1015, 1038–42 (1982). \textit{See also supra} note 208 and accompanying text.

\textsuperscript{240} \textit{See generally} von Moschzisker, supra note 127. \textit{See also supra} note 127 and accompanying text.


\textsuperscript{242} \textit{See} Miller v. Kerr, 20 U.S. (7 Wheat.) 1, 5 (1822).

\textsuperscript{243} \textit{See}, e.g., Federal Equity Rule V (1822) (“The plaintiff may amend his bill before the defendant or his attorney or solicitor hath taken out a copy thereof, or in a small matter afterwards, without paying costs; but if he amend in a material point after such copy obtained, he shall pay the defendant all costs occasioned thereby.”); \textit{id}. Rule XV (“If upon argument the plaintiff’s exceptions shall be overruled, or the defendant’s answer adjudged insufficient, the plaintiff shall pay to the defendant, or the defendant to the plaintiff, such costs as shall be allowed by the court.”).

\textsuperscript{244} \textit{See}, e.g., \textit{id}. Rule XXX (1822) (“The courts, in their sittings, may regulate all proceedings in the office, and may set aside an dismissions, and reinstate the suits on such terms as may appear equitable.”); \textit{id}. Rule XXXII (“The Circuit Courts may make further rules and regulations, not inconsistent with the rules hereby prescribed, in their discretion.”).

\textsuperscript{245} 41 U.S. (16 Pet.) 1 (1842).

substantially revised only twice in the succeeding century—in 1842 and in 1912. The latter revision was a comprehensive reform that modeled many of the provisions of the Field Code, especially those dealing with the joinder of parties.

Meanwhile, the procedure in law cases was controlled by congressional legislation requiring the federal courts to follow state procedure “as near as may be.” The Conformity Act was unpopular and true conformity seemed largely unobtainable. Noting the success of equity procedure, the American Bar Association blamed legislative control of federal practice for the problem and proposed that the power to promulgate federal rules of procedure for law cases be turned over to the United States Supreme Court. After years of debate and struggle, Congress passed a bill providing:

[T]hat the Supreme Court of the United States shall have the power to prescribe, by general rules, for the district courts of the United States and for the courts of the District of Columbia, the forms of process, writs, pleadings, and motions, and the practice and procedure in civil actions at law.

The legislation further provided that “[t]he court may at any time unite the general rules prescribed by it for cases in equity with more in actions at law as to secure one form of civil action and procedure for both . . . .” However, the Court did not rush to the task; an advisory


248. See CLARK, supra note 213, at 23.

249. The text of one of the Conformity Acts appears in the Act of June 1, 1872, ch. 255 §§ 5–6, 17 Stat. 196, 197.

250. Most people agreed that the Conformity Act was a failure. See, e.g., Charles E. Clark & James Wm. Moore, A New Federal Civil Procedure: I. The Background, 44 YALE L.J. 387, 401–11 (1935); Thomas Wall Shelton, Uniformity of Judicial Procedure and Decision, 22 LAW STUDENT’S HELPER 5, 5 (1914). See also supra note 239.

251. See Clark & Moore, supra note 250, at 435 (Federal Equity Rules of 1912 were “the substantial model for the new Federal procedure of the future”).

252. William D. Mitchell, The Federal Rules of Civil Procedure, in REPPY, supra note 214, at 75. Of course, the Supreme Court already possessed the authority with regard to equity procedure. See supra notes 204–05 and accompanying text. Moreover, the Supreme Court had rulemaking authority for actions at law pursuant to 1842 legislation, but the Court never exercised that authority. See Act of August 23, 1842, ch. 188 § 6, 5 Stat. 516, 518.


254. 28 U.S.C. § 723(b) (1934).

255. Id. § 723(c).
Equity and Procedure

committee was appointed the following year. Two years thereafter, a set of uniform rules was promulgated, eliminating the distinction between procedures for cases in equity and in law. “Under the new rules the hideous Conformity Act [was] relegated to the limbo of ‘old unhappy, far off things.’” In his address to the American Law Institute Chief Justice Hughes stated the objective of the new rules:

It is manifest that the goal we seek is a simplified practice which will strip procedure of unnecessary forms, technicalities and distinctions and permit the advance of causes to the decision of their merits with a minimum of procedural encumbrances. It is also apparent that in seeking that end we should not be fettered by being compelled to maintain the historic separation of the procedural systems of law and equity.

Carrying the torch lit by Blackstone 150 years earlier, the reformers argued that procedure had a tendency to be obtrusive, and that it should be restricted to its proper and subordinate role. The Chief Justice transmitted the Rules to Congress over the dissent of Justice Brandeis, and in 1938 the new uniform Federal Rules of Civil Procedure went into effect.

The philosophy and procedures of equity heavily influenced the tenor of the new Federal Rules.

---


258. Armistead Dobie, The Federal Rules of Civil Procedure, 25 VA. L. REV. 261, 262 (1939). See also Edson Sunderland, The New Federal Rules, 45 W. VA. L.Q. 5, 6 (1938) (“I think no tears will be shed by the bar of this country over the fact that the immense body of judicial decisions as to what matters are or are not controlled by the conformity act no longer have any value except for the legal historian.”).

259. Charles Evan Hughes, Address Before American Law Institute, cited in 55 S. Ct. 35 (1935). See also Clark, supra note 23.

260. For a description of Blackstone’s vision, see supra notes 184–88 and accompanying text.

261. 308 U.S. 645 (1938).

262. See Subrin, supra note 23, at 970–82; Marcus, supra note 177, at 725 (“[A]s to the pretrial portion of litigation, equity conquered law.”); Judith Resnik, Managerial Judges, 96 HARV. L. REV. 374, 376 (1982) (the relaxed procedures of equity allow activist judges to take control of litigation throughout the pretrial stage); Weinstein & Hershkovitz, supra note 12, at 278–81 (offering examples of how equity has dominated legal system through “modes of proof and trial,” and the “varied circumstances in which courts today turn to equitable remedies” in mass tort cases); Melissa A. Waters, Common Law Courts in an Age of Equity Procedure: Redefining Appellate Review for the Mass Tort Era, 80 N.C. L. REV. 527, 542–51 (2002) (discussing equity’s triumph in mass tort trial
applicable to all types of cases established a fluid standard of pleading. Parties could plead alternative theories. Plaintiffs were able to pursue novel theories of relief. Related and unrelated claims could be joined in a single action. Judges could hear the counterclaims and cross-claims of parties already joined in the filed action. As in equity, there were numerous specialized devices through which judges could allow the lawsuit to expand further in order to develop a more efficient litigation unit—e.g., impleaders, interpleaders, interventions, and class procedures; Laycock, supra note 12 at 53–54 ("The war between law and equity is over. Equity won."). See also Abram Chayes, The Role of the Judge in Public Law Litigation, 89 HARV. L. REV. 1281, 1292–96 (1976) (discussing in public law litigation the "triumph of equity"); 1 CHARLES E. CLARK, CASES ON PLEADING AND PROCEDURE Pref. (1930) (among other highlights of the volume were "[t]he history of equity, and its triumph over law").

263. See FED. R. CIV. P. 8(a)(2) (1938) ("A pleading which sets forth a claim for relief . . . shall contain . . . a short and plain statement of the claim showing that the pleader is entitled to relief . . . . "). Cf. Federal Equity Rule 25 (1912) ("[H]ereafter it shall be sufficient that a bill in equity shall contain, in addition to the usual caption . . . a short and simple statement of the ultimate facts upon which the plaintiff asks relief").

264. See FED. R. CIV. P. 8(e)(2) (1938) ("A party may set forth two or more statements of a claim or defense alternatively or hypothetically, either in one count or defense or in separate counts or defenses . . . . [A] party may also state as many separate claims or defenses as he has regardless of consistency and whether based on legal or on equitable grounds or on both."); id. 8(a) ("Relief in the alternative or of several different types may be demanded"). Cf. Federal Equity Rule 18 (1912) ("Unless otherwise prescribed by statute or these rules the technical forms of pleadings in equity are abolished."); id. Rule 30 ("The answer may state as many defenses, in the alternative, regardless of consistency, as the defendant deems essential to his defense.").


266. See FED. R. CIV. P. 18(a) (1938) ("The plaintiff in his complaint . . . may join either as independent or as alternate claims as many claims either legal or equitable or both as he may have against an opposing party."). Cf. Federal Equity Rule 26 (1912) ("The plaintiff may join in one bill as many causes of action, cognizable in equity, as he may have against the defendant.").

267. See FED. R. CIV. P. 13(a) (1938) ("A pleading shall state as a counterclaim any claim . . . which at the time of filing the pleader has against any opposing party, if it arises out of the transaction or occurrence that is the subject matter of the opposing party’s claim . . . ."); id. 13(b) ("A pleading may state as a counterclaim any claim against an opposing party . . . ."). Cf. Federal Equity Rule 30 (1912) ("The answer must state in short and simple form any counterclaim arising out of the transaction which is the subject-matter of the suit . . . .").

268. See FED. R. CIV. P. 14 (1938) ("[A] defendant may move . . . for leave as a third-party plaintiff to serve a summons and complaint upon a person not a party to the action who is or may be liable to him or to the plaintiff for all or part of the plaintiff’s claim against him."). The Advisory Committee Note to Rule 14 suggests that the impleader was a modern innovation developed in England. See 1937 Adv. Comm. Note to FED. R. CIV. P. 14 (citing English Rules Under the Judicature Act (The Annual Practice, 1937)).

269. See FED. R. CIV. P. 22(1) (1938) ("Persons having claims against the plaintiff may be joined as defendants and required to interplead when their claims are such that the plaintiff is or may be exposed to double or multiple liability."). See WILLIAM W. DAWSON, PROCEEDINGS OF THE
Equity and Procedure

actions. Complementing the new pleading regime were new liberal rules of discovery, and judges were vested with the authority to “manage” the case through pretrial conferences and special masters.

The Federal Rules reflected a philosophy that the discretion of individual judges, rather than mandatory and prohibitory rules of procedure, could manage the scope and breadth and complexity of federal lawsuits better than rigid rules. Indeed, Rule 1 articulated this

INSTITUTE ON FEDERAL RULES 262 (1938) (tracing the chronology of interpleader from the “old equitable action” through state codes).

270. See FED. R. CIV. P. 24(a) (1938) (“[A]nyone shall be permitted to intervene in an action when the representation of the applicant’s interest by existing parties is or may be inadequate.”); id. 24(b) (“[A]nyone may be permitted to intervene in an action . . . when an applicant’s claim or defense and the main action have a question of law or fact in common.”). Cf. Federal Equity Rule 37 (1912) (“Anyone claiming an interest in the litigation may at any time be permitted to assert his right by intervention . . . .”).

271. See FED. R. CIV. P. 23 (1938) (detailing prerequisites and types of maintainable class actions). Cf. Federal Equity Rule 38 (1912) (“When the question is one of common or general interest to many persons constituting a class so numerous as to make it impracticable to bring them all before the court, one or more may sue or defend for the whole.”).


273. See FED. R. CIV. P. 16 (1938) (“in any action, the court may in its discretion direct the attorneys for the parties to appear before it for a conferences to consider (1) The simplification of the issues; (2) The necessity or desirability of amendments to the pleadings; (3) The possibility of obtaining admissions of fact and of documents which will avoid unnecessary proof; (4) The limitation of the number of expert witnesses; (5) The advisability of a preliminary reference of issues to a master for findings to be used as evidence when the trial is to be by jury; (6) Such other matters as may aid in the disposition of the action.”). Compare the English procedure known as the “summons for directions,” in ENGLISH RULES UNDER THE JUDICATURE ACT (The Annual Practice, 1937).


275. See, e.g., Pound, supra note 23, at 402 (“It should be for the court, in its discretion, not the parties, to vindicate rules of procedure intended solely to provide for the orderly dispatch of business, saving of public time, and maintenance of the dignity of tribunals; and such discretion should be reviewible only for abuse.”); CLARK, supra note 213, at 31 (“The rules of practice should simply point out the purpose to be subserved, leaving the application thereof to the discretion of the trial judge.”). Cf. Charles E. Clark & James William Moore, A New Federal Civil Procedure II: Pleadings and Parties, 44 YALE L.J. 1291, 1323 (1935) (“In fact if the vital provisions for a completely unified procedure with clear specifications as to jury trials and waiver thereof are adopted, and if flexible rules as to pleadings and parties, leaving much to the discretion of the trial court, are drafted, we feel that the reform is assured of success, whatever the detailed provisions may be.”); Roscoe Pound, The Causes of Popular Dissatisfaction with the Administration of Justice, 29 A.B.A. REP. 395, 404–06 (1906); CLARK, supra note 23, at 308; Charles E. Clark, The Challenge of a
very purpose: “[The Federal Rules] shall be construed and administered to secure the just, speedy, and inexpensive determination of every action.”276 Commenting generally on the philosophy and durability of discretionary rules, Professor Carrington mellifluously recites: “Tight will tear. Wide will wear.”277

Like the Field Code, the reforms were directed exclusively to the procedural problem: the 1934 enabling legislation provided that “said rules shall neither abridge, enlarge nor modify the substantive rights of any litigant.”278 The Supreme Court later confirmed that “[t]he Rules have not abrogated the distinction between equitable and legal remedies. Only the procedural distinctions have been abolished.”279 The fundamental substantive characteristics that distinguished the regimes of law and equity remained intact.280 Again, in the event of any substantive conflict between law and equity, the latter was to prevail.281

276. FED. R. CIV. P. 1. See also infra notes 422–27 and accompanying text.


279. Stainback v. Mo Hock Ke Lok Po, 336 U.S. 368, 382 & n.26 (1949). See also Grupo Mexicano de Desarrollo, S.A. v. Alliance Bond Fund, Inc., 527 U.S. 308, 322 (1999); Holtzoff, supra note 238, at 130 (“[A]bolition of the procedural distinction . . . does not extend to abrogating the differentiation and demarcation between the substantive rules of equity and the substantive common law.”); Percy Bordwell, The Resurgence of Equity, 1 U. CIN. L. REV. 741, 750 (1934) (“The abolition of the common-law forms of action was not intended to change the substantive law . . . . There was no desire to destroy a single equity which anyone had had before the reform. There was a desire to abolish a red-tape which seemed intolerable.”); Dobie, supra note 258, at 262 (“Of course, [the Federal Rules of Civil Procedure are] applicable only to procedure. It is still quite proper to speak of equitable rights, equitable remedies and equitable titles.”).


281. See MATILAND, supra note 7, at 16–22 (in matters of conflict or variance, the rules of equity shall prevail); Pound, supra note 23, at 29 (equity should prevail “statute or no statute”); JOHN
Many states, in turn, modeled the federal rules for their state court procedures. In 1960, in the first comprehensive survey of state adoption of the Federal Rules, Professor Charles Alan Wright concluded that, after twenty years of operating under the Federal Rules, state procedural systems were approximately evenly divided among procedural systems modeled on the Federal Rules, the common law and the Field Code. \(^\text{282}\) Decades later, Professor John Oakley detailed “the pervasive influence of the Federal Rules on at least some part of every state’s civil procedure.” \(^\text{283}\)

Although the remainder of this essay focuses exclusively on American courts, it bears mention that the English likewise have effected a procedural merger of law and equity. \(^\text{284}\) Deconstruction of the historic separation between law and equity in England has been traced from Blackstone to an 1828 speech by Lord Brougham, \(^\text{285}\) and a series of royal commissions that led to a partial procedural fusion effected by the Common Law Procedure Act of 1854. \(^\text{286}\) Over the course of the two

\begin{flushleft}
FOREST DILLON, LAWS AND JURISPRUDENCE ENGLAND AND AMERICA 1, ch. 368 (Da Capo Press 1970) (1895); Bordwell, supra note 279, at 742.


284. For additional comparative materials, see Walter Wheeler Cook, Equity in the Canadian Courts, in CASES ON EQUITY 9–10 (4th ed. 1948) (describing origins of law and equity and the successful effort to fuse of law and equity procedures); Raymond Evershed, Is Equity Past Childbearing?, 1 SYDNEY L. REV. 1, 4 (1953) (“the truth of the matter is, I am sure, that the so-called ‘fusion’ of law and equity was a procedural matter and (save incidentally and because procedural matters cannot under our system sensibly be divorced from substantive law) the function of equity in relation to the common law was not thereby changed”).


286. See 1854, 17 & 18 Vict., c. 125 (Eng.). See generally Holdsworth, supra note 150; NEWMAN, supra note 92, at 52; ROSCOE POUND & THEODORE F. T. PLUCKNETT, READINGS ON THE HISTORY AND SYSTEM OF THE COMMON LAW 142 (3d ed. 1937)).
\end{flushleft}
subsequent decades, and coincident with the proliferation of codes in the United States, the English Court of Chancery was abolished and thereafter equity was administered in a division of a single court, the High Court of Justice.\textsuperscript{287} As with the Field Code and the Federal Rules of Civil Procedure, the Judicature Acts of 1873 and 1875, fused only the procedure of law and equity, leaving the substance of equity both intact and predominant: “generally in all matters not herebefore mentioned, in which there is any conflict or variance between the rules of equity and the rules of common law with reference to the same matter, the rules of equity shall prevail.”\textsuperscript{288}

IV. THE IMPAIRMENT OF EQUITY IN THE UNIFIED SYSTEM

The sesquicentennial of the merger of law and equity passed several years ago without honorable mention.\textsuperscript{289} Most lawyers and law students seem to be unfamiliar even with the factual underpinnings, much less the significance of the merger of law and equity.\textsuperscript{290} Nevertheless, equity enjoys a potent, even if unappreciated legacy in our unified procedural system. Many statutes and common law doctrines have incorporated the fundamental equitable principle of individualized justice. This principle is reflected in the evolution of broad principles as opposed to narrow

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{288} 1873, 36 & 37 Vict., c. 66, §. 25 (11) (Eng.); 1875, 38 & 39 Vict., c. 77 (Eng.). \textit{See generally} Mr. Justice Lurton, \textit{The Operation of the Reformed Equity Procedure in England}, 26 HARV. L. REV. 99, 100–01 (1912) (explaining that reforms made procedure distinctly subserviant to the demands of the substantive law); Britain v. Rossiter, 11 Q.B.D. 123, 129 (1882) (the Judicature Act neither creates new rights nor alters existing rights but merely changes procedure); \textit{see also} Hogg, \textit{supra} note 216, at 245; DILLON, \textit{supra} note 281, at 1, c. 368.
\item \textsuperscript{289} I could find no symposia celebrating (or lamenting) the anniversary. \textit{See also} supra note 12 and accompanying text.
\item \textsuperscript{290} I make this assertion with confidence, but with merely anecdotal data. For example, I teach the procedural history of law and equity in my first-year Civil Procedure course. \textit{See} STEPHEN N. SUBRIN, ET AL., \textit{CIVIL PROCEDURE: DOCTRINE, PRACTICE, AND CONTEXT} 274–96 (2000) (recounting historical background of civil procedure). However, my impression is that the students view it as a passion peculiar to my historical perspective, rather than something intimately related to the course. The same pattern holds true in my Remedies course, where I use the historical relationship of law and equity to contextualize the doctrine. \textit{Cf.} DOUGLAS LAYCOCK, \textit{MODERN AMERICAN REMEDIES: CASES AND MATERIALS} 7 (3d ed. 2002) (“The line between law and equity is largely the result of a bureaucratic fight for turf; each set of courts took as much jurisdiction as it could get. Consequently, the line is jagged and not especially functional; it can only be memorized.”).
\end{itemize}
\end{footnotesize}
rules, 291 broad grants of discretionary authority, 292 variable standards of conduct, 293 balancing tests, 294 lee ways of precedent, 295 and the acceptance of legal fictions. 296 Many of our fundamental legal principles originated in equity. 297 That equity intervenes when there is no adequate remedy at law is a most familiar refrain. 298 Courts routinely grant preliminary and permanent injunctions to protect legal rights when there


296 See generally Lon Fuller, Legal Fictions 9 (1967); Louise Harmon, Falling Off the Vine: Legal Fictions and the Doctrine of Substituted Judgment, 100 Yale L.J. 1 (1990); Maine, supra note 3, at 17–36.

297 “[L]et us not forget that the court of chancery was the first to ignore the absence of a seal; the first to recognize the discharge of a specialty by anything less than a specialty; the first to permit a recovery on a lost instrument; the first to enforce a moral view of penalties; the first to treat a mortgage as a mere lien; the first to hold anyone estopped by his falsehoods from proving a contrary truth; the first to recognize any ownership in the beneficiary of a trust; the first to adopt the doctrine of notice now enacted into recording statutes; the first to force parties to testify; the first to ignore form for substance; the first to disregard all contracts and even judgments when procured by fraud; and certainly the first and last court in historic times to devise remedies at all adequate for the redress of wrongs.” Frierson, supra note 32, at 412. See also supra note 85.

is no adequate remedy at law.\textsuperscript{299} Courts frequently exercise their broad discretion to award various equitable remedies such as specific performance,\textsuperscript{300} rescission,\textsuperscript{301} subrogation,\textsuperscript{302} disgorgement,\textsuperscript{303} and restitution.\textsuperscript{304} Occasionally, courts have used the awesome power of equity to create entirely new rights.\textsuperscript{305}

Yet the legacy of equity is unfulfilled in a unified procedural system if the procedural apparatus administering jointly the substantive principles of law and equity is not itself subject to the moderation and correction of the jurisdiction of equity. In this Part I demonstrate, first, that the “all-


\textsuperscript{300} See generally E. ALLAN FARNSWORTH, CONTRACTS § 12.6, at 826 (3d ed. 1999) (contemporary approach is to compare damages and specific performance to determine which is more effective in affording suitable protection to the injured party’s legally recognized interest). See, e.g., Aeronautical Indus. Dist. Lodge 91 v. United Tech. Corp., 87 F. Supp. 2d 116 (D. Conn. 2000).

\textsuperscript{301} See generally EDWARD SAMPSON THURSTON, CASES ON RESTITUTION (1940); Howard W. Brill, Equitable Remedies for Common Law, 1999 ARK. LAW NOTES 1, 7 & n.97 (citing Kennedy v. Strout Realty Agency, Inc., 490 S.W.2d 786 (Ark. 1973) (buyer entitled to rescind purchase of rental property because of misrepresentation as to the water supply; buyer recovered all payments made on the purchase price, together with compensation for improvements made on the property in good faith)). See, e.g., Roberts v. Sears, Roebuck & Co., 617 F.2d 460 (7th Cir. 1980).

\textsuperscript{302} See generally 1 DAN B. DORBS, LAW OF REMEDIES § 4.3(4), at 604 (2d ed. 1993) (“Subrogation is another equitable remedy in which tracing is used to prevent unjust enrichment and to give effective relief to the plaintiff.”); RESTATMENT (THIRD) OF PROP. MORTGAGES § 7.6 cmt. a (1997) (subrogation is an equitable remedy).


\textsuperscript{304} See generally Emily Sherwin, Restitution and Equity: An Analysis of the Principle of Unjust Enrichment, 79 TEX. L. REV. 2083 (2001). See, e.g., Reich v. Cont’l Cas. Co., 33 F.3d 754, 755–56 (6th Cir. 1994) (declaring that restitution is a historical remedy in law and in equity), cert. denied, 513 U.S. 1152 (1995). See also Smith, supra note 20, at 314 (“The remedies of cancellation, rectification and specific performance were created in the Court of Chancery.”).

equity” characterization of the Federal Rules of Civil Procedure may be overstated because, in fact, the Federal Rules increasingly bear resemblance to a strict law model. Second, I demonstrate an unfortunate consequence of a strict law paradigm: procedural insufficiency and hardship. I use recent mass tort cases to illustrate the mischief that can be created by elaborate legislative drafting that fails to contemplate all conceivable applications. Finally, I emphasize that procedural hardships were not tolerated under dual systems of law and equity, because procedural insufficiency in the law courts was a permissible basis for the exercise of equitable jurisdiction.

A. The Complication, Trivialization and Ossification of the Federal Rules of Civil Procedure

Centuries of legal history confirm that flexible and discretionary rules and standards of any form tend to rigidify over the course of time. Professor Bayless Manning recognized this general trend more than two decades ago, and coined the term “hyperlexis” to capture this “pathological condition caused by an overactive law-making gland.” Moreover, “[i]t appears to be of the essence of procedure that it tends to harden and solidify.” Procedural rules may develop in a dialectic, and short simple rules expand as they become “more nuanced, textured and complex.” It is hardly surprising, then, that the discretion and flexibility featured in the Federal Rules could suffer such an ignominious fate at the hands of complication, trivialization and ossification—the familiar pathogens of strict law.

---

306. For sources detailing the equitable origins of the Federal Rules of Civil Procedure, see supra notes 262–67 and accompanying text. See also infra note 323.

307. See MAINE, supra note 3, at 19–21 (equity is a stage in the growth of law, whereby it is expanded and ultimately fossilized). See generally Pound, Justice, supra note 49, at 711 (comparing law to the rules and formulas of the engineer, where the engineer is informed by the wisdom and experience of predecessors); Evershed, supra note 284, at 9–13 (discussing equity and modern legislation).


310. Linda S. Mullenix, Lessons from Abroad: Complexity and Convergence, 46 VILL. L. REV. 1, 2–3 (2001); see also id. at 3 (referring to “textual sprawl”).

The most compelling evidence of this trend toward a model of strict law is the number, pattern, and length of amendments to the Federal Rules. Amendments are not intrinsically maleficent, but their prevalence is consistent with a paradigm of strict law. Indeed, one way of testing the amount of discretion and flexibility that is built into a rule is to test its durability. Presumably, a rule that is written without much technical content and instead incorporates more generalized principles that can be adjusted with each application could evolve over time with a certain amount of resistance to wear and tear. The reverse is true with strict laws because the task of drafting is much more exact; the law must contemplate all conceivable applications. The virtue of “strict law” is certainty in definition and application, creating predictability. Yet the passage of time leads to circumstances that were not initially contemplated, and the law must be modified to some extent to retain its certainty in definition and application.

The number of amendments to the Federal Rules is striking and is increasing. Much, if not most contemporary procedural scholarship focuses on the need for some procedural reform. And all indications suggest that a good portion of these reform efforts have been successful. Indeed, the original Federal Rules of Civil Procedure enacted in 1938...
included eighty-six rules—combinations of which were substantially amended in each of 1948, 1961, 1963, 1966, 1970, 1980, 1983, 1985, 1991, 1993, 2000, 2001 and 2002. Another set of amendments is already in the queue and will likely take effect on December 1, 2003. And the best may be yet to come. Only ten of the original Federal Rules of Civil Procedure have never been amended. Twenty-six of the original rules—nearly one-third of the original 1938 set—have been amended at least five times. Fifty-one rules—approaching two-thirds of the original 1938 set—have been amended at least three times. Trends in the number and frequency of amendments suggest a rapidly decreasing shelf-life for Federal Rules of Civil Procedure. They also suggest that the Federal Rules may not be as discretionary and flexible as many suggest.

A second and related attribute of strictness is the length of the text of the mandate. The task of developing a scientific and complete body of rules that can be applied universally requires elaborate legislative
drafting to contemplate all future applications of that law.\textsuperscript{324} Again the evolutionary pattern of the Federal Rules is consistent with these generalizations about strict law paradigms; the many amendments described above are leading to more and longer rules.\textsuperscript{325} The original 1938 set of eighty-six Federal Rules of Civil Procedure contained approximately 29,000 words.\textsuperscript{326} The current version of the Federal Rules of Civil Procedure includes eighty-nine rules and nearly 45,000 words—more than half again as many words as the 1938 version.\textsuperscript{327} A few of the Federal Rules have become dramatically longer through the course of amendments,\textsuperscript{328} but the trend is certainly not attributable only to a few outliers: the median length of a 1938 Federal Rule was two hundred and forty seven words; the median length of a current Federal Rule is four hundred and six words.\textsuperscript{329}

Of course, these measures do not really even begin to capture the contrast between the original and current procedural infrastructures. For, as Professor Carl Tobias describes:

> Federal civil procedure is now Byzantine . . . . [T]here are strictures in the Federal Rules of Civil Procedure as well as Title 28 of the United States Code and dozens of substantive statutes. A stunning array of local measures—including local rules; general, special, and scheduling orders; individual-judge practices; and mechanisms that courts adopted under the Civil Justice Reform Act of 1990 to reduce cost and delay—also govern cases in all ninety-four districts. Many of the provisions are inconsistent or duplicative, while a significant percentage are difficult to discover, master, and satisfy . . . . The developments mean that federal practice is more

\textsuperscript{324} See generally supra note 5.
\textsuperscript{325} See Tigar, supra note 311.
\textsuperscript{326} My exact count is 28,883 words.
\textsuperscript{327} My exact count for the set of Federal Rules of Civil Procedure effective through (and including) the December 1, 2001 amendments is 44,695 words.
\textsuperscript{328} See, e.g., FED. R. CIV. P. 16 (the current version contains 1,005 words, nearly 4.5 times the length of the original (1938) version); id. 23, 23.1, 23.2 (the current version of these rules combined contain 1,300 words, well over three times the length of the original (1938) version of Rule 23); id. 26 (the current version contains approximately 3,484 words, well over four times the length of the original (1938) version); id. 32 (the current version contains approximately 1,059 words, more than three times the length of the original (1938) version).
\textsuperscript{329} The ratio at the median, then, is 1.65 times the original (1938) length.
fractured than at any time since the Supreme Court prescribed the original federal rules during 1938.\textsuperscript{330} This “bewildering panorama of requirements” exacerbates the hypertechnicality of federal practice.\textsuperscript{331}

Complexity is a characteristic of strict law that is inconsistent with a set of truly equitable rules. Importantly, neither prolixity nor complexity is “inevitable.”\textsuperscript{332} A comparison of the history of the Federal Rules of Civil Procedure to the evolutionary pattern of the Federal Equity Rules, which governed federal procedure in equity cases for well over a century, is instructive. In stark contrast to our experience under the Federal Rules of Civil Procedure, the pattern of the Federal Equity Rules demonstrates substantial periods of durability and resistance to prolixity. The original set of Federal Equity Rules, effective in 1822, contained thirty-three rules and a total of only slightly more than 2,500 words.\textsuperscript{333} Soon thereafter, in 1842, the Equity Rules were substantially revised into a set of ninety-two rules containing a total of 9,202 words. In the succeeding seventy years of practice under the Equity Rules, only seven of these rules were amended, and only one rule was amended more than once.\textsuperscript{334} Only three new rules were added.\textsuperscript{335}

In 1912, the equity rules were reorganized and substantially revised.\textsuperscript{336} The 1912 set of Equity Rules contained only eighty-one rules and a total


\textsuperscript{331} Tobias, Local Federal Civil Procedure, supra note 330, at 533.

\textsuperscript{332} See generally Cooper, supra note 28, at 1795 (“It may be inevitable that a continuing revision process lengthens the rules and adds complexity to them. Doubts grow up around old solutions, and new problems appear. The Civil Rules have not escaped this effect.”). But see id. at 1795 (“It does not seem fair to charge the revision process with a descent into the nagging detail and sterile ossification that have overtaken earlier procedural systems.”).

\textsuperscript{333} My exact count is 2,525 words.

\textsuperscript{334} Rule 13 was amended in 1875; Rules 18 and 19 were amended in 1878; Rule 40 was amended in 1859; Rule 41 was amended in 1871; Rule 67 was amended in 1861, 1869, 1892 and 1892; and Rule 82 was amended in 1894. See generally Hopkins, supra note 129.

\textsuperscript{335} Rules 92, 93 and 94 were added in 1863, 1878 and 1881, respectively. See generally Hopkins, supra note 129.

\textsuperscript{336} Professors Tidmarsh and Transgrud have suggested that the 1912 rules were very different from traditional equity practice, because the 1912 rules had been greatly influenced by common law procedure, code pleading, and notice pleading. See Jay Tidmarsh & Roger H. Transgrud, Complex Litigation and the Adversary System 59 (1998). Such an evolutionary pattern for the Federal Equity Rules would be consistent with the ossification attributed to flexible and
Thus, over the course of these seventy years, the procedure in equity cases experienced a decline in the number of rules and a negligible increase of less than three percent in the total number of words. The Equity Rules of 1912 remained in effect until 1938 when they were replaced by the new Federal Rules of Civil Procedure, which purported to model the procedure of equity. Remarkably, the original set of Federal Rules was more than three times the length of their 1912 mold. The current Federal Rules are nearly five times the length of the Equity Rules of 1912.

To be sure, generalizations about the number of amendments to and the textual girth of the Federal Rules cannot prove the emergence of a strict law paradigm. However, the content of these amendments further substantiates the hypothesis. The contemporary class action is one of the many procedures of the merged system borrowed from equity. In 1842, the Supreme Court enacted Equity Rule 48, which provided in its entirety:

Where the parties on either side are very numerous, and can not, without manifest inconvenience and oppressive delays in the suit be all brought before it, the Court in its discretion may dispense with making all of them parties, and may proceed in the suit, having sufficient parties before it to represent all the adverse interest of the plaintiffs and the defendants in the suit properly before it. But, in

---

discretionary rules generally and to procedural rules in particular. See supra notes 307–10 and accompanying text. And their suggestion that the Federal Rules may not be as “all equity” as we long have thought advances considerably my thesis here.

337. The 1912 Federal Equity Rules contained 81 rules and 9,442 words.

338. See supra notes 240–277 and accompanying text.

339. The ratio is approximately 3.06.

340. The ratio is approximately 4.73.

341. See Richard L. Marcus, The Puzzling Persistence of Pleading Practice, 76 TEX. L. REV. 1749, 1778 (1998) (“For much of this century, the prevailing view has been that change in procedure tends always toward the more relaxed and away from the more rigid. That certainly describes the Federal Rules’ treatment of pleadings. But one may doubt the universal attractiveness of unbridled flexibility. More to the point, what goes up can come down, and change may move toward constraint rather than latitude. In many ways, the last quarter century has produced changes in the rules that sought to constrain rather than to liberate.”).

342. See supra note 271 and accompanying text. See also infra notes 394–97 and accompanying text.
such cases, the decree shall be without prejudice to the rights and claims of all the absent parties.343

This Rule was not amended for seventy years, until it was refined in 1912 by Equity Rule 38, which provided, in full, the following forty words:

When the question is one of common or general interest to many persons constituting a class so numerous as to make it impracticable to bring them all before the court, one or more may sue or defend for the whole.344

The original drafters of the Federal Rules of Civil Procedure “adopted” this device of equity, but drafted a rule of three hundred and ninety three words, nearly ten times longer than its model.345 Many decades and amendments later, Federal Rule 23 and its textual progeny currently bespeak 1,300 words, more than thirty times the length of its primogenitor.346 The evolution has added prerequisites to, a typology for, and limitations upon a trial judge’s discretion to certify a class.347 Further, earlier this year, the Court approved amendments to Rule 23 that

343. Equity Rule 48 (1842). See generally HOPKINS, supra note 129. This codification announced a pre-existing rule of equity that was not codified at all. See, e.g., West v. Randall, 29 F. Cas. 718, No. 17,424, 2 Mason 181 (1820).

344. The rule promulgated in 1842 was not amended prior to this revision in 1912.


would revise and lengthen two of the existing sections and add two new sections.348

B. Procedural Mischief in Mass Tort Cases

In this section I offer an example of procedural mischief occasioned by a textual rule that increasingly may leave too little discretion for judges to reach the fair, just and efficient result. The dynamics of certain mass tort claims can impose extraordinary demands on the judicial system.349 The challenges presented by cases involving millions of plaintiffs can place strains on existing mechanisms to effect consolidation for discovery and pretrial management, and resolution by trial or settlement.350 Included among the Federal Rules of Civil Procedure that are bursting at their seams are Rules 19, 20, 22, 23, 24, 42, 49, 51, and 53.351 I focus on only one of these Rules here, but the problem is—or, at least is becoming—endemic to the broader procedural infrastructure. The problem is this: “a court facing a complex case must choose the most efficient procedure that violates no normative

348. See supra note 318.


350. See generally Paul Niemeyer, Report on Mass Tort Litigation, 187 F.R.D. 293, 304–05 (1999); see also Cimino v. Raymark Indus., 751 F. Supp. 649, 652 (E.D. Tex. 1990) (if the district court could close “thirty cases a month, it would [still] take six and one-half years to try these cases and [due to new filings] there would [still] be pending over 5,000 cases”).

351. See generally supra notes 349–50.
component of adjudication." But the existing adjudicatory system is structured in a way that, for the most part, contemplates the individualized resolution of disputes. To be sure, there are numerous aggregation mechanisms, including class actions, which contemplate a certain collective justice. According to many commentators, the procedural prerequisites drafted to identify those actions where the rule might be most efficiently used are now a major source of inefficiency, unfairness and uncertainty. Indeed, the undesirable outcomes flowing from flaws inherent in the existing structure of Federal Rule 23 have satisfied few if any of the constituencies affected by class action procedure.

In the context of mass tort class actions, the procedural questions are truly extraordinary. Yet no matter how bizarre or unprecedented the phenomenon, empirically or theoretically unsound the technique, or inefficient or unfair the application, the certification of a mass tort class action can focus on the particulars of a Federal Rule to the exclusion of these contextual factors. For example, in Ortiz v.
Fireboard Corp., the Supreme Court for the second time in three years rejected a settlement in a class action case “prompted by the elephantine mass of asbestos cases.” The opening line of Justice Souter’s majority opinion declared: “This case turns on the conditions for certifying a mandatory settlement class on a limited fund theory under Federal Rule of Civil Procedure 23(b)(1)(B).” As it did in Amchem Products, Inc. v. Windsor, the Court methodically and strictly applied the prerequisites for certification, and ultimately rejected the class action settlement, notwithstanding its acknowledgment that the settlement was both substantively and procedurally fair.

Of particular interest, the Ortiz Court acknowledged that “this litigation defie[d] customary judicial administration.” Ironically, custom prevailed nevertheless. Chief Justice Rehnquist joined the Court’s decision and, in a separate opinion, acknowledged the insufficiency of the outcome but demurred to creativity: “Under the present regime, transactional costs will surely consume more and more of a relatively static amount of money to pay these claims. But we are not free to devise an ideal system for adjudicating these claims. . . . Unless and until the Federal Rules of Civil Procedure are revised, the Court’s opinion correctly states the existing law.”

In Amchem and Ortiz, the Court upheld the general principle that Rule 23 requires structural protections for absent class members that class action settlements and their proponents must respect. More specifically, the Court in Ortiz held that for a class action settlement to qualify as a “limited fund,” the settlement proponents must prove the limits of the

363. See Ortiz, 527 U.S. at 821; see also id. at 830 (“The nub of this case is the certification of the class under Rule 23(b)(1)(B).”). See Marcus, supra note 356, at 2028, n.102.
365. See Ortiz, 527 U.S. at 821. See generally Cohen, supra note 362.
366. Ortiz, 527 U.S. at 821; see also id. at 866 (Breyer, J., dissenting) (quoting a Judicial Conference report that referred to the mass of asbestos cases as “having ‘reached critical dimensions’ threatening ‘a disaster of major proportions’”).
367. Id. at 865 (Rehnquist, C.J., concurring).
fund by more than their agreement, protect class members from the class
counsels’ conflict of interest in settling the claims of their clients outside
the class, create separately represented subclasses for serious conflicts of
interest among class members, and justify the failure to exhaust the entire
fund.368 The rule of law exists in the class action setting, the Court
especially says, despite the enormous pressures to bend, skirt, or ignore
the rules.369 But when does attention to the rule of law become a
preoccupation, a fetish?

As the majority in Amchem and Ortiz took pains to note, the issues
before the Court in the two cases arose out of “near-heroic efforts” by the
plaintiffs’ lawyers, defense counsel, and district court judges to resolve
litigation and thereby to make the best of a bad situation.370 The class
action settlement that the district judge had approved in Georgine v. Amchem Products, Inc.,371 would have settled all future asbestos personal
injury claims, except the claims of those who chose to opt out, against
the consortium of asbestos defendants.372 The class-action settlement that
was approved by the district judge in Ahearn v. Fibreboard Corp.373 and
later challenged in Ortiz would have settled virtually all future asbestos
personal injury claims against the defendant Fibreboard on a mandatory
basis.374 Notwithstanding their sympathies with the district courts’
efforts, in both cases the majority concluded that Rule 23 does not permit
the resolutions that the parties had negotiated and the district court
judges had approved.375

The Federal Rules, thus, may constrain judicial inventiveness. Courts
generally accord the Federal Rules of Civil Procedure the binding status
of statutes.376 Judges may apply the letter of the rule notwithstanding
“worthy goals and lofty-stated purposes” supporting a contrary result. Judges may even celebrate an insistently narrow construction. They may distance themselves from—or even apologize for—results that they suggest are foreordained. Or courts may express a preference for a result that “must be obtained by the process of amending the Federal Rules, not by judicial interpretation.” Courts generally seem reluctant to use their equitable powers to correct procedural insufficiencies. The notable exceptions here help to prove the rule.


378. See, e.g., Healy v. Pa. R.R., 181 F.2d 934, 937 (3d Cir. 1950) (“[I]t cannot be gainsaid that certain formalities are indispensable to ‘just, speedy, and inexpensive’ litigation, and these attributes of our federal judicial system are forthcoming only upon adherence to, rather than upon rejection of, the Rules.”), cert. denied, 340 U.S. 935 (1950); Pan Am World Airways v. U.S. Dist. Ct., Cent. Dist. of Cal., 523 F.2d 1073, 1082 (9th Cir. 1975) (Schackne, J., dissenting) (criticizing the majority for an “overly technical” reading of Rule 42 and for “contemplat[ing] the wrong question. The question is not whether some rule permits the action proposed, but whether any rule, statute or logical concept forbids it”).


380. See Tuke, 76 F.3d at 157. See also Amchem, 521 U.S. at 620.

381. Leatherman v. Tarrant County Narcotics Intelligence & Coordination Unit, 507 U.S. 163, 168 (1993) (“Perhaps if Rules 8 and 9 were rewritten today, claims against municipalities under § 1983 might be subjected to the added specificity requirement of Rule 9(b). But that is a result which must be obtained by the process of amending the Federal Rules, and not by judicial interpretation.”).

382. See, e.g., Swierkiewicz v. Sorema N.A., 534 U.S. 506, 508 (2002) (relying on Fed. R. Civ. P. 8(a) concerning denial of motion to dismiss); Guthrie v. Am. Broad. Co., 733 F.2d 634, 637 (4th Cir. 1984) (“Whatever inherent powers the district courts may once have had, they now have no power to issue a deposition subpoena unless expressly or impliedly so authorized by the Rules. Neither plaintiff nor the District Court has identified any authority in support of such an extraordinary notion of ‘inherent power’ exceeding the scope of the Rules.”); Guilford Nat’l Bank v. S. R.R., 297 F.2d 921, 925 (4th Cir. 1962) (“The Federal Rules of Civil Procedure should be liberally construed, but they may not be expanded by disregarding plainly expressed limitations. We are not prepared to depart from the explicit language of Rule 34 when viewed in the context of the entire discovery section.”).

383. See Schuck, supra note 349, at 974 n.148 (speculating that “some judges in mass tort cases, such as Jack Weinstein, Robert Parker, and Thomas Lambros, sometimes compete to be the most innovative”); Waters, supra note 262, at 552–53, n.104 (“By employing an amazing variety of
 Courts are more inclined to wait for Congress or other rulemakers to find solutions. Accordingly, either the insufficiency persists or it leads eventually to an amendment which, in turn, tends to make the rule even longer, more specific, and more likely to work a hardship when it, in turn, is applied in a particular situation. As this pattern of hardship-and-amendment repeats, the Federal Rules of Civil Procedure increasingly resemble a strict law paradigm, and the impairment of equity becomes even more pronounced.

C. The Traditional Model for Correcting Procedural Insufficiencies

In addition to the more conventional forms of relief, traditional equity interfered to offer relief in cases where common law processes were defective or too complicated.

flexible and innovative approaches to both procedural and substantive law, these specialists have become ‘adept at routinizing the extraordinary.’") (quoting Schuck, supra note 349, at 956).

384. See supra notes 376–83 and accompanying text. See also Grupo Mexicano de Desarrollo, S.A. v. Alliance Bond Fund, Inc., 527 U.S. 308, 322 (1999) (‘‘When there are indeed new conditions that might call for a wrenching departure from past practice, Congress is in a much better position than we both to perceive them and to design the appropriate remedy.’’).

385. For a discussion of the doctrine and cases where the cause of action involved the maintenance or protection of an equitable right, estate, or interest, see, e.g., JARIUS PERRY, TRUSTS § 22 (1874) (cestui que trust); 2 LEONARD JONES, A TREATISE ON THE LAW OF MORTGAGES § 1093 (1881) (mortgagor’s equity of redemption); POMEROY, supra note 216, § 124 (rights of assignees of choses in action); CHARLES BEACH, MODERN EQUITY JURISPRUDENCE § 287 (1892) (rights of equitable lienors). There were (and are) also purely equitable remedies, which were (and are) administered by courts of equity, and not by courts of law. These remedies do not necessarily depend upon the estate involved for their equitable character. They are equitable because they can only be obtained in courts of equity. A suit for quieting title, or for removing a cloud upon title by the cancellation of an adverse instrument, may result in the establishment of a legal estate; but the remedy nevertheless is within the exclusive jurisdiction of a court of equity. A suit for the specific performance of a contract is within this exclusive jurisdiction, although an action might be maintained at law for the recovery of damages for the breach of the contract sought to be enforced. See JAMES W. EATON, HANDBOOK OF EQUITY JURISPRUDENCE § 7, 30–31 (1901). See also supra notes 80–85 and accompanying text.

386. In addition to circumstances discussed infra in the text, there are many situations where procedural insufficiency justified equitable intervention. For example, equity would interfere where the plaintiff had a valid defense at law, but it was doubtful whether he could plead it in an action at law. See, e.g., Davis v. Wakelee, 156 U.S. 680, 692 (1895) (sustaining a suit in equity on a void judgment where defendant was estopped to question the judgment, but where it was doubtful whether that issue could be raised by the pleadings in an action at law). When parties were not competent witnesses in common law courts, equity might give relief because of the difficulty of proof caused thereby. See, e.g., Taylor v. Merchs. Fire Ins. Co., 50 U.S. (9 How.) 390 (1850) (addressing a suit to enforce a contract to issue a fire insurance policy after the loss had occurred where proof of the terms of the proposed policy would be difficult at law, though the court also relied on the fact that specific performance would clearly have been decreed before the loss and that
common law were easily amenable to the procedural devices of equity. Even after the crystallization of equity jurisprudence in the nineteenth century, in order to deny the jurisdiction of equity, the remedy at law had to be as plain, certain, prompt, adequate, full, practical, just, final, complete, and efficient as the remedy in equity. For example, equity interfered in the name—and with the imprimatur—of efficiency to avoid the injurious effects of a multiplicity of actions. Describing the contrast between law and equity in these instances, Professor Chafee wrote:

A common-law action soon came to be a two-sided affair, usually with only one plaintiff and one defendant but sometimes with several plaintiffs or defendants tightly bound together as joint obligees or obligors, etc. Except in such joint situations, however, a dispute of one person against many persons usually had to come before the law courts, if at all, in the form of many separate actions. Hence it was far cheaper and more convenient to have a single suit.

---

(event did not defeat equity jurisdiction). Where a purchase of land was, after the conveyance, rescinded for fraud, equity might give relief by allowing the purchaser to recover the price paid, because the common law court had no means of revesting the title in the vendor. See, e.g., Bullard Shoals Mining Co. v. Spencer, 95 So. 1 (Ala. 1922). Equity could also aid in enforcing a common law judgment against equitable or concealed assets. See, e.g., Overmire v. Haworth, 51 N.W. 121 (Minn. 1892); State Bank of Ceresco v. Belk, 94 N.W. 617 (Neb. 1903).

388. See supra notes 112–25.
389. See supra notes 136–46.
390. The reference to a “multiplicity of actions” can be confusing because equity exercised jurisdiction in four types of cases involving a multiplicity of actions—(i) where the nature of the wrong is such that at law it would be necessary for the injured party, in order to obtain complete relief, to bring a number of actions, arising from the same wrongful act against the same wrongdoer; (ii) where a party institutes, or is about to institute, a number of successive or simultaneous actions against another party, all depending upon the same legal questions and similar issues of fact; (iii) where a party claims a common right against a number of persons, the establishment of which would require a separate legal action brought by him against each of such persons, and which are of such a nature that they might be determined in a single suit in equity brought against all of such persons; and (iv) where a number of persons have separate and distinct rights of action against the same party, arising from the same cause, governed by the same legal rule, and involving similar facts, and the circumstances are such that the rights of all may be settled in a single suit. See generally 1 POMEROY, supra note 101, §§ 243–275, at 318–377. References herein to a “multiplicity of actions” refer to group (iv). See also HENRY L. MCCINTOCK, HANDBOOK ON THE PRINCIPLES OF EQUITY § 178 (2d ed. 1948) (“the plight of a defendant at law, subjected to one hundred and ten separate actions arising from the same accident, many of the actions being brought in different counties and some of them set for trial in the different counties at the same time, so that it would be impossible for the witnesses for the defense to attend each trial, is one that calls for some sort of relief if it can be given”) (citing S. Steel Co. v. Hopkins, 47 So. 274 (1908)).
in chancery, which was accustomed to handle polygonal controversies... [I]t was an obvious waste of time to try... common question[s] of law and fact over and over in separate actions at law.... It was much more economical to get everybody into a single chancery suit and settle the common questions once and for all.  

Thus, a court of equity would hear a controversy to prevent a multiplicity of suits, even if the exercise of such jurisdiction called for adjudication on purely legal rights and to confer purely legal relief. Moreover, when the number of plaintiffs or defendants were too numerous to join in a single suit, equity would permit a few of the litigants to represent the many in connection with an equitable bill of peace, the ancestor of the contemporary class action.

Of course, Professor Yeazell’s remarkable historical inquiry suggests that today’s class actions are not true lineal descendents of the English equity devices. Although Chancery intervened to prevent a multiplicity of suits at common law, for the most part, these aggregations were in the

---

393. See POMEROY, supra note 101, § 269, at 367–68 (“[T]he jurisdiction may and should be exercised, either on behalf of a numerous body of separate claimants against a single party, or on behalf of a single party against such a numerous body, although there is no ‘common title,’ nor ‘community of right’ or of ‘interest in the subject-matter,’ among these individuals, but where there is and because there is merely a community of interest among them in the questions of law and fact involved in the general controversy, or in the kind and form of relief demanded and obtained by or against each individual member of the numerous body.”). See generally STEPHEN C. YEAZELL, FROM MEDIEVAL GROUP LITIGATION TO THE MODERN CLASS ACTION (1987); CHAFFEE, supra note 391, at 200. See also Watson v. Nat’l Life & Trust Co., 162 F. 7, 7–12 (1908); United States v. Old Settlers, 148 U.S. 427, 480 (1893); Barnes v. Berry, 156 F. 72, 73 (1907); Am. Steel & Wire Co. v. Wire Drawers’ & Die Makers’ Union Nos. 1 & 3, 90 F. 598, 606 (1898); Soc’y of Shakers v. Watson, 68 F. 730, 741 (1895); McArthur v. Scott, 113 U.S. 340, 404–06 (1885); Ayres v. Carver, 58 U.S. (17 How.) 591, 593–96 (1854); Smith v. Wormstedt, 57 U.S. (16 How.) 288, 302 (1850); Scott v. Donald, 165 U.S. 107, 108 (1897); United States v. Coal Dealers’ Assoc., 85 F. 252, 260 (1898); Bailey v. Tillinghast, 40 C.C.A. 93, 99 F. 801, 807 (1900).
394. See generally YEAZELL, supra note 393.
nature of declaratory judgments, rather than damage actions.\textsuperscript{395} Further, virtually all of the ancient collective litigation cases concerned disputes about some traditional duty owed by one party to the other.\textsuperscript{396} Professor Yeazell thus contends that courts tolerated early group litigation solely because it assisted in reinforcing social norms between and among preexisting social groups.\textsuperscript{397} One might argue, then, that the class suit was developed in English chancery primarily to enable the plaintiff to obtain relief for himself, and not as an all-purpose technique to cure procedural insufficiencies presented in the law courts.

Yet the relevance of the history of equity here is not about finding a specific analogue. Rather we must consider what the equity courts would have done had they been faced with the contemporary crises.\textsuperscript{398} Indeed, the procedural devices in equity were descriptive, not technical, and a precise statement of their scope is impossible.\textsuperscript{399} Most important, to some extent or other, procedural insufficiency was an independent ground for equitable action.\textsuperscript{400} Accordingly, if procedural insufficiency is no longer

\textsuperscript{395} See id. at 148–51.

\textsuperscript{396} See, e.g., Brown v. Booth, 23 Eng. Rep. 720 (Ch. 1690) (suit by vicar against parish miners); How v. Tenants of Bromsgrove, 23 Eng. Rep. 277 (Ch. 1681) (suit by lord of manor against tenants). See also Tribette v. R.R., 70 Miss. 182, 188 (1892) ("There must be some recognized ground of equitable interference, or some community of interest in the subject matter of the controversy, or a common right or title involved, to warrant the joinder of all in one suit; or there must be some common purpose in pursuit of a common adversary, where each may resort to equity, in order to be joined in one suit; and it is not enough that there is a community of interest, merely, in the questions of law or fact involved."); Lehigh Val. R. Co. v. McFarlan, 31 N.J. Eq. 730 (1879); Nat'l Park Bank of N.Y. v. Goddard, 131 N.Y. 494 (1892); Hanstein v. Johnson, 17 S.E. 155 (1893); N. Pac. R.R. v. Amacker, 46 F. 233 (1891).


\textsuperscript{398} In fact, part of Yeazell’s thesis is that the modern-day class action developed in response to changing social and economic issues that accompanied industrialization and the entrepreneurial nature of society. See YEAZELL, supra note 394, at 39–40 (asserting that the evolution of the modern day class action device over the past eight centuries was not a consistent and unified development, but rather a fragmented and broken one). Cf. Robert G. Bone, Personal and Impersonal Litigative Forms: Reconcieving the History of Adjudicative Representation, 70 B.U. L. REV. 213, 222–26 (1990).

\textsuperscript{399} See CHAFFEE, supra note 391, at 149–295. See also Wyman v. Bowman, 127 F. 257, 264 (1904) (equity court’s broad discretion re multiplicity suits to be exercised upon consideration of, \textit{inter alia}, "the substantial convenience of all parties") (emphasis added); Hosmer v. Wy. & I. Co., 129 F. 883, 888 (1904) (citing authorities); Hale v. Anderson, 188 U.S. 56, 77 (1903); Buchanan Co. v. Adkins, 175 F. 692, 791 (1909); Hyman v. Wheeler, 33 F. 629, 630 (1888); De Forest v. Thompson, 40 F. 375, 377 (1889).

\textsuperscript{400} See supra notes 75–78, 146–48 and accompanying text.
an independent ground for invoking the moderating and corrective function of equity, the jurisdiction of equity is impaired.

V. THE RESURRECTION OF EQUITY TO CURE PROCEDURAL INSUFFICIENCIES

The moderating force of equity ensures just results in each application of the strict law and also fulfills an essential role in the dialectic evolution of the law. The legal historian John Millar captured this idea well:

Law and equity should be in continual progress, with the former constantly gaining ground upon the latter. Every new and extraordinary interposition is, by length of time, converted into an old rule. A great part of what is now strict law was formerly considered as equity, and the equitable decisions of this age will unavoidably be ranked under the strict law of the next.

As the procedural infrastructure of our unified system grows increasingly formulaic and prescriptive, judges can and should supply any deficiencies created or ignored by strict applications of that procedural law. Further, the fair, just and efficient application of substantive law should be accommodated by procedural rules that feature broad judicial discretion, rather than elaborate legislative drafting that is undertaken to contemplate all conceivable applications.

---

401. See Beale, supra note 127, at 25 (“In a system which has, separately, law and equity, the doctrines of equity represent the real law.”). See also Mullenix, supra note 310, at 2–3 (referring to the dialectic of complexity and simplicity in procedural reform).

402. 2 JOHN MILLAR, AN HISTORICAL VIEW OF THE ENGLISH GOVERNMENT 358 (1789). See also Munger, supra note 49, at 50–51 (discussing the history and progress of equity “from conscience to precedent”); Bordwell, supra note 279, at 749 (“The equity of today becomes the right of tomorrow.”); FREDERIC R. COUDERT, CERTAINTY AND JUSTICE 1 (1914) (“On the one side is made an appeal to progress, on the other to precedent.”).


404. Cf. Bordwell, supra note 279, at 749 (expressing “hope for a militant equity that will repeat the conquests of former days”).
A. “Mildening” the Application of Procedural Rules With Equity

Since its beginnings, equity has been the extraordinary justice administered to enlarge, supplant, or override strict law that has become too narrow and rigid in its scope. A separate system of equity was free to exert this reforming purpose, and for centuries prior to the merger, “the two systems had been working together harmoniously.” Some commentators have suggested that equity requires structural autonomy to perform this “high office”—to act “as a check upon strict law and in opposition to it.” For example, Pomeroy demonstrated that, in merged systems:

The tendency . . . has plainly and steadily been towards the giving an undue prominence and superiority to purely legal rules, and the ignoring, forgetting, or suppression of equitable notions . . . . In short, the principles, doctrines, and rules of equity are certainly disappearing from the municipal law of a large number of the states, and this deterioration will go on until it is checked either by a legislative enactment, or by a general revival of the study of equity throughout the ranks of the legal profession.

Arguments for a separate system of equity likely are antediluvian, but the virtues of equity should not require the formal establishment of dual systems.

405. The corrective function of equity is a power that the Anglo-Saxons called “mildening law.” See Smith, supra note 53, at 209.
406. See Maftland, supra note 7, at 17.
407. Emmerglick, supra note 67, at 255.
408. Pomeroy, supra note 101, at ix. See also Cooth v. Jackson, 6 Ves. Jr. 39 (Lord Eldon). See generally Frierson, supra note 32, at 411 (expressing doubt about “whether the fullest benefit of the principles of equity has ever been attained under the amalgamated system”).
409. Of course, three states currently maintain separate law and equity courts. The Delaware constitution vests the state's judicial power in several courts, including “a Court of Chancery.” Del. Const. art. 4, §1. The constitution of the State of Mississippi provides for a chancery court with jurisdiction in equity cases. See Miss. Const. art. 6, § 159. And under the Tennessee constitution, the judicial power is vested “in such Circuit, Chancery, and other inferior Courts as the Legislature shall from time to time ordain and establish . . . .” Tenn. Const. art. 6, § 1. In November 2000, Arkansas voters approved Amendment 80, which rewrote virtually the entire judicial article of the state Constitution of 1874. One of the amendment’s “fundamental purposes [was] the merger of law and equity.” In re Implementation of Amendment 80: Admin. Plans Pursuant to Admin. Order No. 14, 345 Ark. Adv. App. (June 28, 2001) (per curiam).

My limited and preliminary inquiry into the systems of these states suggests, however, that none of the three operate dual systems in the traditional sense. See John J. Watkins, Law and Equity in Arkansas—Or, Why to Support the Proposed Judicial Article, 55 Ark. L. Rev. 401, 436, n.421 (2000) (suggesting that, in Delaware, the chancery court is largely but a specialized tribunal for
systems. Query whether a more sophisticated understanding of the origins of equity and the mechanics of the merger could cure some of the insufficiencies that inhere in the unified system. Indeed, we await the “revival of the study” urged by Pomeroy, as doctrinal infidels further subjugate equity within a paradigm of strict procedural law. I advocate that district judges invoke the jurisdiction of equity to avoid the application of procedural rules in those unique circumstances where the outcomes produced by rigid application of the rules are deficient.

More prominent use of the term “equity” may restore the dialectic with strict law that should inhere in the procedural infrastructure. That dialectic has faded, as reformers seem resigned, if not committed to the assumption that all innovation must occur from within the infrastructure of the Federal Rules. Notwithstanding the procedural merger of law and equity, federal judges are vested with the equity jurisdiction that was exercised by the English Court of Chancery at the time the Constitution

corporate law matters; in Mississippi, “the division between law and equity has little jurisdictional significance”; and in Tennessee, “the chancery and circuit courts have concurrent jurisdiction in most civil cases by virtue of a statute that dates to 1877”). The State of New Jersey also maintains a separate chancery division of its superior court. See N.J. CONST. art. IV, § 3. But as in the state of Delaware, the structure is driven by an effort to “develop[ ] special expertise and abilities with regard to complex corporate law matters.” AMERICAN BAR ASSOCIATION, AD HOC COMMITTEE ON BUSINESS COURTS, THE STATUS OF BUSINESS COURTS IN THE UNITED STATES (Nov. 4, 1999), available at http://www.abanet.org/buslaw/buscts/ctsurvey.html.

410. See generally POMEROY, supra note 101, §§ 40–42, 41–43. See also Livingston’s Lesse v. Moore, 32 U.S. (7 Pet.) 469, 547 (1833) (“It is true that the separation of common law from equity jurisdiction is peculiar to Great Britain; no other of the states of the Old World having adopted it.”); DILLON, supra note 281, at 386 (“The separation of what we call equity from law was originally accidental, or at any rate was unnecessary; and the development of an independent system of equitable rights and remedies is anomalous and rests upon no principle.”).

411. For sources suggesting a declining interest in the study of equity, see supra notes 12 and 290. For sources suggesting the trend toward certainty in procedural reform, see supra notes 289–348 and accompanying text. Cf. Procedure in the Federal Courts: Hearing Before House Comm. on the Judiciary on H.R. 2377 and H.R. 90, 67th Cong., 2d Sess. 28 (1922) (Statement of Thomas W. Shelton) (“I want to suggest that one of the great criticisms of our present system is that it is utterly impossible for a client, in many instances, when his case is thrown out on a technicality to understand why. That is an important thing. As I said over in the Senate the other day, when arguing this matter, this is one of the things that is making Bolshevists in this country.”), cited in Subrin, supra note 213, at 320 n.78. But see Cooper, supra note 42, at 1944 (“[I]t may be better to leave judges free to adapt to the challenges without interference from statutes and rules framed for the last war by Congress and the rulemaking committees.”).

412. See Frank, supra note 12, at 895 (The word equity “instills in the judge a different mood, one of elasticity and fairness. Nothing is more absurd than the saying: ‘A rose by any other name is just as sweet.’ If the rose were called the ‘bloody-nose flower,’ it might well be odious.”).
was adopted and the Judiciary Act of 1789 was enacted.\textsuperscript{413} This jurisdiction included the authority to offer relief when the cumbersome procedures of the law courts were defective or when the procedural system itself was administratively overwhelmed.\textsuperscript{414} The mandate permitted equity to consider not only the interests of the plaintiffs, but also the defendants and the judicial system.\textsuperscript{415} The merger of law and equity was only procedural,\textsuperscript{416} and the Federal Rules of Civil Procedure neither abridged nor modified the substantive rights in equity.\textsuperscript{417} Yet the substance of equity, which has pervasive even if subtle influence throughout the substantive law, \textsuperscript{418} is largely quiescent in the procedural context.\textsuperscript{419}

Using equitable discretion to avoid the application of a Federal Rule raises two bundles of structural issues that I can only identify here, but then must take a wide arc around and save for another day. First, there are separation of powers issues presented if a federal court does not defer


\textsuperscript{414} See Barbour, supra note 61, at 858 (equity could overcome defects that inhered in the procedure of the law courts and "dealt easily with situations which baffled the common law"); JOHN MITFORD (LORD REDESDALE), PLEADINGS IN CHANCERY 112 (New York, John S. Voorhies, 3d ed. 1833) (jurisdiction of equity exercised where, \textit{inter alia}, "the principles of law, by which the law courts were guided, give a right, but the powers of those courts were not sufficient to afford a complete remedy, or their modes of proceeding were inadequate to the purpose . . . [or] where the courts of ordinary jurisdiction were made instruments of injustice").

\textsuperscript{415} See, e.g., Polaroid Corp. v. Casselman, 213 F. Supp. 379, 381 (D.N.Y. 1962) ("A law suit is not a game but a search for truth. The ends of justice are served, not by giving one side a vested right to exhaust the other, but by affording both an equal opportunity to a full and fair adjudication on the merits."); Gen. Mill Supply Co. v. S.C.A. Serv., Inc., 697 F.2d 704, 711 (6th Cir. 1982) ("The parties might jointly desire a trial by ordeal or trial by battle, but the public interest would not allow this. The public interest demands a seemly and efficient use of judicial resources towards the just, speedy, and inexpensive remedy spoken for in \textit{FED. R. CIV. P. 1}"); Active Prods. Corp. v. A.H. Choitz & Co., 163 F.R.D. 274 (D. Ind. 1995) (relying upon Rule 1 to implement an electronic filing system called "Complex Litigation Automatic Docket" in complex superfund litigation); \textit{In re Paris Air Crash of March 3, 1974}, 69 F.R.D. 310, 322 (C.D. Cal. 1975) (decision on the claims of plaintiffs on issue of products liability alone would avoid prejudice to any party and was the best way to secure the just, speedy, and inexpensive determination of the whole cluster of controversies and lawsuits). See also supra notes 385–400 and accompanying text.

\textsuperscript{416} See supra notes 211–88 and accompanying text.


\textsuperscript{418} For a discussion of the potent legacy of equity on the development of substantive law, see supra notes 291–305 and accompanying text. See also Bordwell, supra note 279, at 750.

\textsuperscript{419} See generally Deborah A. DeMott, \textit{Foreword}, 56 LAW & CONTEMP. PROBS. 1 (1993) ("Equity, however large its triumphs, has long been trailed by challenges to its legitimacy").
to procedural rules that have been adopted pursuant to the Rules Enabling Act.\textsuperscript{420} Second, there are \textit{Erie} issues presented if, in a diversity case, the source of law upon which a federal court relies to invalidate the operation of a Federal Rule is a general federal common law of equity.\textsuperscript{421}

Neither of these structural issues, however, spoils the role for equity that I propose. Indeed, the very essence of the tradition of equity is incorporated within the mandate of Federal Rule 1: that the Federal Rules of Civil Procedure “be construed and administered to secure the just, speedy, and inexpensive determination of every action.”\textsuperscript{422} Thus the equitable discretion to avoid the application of a Federal Rule is contemplated by the procedural infrastructure itself. Reference to the \textit{construction} of the Federal Rules suggests some flexibility in interpreting the applicable language of the rules. The reference to the \textit{administration} of the Federal Rules invites even more flexibility—suggesting a more fundamental or threshold inquiry into the relevance of a Federal Rule.\textsuperscript{423} Yet courts invoke Rule 1 rather infrequently. Occasionally Rule 1 is cited in the context of resolving a lacunae or

\footnotesize
\textsuperscript{420} See Carrington, supra note 277, at 967 (concluding that Congress’ power over the federal courts is limited both by inherent powers and by the constitutional requirement that they be independent in “[p]erform[ing] the core judicial function of applying law to fact in . . . cases and controversies”) (citing Martin H. Redish, \textit{Federal Judicial Independence: Constitutional and Political Perspectives}, 46 MERCER L. REV. 697 (1995)). See generally Robert J. Pushaw, Jr., \textit{The Inherent Powers of Federal Courts and the Structural Constitution}, 86 IOWA L. REV. 735 (2001). See also Link v. Wabash R.R., 370 U.S. 626 (1962) (Rule 41 did not abrogate the federal courts’ inherent authority to dismiss for failure to prosecute); Linda S. Mullenix, \textit{Unconstitutional Rulemaking: The Civil Justice Reform Act and Separation of Powers}, 77 MINN. L. REV. 1283 (1993); cf. Stephen B. Burbank, \textit{Interjurisdictional Preclusion, Full Faith and Credit and Federal Common Law: A General Approach}, 71 CORNELL L. REV. 733 (1986) (“[W]hen the Supreme Court has exercised the power delegated by Congress to prescribe uniform Federal Rules, we should regard those Rules, if valid, as if they were acts of Congress.”). See also Chambers v. NASCO, 501 U.S. 32 (1991) (recognizing some inherent power of the federal courts but declining to define that power or identify extent to which Congress may limit it).


\textsuperscript{422} FED. R. CIV. P. 1. See also supra notes 3–8, 385–400 and accompanying text.

\textsuperscript{423} See \textit{BLACK’S LAW DICTIONARY} (6th ed. 1990) (defining “administer” as to “manage or conduct . . . discharge . . . [or] execute”).
nonexistent norm.\textsuperscript{424} Rule 1 is also cited intermittently as an additional justification for the straightforward application of another procedural rule.\textsuperscript{425} Similarly, courts also will justify particular applications of procedural rules by referring to the “spirit” of the Federal Rules.\textsuperscript{426} But

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{424} See, e.g., D.L v. Unified Sch. Dist., No. Civ.A. 00-2349-CM, 2002 WL 31296445, at *2 (D. Kan. Oct. 1, 2002) (“Though Rule 59(e) is silent as to whether the court may order such relief on its own initiative, the Eleventh Circuit has interpreted the rule’s silence to be without significance, given the court’s inherent powers . . . . Fed. R. Civ. P. 1”); W. Res., Inc. v. Union Pac. R. Co., No. 00-2043-CM, 2002 WL 1822432, at *2 (D. Kan. Jul. 23, 2002) (“given the textual ambiguity of Rule 45 combined with the repeated attempts of the Plaintiff to effectuate personal service, and the cost and delay that would result by requiring further attempts at such service, this Court thus joins those holding that effective service under Rule 45 is not limited to personal service”) (citing Fed. R. Civ. P. 1); United States v. Star Scientific, Inc., 205 F. Supp. 2d 482, 484–85 (D. Md. 2002) (“The language of Rule 45 clearly contemplates that the court enforcing a subpoena will be the court that issued the subpoena. However, this language must be read in light of the underlying purposes of the rule, which include ‘protect[ing] . . . persons who are required to assist the court by giving information and evidence . . . . ’”) (quoting Fed. R. Civ. P. 45, Advisory Committee Notes, 1991 Amendment, \textit{citing} Fed. R. Civ. P. 1).
\item \textsuperscript{426} See, e.g., United States on behalf of Mar. Admin. v. Cont’l Ill. Nat’l Bank & Trust Co., 889 F.2d 1248, 1254 (2d Cir.1989) (discretion of the Court with regard to motions seeking leave to amend “must be exercised in terms of a justifying reason or reasons consonant with the liberalizing ‘spirit of the Federal Rules’”); Nieto v. Kapoor, 210 F.R.D. 244, 246 (D.N.M. 2002) (“outright refusal to grant the leave [to amend] without any justifying reason appearing for the denial is not an exercise of discretion; it is merely abuse of that discretion and inconsistent with the spirit of the Federal Rules”) (quoting Foman v. Davis, 371 U.S. 178, 182 (1962)); Hurburt v. Zaunbrecher, 169 F.R.D. 258, 260 (N.D.N.Y. 1996) (putting defendant on notice “that, although she is technically entitled to insist upon service in strict compliance with Rule 4(e), in the opinion of the court such insistence violates the spirit of the Federal Rules as expressed in Rules 1 and 4(d)”) (); Hartley & Parker, Inc. v. Fl. Beverage Corp., 348 F.2d 161, 163 (5th Cir. 1965) (court refused to reverse judgment on ground that failure to verify answers to admissions under Rule 36 resulted in technical admission of truth of statements, particularly when proof clearly refuted truth of any such admissions; “the spirit of the federal practice [is] to accord substantial justice over mere technical contentions”; Ford Motor Credit Co. v. Beard, 45 F.R.D. 523, 525 (D.S.C. 1968) (“spirit of the federal rules” contemplates avoidance of circuitry or multiplicity of litigation); Gonzales v. Sec. of the Air Force, 824 F.2d 392, 396 (5th Cir.), \textit{cert. denied}, 485 U.S. 969 (1987). (Brown, J., dissenting) (stating that it is “contrary to the spirit of Federal Rule 1” to require the plaintiff to file
\end{itemize}
\end{footnotesize}
very rarely will courts avoid the straightforward or even rigid application of a procedural rule, notwithstanding Federal Rule 1 and the proverbial “spirit” of the Federal Rules.\textsuperscript{427}

Unfortunately, the historical understanding of equity is freighted with the baggage of “natural law.”\textsuperscript{428} In the tradition of separate systems of law and equity, the strict law was subordinate to principles of reason and conscience as determined by the law of God.\textsuperscript{429} With common philosophical and religious systems of morals, universal standards of justice could be divined in earlier times.\textsuperscript{430} In contemporary discourse, however, consensus as to the content of those standards is unlikely, and thus any invocation of morality can be highly controversial and analytically suspect.\textsuperscript{431} Implementing the principles of equity need not, however, introduce a system wedded to natural law. Since Blackstone’s scientific explication of their respective roles two centuries ago, and serve the defendant within the 30 day statutory time limit when the statute makes no mention of service).

\textsuperscript{427} See supra note 383 and accompanying text. See also In re Simon II Litigation, 211 F.R.D. 86 (E.D.N.Y. 2002) (Weinstein, J.) (using FRCP 1 in conjunction with equitable maxim that ensures where there is a wrong there is a remedy); Tyson v. City of Sunnyvale, 159 F.R.D. 528 (D. Cal. 1995) (just, speedy, and inexpensive determination requires that court exercise its discretion to extend the time period for serving process when process was served one day late); TPI Corp. v. Merch. Mart of S.C., Inc., 61 F.R.D. 684, 692 (D.S.C. 1974) (notwithstanding considerable contrary authority, court permitted permissive intervention because justice required that the party requesting intervention be granted it); Rollerblade, Inc. v. Rappelfeld, 165 F.R.D. 92, 95 (D. Minn. 1995) (extending time for service of process under Rule 4(m)).

\textsuperscript{428} See Walter Wheeler Cook, Equity, in ENCYCLOPEDIA OF THE SOCIAL SCIENCES 582 (1931) (“The most common of the non-technical meanings of equity, one in which lawyers themselves not infrequently use the word, is as a synonym for ‘natural justice.’”); Pound, Justice, supra note 49, at 702 (characterizing equity as “justice without law”).


\textsuperscript{430} See Roscoe Pound, Juristic Science and the Law, 31 HARV. L. REV. 1047, 1060 (1918).

procedure has been conceptually divorced from substance.\textsuperscript{432} Within this
time-honored framework, procedure was separated from and
subordinated to substance.\textsuperscript{433} The procedural infrastructure is but an
ancillary set of rules of “etiquette” designed to ensure the application of
substantive law.\textsuperscript{434} The application \textit{vel non} of a set of purely functional
procedural rules should not implicate the philosophical and religious
values that make natural law and equity problematic. Indeed, if those
values were implicated in the decision whether to apply the procedural
rule, this would suggest that there is substantive content beyond the
scope of procedure and within the scope of substance where,
incidentally, the principles of equity enjoy greater influence.\textsuperscript{435}

Equity’s charge to deliver “individualized justice” also incites fears of
unabashed and unprincipled judicial activism.\textsuperscript{436} In this regard, Professor
Chafee once remarked of the authority of chancellors: “O, it is excellent
[t]o have a giant’s strength; but it is tyrannous [t]o use it like a giant.”\textsuperscript{437}
But equity is no tyrant. Equity was, and remains, a supplementary law; it

\textsuperscript{432} Toran, \textit{supra} note 195, at 378–79, n.165, n.169 (describing procedure as “an unclogged
artery through which substantive rights could flow”).

\textsuperscript{433} \textit{See supra} notes 183–93 and accompanying text.

\textsuperscript{434} \textit{See} Roscoe Pound, \textit{The Etiquette of Justice}, 3 \textit{PROC. NEB. ST. B.A.} 231 (1908). \textit{See also}
Charles Clark, \textit{Fundamental Changes Effected by the New Federal Rules} \textit{(Pt. 1)}, 15 \textit{TENN. L. REV.}
551, 551 (1939) [hereinafter, Clark, \textit{Fundamental Changes}]; \textit{see also} Clark, \textit{The Challenge, supra}
ote 275.

\textsuperscript{435} \textit{See supra} notes 291–305 and accompanying text.

\textsuperscript{436} Charges of judicial activism used to be the province of conservatives. Liberals have joined
that chorus as congressional legislation becomes increasingly vulnerable to a more conservative
Supreme Court. For scholarship addressing this topic of judges exceeding their proper authority, see
generally Stephen F. Smith, \textit{Taking Lessons From the Left?: Judicial Activism on the Right}, 2002
\textit{GEO. J.L. & PUB. POL’Y}, 57–80 (2002); David P. Bryden, \textit{A Conservative Case for Judicial Activism},
111 \textit{PUB. INTEREST} 73 (1993); Gary Minda, \textit{Jurisprudence at Century’s End}, 43 J. LEGAL EDUC.
27, 32–36 (1993); \textit{Christopher Wolfe, Judicial Activism: Bulwark of Freedom or Precarious
Security} (1991); Earl Maltz, \textit{The Prospects of a Revival of Conservative Activism in Constitutional
Jurisprudence}, 24 \textit{GA. L. REV.} 629 (1990); Lino A. Graglia, \textit{Judicial Activism: Even on the Right
It’s Wrong}, 95 \textit{PUB. INTEREST} 57 (1989); \textit{Steven C. Halpern & Charles M. Lamb, Supreme
Court Activism and Restraint} (Lexington Books 1982); \textit{David Forte, The Supreme Court: Judicial
Activism Versus Judicial Restraint} 17 (D.C. Heath, 1972); Alpheus T. Mason, \textit{Judicial Activism: Old and New},
55 VA. L. REV. 385 (1969); and \textit{Alexander M. Bickel, The Least Dangerous Branch: The Supreme Court at the Bar of Politics} (1962). \textit{See also} Jack
that the approach of the nineteenth century French judges “who abandoned rules of law completely
and instead engaged in ad hoc decision-making according to the equity of the cases” was
“individualism run riot”).

\textsuperscript{437} CHAEE, \textit{supra} note 391, at 303.
was designed “not to destroy the law but to fulfill it.” If a Federal Rule requires a particular result, it should, in most instances, be applied as written. But procedural rules are “but an aid to an end and not an end in themselves.” Accordingly it must be more important that fair and efficient applications of substantive law be facilitated than that predictability and uniformity be assured through rigid adherence to existing procedural rules.

Importantly, the scope of the exercise of equity is limited to the individual case. Faced with procedural insufficiencies, courts will often disclaim responsibility and authority for “amending” the procedural rules. In Amchem, for example, Justice Ginsburg wrote that “The text of a rule thus proposed and reviewed limits judicial inventiveness. Courts are not free to amend a rule outside the process Congress ordered.” But equity does not require judicial amendment; indeed equity does not accommodate it. The purpose of equity “was to provide a tribunal where the hardship of particular cases might be relieved; the purpose was not to provide general rules of law.” The introduction of equity into the procedural context thus is not an invitation for judges to effect wholesale revisions to the applicable Rules. For example, judges could not, under

---

438. See MAITLAND, supra note 7, at 17. See also CLARK, supra note 191, at 1 (equity “supplements” existing rules of law by reference to current standards of morality); Manuel Rodriguez Ramos, Equity in The Civil Law: A COMPARATIVE ESSAY, 44 Tul. L. Rev. 720, 724 (1970) (“Equity . . . is nothing more than reason or natural justice, that is, a supplement to the written law.”).

439. Clark, Fundamental Changes, supra note 434, at 551.


441. This scope thus may be more limiting than the type of “equitable” review suggested by Professor Bauer and by Judge Moore. Their arguments for broadening the interpretation of Federal Rules stem primarily from the Congressional delegation to the courts of the authority to make the Federal Rules in the first instance. See Joseph P. Bauer, Schiavone: An Un-Fortune-ate Illustration of the Supreme Court’s Role as Interpreter of the Federal Rules of Civil Procedure, 63 Notre Dame L. Rev. 720, 720 (1988) (arguing that because the Supreme Court promulgates the Rules, federal courts are “fully justified in taking an expansive view of the Federal Rule under scrutiny, giving it a liberal reading if that is required to fulfill the purposes of the Rule or to do justice between the parties before the court”); Karen Nelson Moore, The Supreme Court’s Role in Interpreting the Federal Rules of Civil Procedure, 44 Hastings L.J. 1039, 1093 (1993) (“Given substantial, although largely unexercised, powers of the Court in the promulgation process, a more activist role in the interpretative stage, one that considers purpose and policy, is appropriate.”).

442. See supra notes 376–84 and accompanying text.


444. Campbell, supra note 4, at 111. See Smith, supra note 20, at 310 (“Law, like surgery, loses a certain number of patients. It has been the function of equity to cut down this loss, not by an entire abrogation of the general rule . . . but by a modification of the rule in the particular case.”).
the authority of equity and in contravention of Federal Rule 8(a), impose heightened pleading standards in all civil rights cases simply because to do so would further the efficient administration of justice. Nor could the Federal Rule 23 prerequisites to certification of a class action be summarily ignored. Rather, equity could offer relief from hardship or mischief created by a particular application of a procedural rule.

The historical continuity of equity should be channeled through moderate and restrained interpretations of the Federal Rules of Civil Procedure. No principle dictates that a court enforce to the outermost limit the mandate of a particular Federal Rule. An unforgiving interpretation of a Federal Rule is inconsistent with the broad judicial discretion that is the principal reported accomplishment of the Federal Rules. Such an interpretation is also inconsistent with Federal Rule 1. Procedural insufficiencies typically will present a conflict between the principles of equity, on one hand, and a particular Federal Rule on the other. As with any conflict analysis, these interests should be balanced. Thus, when the application of a Federal Rule is unfair,


446. See FED. R. CIV. P. 23(a) (identifying prerequisites of numerosity, commonality, typicality and adequacy).

447. See supra notes 75–78, 86–90, 146–48 and accompanying text.


449. See supra notes 422–27 and accompanying text.

450. See, e.g., Tyson v. City of Sunnyvale, 159 F.R.D. 528, 530 (D. Cal. 1995) (just, speedy, and inexpensive determination requires that court exercise its discretion to extend the time period for serving process when process was served one day late); TPI Corp. v. Merch. Mart of S.C., Inc., 61 F.R.D. 684, 692 (D.S.C. 1974) (notwithstanding considerable contrary authority, court permitted permissive intervention because justice required that the party requesting intervention be granted it); Rollerblade, Inc. v. Rappelfeld, 165 F.R.D. 92, 95 (D. Minn. 1995) (extending time for service of process under Rule 4(m)); Rashidi v. Albright, 818 F. Supp. 1354, 1357 (D. Nev. 1993) (postponing responsive pleadings until motion for summary judgment was considered in order to ensure a just, speedy, and inexpensive determination of action); Lawhorn v. Atlant Ref. Co., 299 F.2d 353, 357 (5th Cir. 1962) (“If one hauled into court as a defendant has a claim but the adversary plaintiff has not, the nominal defendant ought to be allowed to name the time and place to assert it . . . . The Rules should be construed in such a manner as to do substantial justice. Under the pain of foregoing permanently a ‘valid counterclaim’ by such a putative defendant, the rules ought not to be construed in any such barratrous fashion.”).

inefficient, too complicated, or otherwise deficient, the applicability of that rule should be re-examined with a view to a more moderate and restrained interpretation.  

Within this framework, correction is a product of either the policy and purpose of the traditional principles of equity, or their proxy, Federal Rule 1. The ability of equity to correct problems stemming from application of strict law modernizes and reforms the legal doctrine while also boosting its societal legitimacy. Equity is a fundamental method by which the law has sought to meet changing conditions. Legislative amendment cannot, of course, correct the mischief created by changed conditions; at best, amendment avoids the hardship in subsequent cases, and only then if the mischief were to repeat. Moreover, the legislative amendment may itself create problems upon changed circumstances. Equity thus plays an important role in the growth of the law, and without that engine, "our law will be moribund, or worse."

Equity was a court of vast jurisdiction, and occasionally throughout history was eminently progressive in increasing its authority. Hardship and mischief created by the procedural infrastructure need not be tolerated. On behalf of the state, equity has stood in this breach since time far beyond memory. Why not now? Our courts are not inferior to those of our ancestors. The jurisprudence of the twenty-first century is not beneath that of the 14th century. The administration of equity

---

452. See, e.g., Smith v. Morrison-Knudsen Co., 22 F.R.D. 108, 113 (D.N.Y. 1958) ("Rules of procedure, like principles of substantive law, should be interpreted to meet the challenge of changing conditions of life and litigation . . . . Problems created by [the mobility of Americans to travel abroad] can be solved by our law’s inherent flexibility."); United States v. U.S. Cartridge Co., 6 F.R.D. 352, 354 (D. Mo. 1946) (court was "loathe" to give strict construction to Rule 33 "absent the showing that such is necessary to promote the ends of justice").

453. The traditional principles of equity may be more limited after Grupo Mexicano, but that case should usually be distinguishable since the relief sought there was not the type of relief "traditionally accorded by courts of equity" and because it had been "specifically disclaimed by longstanding judicial precedent." 527 U.S. at 322.

454. See Smith, supra note 53, at 209 (crediting Sir Henry Sumner Maine for the famous dictum that there are three methods by which the law has sought to meet changing conditions: (i) fictions; (ii) legislative amendment; and (iii) equity). See also Johnson, supra note 19, at 352–53.

455. See generally Manning, supra note 308, at 767.

456. See Cooper, supra note 42, at 1944 ("[I]t may be better to leave judges free to adapt to the challenges without interference from statutes and rules framed for the last war by Congress and the rulemaking committees."). See also supra note 312.

457. Bordwell, supra note 279, at 749.

458. See supra notes 99–104 and accompanying text. See also Simkins, supra note 177, at 6.

459. See Campbell, supra note 4, at 112.
requires skill, nerve, good judgment, detachment, compassion, ingenuity, and the capacity to sustain confidence.\textsuperscript{460} But such is the task of judging.\textsuperscript{461} Judges routinely administer the substantive principles of equity in the application of substantive law.\textsuperscript{462} Procedure, too, outside the context of the Federal Rules has been dramatically influenced by the substantive principles of equity.\textsuperscript{463} Formalistic interpretations of Federal Rules of Civil Procedure are inconsistent with and unnecessary within a unified system of law and equity.

B. Accommodating Equity in the Construction of Procedural Rules

Application of the substantive principles of equity is best accommodated by succinct and generalized procedural rules that accord judges broad discretion. A proposal that our unified system of law and equity be administered by modern, elastic and sleek procedural rules should be relatively uncontroversial. Of the various fronts where the rules-standards war have been fought,\textsuperscript{464} the rhetoric of discretionary standards has clearly prevailed on the civil procedure battlefield.\textsuperscript{465} Indeed, the complication, trivialization and ossification of the Federal Rules seems to have commenced in spite of our allegiance to a set of rules that despises formality.\textsuperscript{466} My effort here is to urge a commitment and return to more flexible and discretionary rules of procedure that reflect the rhetoric and the common perception that the Federal Rules are “all equity.”\textsuperscript{467}

\textsuperscript{460} See GEOFFREY HAZARD, JR., ETHICS IN THE PRACTICE OF LAW 65 (1978).

\textsuperscript{461} See David Luban, Heroic Judging in an Antiheroic Age, 97 COLUM. L. REV. 2064, 2090 (1997) (broad discretion in the hands of judges may be “a risky and uncomfortable state of affairs—but it is always risky and uncomfortable when our well-being lies in the hands of heroes”).

\textsuperscript{462} See supra notes 291–305 and accompanying text.

\textsuperscript{463} See generally Weinstein & Hershenov, supra note 12; Subrin, supra note 23, at 970–82; Marcus, supra note 177, at 725; Resnik, supra note 262, at 376; Waters, supra note 262, at 542–51; Laycock, supra note 12, at 53–54. See also Chayes, supra note 262, at 1292–96 (discerning in public law litigation the “triumph of equity”). See also supra note 323.


\textsuperscript{466} See supra notes 307–48 and accompanying text.

\textsuperscript{467} See supra notes 307–48 and accompanying text.
To be sure, strict definite rules have their virtues. Some interests may require the protection of more comprehensive, predictable and uniform mandates. In criminal or commercial law contexts, for example, one can argue that it is better to have a bad rule that everyone knows than a new and better rule of which no one is certain. Similarly, a formalist procedure may advance certain values. Indeed, some argue that judicial discretion threatens the internal morality and general legitimacy of law by undermining the notice and publicity requirement of rules, which are fundamental elements of a legal system in the formalist tradition. Moreover, even in the context of procedural rules, strictness and definiteness have an illustrious pedigree. The common law courts embraced certainty, predictability and uniformity as primary norms and celebrated the technicality that engineered that system.

But the merger of law and equity changed our ability to (micro)manage procedure. An autonomous system of equity enjoyed the

---

468. See supra notes 91–98 and accompanying text. See generally Schlag, supra note 464, at 400–01 (cataloging the superficial virtues and vices of rules and standards).

469. See Johnson, supra note 19, at 355. Cf. Burnet v. Coronado Oil & Gas Co., 285 U.S. 393, 406 (1932) (Brandeis, J., dissenting) (“in most matters it is more important that the applicable rule of law be settled than that it be settled right”).


472. See LON L. FULLER, THE MORALITY OF LAW 35 (2d ed. 1969) (stating that it is “very unpleasant to have one’s case decided by rules when there was no way of knowing what those rules were”).

473. See supra notes 156–72 and accompanying text.
flexibility and discretion to prevent hardship caused by procedural insufficiency and to fulfill the curative purpose of equity.\textsuperscript{474} A unified system of law and equity seems less capable of accommodating individualized justice.\textsuperscript{475} And more and longer procedural rules will never anticipate all of the eccentricities that fate or human ingenuity are "virile enough to devise."\textsuperscript{476} Roscoe Pound recognized this and favored a flexible procedural system that granted judges liberal discretion “despite his emphasis on certainty.”\textsuperscript{477} Similarly, Charles Clark insisted on flexibility and discretion “even when pragmatism dictated the need for more detail and complexity.”\textsuperscript{478}

The potent legacy of equity in various non-procedural contexts demonstrates that judges can be more than umpires.\textsuperscript{479} Procedural rules could likewise be drafted with more discretion to allow and encourage judges to apply their judicial skill.\textsuperscript{480} The solution, then, is to avoid elaborate amendments tailored specifically to address particular procedural insufficiencies.\textsuperscript{481} Such amendments perpetuate a cycle that leads to the creation of further procedural insufficiencies that, in turn, require still more elaboration.\textsuperscript{482} That cycle could be broken with broader and more discretionary forward-looking rules that facilitate vigor and common sense and efficiency and fairness in their application.

One need not be especially creative to develop talking points for equity-based reforms to the Federal Rules. Prior to the merger of law and equity in 1938, the federal courts operated courts of equity for more than a century under the Federal Equity Rules. A comparison of the current procedural rules to their professed source, would be an obvious

\begin{thebibliography}{99}
\bibitem{474} See supra notes 86–90 and accompanying text.
\bibitem{475} See supra notes 307–84 and accompanying text.
\bibitem{476} See Campbell, supra note 4, at 113.
\bibitem{477} Toran, supra note 195, at 374 (citing Bone, supra note 161, at 98–100 and Subrin, supra note 23, at 946–48).
\bibitem{479} For a discussion of the potent, if unappreciated legacy of equity in contemporary jurisprudence, see supra notes 291–305 and accompanying text.
\bibitem{480} See Frierson, supra note 32, at 404.
\bibitem{481} For a discussion of the increasing number and length of amendments to the Federal Rules, see supra notes 307–84 and accompanying text.
\bibitem{482} See supra notes 307–10 and accompanying text.
\end{thebibliography}
Equity and Procedure

baseline. In the context of class actions, for example, consider the old Federal Equity Rule which provided in its entirety that:

When the question is one of common or general interest to many persons constituting a class so numerous as to make it impracticable to bring them all before the court, one or more may sue or defend for the whole.

Would we not trust the reasoned, intelligent discretion of judges trained for their task to apply this rule and to satisfy as many competing demands as that particular application of the rule presents? The certification of class actions can, of course, destroy individual rights, and I do not intend to devalue the many difficulties presented by mass solutions. However, in situations like Amchem or Ortiz, for example, a district judge applying this broad discretionary rule would be permitted to balance all of the difficult issues and competing interests of the various parties to determine whether litigation as a settlement class would be appropriate and fair and efficient. Indeed, the Supreme Court suggested in both of these cases that the proposed settlement was both substantively and procedurally fair and that the settlement proposed an efficient resolution of a phenomenon that defied customary judicial administration. Yet the classes in those cases could not be certified, the


484. Federal Equity Rule 38 (1912).

485. See generally Johnson, supra note 19, at 354.


487. See supra notes 368–75 and accompanying text.
Court explained, because the elaborate Federal Rule 23 imposed substantive, err “structural” protections of its own. 488

The ultimate result in Amchem or Ortiz might have been the same even under a broad discretionary rule. Even so, it should not have been a procedural rule that dictated the result. 489 A broad and discretionary rule would not impose structural requirements and would instead bring into relief the competing interests at stake. Procedural rules should not, indeed cannot, enlarge, abridge or modify substantive rights. 490 The primary and ultimate issue in those cases was protection of the due process and other substantive rights implicated by the settlement class. 491 Interpretation and application of strict definite procedural rules instead dominated the Court’s attention. 492

In light of the increasing strictness of the procedural infrastructure, it is perhaps not surprising that many proposed reforms for the resolution of mass tort claims are looking to bankruptcy courts as a forum and model. 493 Bankruptcy courts, of course, are courts of equity 494 but there,
too, the traditional flexibility and discretion of equity seem to be mired within a very detailed code and an extensive set of procedural rules. Indeed, bankruptcy courts seem to be equity courts largely in name only. But the successful—and equitable—resolution of mass tort claims should not require escape to an alternative system, as the substantive principles of equity should inhere in the unified system of law and equity administered pursuant to the Federal Rules of Civil Procedure.

Procedural rulemakers can have their cake and eat (at least some of) it too. Future amendments to the Federal Rules should feature flexibility and discretion. Elaborate legislative drafting to contemplate all conceivable applications can create hardship and mischief. Thus we should be skeptical of strictness, and tolerate rigidity only in circumstances where the norms of certainty, predictability and uniformity are particularly compelling. In circumstances where those norms are present, but are not particularly compelling, suggested limitations or applications could be included in the rule but relegated to a secondary, illustrative or non-binding status.

Specifically, I recommend three techniques that procedural rulemakers could use to suggest, without imposing constraints on the discretion and flexibility of judges. First, rulemakers can signal suggested applications of procedural rules through more elaborate promulgation of Committee Notes. The Rules Enabling Act requires that amendments include “an explanatory note . . . and a written report explaining the body’s

---

496. See F. R. BANKR. P. 1–9032.
497. See Krieger, supra note 494, at 310 (“From historical, procedural, jurisprudential and practical perspectives the bankruptcy court is not a court of equity. It is, instead, a specialized court of limited jurisdiction applying statutory law that embodies a particular, often changing, social objective.”); Steve H. Nickles, Another Way of Thinking About Section 105(A) and Other Sources of Supplemental Law Under the Bankruptcy Code, 3 CHAP. L. REV. 7, 16 (2000) (“Equitable principles and rules, however, are not a source of general authority to act beyond or different from the Bankruptcy Code. So, even if there is a real and lawful basis for bankruptcy judges to assume the role of equity chancellors, this role gives them little reason or room to add substantive, supplemental law to the Bankruptcy Code.”); Jason A. Rosenthal, Courts of Inequity: The Bankruptcy Laws’ Failure to Adequately Protect the Dalkon Shield Victims, 45 FLA. L. REV. 223, 226–32 (1993).
action." The Advisory Committee frequently articulates the purpose of a rule or an amendment in these Committee Notes, and offers guidance on future interpretations. "Over the years, the Notes have increased in significance, and they now play an integral role in the rulemaking process." Thus in many circumstances the procedural norms of predictability, certainty and uniformity might be satisfied by articulating the desired applications in Committee Notes. This technique would allow rulemakers to signal particular desired outcome, but would not rigidly constrain district judges.

Second, procedural rulemakers could recognize a hierarchy of rules within the Federal Rules of Civil Procedure, subordinating procedural rules of particular application to primary meta-rules containing broad discretionary mandates. For example, a broad discretionary class action rule (perhaps similar to the old Federal Equity Rule or to the current Federal Rule 23(a)) could be adopted as a primary and dominant rule. Suggested applications of or limitations upon the general

499. See Catherine T. Struve, The Paradox of Delegation: Interpreting the Federal Rules of Civil Procedure, 150 U. Pa. L. Rev. 1099, 1112 (2002) (citing FED. R. CIV. P. 4, Advisory Committee Note (1993)) ("The general purpose of this revision is to facilitate the service of the summons and complaint."); FED. R. CIV. P. 26 Advisory Committee Note (1993) ("[a] major purpose of the revision [to Rule 26(a)] is to accelerate the exchange of basic information about the case"); id. Advisory Committee Note (2000) (stating that the amendments "restore national uniformity to disclosure practice [and] to other aspects of discovery"); FED. R. CIV. P. 30 Advisory Committee Note (1993) (explaining that the aims of new Rule 30(a)(2)(A) are to assure judicial review before any side takes more than ten depositions without consent of other parties and "to emphasize that counsel have a professional obligation to develop a Mut. cost-effective plan for discovery"); FED. R. CIV. P. 33 Advisory Committee Note (1993) ("The purpose of this revision is to reduce the frequency and increase the efficiency of interrogatory practice."); FED. R. CIV. P. 45 Advisory Committee Note (1991) (listing five purposes for the amendment); FED. R. CIV. P. 53 Advisory Committee Note (1991) ("The purpose of the revision is to expedite proceedings before a master."); FED. R. CIV. P. 77 advisory committee's note (1991) ("The purpose of the revisions is to permit district courts to ease strict sanctions now imposed on appellants whose notices of appeal are filed late because of their failure to receive notice of entry of a judgment"); FED. R. CIV. P. 50 Advisory Committee Note (1991) ("Paragraph (a)(1) articulates the standard for the granting of a motion for judgment as a matter of law. It effects no change in the existing standard."); FED. R. CIV. P. 50 Advisory Committee Note (1993) (reaffirming that "the 1991 revision . . . was not intended to change the existing standards under which directed verdicts could be granted").

500. Struve, supra note 499, at 1112. Indeed, the article suggests that, because of the uniqueness of the Congressional delegation to the Supreme Court of the rulemaking authority, the Courts should accord the Notes "authoritative effect." Id. at 1103. If Professor Struve’s proposal proves to be as persuasive as it is fascinating, my recommendation that the Notes be used as secondary, illustrative and non-binding signaling would obviously be ineffectual.

501. Of course, I suggest in Part V.A., supra, that Federal Rule 1 outlines such a dominant purpose.
principle(s) could be expressed separately. Such a framework would avoid the hardship created by strict applications, either because the particularized rule could be narrowly construed in light of the broader mandate or because the particularization could be expressly subordinate to the primary mandate. Although none of the current Federal Rules of Civil Procedure reflect this organization, one might consider the Federal Rules of Evidence as a possible model. Federal Rule of Evidence 403, for example, provides broad discretionary prerogative for a judge to exclude relevant evidence “if its probative value is substantially outweighed by the danger of unfair prejudice . . . .” Rules 404 through 415 then target specific circumstances that may curb that discretion in particular circumstances. A similar structure of meta- and subordinate procedural rules could allow procedural rulemakers to limit discretion while allowing the principles of hierarchy to avoid hardship and mischief that could be created by strict and technical applications of subordinate rules.

Finally, procedural rules could simply be drafted with more hortatory and less mandatory language. A renewed commitment to discretionary and flexible rules could provide a durable corpus of procedural rules that would not require constant tinkering by amendment. Although a new structural framework for procedure may be unnecessary, rulemakers might consider something entirely new. The Restatements of Laws, for example, tend to follow a model where the governing rules are stated broadly with certain finer points relegated to a secondary and often non-

---

502. Suggested applications could refer to the current typology of actions contained in Federal Rule 23(b); suggested limitations could be those currently detailed in Federal Rule 23(c).

503. I thank and credit Professor Peter Nicolas for providing me with this interesting analogue.

504. FED. R. EVID. 403 (“Although relevant, evidence may be excluded if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury, or by considerations of undue delay, waste of time, or needless presentation of cumulative evidence.”).

505. See FED. R. EVID. 404 (“Character Evidence Not Admissible to Prove Conduct; Exceptions; Other Crimes”); id. 405 (“Methods of Proving Character”); id. 406 (“Habit; Routine Practice”); id. 407 (“Subsequent Remedial Measures”); id. 408 (“Compromise and Offers to Compromise”); id. 409 (“Payment of Medical and Similar Expenses”); id. 410 (“Inadmissibility of Pleas, Plea Discussions; and Related Statements”); id. 411 (“Liability Insurance”); id. 412 (“Sex Offense Cases; Relevance of Alleged Victim’s Past Sexual Behavior or Alleged Sexual Predisposition”); id. 413 (“Evidence of Similar Crimes in Sexual Assault Cases”); id. 414 (“Evidence of Similar Crimes in Child Molestation Cases”); id. 415 (“Evidence of Similar Acts in Civil Cases Concerning Sexual Assault or Child Molestation”). See Frank v. County of Hudson, 924 F. Supp. 620, 624 (D.N.J. 1996) (“FRE 413–415 ‘are permissive rules of admissibility, not mandatory rules of admission’”)(quoting Memorandum from Department of Justice to United States Attorneys 3 (July 12, 1995)).
binding status as “comments” or “illustrations.” Similarly, the *Model Code of Professional Responsibility* provides layers of Canons, Ethical Considerations and specific Disciplinary Rules. In both of these examples, the multiple layers of authority accommodate the overactive lawmaking gland with minimal interference to the controlling fundamental principles. These models could provide a way for procedural rulemakers to signal certain outcomes, without risk of creating procedural hardship and mischief.

All three proposals for modifying the construction of procedural rules underscore the significance of discretion and flexibility. In a unified system of law and equity, the procedural infrastructure must allow the substance of equity to flourish. Until judges are inclined to invoke equity as a justification for avoiding the application of a procedural rule that would create hardship, it is especially important that the moderating force of equity be incorporated into the procedural infrastructure through discretion and flexible rules.

VI. CONCLUSION

The wisdom of dual systems of law and equity was the ability of the latter to correct the substantive and procedural deficiencies of the former. The genius of a unified system of law and equity was the notion that the substance of law and equity could be merged procedurally without affecting the substance of either law or equity. But a problem is presented when the procedural infrastructure of the unified system begins to interfere with the traditional jurisdiction of equity. This phenomenon is underway. And unless the principles of equity are allowed to trump the rules of the procedural infrastructure charged with the administration of its substantive principles, traditional equity is impaired.

---

506. See, e.g., *Restatement (Second) of Property* § 4.1 (1981) (including restatement of Validity of Disabling Restraint, with comments and illustrations); *Restatement (Second) of Judgments* § 73 (including restatement of Changed Conditions supporting relief from judgment, with comments and illustrations).

507. See generally Nancy B. Rapoport, *Turning & Turning in the Widening Gyre: The Problem of Potential Conflicts of Interest in Bankruptcy*, 26 Conn. L. Rev. 913, 941 (1994) (explaining that Canons define what conduct is expected of attorneys; Ethical Considerations are guidelines for attorneys to act ethically; and Disciplinary Rules are minimum standards required of attorneys).

508. See supra note 455 and accompanying text.

509. See supra Part V.A.
WHAT IS THE RULE OF LAW? PERSPECTIVES FROM CENTRAL EUROPE AND THE AMERICAN ACADEMY*

Louis E. Wolcher†

It is an honor to be invited to address you at this celebration of the silver anniversary of the Federal Bar Association for the Western District of Washington. Please accept my congratulations on the health, longevity, and high importance that your organization has achieved during the past quarter century. I want to thank in particular your President, Kevin Swan, for inviting me. He had the kindness (and I hope not the bad judgment) to remember his old Federal Courts professor at the University of Washington Law School, and to imagine that I might have something of interest or value to say to you tonight.

The title of my talk is “What is the Rule of Law?”—and its subtitle is “Perspectives from Central Europe and the American Academy.” I represent the “American Academy” part, and as I will make clear in a little while, the other part comes from my sustained engagement, over the past ten years, with legal theorists and political philosophers in the Republic of Slovenia.1 Slovenia, by the way, is a nation that was created twelve years ago as the northernmost of those “breakaway” republics of the former Yugoslavia.

Before I begin developing my theme, permit me to say a relevant word or two about my own personal background. For nearly a decade I was a partner in the litigation department of a large San Francisco law firm, before rising like a Phoenix, seventeen years ago, from the ashes of law practice and into the rarefied air of the Ivory Tower. Now you will recall that the Phoenix is a kind of bird, and some of you may even be tempted to say that what we professors do in academic law is often a bit bird-brained. To those of you who think this I will even admit that I specialize in what you might be tempted to call the most bird-brained and useless academic field of all—the philosophy of law. Nevertheless, if the Phoenix has a bird’s brain it also

---

* Keynote Address at The Federal Bar Association of the Western District of Washington Nineteenth Annual Dinner & Twenty-Fifth Anniversary Celebration (December 4, 2002).
† Professor of Law, University of Washington School of Law. Telephone: (206) 543-0600. E-mail address: wolcher@u.washington.edu.
1. For more about this collaboration, see Marijan Pavčnik & Louis Wolcher, A Dialogue on Legal Theory Between a European Legal Philosopher and His American Friend, 35 TEX. INT’L L.J. 335 (2000).
has the ability to fly up for a bird’s eye view, and I hope that this means we can profitably think together about the Rule of Law from a “bird’s eye view,” at least for the next twenty minutes or so.

At the time of our nation’s founding, John Adams declared, in a very famous epigram from his Novanglus Papers,² that the most important characteristic of a democratic republic consists in its being “a government of laws and not men.”³ Since men are not the only kind of people who can acquire power these days (thankfully), I will take the liberty of updating Adams’ saying by deleting the word “men” in favor of the gender-neutral term “people.” Having done so, I surmise that a lot of people in this room would probably say that they know exactly what Adams’ epigram means. But I have to confess to you that I am not one of them. Taking Adams’ saying as our initial indication of the Rule of Law, I for one am prone to wonder, What on earth does this “Rule of Law” idea mean? And does it mean the same thing everywhere? I will take these as my two guiding questions tonight.

The habit of everyday life is usually not to question things too often or too rigorously. But the utter questionability of the so-called “Rule of Law” was brought home to me vividly in the early nineties, shortly after the fall of the Berlin Wall, and after the breakup of the Soviet Union and the former Yugoslavia. Due to a series of international exchanges and a Fulbright award, I was able to acquire close professional and personal ties with scholars at the University of Ljubljana School of Law, in Slovenia. This small country achieved its independence in 1990 (for the first time in one thousand years, by the way), and its intellectuals were (and still are) struggling to think through and implement the meaning of “Democracy” and the “Rule of Law” in a country that had officially known nothing but the social and economic system of Communism and its official ideology for half a century. Although theirs is a genuine struggle for democracy and the rule of law—I am sure of that—it is still worth asking what kind of democracy and what kind of rule of law they are struggling to achieve. One fact that made me really think about the ambiguity of the meaning of the Rule of Law came from a survey of Slovene public opinion that was conducted in 1996: Only twenty-four percent of those surveyed said that they trusted the law courts and lawyers in Slovenia—indeed, the study showed that the Slovene president, their police, their army, their educational

³. Id.
The Rule of Law

institutions, and even their banks were all trusted at a significantly higher level than were the courts. What was going on? And what does this apparent contempt for law entail for the meaning of the Rule of Law in their country and in ours?

Before I attempt to answer this question, let me say something about John Adams’ epigram that many of you may find rather shocking. There is a very important sense in which the idea that democracy entails a “government of laws and not of men” is complete and utter fantasy and nonsense. I am referring explicitly to the idea that legal texts—rules, decisions, principles—somehow automatically determine the actions of those who apply them. I say that this idea is complete and utter fantasy and nonsense because it incorrectly imagines that “laws” and “humans” reside in two utterly different logical spaces—that THE LAW represents a realm in which musty documents from the past push or can push people in the present like so many levers on a complicated machine, and that humans, who are fated always to live their lives in the present, are or can be miserable slaves of the documents that they or their ancestors deposited in the past.

In the curious dreamworld depicted in Adams’ epigram, a great reversal is enacted: Our own previous creations—legal texts and practices—come to rule over us and dominate us like the multitude of self-replicating brooms that are brought to life by the sorcerer’s apprentice (played by Mickey Mouse) in the famous Disney cartoon. In this picture THE LAW dominates humans, rather than the other way around. This dehumanized idea of law, wherein the human subject “becomes the object of his object” (to quote the Chilean legal philosopher Rolando Gaete⁵) is brilliantly and vividly depicted in Franz Kafka’s great parable “Before the Law.”⁶ In that parable a man from the country comes to the Door of the Law seeking admittance—just like everyone else, he too wants to receive justice. But the doorkeeper will not let him in, and so the man sits down and waits to be admitted later. In fact, as he waits and waits the days grow into months, and the months into years, until finally the man nears the end of his own life, without ever once having been allowed admittance to the Law. As the man is dying, he asks the doorkeeper just one last question: In all my years of waiting, he

---

wonders, why is it that no one else has ever come seeking admittance to the Law? At the very moment of the man’s demise, the doorkeeper at last deigns to answer his question; the doorkeeper bends down and whispers: “No one but you could gain admittance through this door, since this door was intended for you. I am now going to shut it.”

Kafka’s parable appears in his novel *The Trial*. The book as a whole depicts a bureaucratic nightmare-world corresponding to the fairy-tale world that is projected by John Adams’ sunny phrase “a government of laws and not of men.” In Kafka’s world dehumanized legal institutions relentlessly grind people down, and human freedom and spontaneity are cruel illusions. But I have always thought that the good news in this otherwise depressing novel comes from the parable “Before the Law,” for in its deepest meaning this parable says that legal domination and emancipation are both absolutely impossible without constant human participation. It took the gatekeeper’s presence and implicit threats to keep the man outside the Law, just as the man also kept himself outside the law by failing to attempt an unauthorized entry. As for the man’s being before the Law in the temporal sense of the word “before,” the primordial phenomenon of time assures that he could never in principle have entered the door anyway! For there is no pause or gap in time—no stasis of time between the past and the present—that could ever be filled by such a thing as THE LAW before it gets received and applied in just the way that it is received and applied. Therefore humans must always make their appearance in time before the Law can appear, and never the other way around.

All of this implies that neither Kafka’s nor Adams’ worlds are even remotely possible, metaphysically speaking. This is because legal texts always require interpretation, and interpretation is always a human act. The irruption of the human being in the act of interpretation occurs even in the so-called “easy” case, where just this unique set of facts must be subsumed (by a human!) under a general rule that never explicitly refers to these particular facts as such. Even the quintessentially clear rule of chess that “Bishops can move only on the diagonal” contains nothing within its four corners about this particular move in this particular game. Someone—a human—must make the rule apply in every instance of its application. It was Plato, in the *Phaedrus*, who first drew attention to the fact that written

7. Id. at 269.
8. Id. at 267–69.
words can never “speak” what they mean, since they themselves have no understanding. Only the one being who is endowed with logos (the power of speech) has understanding, and that being is none other than the human being.

It seems to me that this obvious but nonetheless profound truth shows that each of us in this room is morally responsible, every day and every minute, for keeping alive all aspects—good and bad—of the social world that we inhabit. We do law—it does not do us. It follows that the most we could ever achieve is “a government of laws interpreted by humans,” and that this state of affairs could never fairly be put into opposition to Adams’ so-called lawless “government of men.” I know that it may sound a bit sacrilegious to say it, but both tyrannies and democracies happen to be governments “of men (and women)”—it’s just that they are different kinds of human ordering. From this point of view, even the practice of so-called judicial “strict constructionism” is not exempt from the fundamental truth that humans apply law, and not the other way around. To paraphrase von Clausewitz’s famous remark to the effect that war is the continuation of politics by other means,11 it seems to me that “strict constructionism,” “plain meaning,” and the jurisprudence of “original intent” are, notwithstanding their pretensions, really just the continuation of interpretation by other means. Indeed, these practices may actually reverse von Clausewitz’s epigram: for legal interpretation and application (considered as quintessentially “political” acts) are, as Foucault suggests, really the continuation of war (between classes and individuals) by other means.12 Is it not possible to gain an insight into the nature of law by viewing the venue of the legal decision as a battlefield that is littered with the metaphorical corpses of the losers? Indeed, the corpses are no longer metaphorical in the case of the administration of the death penalty. In exercising the power over life and death, the judicial system reveals what Walter Benjamin calls “something rotten in law”—namely, the secret that law is human violence
through-and-through, and that it reaffirms itself as violence each time it hands down a judgment.

When I was practicing law—and I conducted a dozen or so full-blown trials and countless settlements in my years in the trenches—it became manifestly clear to me how little “law” there is in the Law. The skill and intelligence of the lawyers, the politics and peccadillos of judges, the prejudices of juries, the sheer manipulability of legal texts and emotional responses, not to mention the constructibility of the “facts,” the brutal reality that wealth matters enormously to the ability to gain access to the law and to be successful at it: all of these factors, repeated countless times in my practice, showed me that the “Rule of Law” names a particular kind of rule of people, and not some radically pure kind of social arrangement. The Legal Realists of the 1920s and 30s taught us all a valuable lesson when they demonstrated, in field after field of law, the fundamental indeterminacy of legal materials considered as mere “texts,” and the ability of the wealthy and the powerful to have a disproportionate share of influence in those important interpretive moments when human beings use the law as a justification for violence. As the great and controversial German legal theorist Carl Schmitt wrote, in his 1922 book Political Theology: “Like every other order, the legal order rests on a decision, and not on a norm.” The Slovene philosopher Slavoj Žižek recently gave us this useful formulation of the unavoidable implication of Schmitt’s decisionism for the meaning of the Rule of Law: “The rule of law ultimately hinges on an abyssal act of violence (violent imposition) grounded only in itself: every positive statute to which this act refers in order to legitimate itself is self-referentially posited by this act itself.” In speaking about the insight that there is an important sense in which people’s actions are law, I would be remiss if I did not also mention America’s own John Chipman Gray, whose important book, The Nature and Sources of the Law, attempted to demonstrate that in a democracy the ultimate sovereign power to declare “what the law is” does not really reside in the people or their legislature, but

15. CARL SCHMITT, POLITICAL THEOLOGY: FOUR CHAPTERS ON THE CONCEPT OF SOVEREIGNTY 10 (George Schwab trans., 1988) (1st German ed. 1922).
in the judiciary, whose interpretations of legal texts in every real sense are the “Rule of Law.” 17

Or so I thought, until I got to know my friends in Slovenia who are trying to build their own version of the Rule of Law. They confront a national environment in which the people at large distrust and disdain law and lawyers, owing largely to the practices of party bosses in the former Yugoslav era. In those days, the political officials of the state and the party routinely interfered and intervened in legal disputes to dictate outcomes, operating under Marx’s thesis that bourgeois law is but a temporary expedient that is destined to wither away when socialism becomes fully established in the minds and hearts of the people. The result of this practice in the old Yugoslavia included the arrests of countless dissidents, as well as show trials in which there was only one permissible outcome—conviction and, in some cases, execution.

For my friends in Slovenia the “Rule of Law” did not mean the metaphysically impossible capacity of legal texts to “constrain” and “determine” the behavior of those judges who apply them. It meant, rather, somehow obtaining a judiciary that actually felt itself to be independent of the more overtly political branches of government, to such an extent that judges would actually stand up to the executive and legislative branches—as Chief Justice Marshall did in Marbury v. Madison 18—and hold them accountable for failing to abide by fundamental norms of human rights. In short, I came to learn that in Slovenia the “Rule of Law” meant, first and foremost, a separation of powers in which the judicial power is real and concrete, and not merely formal.

From their point of view, my friends told me, it would be an incredible luxury to be in a position to worry and fret, as some American politicians do today, about whether or not a particular judge is an “activist.” Such a problem, they told me, could only be perceived as a problem in a social context in which civil respect for law and legal institutions is already well-established—in fact, where it is ingrained in the respect that people show for judicial authority. Respect for authority of any kind is not something that occurs overnight, and once lost it is very difficult to regain. For Slovene intellectuals the decision in favor of the Rule of Law has been a decision for the formal principle of judicial ordering as such; the contingent content of that ordering is for them in many ways beside the point (at least

18. 5 U.S. (1 Cranch) 137 (1803).
for the time being). They would love to have judges, activist or not, who are not crushingly beholden to political actors—judges who feel themselves able and willing to stand up to the executive and legislative branches of government, as well as to the multi-national corporations who would like to exploit Slovenia’s resources by “special deals” that benefit only the few.

This insight into the problems that my Slovene colleagues face has made an enormous impact on my own thinking about the meaning of the Rule of Law. On the one hand, I feel even more sure than ever that there exists no golden “metaphysical certainty” in law, only the base metal of what I will call social certainty. The only reason judges think that THE LAW controls them is that they belong to a social group (lawyers and judges) whose members respond pretty much in similar ways to similar problems of application. That kind of relatively uniform social response is wired into us by our legal and social training. It shows, for example, why most people just stop at a stop sign, without first thinking, “Does this rule really apply to me in just this context?” In other words, much (even most) of law consists of routinized behavior—almost as if people were mindlessly plodding through a great deal of their lives like mules in harness. The idea that law is in many respects socially (but not metaphysically) determinate means that John Adams’ “government of laws and not of men” really describes a government of social tendencies about which we rarely think and of which we are rarely aware. The danger that this unthinking process of social interpretation and enforcement can become a tool of special interests, power, prejudice, and injustice is very real. We need only consult our own shameful history of legalized slavery and apartheid to confirm this. But—and this is a big but—the “capture” of law by injustice is possible only if those who apply the law are respected sources of authority in the first place!

As I see it, social power is social power, whether it is called “public” or “private.” The humiliations that an overweening and obnoxious boss in the private sector inflicts on a secretary who badly needs the job are every bit as hurtful as a similar round of mistreatment dealt by an uncaring state official to a needy supplicant. Whether it is public or private, power always craves to reinforce and extend itself wherever and however it can. In this respect I side with Nietzsche, for experience teaches me (as it did him) that the primordial instinct of Will-to-Power is a ubiquitous and overweening “force

of all forces” in social life. In the former Yugoslavia, social power attempted to reinforce itself by means of more-or-less direct command and control by a cadre of leaders at the top. In our society, congealed units of social power (e.g., large corporations) are always attempting to enhance themselves by means of the law—by using the law to their advantage. The main difference is that the social institution of judicial independence in our country tends to open up a larger space for what I will call “counter-hegemonic struggle.”

Don’t get me wrong!—I am no Pollyanna when it comes to thinking about the potential of our system to radically emancipate human beings. It is quite clear to me that in the current political war against so-called “frivolous lawsuits” and against publicly funded legal services for the poor, in habeas corpus “reform” and in immigration laws that cut back on judicial review, in the conservative drum-beat for universal “tort reform,” in recent drastic curtailments of civil liberties and access to courts in the interest of “security from terrorism,” and in countless other ways, today’s bastions of congealed social power have mounted a broad and effective counter-attack on the use of law by the powerless. But be that as it may, this battle against the powerless must still be waged on a legal terrain that is at least somewhat unpredictable, owing to the diffusion of power that is represented by the institution of judicial independence and widespread respect for that independence. If Legal Realism taught us how the rich and powerful get to stack the deck of the law in their favor, the story I have told you tonight about Slovenia’s struggles to achieve democracy ought to teach us to be grateful that at least there is a legal game to play in our country—a place in law where people have at least some chance to resist abuse by the powerful, and where judges just might agree with them and do something about it. As I see it, the possibility of access to an institutional place where power does not always get what it wants is none other than the possibility of Justice through law.

What all of this says is that the Rule of Law means something quite a bit more realistic than John Adams’ trite expression would have it mean. It seems to me that if the Rule of Law sometimes (or even often) can be used as a tool by the powerful, it can also occasionally be wielded as a force to hold power accountable. Its proper home lies within institutions staffed by men and women who do not reflexively do the bidding of the elite, but who actually delve into the past for documents and practices that might be

---

interpreted (always in the present!) to the disadvantage of the elite. From the standpoint of Justice, the Rule of Law is the chance to use social power to thwart social power—a brutally realistic but refreshingly honest realization that my Slovene colleagues came to more than a decade ago.

Viewed in its proper light, then, and shorn of all its silly metaphysical pretensions, the realistic idea of the Rule of Law that I have attempted to describe to you tonight strikes me as an important cultural achievement, at least for those who value the moral worth of the individual as an ultimate value. In the marketplace we are constantly required to prove our worth by our value to those with the means to pay us in money and prestige. But the Rule of Law is the possibility—however remote—that our worth as human beings is not solely determined by our market price, or by the powerful people that we happen to know. But if the Rule of Law in this sense is an achievement, it is also an achievement that must be continually won, over and over again, in concrete struggles before fiercely independent and courageous judges. I speak here of judges who know that the violence of law is always their doing, manifested in the interpretations they give. These are judges who do not think of themselves as cogs in a soul-less legal machine, but as creative and responsible actors in an unfolding and uncertain historical drama. It strikes me that such a humanized version of the Rule of Law might be well worth fighting for.

Thank you very much for listening to me so patiently, and good night to you all.
THE FORESEEABILITY OF TRANSFERENCE: 
EXTENDING EMPLOYER LIABILITY UNDER 
WASHINGTON LAW FOR THERAPIST SEXUAL 
EXPLOITATION OF PATIENTS

Timothy E. Allen, M.A.

Abstract: Transference, or the idealization of therapists, is a phenomenon that is foreseeable in every relationship between a therapist and a patient, and makes patients uniquely vulnerable to sexual exploitation by therapists. Transference has been recognized as a basis for finding therapists directly liable for harm resulting from sexual relations with patients. However, limitations on damages directly available from therapists lead patients to seek redress from therapists’ employers under theories of employer liability. Washington courts generally deny victimized patients relief from the employers of sexually exploitative therapists. This Comment argues that Washington courts should impose employer liability when therapists sexually exploit their patients, due to the foreseeability of transference. Employer liability is consistent with three existing theories of Washington agency law: (1) the exploitation of transference arises out of pursuit of the employer’s business, subjecting employers to respondeat superior liability; (2) patients qualify for a special relationship with their therapist’s employer, which imposes a direct duty on the employer to protect those patients; and (3) the risk of a therapist’s exploitation of transference is sufficiently foreseeable to subject employers to liability on the basis of negligent supervision.

S.H.C. became a follower of a prominent Buddhist priest. After years of devotion, she sought the priest’s counsel for curing her headaches. The priest informed S.H.C. that her headaches should be the least of her worries, as he sensed she would soon die. However, the priest offered to both save her life and cure her headaches by treating her with the “Twin Body Blessing.” After three years of the priest’s periodic “cure” in the form of sexual intercourse, S.H.C. sued the priest’s temple under various theories of employer liability. Division I of the Washington Court of Appeals upheld the trial court’s dismissal of all claims against the temple, interpreting Washington agency law to require a “special relationship” between S.H.C. and the temple before the temple could be held directly liable for any harm caused by the priest’s intentional acts. After reviewing available Washington precedent, the court held that

2. Id.
3. Id.
4. Id.
5. Id. at 515, 54 P.3d at 176.
6. Id. at 531, 54 P.3d at 184.
7. Id. at 525, 54 P.3d at 180.
S.H.C. did not have a special relationship with the temple because her status as an adult made her “unlike the victims who have been found to be vulnerable in other cases.” The appellate court also refused to impose direct liability on the temple under a theory of negligent supervision.

The case of S.H.C. highlights the shortcomings in Washington law regarding tort liability for employers of sexually exploitative therapists. Therapists’ employers face minimal liability in Washington courts for therapist sexual misconduct. Some courts have found employee sexual misconduct necessarily beyond the scope of employment. Other courts, although willing to entertain theories of direct employer liability, have found that employers could not have reasonably foreseen the intentional tortious misconduct of their therapist employees.

However, Washington courts recognize that therapists may be personally liable for harm caused by sexual relationships with patients on the basis that such harm is foreseeable due to patients’ vulnerability. Washington common law considers patients presumptively incapable of consenting to sexual relationships with their current therapists, and a state criminal statute imposes the burden of proving patient consent on....

8. Id. at 525–26, 54 P.3d at 181.
9. Id. at 523–24, 54 P.3d at 180. While recognizing that genuine issues of fact existed as to whether the temple knew of the priest’s sexual activities, the appellate court reasoned that the Free Exercise Clause of the First Amendment bars a civil court from imposing a duty of care on a religious authority, even where an employee’s sexual misconduct was known to the authority. Id. The relationship of the First Amendment to employer liability for sexual exploitation by religious counselors is beyond the scope of this Comment. Jurisdictions are divided over the extent to which liability can be imposed on religious institutions that employ sexually exploitative clergy. See generally Marjorie A. Shields, Annotation, Liability of Church or Religious Organization for Negligent Hiring, Retention, or Supervision of Priest, Minister, or Other Clergy Based on Sexual Misconduct, 101 A.L.R. 5th 1 (2002); Janice D. Villiers, Clergy Malpractice Revisited: Liability for Sexual Misconduct in the Counseling Relationship, 74 DENV. U. L. REV. 1 (1996); James T. O’Reilly & JoAnn M. Strasser, Clergy Sexual Misconduct: Confronting the Difficult Constitutional and Institutional Liability Issues, 7 ST. THOMAS L. REV. 31 (1994).
10. This Comment follows the Supreme Court of Minnesota by using the term “therapist” to refer to “a psychologist or psychiatrist, or, at times, to other professionals who hold themselves out as engaging in psychotherapy or counseling therapy for marital, family and sexual problems.” St. Paul Fire & Mar. Ins. Co. v. Love, 459 N.W.2d 698, 700 n.1 (Minn. 1990).
11. See infra notes 122–131 and accompanying text.
12. See infra notes 187–197 and accompanying text.
Employer Liability for Therapist Sexual Misconduct

therapists who raise it as an affirmative defense to charges of sexual misconduct.\textsuperscript{15} Courts and legislatures are rightly suspicious of consent defenses because the ability of a patient to consent to a sexual relationship with her\textsuperscript{16} therapist has long been questioned in the literature of psychoanalysis, in codes of professional ethics, and in law.\textsuperscript{17} The idealization of therapists by patients, also known as the \textit{transference} phenomenon,\textsuperscript{18} is a foreseeable occurrence in a normal counseling relationship.\textsuperscript{19} Transference makes patients particularly vulnerable to therapist suggestion and imposes upon therapists a heightened duty of care.\textsuperscript{20}

Employers of exploitative therapists in Washington should be liable for the devastating harms that victims often suffer. Many studies document the extensive damage caused to patients by therapist sexual exploitation.\textsuperscript{21} Yet the availability of direct damages from therapists is limited, leading plaintiffs to seek redress instead from the tortious therapists’ employers.\textsuperscript{22} However, under existing Washington law, plaintiffs lose suits against employers, whether pursued under theories of

\begin{itemize}
\item \textsuperscript{15} See WASH. REV. CODE § 9A.44.050 (2002).
\item \textsuperscript{16} While recognizing that exceptions exist to the general rule, this Comment will use male pronouns when referring to therapists and female pronouns when referring to patients, as the vast majority of sexual relationships arising from transference are between persons of those biological sexes. See St. Paul Fire & Mar. Ins. Co. v. Love, 459 N.W.2d 698, 700 n.1 (Minn. 1990); Malkah T. Notman & Carol C. Nadelson, \textit{Psychotherapy With Patients Who Have Had Sexual Relations With A Previous Therapist, in PHYSICIAN SEXUAL MISCONDUCT} 248 (Joseph D. Bloom et al. eds., 1999).
\item \textsuperscript{17} See infra Part I.B (explaining that transference makes patients vulnerable to sexual exploitation by therapists). See also Morgan v. Psychiatric Inst. of Wash., 692 A.2d 417, 421–22 (D.C. Cir. 1997) (holding therapist exploitation of transference denied patient ability to consent to sexual relationship with therapist); Ferguson v. People, 824 P.2d 803, 812–13 (Colo. 1992) (holding elimination of consent defense in state statute criminalizing psychotherapist-patient sexual relations not an unconstitutional violation of Due Process); Missouri v. Cone, 3 S.W.3d 833, 842 (Mo. 1999) (upholding psychiatrist’s conviction for sexual assault of two patients where plaintiffs unable to appreciate their actions due to manipulation of transference and patients’ severe mental incapacitation). \textit{But see} L.A.B. v. P.N., 533 N.W.2d 413, 417 (Minn. Ct. App. 1995) (holding loyalty to therapist caused by transference phenomenon insufficient to toll running of statute of limitations for medical malpractice under provision for “insanity”); State v. Leiding, 812 P.2d 797, 799–800 (N.M. Ct. App. 1991) (holding transference phenomenon cannot satisfy element of state statute criminalizing sexual penetration where victim suffers from “mental condition” rendering victim “incapable of understanding the nature or consequences of the act”).
\item \textsuperscript{18} See infra notes 28–39 and accompanying text.
\item \textsuperscript{19} See Doe v. Finch, 81 Wash. App. 342, 352, 914 P.2d 756, 762 (1996) (quoting statement of plaintiff’s expert witness that it “is well known that patients commonly become infatuated with their therapists”).
\item \textsuperscript{20} See infra notes 49–72 and accompanying text.
\item \textsuperscript{21} See infra Part I.B.
\item \textsuperscript{22} See infra Part II.A.
\end{itemize}
vicarious or direct liability. This leaves many victims of foreseeable therapist exploitation without an effective remedy.

This Comment argues that employer liability for a therapist’s exploitation of the transference phenomenon satisfies existing Washington tests for both vicarious and direct employer liability. The transference phenomenon is inherent to the therapeutic relationship, making the risk of therapist exploitation of transference foreseeable to employers. Because the dynamics of transference arise out of the employer’s pursuit of business, Washington courts should recognize vicarious employer liability under the theory of respondeat superior. Further, therapists’ patients satisfy the current definition of parties that deserve a special relationship with an employer, thus warranting direct employer liability. Finally, employers should face liability when they are negligent in supervising therapists, because the risk of exploiting transference is foreseeable.

This Comment begins with a description of the transference phenomenon. Part I defines transference, explores the potential harm patients suffer due to sexual relations with current or former therapists, and summarizes the reasons why therapists’ employers should reasonably foresee the risk that therapists might exploit transference. Part II reviews existing legal tests in Washington for holding employers liable for intentional employee misconduct and describes how Washington courts have not held employers liable for the foreseeable harms caused by therapist employees who sexually exploit patients by mishandling transference. In response to these shortcomings, Part III argues that the foreseeability of transference supports the extension of tort liability to employers of sexually exploitative therapists.

I. THE TRANSFERENCE PHENOMENON POSES A FORESEEABLE RISK THAT THERAPISTS MIGHT SEXUALLY EXPLOIT PATIENTS

The transference phenomenon is a regular and foreseeable aspect of therapy. Although it is a normal part of successful therapy, transference puts therapists in positions of power and authority that are subject to

23. See infra Parts II.B.–II.D.
24. See Sigmund Freud, Transference, in INTRODUCTORY LECTURES ON PSYCHOANALYSIS 431, 442–43 (James Strachey trans. & ed., 1966) (noting that “transference is intimately bound up with the nature of the illness itself” and “is present in the patient from the beginning of the treatment”); infra notes 28–48 and accompanying text.
Employer Liability for Therapist Sexual Misconduct

abuse and mishandling. When therapists exploit transference for sexual gratification, patients suffer a variety of harms. Courts find that the foreseeability of patient harm is a basis for imposing liability on both therapists and their employers when the transference phenomenon is exploited.

A. Transference is a Foreseeable Phenomenon in the Therapeutic Relationship

Therapists regularly probe their patients’ minds and learn the intimate details of their patients’ lives. A zone of intimacy called transference develops during this process, which makes it common for a psychotherapy patient to “fall in love with” or “idolize” the therapist. Transference has been defined as “the process whereby the patient displaces on to the therapist feelings, attitudes and attributes which properly belong to a significant attachment figure of the past . . . and responds to the therapist accordingly.”

As a common and expected occurrence, transference is a foreseeable phenomenon in therapy. Practitioners note that “[f]or any patient to experience an erotic transference is . . . entirely normal and expectable in the course of psychotherapy.” Mental health professionals expect and even elicit transference as a regular and accepted part of treatment. The proper instigation and response to transference is considered a basic psychoanalytic technique.

25. See infra notes 49–52 and accompanying text.
26. See infra notes 57–62 and accompanying text.
27. See infra notes 63–72 and accompanying text.
32. See supra note 24 and accompanying text.
34. Iwanski v. Gomes, 611 N.W.2d 607, 615 (Neb. 2000).
35. Zipkin v. Freeman, 436 S.W.2d 753, 755 n.1 (Mo. 1968) (quoting BLAKISTON’S NEW GOULD MEDICAL DICTIONARY 1260 (2d ed. 1956)).
and memories, the patient must often project those feelings upon someone trained to channel the feelings in a healthy direction.36

The appropriate handling of transference is defined for therapists by the standard practices of their profession. Therapists are trained to anticipate and deal appropriately with transference when it emerges in the therapeutic relationship.37 Emergent transference is recognized by inappropriate emotions directed toward the therapist.38 Once recognized, a therapist should “reject the patient’s erotic overtures and explain to the patient the true origin of her feelings” in emotions felt toward other important figures in her life.39

Therapists are also trained to avoid the mishandling of transference. A therapist may be tempted to reciprocate the intimate feelings proffered by a patient caught in the grips of transference.40 The proper response is counter-transference, requiring the therapist to self-monitor for developing feelings of intimacy41 and to avoid emotional involvement with the patient.42 Should a therapist find he is becoming inappropriately involved with a patient, he should terminate treatment and refer the patient to another therapist.43

Transference is foreseeable in every therapeutic relationship. While some courts restrict the duty to properly handle transference to licensed mental health professionals on the theory that only they “offer a course of treatment and counseling predicated upon handling the transference phenomenon,”44 other courts have also imposed this duty on other types of counselors.45 The transference phenomenon is not confined to mental health therapy46 but can arise in any counseling relationship,47 even pastoral counseling.48

38. Id. at 401–02.
40. See, e.g., Stuart W. Twemlow & Glen O. Gabbard, The Lovesick Therapist, in SEXUAL EXPLOITATION IN PROFESSIONAL RELATIONSHIPS, 71, 73–83 (Gabbard ed., 1989) (exploring the motivations of therapists who fall in love with their patients).
41. Id. at 85–87; see Notman & Nadelson, supra note 16, at 257–58.
43. Love, 459 N.W.2d at 700.
44. Simmons v. United States, 805 F.2d 1363, 1366 (9th Cir. 1986).
B. The Abuse or Negligent Mishandling of Transference Results in Foreseeable Harm to Patients

Transference lends itself to abuse by therapists. Transference puts therapists in positions of power and authority that are easily exploited or mishandled, because patients come to perceive therapists as powerful, benevolent figures. Studies of therapists indicate the temptation to take advantage of their position can prove overwhelming: therapists who have had sexual relationships with patients admit that the sexual contact was for their own gratification, and the proportion of therapists self-reporting sexual contact with patients is as much as twelve percent.

Most patients lack the ability to give meaningful consent to a sexual relationship with a therapist. Patients who become involved in such relationships tend to be unusually vulnerable. Studies show that the most reliable predictor of such sexual involvement is prior sexual victimization, usually in the form of childhood incest. One commentator notes that, in cases where therapist sexual exploitation is alleged, expert witnesses often testify that therapist-patient sexuality is analogous to parent-child incest. Endorsing this analogy, the Ninth Circuit has recognized that a sexual relationship between a therapist and a patient “replicat[es] at a symbolic level the situation in which a parent would be sexual with a child.”

Sexual relationships premised upon the exploitation of transference pose the risk of serious harm to patients. Common effects include

---

49. BISBING, supra note 28, at 210.
50. Simmons v. United States, 805 F.2d 1363, 1365 (9th Cir. 1986).
52. BISBING, supra note 28, at 199. See also Gary C. Hankins et al., Patient-Therapist Sexual Involvement: A Review of Clinical and Research Data, 22 BULL. AM. ACAD. PSYCHIATRY L. 109, 123 (1994) (“At least 10 percent of male and 3 percent of female psychotherapists will acknowledge being sexually involved with patients, and a sizeable portion of these professionals are involved with multiple patients.”).
54. Id.
56. Simmons v. United States, 805 F.2d 1363, 1365 (9th Cir. 1986) (quoting testimony of plaintiff’s expert witness).
increased depression, loss of motivation, impaired social adjustment, significant emotional disturbance, suicidal feelings or behavior, and increased drug or alcohol use.”57 Victimized patients experience feelings of abandonment, humiliation, and anger, and suffer from decreased self-esteem.58 The harm to the patient may also be aggravated after the relationship has ended because a therapist who exploits transference seriously damages his patient’s ability to trust future therapists.59 Studies indicate that over ninety percent of patients who are sexually victimized by therapists suffer serious residual harm.60 The compensatory damages available in actions against sexually exploitative therapists include awards for past and future pain and suffering, medical expenses, diminution in earning capacity, lost wages, loss of consortium or services, and the reasonable value of the therapists’ services.61 Sexual relationships between other professional counselors and their clients pose the risk of similar harm.62

The exploitation of transference makes certain relationships between therapists and patients susceptible to legal intervention. Not every personal relationship between therapists and patients reflects the exploitation of transference.63 If a client has consented to a sexual relationship with her therapist in the absence of the power imbalance that is characteristic of transference, the relationship is no more suspect than any other relationship between consenting adults.64 However, because of

57. Bouhoutsos et al., supra note 55, at 190.
58. See Shirley Feldman-Summers & G. Jones, Psychological Impacts of Sexual Contact Between Therapists or Other Health Care Practitioners and Their Clients, 52 J. CONSULTING & CLINICAL PSYCHOL. 1054, 1055 (1984); Zelen, supra note 51, at 181–82.
59. Twemlow & Gabbard, supra note 40, at 84.
60. Feldman-Summers & Jones, supra note 58, at 1055.
61. Elizabeth Williams, Cause of Action for Negligence or Malpractice of Psychiatrist, 13 CAUSES OF ACTION 2d 453, § 43 (1999).
62. See, e.g., Shirley Feldman-Summers, Sexual Contact in Fiduciary Relationships, in SEXUAL EXPLOITATION IN PROFESSIONAL RELATIONSHIPS 193, 205–09 (Gabbard ed., 1989) (arguing that the risk of harm due to therapist-client sexual contact is present in other fiduciary-client relationships); Joel Friedman & Marcia Mobilia Boumil, Betrayal of Trust: Sex and Power in Professional Relationships 30–43 (1995) (arguing that sexual relations with clergy, attorneys, professors and physicians are harmful to their clients).
63. See, e.g., Doe v. Swift, 570 So.2d 1209, 1213 (Ala. 1990) (noting transference requires period of time for trust in therapist to develop); Sisson v. Seneca Mental Health Council, 404 S.E.2d 425, 429 (W. Va. 1991) (holding transference requires both period of time for trust to develop as well as actual therapy sessions).
64. Cf. Jacobsen v. Muller, 352 S.E.2d 604, 607 (Ga. Ct. App. 1986) (holding summary judgment for psychologist appropriate where plaintiff was aware that sexual advances were beyond the duties of psychologist yet responded positively).
the high risk of harm to clients who enter such relationships under the influence of transference, many states impose some form of liability on therapists who exploit transference by engaging in sexual relations with their clients. Similarly, therapist organizations uniformly proscribe therapist-patient sexual relations in their codes of professional ethics.

Courts have held that the foreseeability of transference can serve as a basis for both therapist and employer liability for the harm caused to patients by sexual relations with therapists. In Simmons v. United States, the Ninth Circuit invoked the established nature of the transference phenomenon to distinguish between the heightened duty imposed on therapists to refrain from sexual relations with clients and the lesser duty owed by other professionals to clients. The court found the transference phenomenon to be a normal professional technique whose mishandling would constitute malpractice, stressing that "the same kind of authority power" held by the therapist is also held by any "powerful, benevolent parent figure." More recently, a federal district court in the Eastern District of Virginia ruled that the widespread public awareness of the problem of therapist sexual exploitation justified imposing employer liability for therapist sexual misconduct. The failure to explore the role of transference in an allegation of therapist sexual exploitation may even constitute attorney malpractice.

65. Many states criminalize therapist-patient sexual contact. See, e.g., ARIZ. REV. STAT. ANN. § 13-1418 (West 2001); CAL. BUS. & PROF. CODE § 729 (West 2003); COLO. REV. STAT. ANN. §§ 12-43-222(1)(c), 12-43-226(2) (West 2002); CONN. GEN. STAT. ANN. § 53a-71(a)(6) (West 2001); FLA. STAT. ANN. § 491.0112 (West 2001); GA. CODE ANN. § 16-6-5.1(c)(2) (Harrison 1999); IOWA CODE ANN. § 709.15 (West 1995); MISS. STAT. ANN. § 609.345 (West 2002); N.D. CENT. CODE § 12.1-20-06.1 (1997); N.H. REV. STAT. ANN. § 632-A:2(f)(g) (2002); S.D. CODIFIED LAWS § 22-22-29 (Michie 1998); TEx. PENAL CODE ANN. § 21.011(b)(9) (Vernon 2002); WIS. STAT. ANN. § 940.22(2) (West 2002). Additionally, some states impose statutory civil liability. See, e.g., CAL. CIV. CODE § 43.93(b)(2) (West 2003); ILL. COMP. STAT. ANN. § 140/2(2) (West 2002); MAss. STAT. ANN. § 148a.02 (West 1989); TEx. CIV. PRAC. & REM. CODE ANN. § 81.002 (Vernon 1997); WIS. STAT. ANN. § 895.70(2) (West 2002).


67. 805 F.2d 1363 (9th Cir. 1986).

68. Id. at 1366.

69. Id. at 1365.

70. Id. (quoting testimony of plaintiff’s expert witness).


72. See Jacobsen v. Boyle, 397 S.E.2d 1, 2–3 (Ga. Ct. App. 1990) (holding a question of fact whether attorney committed malpractice by failing to introduce expert testimony on the phenomenon of transference in action by plaintiff against sexually exploitative therapist).
In sum, transference is a normal and foreseeable phenomenon in effective therapy. Therapists are trained to anticipate transference and respond to it in a manner that furthers their patients’ well-being. However, transference also provides an opportunity for therapists to sexually exploit their patients. Courts have recognized that the foreseeability of transference may give rise to liability for therapists and their employers when patients are harmed by such sexual exploitation.

II. WASHINGTON LAW CURRENTLY LIMITS PLAINTIFFS’ ABILITY TO OBTAIN RELIEF FOR THERAPIST SEXUAL EXPLOITATION

Therapists can be held directly liable for harm caused by sexual relations with their patients. However, the limitations on damages recoverable from therapists lead plaintiffs to seek compensation from their therapists’ employers. Employers of exploitative therapists may be found liable under one form of vicarious and two forms of direct employer liability. One form of vicarious liability, respondeat superior, holds an employer liable for the torts of an employee. Respondeat superior liability, unlike direct liability, does not require the injured party to prove fault on the part of the employer. The two forms of direct employer liability require plaintiffs to prove either (1) the employer breached a duty owed by the employer directly to third parties, or (2) the employer’s own negligence in supervising its employee. Plaintiffs in Washington courts have sued employers of sexually exploitative therapists under all three theories of liability with relatively little success.

73. See supra notes 28–36 and accompanying text.
74. See supra notes 37–43 and accompanying text.
75. See supra notes 49–52 and accompanying text.
76. See supra notes 67–72 and accompanying text.
77. See infra notes 84–92 and accompanying text.
78. See infra notes 100–106 and accompanying text.
79. BISBING, supra note 28, at 181.
81. See id. at 906.
82. BISBING, supra note 28, at 181–82.
A. Washington Law Imposes Direct Liability on Therapists, but the Limited Damages Available to Victims Leads Them to Seek Further Redress from the Therapists’ Employers

Washington law imposes liability on therapists for harm resulting from sexual relations with patients. In Omer v. Edgren,84 Division III of the Washington Court of Appeals held that a fiduciary relationship85 exists between a psychiatrist and patient.86 Analogizing the relationship to that between a guardian and ward, the court held that sexual relations between a psychiatrist and patient may cause actionable harm due to the “malpractice, deceit, assault, and coercion by a person in a position of overpowering influence and trust.”87 In Omer, the plaintiff had an ongoing sexual relationship with her psychiatrist during her fifteen years of treatment.88 The court noted that permitting a patient who had been the victim of “deliberate and malicious abuse of power and breach of trust by a psychiatrist” to pursue a civil remedy, vindicated not only “a wrong against her” but also a wrong against “the public interest as well.”89

Therapists are subject to malpractice liability where they mishandle the transference phenomenon. Therapist liability is premised upon the unique characteristics of psychotherapy as a profession.90 Mishandling transference by engaging in sexual relations with a patient has been recognized as a basis for imposing liability on therapists.91 The

---

85. Fiduciary relationships arise “when one person places trust in the faithful integrity of another, who as a result gains superiority or influence over the first . . . .” BLACK’S LAW DICTIONARY 640 (7th ed. 1999). Fiduciary relationships entail “a duty to act with the highest degree of honesty and loyalty toward another person and in the best interests of the other person . . . .” Id. at 523. See also Flannigan, The Fiduciary Obligation, 9 OXFORD J. LEGAL STUD. 285, 286-97 (1989) (analyzing the trust-based nature of the fiduciary relationship).
87. Omer, 38 Wash. App. at 378, 685 P.2d at 636 (summarizing the rationale of the New York Court of Appeals in Roy v. Hartogs, 366 N.Y.S.2d 297, 299–301(N.Y. 1975)).
88. Id.
89. Id. at 379, 685 P.2d at 637 (quoting the New York Court of Appeals in Roy, 366 N.Y.S.2d at 301).
90. Benavidez v. United States, 177 F.3d 927, 930 (10th Cir. 1999).
foreseeability of transference transforms what would otherwise be an intentional tort into an action for negligence. Division I of the Washington Court of Appeals explained in an unpublished opinion that intentional sexual misconduct by therapists constitutes professional negligence subject to coverage by malpractice insurance, “even though similar contact between other professionals and their clients would be excluded by malpractice insurance as an intentional tort . . . because therapists, unlike doctors and other professionals, have to deal with transference as an aspect of the course of treatment.”

Washington courts have relied on the distinctive characteristics of transference to distinguish between imposing liability on malpractice insurers of therapists who have sexual relationships with their patients and insurers of other professionals who have sexual relationships with their clients. In Washington Insurance Guaranty Ass’n v. Hicks, Division I of the Washington Court of Appeals relied on the distinction made in Simmons between the duty of therapists toward clients and that of other professionals, in holding that a malpractice policy did not cover an insured chiropractor for sexual contact with the plaintiff during a visit to the chiropractor’s office for treatment. The court noted that the lack of transference in normal chiropractic care placed the chiropractor’s sexual misconduct beyond the norms of standard professional practice, and therefore outside the realm of insured acts. Similarly, in Standard Ins. Co. v. Love, 459 N.W.2d 698, 701 (Minn. 1990) (holding transference gives rise to therapist duty to refrain from personal relationship with patient during or outside therapy); MacClements v. LaFone, 408 S.E.2d 878, 880 (N.C. Ct. App. 1991) (noting sexual intimacy between therapist and patient constitutes malpractice or gross negligence); Darnaby v. Davis, 57 P.3d 100, 103 (Okla. Civ. App. 2002) (noting many courts recognize a cause of action against therapists who engage in sexual acts with patients); Sisson v. Seneca Mental Health Council, 404 S.E.2d 425, 429 (W. Va. 1991) (listing jurisdictions where therapist sexual contact with patient constitutes malpractice); see also C. Katherine Mann & John D. Winer, Medical Negligence: Psychotherapist’s Sexual Contact With Client, in 14 AM JUR. PROOF OF FACTS 3d 319 (2002); Brendan de R. O’Byrne, Annotation, Civil Liability of Doctor or Psychologist for Having Sexual Relationship With Patient, 33 A.L.R. 3d 1393 (1970).


94. See supra note 67–68 and accompanying text.

95. Hicks, 49 Wash. App. at 627, 744 P.2d at 627.

96. Id. at 627–28, 744 P.2d at 627.
Fire Insurance Co. v. Blakeslee,97 Division II of the Washington Court of Appeals held that a dentist’s intentional sexual assault of a patient while the patient was in a semiconscious state induced by nitrous oxide was not covered under the dentist’s professional liability policy.98 The court followed Hicks’ lead by excluding from malpractice coverage sexual contact not necessitated by the particular course of medical treatment.99

Although therapists may be found liable for the exploitation of transference, victims of sexually exploitative therapists face difficulty recovering damages directly from therapists because of exclusionary clauses in therapists’ malpractice policies. A patient’s ability to initiate a suit against a sexually exploitative therapist generally depends on finding an attorney willing to undertake such a suit on a contingent fee basis.100 However, the availability of monetary relief is limited. Uninsured therapists are virtually judgment-proof.101 Although courts have held that sexual exploitation of transference constitutes a professional breach of duty that is subject to coverage under malpractice insurance,102 many insurance companies specifically exclude coverage for sexual misconduct from their professional malpractice policies.103 Courts in other jurisdictions have upheld these exclusionary clauses for sexual misconduct and denied relief to plaintiffs that alleged the misuse of transference in malpractice claims,104 and Washington courts have

---

98. Id. at 11, 771 P.2d at 1177.
99. Id. at 9, 771 P.2d at 1176.
followed suit. As a consequence, plaintiffs have sought redress from therapists’ employers.

B. Washington Courts Use the “Scope of Employment” Test to Limit Vicarious Employer Liability for the Sexual Misconduct of Tortious Therapists

Under the vicarious liability theory of respondeat superior, employers are held strictly liable for the tortious act of their employees. The nature of respondeat superior liability varies by jurisdiction. Generally, respondeat superior liability makes employers “stand good” for the wrongs committed by their employees. To ensure that liability is imposed fairly, courts have limited the sphere of employer liability to harm posed by an employee’s conduct when acting “within the scope of . . . employment.” Defining what conduct falls within the scope of employment is notoriously difficult, and jurisdictions are divided as to


106 Jorgenson et al., supra note 103, at 597, 600 n.33.

107 See DOBBS, supra note 80, at 905.

108 See BISBING, supra note 28, at 181–82.

109 DOBBS, supra note 80, at 906.

110 Id. at 910. An alternative approach to respondeat superior liability is the theory of job-created authority, which supports employers absorbing as a foreseeable business expense any harm made peculiarly possible by the nature of the employment. See id. at 914. See also Rochelle Weber, Note, Scope of Employment Redefined: Holding Employers Vicariously Liable for Sexual Assaults Committed by Their Employees, 76 MINN. L. REV. 1513, 1533 (1992):

People are rarely employed to commit torts, but employers may put employees in positions where they can do so. For example, employing a therapist who will use transference, which places the patient in a vulnerable position, necessarily entails the danger of misuse and therefore should be considered a cost of doing business.


111 See WILLIAM L. PROSSER, TORTS § 70, at 460 (4th ed. 1971) (“This highly indefinite [language] which sometimes is varied with ‘in the course of the employment,’ is so devoid of
whether intentional sexual misconduct falls within a therapist’s scope of employment.\textsuperscript{112}

In Washington, the doctrine of respondeat superior provides that the employer\textsuperscript{113} is liable for the acts of an employee committed within the scope or in the course of employment.\textsuperscript{114} The general test for determining whether an employee is acting in the scope of employment is whether the employee is engaged in the furtherance of the employer’s interest.\textsuperscript{115}

meaning in itself that its very vagueness has been of value in permitting a desirable degree of flexibility in decisions.


113. Employers are “masters” and employees “servants” for purposes of the traditional terms of respondeat superior. See DOBBS, supra note 80, at 905.


115. Dickenson, 105 Wash. 2d at 467, 716 P.2d at 819. The issue of whether a particular employee is engaged in furtherance of the employer’s interest is often one for the trier of fact, as in the example of a messenger who negligently injures someone while driving back to work from a baseball game. See DOBBS, supra note 80, at 912.
Thus, an employer can be liable for an employee’s intentional sexual misconduct if it is committed in pursuit of the employer’s business.116

Washington courts have been reluctant to find that an employee’s intentional sexual misconduct was motivated by a desire to further the employer’s interest, and therefore have held that the misconduct was outside the employee’s scope of employment.117 This position is in opposition to the Ninth Circuit Court of Appeals’ earlier interpretation of Washington law.118 Applying Washington law, the Ninth Circuit held in Simmons v. United States119 that a counselor was acting within the scope of employment when he engaged in a long-term sexual relationship with a client.120 The court reasoned that because the sexual relationship arose from the counselor’s mishandling of transference, a phenomenon arising within the scope of the employee’s duties, the employer was liable under Washington agency law.121

However, Washington courts have rejected Simmons as a basis for imposing liability on the employers of tortious employees. Division I of the Washington Court of Appeals expressly rejected Simmons in Thompson v. Everett Clinic.122 In Thompson, the plaintiff sued a clinic where he had been sexually assaulted by a physician during a physical examination, bringing his claim under a theory of respondeat superior.123 The Thompson court applied the intentionality rule of Kuehn v. White,124 that a servant’s intentionally tortious or criminal acts will not be attributed to the employer “even though the employment situation provided the opportunity for the servant’s wrongful acts or the means for carrying them out.”125 The Thompson court interpreted Kuehn to mean that “a tort committed by an agent, even if committed while engaged in the employment of the principal, is not attributable to the principal if it emanated from a wholly personal motive of the agent and was done to

117. See infra notes 122–131 and accompanying text.
119. 805 F.2d 1363 (9th Cir. 1986).
120. Id. at 1369.
121. Id. at 1369 (citing Smith v. Leber, 34 Wash. 2d 611, 623, 209 P.2d 297, 303 (1949)).
123. Id. at 550, 860 P.2d at 1056.
125. Id. at 278, 600 P.2d at 682.
Employer Liability for Therapist Sexual Misconduct

gratify solely personal objectives or desires of the agent.”126 Thus, the court held that the clinic was not liable for the physician’s misconduct because it was a wholly personal and intentional act that was not done in furtherance of the clinic’s business.127

In Lavalley v. Ritchie,128 an unpublished opinion of Division I of the Washington Court of Appeals, the court relied on Kuehn and rejected Simmons in holding that a therapist’s sexual relationship with a client initiated during therapy was beyond the scope of employment because it was not furthering the purpose of the employer.129 In reaffirming its earlier rejection of Simmons, the court noted that the therapist’s “wholly personal motive” in the relationship with his patient was “too far removed from his counseling duties” to warrant imposing respondeat superior liability on his employer.130 Washington courts have focused on the presence of a personal motivation to generally exclude employers from liability for the sexual misconduct of their employees.131

While Washington courts have not distinguished between therapist employees and other employees in determining employer liability for sexual misconduct, courts in other jurisdictions have held employers of sexually exploitative therapists liable under a scope of employment test on the theory that the exploitation was a foreseeable risk in the nature of therapy. The Supreme Court of Minnesota held in Marston v. Minneapolis Clinic of Psychiatry132 that the test for whether an employee’s act is within the scope of employment should be a factual one, based on both foreseeability and whether the act bore a relation to or was connected with acts otherwise within the scope of employment.133 In

126. Thompson, 71 Wash. App. at 553, 860 P.2d at 1058.
127. Id. at 554, 860 P.2d at 1058. Accord Doe v. Swift, 570 So.2d 1209, 1213 (Ala. 1990) (holding personal motivation placed therapist’s sexual acts beyond scope of employment); Dausch v. Rykse, 52 F.3d 1425, 1428 (7th Cir. 1994) (holding sexual acts for therapist’s personal benefit and thus beyond scope of employment).
129. Id. at *10.
130. Id. at *8–10.
132. 329 N.W.2d 306 (Minn. 1983).
133. Id. at 311.
Doe v. United States, a federal district court in the Eastern District of Virginia reasoned that a psychologist’s sexual exploitation of a patient under hypnosis was foreseeable and therefore within the psychologist’s scope of employment, given the increasing public awareness of the ethical edicts prohibiting therapist sexual exploitation of clients. And the Supreme Court of California recognized, in Lisa M. v. Henry Mayo Newhall Memorial Hospital, that sexual relations resulting from the mishandling of transference may be a foreseeable intentional tort for which an employer may be liable under a theory of vicarious liability.

In sum, Washington courts have rejected imposing vicarious liability on employers for an employee’s intentional sexual misconduct. This distinguishes Washington courts from courts in other jurisdictions that have determined the scope of employer liability based on the foreseeability of the risk of an abuse of transference. However, in the absence of respondeat superior liability, employers may be held liable for acts of employee sexual misconduct under two theories of direct employer negligence. The sources of direct employer liability for the sexual misconduct of employees are duties that arise out of (1) a special relationship with a party whose well-being has been entrusted to the employer, or (2) the employer’s negligent supervision of the employee. An employer’s liability for the breach of these duties is “analytically distinct and separate from” respondeat superior liability, and is not limited by the scope of employment test.

C. Washington Courts Have Been Unwilling to Recognize a “Special Relationship” Between Patients and Therapists’ Employers

Under Washington law, an employer has a duty to make reasonable efforts to protect parties with whom it stands in a “special relationship” from foreseeable harm caused by employees’ intentional or criminal

135. Id. at 194–95.
136. 907 P.2d 358 (Cal. 1995).
137. See id. at 365.
138. See supra notes 122–131 and accompanying text.
139. See supra notes 132–137 and accompanying text.
141. Id. at 48, 929 P.2d at 426.
142. Id.
Employer Liability for Therapist Sexual Misconduct

This duty, based on the Restatement (Second) of Torts § 315, represents an exception to the general rule that one has no legal duty to protect others from the intentional or criminal acts of third parties. In the case of the patients of therapists, a special relationship can arise from either the employer’s duty to control therapists’ actions toward patients, or from a right of protection owed patients by employers.

Washington courts have recognized the existence of a special relationship when one party has entrusted its well-being to another and as a consequence has become particularly vulnerable. In *Funkhouser v. Wilson*, Division I of the Washington Court of Appeals noted that the special relationship between a hospital or nursing home and its patients is based on the vulnerability of the physically or mentally ill and their inability to care for themselves. The court identified multiple factors to consider when determining whether the vulnerability created by entrustment gives rise to a legal duty of protection:

> [T]he balancing of the societal interests involved, the severity of the risk, the burden upon the defendant, the likelihood of occurrence, the relationship between the parties, the temptation presented by the act or failure to act, the gravity of the harm that may result, and the possibility that some other person will assume the responsibility for preventing the conduct or the harm, together with the burden of the precautions which the actor would be required to take.

The special relationships recognized to date by Washington courts include those between a school and its students, an innkeeper and its guests, a common carrier and its passenger, an employer and its

---

145. RESTATEMENT (SECOND) OF TORTS § 315. General Principle: There is no duty so to control the conduct of a third person as to prevent him from causing physical harm to another unless (a) a special relation exists between the actor and the third person which imposes a duty upon the actor to control the third person’s conduct, or (b) a special relation exists between the actor and the other which gives to the other a right to protection.
147. Id. at 660, 950 P.2d at 509.
148. Id. at 661, 950 P.2d at 509.
employee, a hospital and its patient, a business establishment and the customer, a group home and its developmentally disabled residents, and a church and vulnerable persons within the custody of its workers or volunteers.

The duty imposed on an employer by its special relationship with third parties is broader than the duty imposed by the scope of employment test. The special relationship that arises from entrustment gives potential victims a right to protection from third parties even in the absence of a custodial or continuous relationship between the caregiver and the victim. The caregiver’s duty to protect potential victims extends to foreseeable injuries, even if perpetrated off the premises or outside working hours. Furthermore, because the “duty is limited only by the concept of foreseeability,” it includes intentional or criminal sexual misconduct. Under Washington law, intentional or criminal sexual misconduct “may be foreseeable unless it is ‘so highly extraordinary or improbable as to be wholly beyond the range of expectability’” and is “within the general field of danger which should have been anticipated.”

Washington courts have held employers liable for their employees’ sexual misconduct where a special relationship existed between the victim and the employer. In *Niece v. Elmview Group Home*, the plaintiff, a developmentally disabled woman, sued a nursing home where she had suffered a sexual assault by a male employee. The Supreme


158. C.J.C., 138 Wash. 2d at 727, 985 P.2d at 277.

159. Niece, 131 Wash. 2d at 50, 929 P.2d at 427.

160. Compare id. at 50–52, 929 P.2d at 426–48, with supra notes 122–126 and accompanying text (discussing cases that preclude employer liability for an employee’s intentional or criminal misconduct when issuing from wholly personal motives).


162. Id.

163. Id. at 41, 929 P.2d at 422.
Court of Washington held that the sexual assault was foreseeable within the “general field of danger which should have been anticipated.” The court determined that the group home had a duty to prevent sexual assault because of the special relationship between the group home and its vulnerable residents. Intentional or criminal sexual misconduct satisfied the test under Washington law for foreseeable harm because it was not extraordinary or improbable enough to be wholly beyond the range of expectability. In a later case, the court imposed liability on a defendant church for sexual misconduct that occurred beyond the church premises and outside working hours because of the special relationship between the church and the children entrusted to its care.

However, Division I of the Washington Court of Appeals declined to find a special relationship between an employer and the victim of an employee counselor’s sexual misconduct in S.H.C. v. Lu. The plaintiff in S.H.C. claimed that the defendant temple breached its fiduciary duty to her when its priest used sexual therapy. Although S.H.C. asserted that she was vulnerable because she was devoted to the sexually exploitative priest, the court reasoned that S.H.C. was not a “particularly vulnerable” victim as required to invoke a special relationship. The court did not consider the phenomenon of transference, nor did it explain why S.H.C. was not sufficiently vulnerable to merit a special relationship; it resolved the issue on the basis of a lack of available Washington precedent.

Conversely, courts in other jurisdictions have held that a special relationship exists between the employer of a sexually exploitative therapist and his victim. In Grzan v. Charter Hospital of Northwest Indiana, an Indiana Court of Appeals held that the victim of a therapist’s alleged mishandling of transference could seek redress from the employing hospital. The court determined that the employer owed the victim a direct duty of care where the therapist was acting outside the

---

165. Id. at 50, 929 P.2d at 427.
166. Id.
167. Id.
169. Id. at 524–27, 54 P.3d at 180–81.
170. Id. at 524–27, 54 P.3d at 180–81.
171. Id. at 525–26, 54 P.3d at 181.
172. Id.
174. Id. at 793–94.
In Richard H. v. Larry D., a California appellate court held that a sexually exploitative therapist’s breach of professional and fiduciary duties toward his patient was also actionable against the hospital where he was employed. According to one legal commentator, the court implied in dicta that the hospital owed the patient a direct fiduciary duty because the therapist was acting on its behalf.

In sum, the special relationship doctrine imposes upon employers a direct duty to protect from sexual misconduct vulnerable parties who have entrusted their well-being to the employers. While courts in other jurisdictions have been willing to impose a direct duty of patient care on employers of sexually exploitive therapists, the S.H.C. court declined to hold a religious therapist’s employer liable on the basis of the special relationship doctrine.

D. Washington Courts Limit Direct Employer Liability for Negligent Supervision to Employers’ Ability to Reasonably Foresee the Risk of Harm Posed to Patients by Particular Therapists

Patients can bring suit against employers for negligent supervision of sexually exploitative therapists independent of any direct duty owed them by virtue of a special relationship. Under the theory of negligent supervision, employers may be liable for acts committed outside an employee’s scope of employment. In Washington, the theory of negligent supervision for acts committed outside an employee’s scope of employment is based on the Restatement (Second) of Torts § 317.

---

175. Id.
177. Id. at 810.
179. See supra notes 143–62 and accompanying text.
180. See supra notes 173–78 and accompanying text.
181. See supra notes 169–72 and accompanying text.
182. See infra note 184.
183. See infra note 184.
184. RESTATEMENT (SECOND) OF TORTS, § 317 (1965). Duty of Master to Control Conduct of Servant:

A master is under a duty to exercise reasonable care so to control his servant while acting outside the scope of his employment as to prevent him from intentionally harming others or from so conducting himself as to create an unreasonable risk of bodily harm to them, if (a) the servant (i) is upon the premises in possession of the master or upon which the servant is
Washington employers have a limited duty to supervise an employee’s conduct outside the scope of employment where the employer knows or should know of “the dangerous tendencies of the particular employee.” Generally, Washington courts have been reluctant to impose liability on employers for the intentionally tortious acts of their employees where those acts involve sexual misconduct.

Washington courts thus far have held that any harm caused in non-therapeutic settings by an employee’s sexual relationship could not have been foreseen in the normal performance of the employee’s duties. In *Scott v. Blanchet*, plaintiffs appealed the trial court’s grant of summary judgment in a suit against a teacher, high school, and archdiocese for damages resulting from a sexual relationship between the plaintiffs’ daughter and her high school teacher. Division I of the Washington Court of Appeals held that there was no evidence that the teacher was acting in his official capacity at the time of the sexual activities, and that the school lacked any advance knowledge of the particular teacher’s proclivities. Consequently, the parents could not recover against the high school or archdiocese on the basis of respondeat superior or for negligent hiring and supervision of the teacher. Similarly, in *Peck v. Siau*, Division II of the Washington Court of Appeals held that a school was not responsible for a teacher’s sexual relations with a student, even though another teacher knew about the relationship. The court reasoned that the school was not liable because the second teacher had no hiring or supervision authority over the exploitative teacher.

Similarly, Washington courts to date have refused to hold the employers of sexually exploitative therapists liable under a theory of negligent supervision. In an unpublished opinion, Division I of the Washington Court of Appeals declined in *Lavalley v. Ritchie* to impose

privileged to enter only as his servant, or (ii) is using a chattel of the master, and (b) the master (i) knows or has reason to know that he has the ability to control his servant, and (ii) knows or should know of the necessity and opportunity for exercising such control.

186. See infra notes 187–97 and accompanying text.
188. Id. at 38, 747 P.2d at 1125.
189. Id. at 41–44, 747 P.2d at 1127–28.
190. Id. at 43–44, 47, 747 P.2d at 1128, 1130.
192. Id. at 287, 289–90, 827 P.2d at 1109, 1110–11.
193. Id. at 290–92, 827 P.2d at 1111–12.
liability on the employer of a sexually exploitative therapist even though the employer had some knowledge of the rumored affair, the therapist was living with the patient, and the therapist listed the patient’s address and phone number as his own. The court held that the therapist’s workplace discussions about his relationship did not identify the client in sufficient detail to inform the employer of the existence of an improper relationship. Recently, when faced with the question whether employer knowledge of employee sexual misconduct was sufficient to establish employer liability, Division I of the Washington Court of Appeals refused to rule on the issue in *S.H.C.*, holding that on the facts of that case it was barred by the Free Exercise Clause of the First Amendment. Given the questionable precedential value of the unpublished *Lavalley* opinion and the court’s recent refusal to reach the merits of the negligent supervision claim, it is unclear how other Washington courts would decide the issue.

However, courts in other jurisdictions have imposed a duty on employers to supervise against therapist sexual misconduct, based on the foreseeability of harm arising from the transference phenomenon. The Court of Appeals in the District of Columbia, in *Morgan v. Psychiatric Institute of Washington*, reversed a grant of summary judgment to an employer in an action for negligent supervision, hiring and training. The court held that the victim’s damages from the employee therapist’s exploitation of transference were foreseeable and the evidence of injury was sufficient to withstand summary judgment. Similarly, in *Nelson v. Gillette*, the Supreme Court of North Dakota held that “the known risk of improperly handling the occurrence of transference,” when coupled with knowledge of the promiscuous sexual history of a developmentally disabled victim, may constitute a cause of action for negligent supervision against the employer of a therapist. The court noted that “[a]ll counseling relationships can be affected by the transference phenomenon.”

195. Id. at *12–16.
196. Id. at *14.
199. Id. at 423.
200. Id.
201. 571 N.W.2d 332 (N.D. 1997).
202. Id. at 342.
203. Id. at 341.
In sum, patients of sexually exploitative therapists have a difficult time receiving redress directly from their therapists. As a result, patients seek compensation from the therapists’ employers. However, Washington courts have not held employers liable for the conduct of sexually exploitative therapists, finding the therapists’ sexual misconduct outside the scope of employment. Similarly, Washington courts have refused to impose a duty of protection on employers arising from the special relationship between the employer and the therapists’ patients. Finally, Washington courts have held employers to a heightened duty of supervision only where the sexual misconduct of a particular therapist has been foreseeable to the employer.

III. EMPLOYER LIABILITY FOR THERAPIST SEXUAL MISCONDUCT SHOULD BE EXPANDED BECAUSE OF THE FORESEEABLE RISKS POSED TO PATIENTS BY THE TRANSFERENCE PHENOMENON

The foreseeability of transference places therapist sexual misconduct within existing Washington tests for employer liability. Respondeat superior liability is satisfied because the risk of exploitation of transference is foreseeable within the scope of a therapist’s employment. Further, a special relationship exists between therapists’ employers and patients because the risk of therapist exploitation of transference satisfies the factors from *Funkhouser*, is foreseeable within the general field of danger, and is not wholly beyond the realm of expectability. Finally, the general foreseeability of sexual misconduct should impose a duty on employers to supervise against therapist exploitation of the transference phenomenon.

---

204. See supra notes 100–05 and accompanying text.
205. See supra note 106 and accompanying text.
206. See supra notes 122–31 and accompanying text.
207. See supra notes 169–72 and accompanying text.
208. See supra notes 187–97 and accompanying text.
209. See infra notes 211–30 and accompanying text.
210. See infra notes 211–30 and accompanying text.
211. See infra notes 252–64 and accompanying text.
A. The Abuse or Negligent Mishandling of Transference Satisfies the Scope of Employment Test Because It Is a Foreseeable Risk of Furthering an Employer’s Business

Full appreciation of the transference phenomenon provides a basis under existing Washington law for holding employers liable for the tortious sexual misconduct of their therapists. Washington courts should find intentional sexual misconduct within the scope of employment if such acts are performed in furtherance of the employer’s business.212 Other courts have found therapist exploitation of the transference phenomenon sufficiently related to the pursuit of an employer’s business to fall within the scope of employment.213 The Ninth Circuit’s decision in Simmons is particularly instructive, given that it interprets Washington agency law to impose employer liability for a therapist’s sexual misconduct.214 Confronting the issue of whether an abusive counselor’s actions were in furtherance of the employer’s interest, the court noted that “the centrality of transference to therapy renders it impossible to separate an abuse of transference from the treatment itself.”215 The court also noted that “a sexual relationship between therapist and patient cannot be viewed separately from the therapeutic relationship that has developed between them.”216 Where the employee’s unauthorized acts are performed in conjunction with acts that the employee is instructed to perform, Washington agency law imposes respondeat superior liability.217

In both its Thompson and Lavalley opinions, Division I of the Washington Court of Appeals neglected to consider the unique dynamics of transference in failing to apply the theory of vicarious employer liability of Simmons.218 In Thompson, the court ruled on allegations of sexual abuse by a physician whose motive it held to be wholly personal.219 In the unpublished Lavalley opinion, the court held that a

212. See supra note 116 and accompanying text.
214. Simmons v. United States, 805 F.2d 1363, 1368–69 (9th Cir. 1986).
215. Id. at 1370.
216. Id. (quoting L.L. v. Med. Protective Co., 362 N.W.2d 174, 178 (Wis. App. 1984)).
218. See supra notes 127–31 and accompanying text.
219. See supra note 127 and accompanying text.
Employer Liability for Therapist Sexual Misconduct

The therapist’s sexual relationship with a patient during therapy was beyond the scope of employment because it determined the therapist’s motive to be wholly personal. But Simmons premised employer liability on a careful analysis of the dynamics of transference in the context of therapy. Neither the Thompson nor Lavalley courts adequately examined the intricate interrelation between a therapist’s normal duties and the abuse of transference. Yet, when imposing direct liability on tortious therapists, Washington courts have recognized that transference poses unique and foreseeable risks of harm to the therapists’ patients. Thus, the Thompson and Lavalley courts should have considered how personal motivation is related to an employee’s expected job duties in the unique context of transference. The Thompson and Lavalley courts’ summary rejection of employer liability in the presence of an employee’s personal motive is in conflict with those courts in other jurisdictions that have held employers of tortious therapists liable under the theory of respondeat superior regardless of personal motive.

Because abuse or negligent mishandling of transference is a risk inherent to the therapeutic relationship, such conduct should qualify as a foreseeable risk in the normal furtherance of the employer’s business. The abuse or negligent mishandling of transference originates in the power created by the therapeutic relationship. It is the transference phenomenon itself that renders the patient particularly vulnerable. When conducting therapy, a therapist employee is seeking to further the interest of the employer; sexual misconduct that occurs in the therapeutic setting is “best understood as inextricably related to the dynamics of the therapeutic relationship.” The foreseeability of the abuse or mishandling of transference should be considered a predictable result of

220. See supra note 129 and accompanying text.
221. See supra notes 119–121 and accompanying text.
223. See supra note 110.
224. See supra notes 90–99 and accompanying text.
225. See supra notes 49–52 and accompanying text.
the “relationship between the nature of the work involved and the type of

tort committed . . . .” Transference inherently arises during a
therapist’s pursuit of the employer’s business, and it is therefore within
the scope of the therapist’s employment. Washington courts should
follow the lead of other jurisdictions in recognizing that sexual
misconduct with a patient is an act that, for purposes of respond recep
tior superior liability, falls within the scope of a therapist’s employment.

B. The Foreseeable Risk that a Therapist will Abuse Transference
Should Qualify Patients for a Special Relationship with the
Employer

The risks posed to victims of exploitative therapists are sufficiently
foreseeable to warrant a duty arising from a special relationship between
victims and the employers or therapists. The unique vulnerability of
patients caught in the grips of transference satisfies the existing factors
for judicial recognition of a special relationship. Additionally, the
relationship between patients and the employers of therapists is directly
analogous to special relationships already recognized by Washington
courts.

Washington courts should recognize that a special relationship exists
between employers and therapists’ patients under the Funkho seur
factors. The risk of mishandling transference is severe enough to make
such mishandling likely to occur, as indicated by empirical studies and
professional prohibitions on sexual misconduct. The same studies and
self-reports of therapists demonstrate that the temptation to exploit

229. Love, 459 N.W.2d at 702 (noting that “sexual contact between therapist and patient arising
from the phenomenon may be viewed as the consequence of a failure to provide proper treatment of
sexual misconduct within scope of employment because represents “misguided effort” to serve employer).
liability premised on inseparability due to transference of patient treatment and therapist sexual
misconduct).
230. See supra notes 132–137 and accompanying text.
231. See infra notes 234–243 and accompanying text.
232. See infra notes 244–251 and accompanying text.
146–148 and accompanying text.
234. See supra notes 51–52 and accompanying text.
235. See supra note 66 and accompanying text.
transference is great.\textsuperscript{236} The harm that may result from the abuse or mishandling of transference can be devastating.\textsuperscript{237} Furthermore, there is a social interest in enforcing the trust relationship\textsuperscript{238} between therapists and their patients through employer liability. No other party is effectively assuming the responsibility for compensating victims for their injuries.\textsuperscript{239} Employers are in a position to directly allocate to the beneficiaries of the employer’s services the business expense of damage awards,\textsuperscript{240} and employers can be powerful agents in the prevention of therapist sexual abuse.\textsuperscript{241} Heightened employer liability creates a strong incentive for an employer to exercise care in the training and supervising of its employees.\textsuperscript{242} Finally, heightened employer liability is not so burdensome as to significantly decrease the delivery of counseling services, as demonstrated by the experience of those states that impose employer civil liability by statute.\textsuperscript{243}

The relationship between employers and their employee therapists’ patients is analogous to other relationships that Washington courts have recognized as special relationships. The \textit{Funkhouser} court noted that the special relationship between a hospital or nursing home and its patients is based on the heightened vulnerability of the physically or mentally ill.\textsuperscript{244} Washington courts should recognize that patients of sexually exploitative therapists possess a special relationship with the therapist’s employer for the same reason, because the \textit{Omer} court recognized that a mental health patient’s ability to consent to a sexual relationship with a therapist is necessarily illusory.\textsuperscript{245} Although the \textit{S.H.C.} court rejected this position where an adult female was sexually exploited by her religious counselor,\textsuperscript{246} the court did not consider the plaintiff’s vulnerability from the standpoint of transference. Therapists’ patients are vulnerable

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{236} See supra notes 51–52 and accompanying text.
\item \textsuperscript{237} See supra notes 57–62 and accompanying text.
\item \textsuperscript{238} See supra note 87 and accompanying text.
\item \textsuperscript{239} See supra notes 100–105 and accompanying text.
\item Weber, supra note 110, at 1537.
\item BISBING, supra note 28, at 227.
\item Weber, supra note 110, at 1537.
\item See BISBING, supra note 28, at 223–27. Both Minnesota and Illinois impose statutory employer civil liability if the employer knows or has reason to know that a therapist has sexually exploited a current or former patient and fails to take appropriate and timely remedial action. Ill. Comp. Stat. Ann. ch. 740, § 140/3 (West 2002); Minn. Stat. Ann. § 148A.03 (West 1998).
\item See supra note 147 and accompanying text.
\item See supra note 13 and accompanying text.
\item See supra notes 171–172 and accompanying text.
\end{itemize}
\end{footnotesize}
because the idealization of their therapists is inherent in effective therapy.\textsuperscript{247}

Washington courts confronted with therapist sexual misconduct should follow the example of the \textit{Niece} court, which held that sexual assault is foreseeable within the general field of danger that a group home should anticipate when caring for vulnerable residents.\textsuperscript{248} Similarly, courts should recognize that the inherent nature of transference and the frequency of its abuse make therapist sexual misconduct foreseeable within the general field of danger that therapists’ employers should anticipate.\textsuperscript{249}

Moreover, once Washington courts recognize the special relationship between employers and therapists’ patients, they should extend the duty of protection arising from the special relationship beyond the normal time and space limitations of the employer’s premises.\textsuperscript{250} The vulnerability of patients in the grips of transference warrants courts extending the duty of protection employers owe to patients to harm caused by sexual relations with therapists outside therapy sessions or off the employer’s premises.\textsuperscript{251}

\textbf{C. Employers Should Assume a Duty of Care For Purposes of Negligent Supervision Because the Risk of Abuse or Negligent Mishandling of Transference by Therapists is Generally Foreseeable to Employers}

Transference is a foreseeable phenomenon for any therapist because transference is inherent in effective therapy.\textsuperscript{252} Therapy patients are uniquely vulnerable to therapist manipulation of transference.\textsuperscript{253} The risk of potential harm to patients is sufficiently foreseeable to make the proper handling of transference a professional expectation of therapists\textsuperscript{254} and for professional organizations to proscribe the exploitation of transference in their ethical codes.\textsuperscript{255}

\begin{footnotesize}
\begin{enumerate}
\item See supra notes 32–36 and accompanying text.
\item See supra notes 163–166 and accompanying text.
\item See supra notes 28-52 and accompanying text.
\item See supra note 158 and accompanying text.
\item See supra note 168 and accompanying text.
\item See supra note 169 and accompanying text.
\item Simmons v. United States, 805 F.2d 1363, 1365 (9th Cir. 1986).
\item See supra notes 53–56 and accompanying text.
\item See supra notes 37–43 and accompanying text.
\item See supra note 66 and accompanying text.
\end{enumerate}
\end{footnotesize}
While the *S.H.C.* court held that imposing such a heightened supervisory duty on a religious organization was barred by the First Amendment, the duty should be imposed on employers of nonreligious therapists. The *Lavalley* court refused to impose liability on the employer of a sexually exploitative employee because the employer did not have particular knowledge of the employee’s misconduct. However, Washington courts have formulated the requirement of particular knowledge primarily within the context of relationships where transference is not encouraged or expected to occur. The *Lavalley* court did not consider how the foreseeability of transference might warrant imposing upon employers of therapists a heightened duty of supervision. Further, Division I of the Washington Court of Appeals declined to publish the decision, thus it has questionable precedential value.

Washington courts confronting the issue in the future should follow other jurisdictions in recognizing that the risk of sexual misconduct posed by transference is sufficiently foreseeable to impose upon employers a heightened duty of supervisory care. The abuse of transference by therapists is foreseeable. As courts in other jurisdictions have recognized, the generalized risk of therapist sexual misconduct is sufficient to impose a heightened duty of supervision on employers, even in the absence of knowledge of any risk posed by a particular employee. Where the risk of foreseeable harm to a patient is coupled with recognition of the patient’s vulnerability, other jurisdictions have held that employers of therapists have a duty to effectively supervise their employees. As an inherent phenomenon in therapy, transference can be easily abused or mishandled. Thus, Washington

256. See supra note 9.
258. Both Scott and Peck involved teacher-student relationships. *See supra* notes 187–97 and accompanying text. But cf. *Bisbing*, supra note 28, at 210 (“Similarities exist in the relationship between a therapist and patient and that of a teacher and student. Both exercise power and authority over, and are able to manipulate and control, vulnerable people in their care.”).
259. *Lavalley*, 1999 Wash. App. LEXIS 933, at *12–16. In upholding summary judgment for the employer on the negligent supervision claim, the *Lavalley* court only considered whether the foreseeability of transference to the patient upon learning of the relationship after the patient had ceased therapy might have mitigated her damages. *Id.* at *15–16.
260. *Id.* at *1.
261. *See supra* notes 90–99 and accompanying text.
262. *See supra* notes 198–203 and accompanying text.
263. *See supra* notes 201–03 and accompanying text.
264. *See supra* notes 49–52 and accompanying text.
courts should impose a heightened duty on therapists’ employers to supervise against therapist sexual exploitation, even in the absence of knowledge of a particular employee’s misconduct.

IV. CONCLUSION

As modern society becomes increasingly characterized by professional-client relations, the risk of harm caused by breaches of professional duty increases proportionately. Unfortunately, the law has been slow to accommodate the evolution of an increasingly fiduciary society. In Washington, this is particularly apparent in the case of the victims of therapist sexual exploitation.

The transference phenomenon is well-recognized in the fields of psychoanalysis and law. The potential harm posed by transference to clients of therapists is sufficiently established to serve as the basis for direct liability on the part of sexually exploitative therapists. But despite the inherent character of transference, courts have not applied the doctrine of respondeat superior to the employers of therapists who sexually exploit patients. Despite the unique vulnerability of patients experiencing transference, Washington courts have held that there is no employer duty to protect patients from sexual assault arising from the special relationship between patients and therapists’ employers. Further, employers are not liable for negligent supervision unless they have real or constructive knowledge of the danger posed by a particular therapist, even though transference and its attendant risks are foreseeable.

The failings in existing Washington agency law reflect a failure to appreciate the foreseeability of the transference phenomenon. A proper appreciation of the foreseeable risks posed by transference would extend employer liability under existing legal tests for respondeat superior, an employer’s duty of protection arising from a special relationship, and negligent supervision. Washington courts should adjudicate future claims of employer liability for sexually exploitative therapists by recognizing the foreseeability of transference.
A IS NOT A: WASHINGTON’S UNCONSTITUTIONAL LAW OF SINGLE-COUNT, SINGLE-DEFENDANT INCONSISTENT VERDICTS IN STATE V. GOINS

Natasha Shekdar Black, M.S.

Abstract: In State v. Goins, Division I of the Washington State Court of Appeals upheld inconsistent general and special verdicts on the same charge, even though the special verdict finding negated an element of the crime. The Goins court reasoned that the United States Supreme Court and the Washington State Supreme Court had previously upheld inconsistent verdicts in various contexts because the verdicts could have been the result of jury lenity. Therefore, overruling existing precedent, the Goins court upheld the inconsistent verdicts on the ground that distinguishing the Goins context would be elevating form over substance. This Note argues that the Goins decision is incorrect because the court failed to follow prior precedent and instead extended inapplicable case law beyond its reach, thereby depriving the defendant of his due process right to have the state prove all elements of the crime charged beyond a reasonable doubt. Further, this Note proposes remedial measures that would direct at least a mistrial in this context.

In the early morning of May 18, 2000, Matthew Goins went to Angela Z’s apartment. While there he tried to kiss Z, but she rebuffed him. They later fought. Goins claimed that Z attacked him, while Z said that Goins tried to rape her. Z called the police, and Goins was arrested and charged with second-degree assault with intent to commit indecent liberties. The crime of second degree assault with intent to commit indecent liberties is not a statutory sex offense in Washington. However, a felony committed with sexual motivation is a sex offense. Thus, in order to make the crime a sex offense the prosecutor had to, and did,

1. “Aristotle posited three laws as basic to all valid thought: the law of identity, A is A; the law of contradiction, A cannot be both A and not A; and the law of the excluded middle, A must be either A or not A.” THE COLUMBIA ENCYCLOPEDIA, Logic 1603 (5th ed. 1993). This Note explores an example of what Aristotle might label “contradiction.”


4. Id. at 726, 54 P.3d at 724–25.

5. Id. at 727, 54 P.3d at 725.

6. Id.

7. Id. at 727–28, 54 P.3d at 725.


9. Id. § 9.94A.030(38)(c).
separately charge that Goins acted with sexual motivation, even though sexual motivation is a recognized element of second-degree assault with intent to commit indecent liberties. The jury found Goins guilty of second-degree assault with intent to commit indecent liberties, but by special verdict found that he did not act with sexual motivation.

On appeal, Goins argued that these verdicts were irreconcilably inconsistent—no jury could have found that he acted both with and without sexual motivation. In State v. Goins, Division I of the Washington State Court of Appeals agreed. Nevertheless, it upheld his conviction. Citing cases that protect the unreviewable exercise of jury lenity, the court held that Goins could only have his guilty verdict reviewed for sufficiency of the evidence. This remedy is available to all criminal defendants, without regard to any inconsistency in the verdict, and presents a low standard: the reviewing court need only find sufficient evidence presented at trial to permit a trier of fact to rationally reach a guilty verdict. Because the appellate court concluded that the jury could have believed Z’s testimony, Goins’ inconsistent verdicts stood.

This Note argues that Division I of the Washington State Court of Appeals incorrectly decided Goins because it failed to follow applicable precedent and inappropriately extended prior case law beyond its scope, thereby denying the defendant his constitutional right to have every element of the crime charged proven beyond a reasonable doubt. Part I sets forth the due process requirement of proof beyond a reasonable doubt, presents an overview of jury verdicts, and explains the extent to which criminal jury verdicts are reviewable. Part II summarizes U.S. Supreme Court and Washington case law on inconsistent criminal verdicts in the various contexts in which they appear. Part III details the Goins decision. Part IV describes the Washington statute on inconsistent verdicts in the civil context. Part V argues that the Goins decision was

11. Id. at 729–30, 54 P.3d at 726.
12. Id. at 729, 54 P.3d at 726.
13. Id. at 726, 54 P.3d at 724.
14. Id. at 725, 54 P.3d at 724.
16. Id. at 730, 54 P.3d at 726.
17. Id. at 725, 54 P.3d at 724.
18. Id. at 742, 54 P.3d at 733.
19. See infra Part I.C.
both unconstitutional and out of line with precedent, and recommends a remedial framework. This Note concludes that under the due process requirement of proof beyond a reasonable doubt, and applicable Washington case law on inconsistent verdicts, defendants in cases such as Goins are entitled to at least a mistrial, and at most a judgment of acquittal.

I. CONVICTIONS IN WASHINGTON JURY TRIALS MUST RESULT FROM PROOF BEYOND A REASONABLE DOUBT AND ARE SUBJECT TO LIMITED REVIEW

In order to obtain a conviction at trial, the state is required to prove each element of the crime charged beyond a reasonable doubt. Juries in Washington criminal trials return their verdicts through general and special verdict forms. Once the jury returns its verdict, the verdict is subject to limited review. The defendant may challenge the sufficiency of the evidence for any conviction. However, because of the constitutional guarantee against double jeopardy, the state may not seek review of an acquittal.

A. The State Must Prove All Elements of a Crime Charged Beyond a Reasonable Doubt

In In re Winship, the U.S. Supreme Court held that the Due Process clause of the Fourteenth Amendment “protects the accused against conviction except upon proof beyond a reasonable doubt of every fact necessary to constitute the crime with which he is charged.” The Court cited two key considerations in upholding the requirement of proof beyond a reasonable doubt in criminal proceedings. First, the interests of the accused are of “immense importance” because conviction may result in loss of liberty and social stigma. A high standard of proof reduces

22. WASH. SUP. CT. CRIM. R. 6.16(c).
23. See infra Part I.C.
25. See infra Part I.C.
29. Id. at 363.
the risk of convictions based on factual error.\textsuperscript{30} Second, “use of the reasonable-doubt standard is indispensable to command the respect and confidence of the community in applications of the criminal law.”\textsuperscript{31} Without this respect and confidence, the “moral force of the criminal law” would be diluted by doubt.\textsuperscript{32} Following the lead of \textit{Winship}, the Washington State Supreme Court and the Washington State Legislature both require proof of every element of a crime charged beyond a reasonable doubt.\textsuperscript{33} Further, the state must also prove all allegations on special verdict forms beyond a reasonable doubt.\textsuperscript{34}

B. Special and General Jury Verdicts in Washington

Washington juries may return both general and special verdicts. A verdict is a jury’s finding or decision on the factual issues of a case.\textsuperscript{35} A general verdict is a verdict by which the jury finds in favor of one party or the other,\textsuperscript{36} such as when a jury finds the defendant guilty or not guilty.\textsuperscript{37} A special verdict\textsuperscript{38} is a verdict in which a jury gives a written finding for a specific question of fact.\textsuperscript{39} For example, when a prosecutor in Washington charges a defendant with a crime and there is sufficient evidence for a jury to find that the crime was committed with sexual motivation, the prosecutor is statutorily required to file a special allegation of sexual motivation.\textsuperscript{40} In such cases, if the jury finds the defendant guilty of the principal offense, it must return a special verdict as to whether the defendant committed the crime with a sexual motivation.\textsuperscript{41} Under Washington law, any felony with a finding of sexual

\begin{thebibliography}{99}
\bibitem{30} \textit{Id.} at 364.
\bibitem{31} \textit{Id.}
\bibitem{32} \textit{Id.}
\bibitem{35} \textit{BLACK’S LAW DICTIONARY} 1554 (7th ed. 1999).
\bibitem{36} \textit{Id.} at 1555.
\bibitem{37} See \textit{WASH. SUP. CT. CRIM. R.} 6.16(c).
\bibitem{39} \textit{BLACK’S LAW DICTIONARY, supra} note 35, at 1554.
\bibitem{40} \textit{WASH. REV. CODE \$ 9.94A.835(1)} (2002). The special allegation is not required for crimes that are inherently sexual offenses. \textit{Id.}
\bibitem{41} \textit{Id.} \textit{\$ 9.94A.835(2)}.
\end{thebibliography}
motivation is a sex offense, and anyone convicted of a sex offense must register as a sex offender. Indeed, sexual motivation is one of the primary situations in which a criminal jury must decide a special verdict in Washington.

C. Permissible Scope of Review of Jury Verdicts

Criminal defendants can seek to have guilty verdicts reviewed for sufficiency of the evidence. For this review, the appellate court determines “whether the evidence adduced at trial could support any rational determination of guilt beyond a reasonable doubt,” viewing the evidence in the light most favorable to the state. The court reviews the sufficiency of the evidence independently for each count on which the defendant was convicted. Courts seldom examine jury verdicts beyond sufficiency of the evidence because of the need for finality in the criminal process.

The state, however, may not seek review of an acquittal. The Double Jeopardy clause of the Fifth Amendment provides that no person shall “be subject for the same offence to be twice put in jeopardy of life or limb.” A defendant is “in jeopardy” once the jury is impaneled and sworn, or when the court begins to hear evidence in a bench trial. Since the U.S. Supreme Court decision in United States v. Ball in 1896, it has been clear that the state may not prosecute a defendant twice for

42. Id. § 9.94A.03(38)(c).
43. Id. § 9A.44.130(1).
44. Another common situation when the court will give the jury a special verdict form is when the state alleges that a defendant or an accomplice was armed with a deadly weapon during the commission of the principal crime. Id. § 9.94A.602. If the jury finds that the defendant was armed, the defendant will face a sentence enhancement. Id. § 9.94A.510(3) (for firearms); id. § 9.94A.510(4) (for deadly weapons other than firearms).
45. United States v. Powell, 469 U.S. 57, 67 (1984). See State v. Salinas, 119 Wash. 2d 192, 201, 829 P.2d 1068, 1074 (1992). There are other questions that criminal defendants may raise on appeal, but only sufficiency of the evidence review is relevant to this Note.
46. Powell, 469 U.S. at 67.
47. Salinas, 119 Wash. 2d at 201, 829 P.2d at 1074.
49. Id.
51. U.S. CONST. amend. V.
54. 163 U.S. 662 (1896).
the same offense. Therefore, once the jury has rendered a verdict of acquittal, the state cannot appeal or challenge the verdict.

II. INCONSISTENT VERDICTS: CAUSES, CONTEXTS, AND RESOLUTIONS

When one of the jury’s general or special verdicts contradicts another of its general or special verdicts, the verdicts are inconsistent. If the verdicts can be liberally construed in a manner that makes them consistent, courts will hold that there is no irreconcilable inconsistency. The commonly recognized causes of inconsistent verdicts are mistake or confusion, jury lenity, and compromise. Inconsistent verdicts arise in three contexts. First, inconsistent verdicts can occur when multiple defendants are tried jointly, and the acquittal of one defendant is inconsistent with the conviction of the other. Second, inconsistent verdicts can arise when one defendant is tried on multiple charges in a single trial and there is an inconsistency between the general verdicts, or between a special verdict on one charge and a general verdict on another charge. In both of these situations, courts will uphold the inconsistent verdicts as long as the conviction is supported by sufficient evidence because in each case the jury may have acquitted the defendant by exercising jury lenity. In the third context, inconsistent verdicts arise when a single defendant is tried on a single charge, and the special verdict on that charge is inconsistent with the general verdict on the same charge. Prior to State v. Goins, Washington courts facing this last situation had reversed the conviction that was inconsistent with a special verdict on the same count and granted a mistrial.

55. Id. at 669.
57. See infra Part II.A.
58. See infra Part II.A.
59. See infra Part II.B.
60. See infra Part II.C.1.
61. See infra Part II.C.2.
62. See infra Parts II.C.1, II.C.2.
63. See infra Part II.C.3.
64. A year prior to the Goins decision, Division II of the Washington Court of Appeals decided State v. Holmes, which, in dicta, explicitly questioned granting a mistrial when the conviction was inconsistent with a special verdict on the same count. See 106 Wash. App. 775, 781 n.2, 24 P.3d 1118, 1122 n.2 (2001); see also infra notes 165–170 and accompanying text. It was not until the Goins decision that there was a contrary holding. See infra Part II.C.3.
Inconsistent Verdicts

A. *Potentially Inconsistent Verdicts are Deemed Consistent Unless They Cannot be Reconciled*

In Washington, inconsistent verdicts exist only when the verdicts cannot be construed to be consistent. When a special verdict is amenable to multiple interpretations, the court will adopt the interpretation that supports the general verdict. Yet, even when two verdicts are logically inconsistent, it may be possible to reconcile them by liberal construction. Once reconciled, the two potentially inconsistent verdicts are deemed consistent.

For example, in *State v. Peerson*, Division I of the Washington State Court of Appeals upheld a conviction because it was able to reconcile arguably inconsistent verdicts through liberal construction. In *Peerson*, the defendant was charged with two counts of aggravated first-degree murder. For each murder charge, the court gave the jury an aggravating circumstance special verdict form that asked if the murder was committed in the course of a common scheme or plan. By definition, a common scheme or plan must result in more than one death in order for the jury to find that the aggravating circumstance existed. However, although the jury found that only one of the murders was committed as part of a common scheme or plan, it convicted the defendant of both counts of aggravated first-degree murder. The *Peerson* court was able to reconcile these verdicts in two ways. First, it reasoned that the jury could have understood the instructions to mean that a common scheme or plan was not technically complete until both the victims were dead. Second, the court noted that the instructions allowed the jury to convict

---

66. *Robinson*, 84 Wash. 2d at 45, 523 P.2d at 1195 (citing McCorkle v. Mallory, 30 Wash. 632, 71 P. 186 (1903)).
67. *Id.*
69. *Id.* at 766, 816 P.2d at 50.
70. *Id.* at 758, 816 P.2d at 46. *Peerson* was also charged with two counts of first-degree assault, which are not relevant to this Note. *Id.*
71. *Id.* at 764, 816 P.2d at 49.
72. *Id.* at 765, 816 P.2d at 50.
73. *Id.* at 764, 816 P.2d at 49–50.
74. *Id.* at 758, 816 P.2d at 46.
75. *Id.* at 766, 816 P.2d at 50.
76. *Id.*
on an alternative aggravating circumstance. Therefore, the Peerson court upheld both convictions through liberal construction.

B. Causes of Inconsistent Verdicts

Inconsistent verdicts typically arise from mistake or confusion, jury leniency, or compromise. In cases involving potential jury mistake or confusion, the court will find the inconsistent verdicts reconcilable. In State v. Burke, for example, the state alleged accomplice liability for two co-defendants in a crime involving only one weapon. The trial court instructed the jury that if either defendant possessed a weapon, both were legally deemed to have possessed the weapon. The jury found by special verdict that one of the defendants, but not both, was armed with a weapon. To reconcile the verdicts, the appellate court reasoned that the jury might not have understood the legal fiction of constructive possession. Because the jury may have misunderstood the court’s instruction, the court construed the special verdict to be consistent with the general verdict.

Another cause of inconsistent verdicts is the exercise of jury leniency. Jury leniency is the jury’s unreviewable power to refuse to enforce the law. Courts and commentators refer to this power alternatively as jury nullification, jury mercy, jury lawlessness, jury justice, and jury veto power. The U.S. Supreme Court has endorsed the exercise of jury leniency as “recognition of the jury’s historic function, in criminal trials, as a check against arbitrary or oppressive exercises of power by the Executive Branch.” Although the Court has recognized that through jury leniency the

77. Id.
78. Id.
80. Id. at 382, 952 P.2d at 621.
81. Id. at 383, 952 P.2d at 621.
82. Id.
83. Id. at 387, 952 P.2d at 623.
84. Id. at 388, 952 P.2d at 623.
86. Id.
88. Powell, 469 U.S. at 65. Despite this judicial characterization of jury leniency as a benign power, the jury has historically been able to use this power to the detriment of criminal defendants. See Eric L. Muller, The Hobgoblin of Little Minds? Our Foolish Law of Inconsistent Verdicts, 111 HARV. L. REV. 771, 803–06 (1998).
Inconsistent Verdicts

jury is exercising a power it does not have, and is “return[ing] a verdict of not guilty for impermissible reasons,” the Court has maintained that inconsistent verdicts possibly resulting from jury lenity are not subject to review on that ground. The exercise of jury lenity has led one observer to note that the phenomenon is “not only unpunishable, but irrevievable and absolute.”

Jury lenity has a similar effect in Washington. In State v. Ng, for example, the jury acquitted the defendant of thirteen counts of felony-murder, but convicted him of the thirteen underlying robbery felonies. The court decided that because the defendant had admitted to the killings, the jury was inconsistent in acquitting him of the felony-murders while convicting him of the robberies. The court attributed the inconsistency to jury lenity.

Finally, inconsistent verdicts can arise through compromise when a deadlocked jury negotiates some mix of convictions and acquittals. For example, in a conspiracy charge, one half of the jury may think both defendants are guilty beyond a reasonable doubt while the other half may think both are innocent. Instead of deadlocking, the jury may decide to compromise by convicting one defendant and acquitting the other. According to one commentator, compromise is not within the power of the jury and undermines the integrity of the judicial system. However, because it is not possible to accurately separate inconsistent verdicts that result from compromise and those that result from jury lenity, the same umbrella that protects jury lenity also shades compromise verdicts from appellate review. Therefore, although compromise verdicts potentially eviscerate the requirement of proof beyond a reasonable doubt and the

89. Powell, 469 U.S. at 66 (citing Dunn v. United States, 284 U.S. 390 (1932)).
90. Id. at 63 (quoting Harris v. Rivera, 454 U.S. 339, 346 (1981)) (internal quotations omitted).
91. Id.
92. CONRAD, supra note 87, at 9.
94. Id. at 36, 750 P.2d at 634.
95. Id. at 45, 750 P.2d at 639.
96. Id. at 48, 750 P.2d at 640. In Washington, the jury’s power of nullification is intact although the defendant is not entitled to an instruction informing the jury of its power. State v. Bonisisio, 92 Wash. App. 783, 794, 964 P.2d 1222, 1229 (1998).
97. Hypothetical created by author.
98. See Muller, supra note 88, at 784.
99. See, e.g., Steckler v. United States, 7 F.2d 59, 60 (2d Cir. 1925).
100. See supra Part I.A.
requirement of unanimity,\textsuperscript{101} they are not reviewable when the verdicts could also be interpreted as the product of jury lenity.\textsuperscript{102}

C. Inconsistent Verdicts Arise in Three Contexts

Inconsistent verdicts can arise whenever the jury is asked to return more than a single verdict. This happens in three contexts. First, an inconsistent verdict can arise when two defendants are tried together, and one of the defendants is acquitted. In the second context, there is only one defendant, but an inconsistency exists on the multiple counts charged against the defendant.\textsuperscript{103} In these two contexts, both federal and Washington jurisprudence dictate that sufficiency of the evidence review is adequate to determine the validity of the verdicts, and that considerations of jury lenity limit further review.\textsuperscript{104} In the final context, there is only one count charged against one defendant, but the inconsistency arises when the general and special verdicts conflict.\textsuperscript{105} Prior to \textit{Goins}, Washington courts reversed the conviction in this context and granted a mistrial.\textsuperscript{106}

1. Multiple-Defendant Inconsistent Verdicts

A defendant cannot challenge his or her conviction on the ground that it is inconsistent with the acquittal of a co-defendant.\textsuperscript{107} In multiple-defendant cases, an inconsistent verdict may arise when the jury acquits one defendant but convicts the other in a case where logic requires that all defendants receive the same verdict.\textsuperscript{108} Appellate review of multiple-defendant inconsistent verdicts is limited to determining whether the conviction was supported by sufficient evidence and resulted from a fair trial.\textsuperscript{109}

\begin{footnotesize}
\begin{enumerate}
\item See United States v. Dotterweich, 320 U.S. 277, 279 (1943) (citing Dunn v. United States, 284 U.S. 390 (1932)) ("Whether the jury’s verdict was the result of carelessness or compromise or [nullification] . . . is immaterial. Juries may indulge in precisely such motives or vagaries.").
\item See infra Parts II.C.2.
\item See infra Parts II.C.I., II.C.2.
\item See infra Part II.C.3.
\item See infra Part II.C.3.
\item See, e.g., \textit{id.}; United States v. Dotterweich, 320 U.S. 277, 279 (1943).
\item Rivera, 454 U.S. at 348.
\end{enumerate}
\end{footnotesize}
In *Harris v. Rivera*, three defendants were each charged with crimes arising from an apartment robbery. The U.S. Supreme Court conceded that either all three defendants should have been acquitted or all three defendants should have been convicted. However, the Court held that a convicted defendant cannot challenge the conviction on the ground that it is inconsistent with the acquittal of a co-defendant. Citing considerations of lenity, the Court held that such a conviction is constitutional as long as it is supported by sufficient evidence and resulted from a fair trial.

2. *Multiple-Count, Single-Defendant Inconsistent Verdicts*

Most jurisprudence on inconsistent verdicts has developed in the multiple-count, single-defendant context. Such inconsistencies can arise between general verdicts on multiple counts, or between the general verdict on one count and a special verdict on another count. There is no federal constitutional right to a consistent verdict in a criminal trial involving inconsistent verdicts for a single defendant charged with multiple counts. Similarly, there is also no such state constitutional right, as Washington courts have followed the U.S. Supreme Court and held that such inconsistent verdicts may have resulted from jury lenity. Therefore, Washington courts have reasoned that inconsistent multiple-count, single-defendant verdicts are not subject to reversal simply because they are inconsistent. Both U.S. Supreme Court and Washington precedents limit review of multiple-count, single-defendant inconsistent verdicts to the sufficiency of the evidence on the count for which the defendant was convicted.

111. Id. at 340.
112. Although this was a non-jury trial, the Court found the reasoning was parallel as between jury and non-jury trials, and that the judge, like the jury, may exercise lenity. Id. at 346–48.
113. Id. at 348.
114. Id. at 344.
117. Powell, 469 U.S. at 69.
118. See, e.g., Ng, 110 Wash. 2d at 48, 750 P.2d at 640.
119. Id.
120. Powell, 469 U.S. at 67; Ng, 110 Wash. 2d at 48, 750 P.2d at 640.
In *United States v. Powell*, the U.S. Supreme Court held that defendants who challenge inconsistent multiple-count verdicts are entitled only to a sufficiency of the evidence review. The inconsistency in *Powell* arose when the jury acquitted the defendant of an underlying felony—conspiracy to possess cocaine with intent to distribute—but convicted her of using a telephone to facilitate that felony. Because proof of possession was an element of the facilitation charge, the defendant’s conviction on facilitation was inconsistent with her acquittal on possession. Although the Court noted that the potentially conflicting results might indicate that the jury failed to express its true conclusions, it reasoned that such results do not necessarily show that the jury did not find guilt with respect to the conviction.

While the Court acknowledged that some error must have occurred to cause the inconsistent verdicts, it reasoned that the inconsistency could favor either the state or the defendant. As the Court explained, although it was possible that the *conviction* was the incorrect result—as the defendant claimed—the *acquittal*, instead, could have been the incorrect result. The jury could have properly reached its verdict of conviction on the facilitation, and then improperly—through mistake, confusion, or lenity—reached its result on the underlying felony. The Court noted that the constitutional guarantee against double jeopardy prohibited the state from challenging an acquittal, and reasoned that it would be “hardly satisfactory” to allow the defendant to challenge the inconsistency when the state could not. The Court held that the possibility of jury lenity, along with the government’s inability to invoke review of the acquittal on the ground that it favored the defendant, made the verdicts unreviewable. Although sufficiency of the evidence review determines only whether the evidence presented at trial could support each conviction when considered independently from the verdicts on other counts, the Court cited the need for finality in criminal

122. *Id.* at 67.
123. *Id.* at 60.
124. *Id.* The government conceded the inconsistency for the purposes of review. *Id.* at 61 n.5.
125. *Id.* at 64–65 (citing Dunn v. United States, 284 U.S. 390 (1932)).
126. *Id.* at 65.
127. *Id.*
128. *Id.* See supra notes 50–56 and accompanying text; see also CONRAD, supra note 87, at 7–9.
130. *Id.* at 66.
matters and concluded that such review was a sufficient safeguard against jury irrationality.131

Washington law similarly requires only that sufficient evidence support a single defendant’s guilty verdict when the inconsistency arises between an acquittal and a conviction because such an inconsistency can be attributed to jury lenity.132 In Ng, for example, an inconsistency arose when the jury convicted the defendant of first-degree robbery on thirteen counts but acquitted him of felony murder on thirteen counts.133 Because the defendant did not dispute that the killings occurred,134 he argued that the guilty verdicts on the robberies required that the jury also find him guilty of the felony murders.135 Therefore, the defendant reasoned that the convictions on first-degree robbery were void because they were inconsistent with the acquittals on felony murder.136 The Washington State Supreme Court, however, upheld the verdicts, holding that concerns with second-guessing a jury’s verdict of acquittal and considerations of jury lenity dictated that the verdicts be affirmed.137 In such cases, the Ng court held that sufficiency of the evidence review is an adequate remedy.138

The Washington State Supreme Court has also held that consistency between multiple convictions is not required as long as sufficient evidence supports each guilty verdict.139 In State v. McNeal,140 the defendant was charged with vehicular homicide and vehicular assault after he caused a fatal car accident.141 A blood test showed evidence of methamphetamine.142 To prove vehicular assault, the state had to show

131. Id. at 67.
133. Id. at 36, 750 P.2d at 634. The court “instructed the jury that it could find Ng guilty of first degree robbery as a lesser included offense to felony murder.” Id. An offense is a lesser-included offense when “each of the elements of the lesser offense [is] a necessary element of the offense charged” and “the evidence in the case . . . support[s] an inference that the lesser crime was committed.” State v. Berlin, 133 Wash. 2d 541, 550, 947 P.2d 700, 704 (1997) (citing State v. Workman, 90 Wash. 2d 433, 447–48, 584 P.2d 382, 385 (1978)) (internal quotations omitted) (emphasis in original).
134. Ng, 110 Wash. 2d at 47–48, 750 P.2d at 640.
135. Id. at 45, 750 P.2d at 639.
136. Id.
137. Id. at 48, 750 P.2d at 640.
138. Id.
140. 145 Wash. 2d 352, 37 P.3d 280 (2002).
141. Id. at 355, 37 P.3d at 282.
142. Id. at 356, 37 P.3d at 282.
that the defendant operated his car while “under the influence of drugs.” The jury was asked by special verdict to find whether the defendant was “operating the motor vehicle . . . while under the influence of drugs.” The jury answered no to the special verdict form, yet convicted the defendant on all counts. In effect, the jury found that the defendant both was and was not under the influence of drugs at the time of the accident. The McNeal court held that where there is an inconsistency between guilty verdicts on two different counts, and the inconsistency arises between general and special verdicts or between two general verdicts, the court will review each conviction independently for sufficiency of the evidence. Therefore, the Washington State Supreme Court in McNeal upheld the vehicular assault conviction because there was sufficient evidence to support it.

Justice Sanders’ dissent in McNeal, however, stressed the distinction between the inconsistent verdicts in McNeal, arising between multiple convictions, and the inconsistent verdicts in Powell and Ng, arising between convictions and acquittals. The dissent argued that jury lenity cannot play a role when the jury convicts on all counts. The McNeal majority disagreed, noting that the defendant would have faced an increased standard sentence had the jury found by special verdict that the defendant was under the influence of drugs. Further, the McNeal majority reasoned that the jury may have understood the jury instructions to mean that it could find guilt of vehicular assault on two separate grounds: either that the defendant was under the influence of drugs, or that the defendant operated a motor vehicle with disregard for the safety of others. The McNeal court reasoned that the jury could have thought that finding guilt on the latter ground implied less culpability than finding that he was under the influence of drugs. Therefore, the court

143. Id. A special finding that the defendant was under the influence of drugs increases the seriousness level of the underlying offense. Id. at 359 n.3, 37 P.3d at 284 n.3.
144. Id.
145. Id. at 359, 37 P.3d at 283–84.
146. Id. at 363, 37 P.3d at 286.
147. Id. at 365, 37 P.3d at 286 (Sanders, J., dissenting).
148. Id. at 366, 37 P.3d at 287 (Sanders, J., dissenting).
149. Id. at 359 n.3, 37 P.3d at 284 n.3.
150. Id. at 361, 37 P.3d at 285.
Inconsistent Verdicts

held that considerations of jury lenity can exist even when the jury
convicts the defendant on all counts.153


Prior to the Washington State Court of Appeals Division I decision in
*Goins*,154 Washington courts reversed the conviction of a single
defendant with a single charge when a jury returned a special verdict
negating an element of the crime charged.155 Single-count, single-
被告 inconsistent verdicts arise where there is an inconsistency
between a general verdict and a special verdict on a single count. In *State
v. Hurley*,156 for example, a guilty robbery verdict was inconsistent with a
special verdict finding that the defendant was not armed with a deadly
weapon.157 The trial court’s robbery instruction stated that to convict the
defendant, the jury had to find that the defendant was armed with a
deadly weapon at the time the robbery was committed.158 Division III of
the Washington State Court of Appeals reasoned that under the trial
court’s instructions, being armed with a deadly weapon was an element
of the robbery charge.159 Thus, the appellate court reversed the
defendant’s robbery conviction because the internal inconsistency
between the general and special verdicts was irreconcilable.160

Similarly, in *State v. Wedner*,161 Division I of the Washington State
Court of Appeals reversed a defendant’s conviction for second-degree
assault because the jury found by special verdict that the defendant was
not armed with a deadly weapon.162 The jury had been instructed that to
convict the defendant of assault, the state had to prove that he was
armed.163 Because of the irreconcilable internal inconsistency between

153. *Id.* at 359 n.3, 37 P.3d at 284 n.3.
154. *See infra* Part III for a discussion of the *Goins* decision.
157. *Id.* at 783, 483 P.2d at 1275.
158. *Id.* at 782, 483 P.2d at 1275. The court noted that this was an inaccurate statement of the law,
but became the law of the case because there was no objection to the instruction. *Id.* at 783, 483 P.2d
at 1275.
159. *Id.* at 783, 483 P.2d at 1275. The court distinguished three prior cases in which being armed
with a deadly weapon was not an element of the crime charged. *Id.*
160. *Id.* at 784, 483 P.2d at 1276.
162. *Id.*
163. *Id.* at 347, 601 P.2d at 951.
the special verdict and the general verdict, the court voided the judgment and granted a new trial.164

Recently, Division II of the Washington State Court of Appeals explicitly questioned the continuing vitality of Hurley and Wedner in State v. Holmes.165 The jury in Holmes convicted the defendant of first-degree robbery, but found by special verdict that he was not armed with a deadly weapon at the time of the crime—an element of robbery.166 The Holmes court upheld the conviction, holding that the verdicts were not irreconcilably inconsistent.167 The court explained that the trial court had given the jury two different definitions of “deadly weapon” which made it possible for the jury to find that the defendant was armed as defined in the robbery instruction, but not armed for the deadly weapon enhancement.168 Although the court cited Hurley and Wedner as contrary authority,169 it did not reconcile them. The court simply noted that Hurley and Wedner were decided when “inconsistent verdicts constituted reversible error,”170 thus questioning their authority and citing Ng for support.

In sum, the U.S. Supreme Court has held that inconsistent verdicts involving two defendants are attributable to considerations of jury lenity and are not reversible purely because they are inconsistent. In the multiple-count, single-defendant context, Washington law parallels federal law and inconsistent verdicts are upheld out of concerns of jury lenity. Washington courts will uphold convictions in both of these situations if the convictions are supported by sufficient evidence—the same level of review available for all criminal defendants. Prior to Goins, however, a single defendant convicted inconsistently on a single count was granted a mistrial.

164. Id. at 347, 601 P.2d at 950.
166. Id. at 778, 24 P.3d at 1120.
167. Id. at 780, 24 P.3d at 1121.
168. Id.
169. Id. at 780–81, 24 P.3d at 1121–22.
170. Id. at 781 n.2, 24 P.3d at 1122 n.2.
III. IN STATE V. GOINS, DIVISION I OF THE WASHINGTON STATE COURT OF APPEALS UPHeld IRRECONCILABLY INCONSISTENT VERDICTS IN THE SINGLE-COUNT, SINGLE-DEFENDANT CONTEXT

Matthew Goins and his friend Steve Haworth went out drinking in May, 2000 and visited the victim’s (Z) apartment. Haworth went into the bathroom and passed out. Z testified that Goins approached her and tried to kiss her twice. She said he retreated the first time when she pushed him away, but the second time he forced her into the bedroom, pinned her on the bed, and tried to lift up her shirt and touch her. She fought back and screamed. Her screams roused Haworth who entered the bedroom, at which point Goins fled. Later, Goins admitted that he tried to kiss Z, and that he might have touched her shoulder. But, he said that Z unexpectedly punched him and he had to grab her to stop her attack. Although he admitted to pushing her on the bed, he said it was an accident.

The state charged Goins with one count of second-degree assault with intent to commit indecent liberties. In addition, the state charged that Goins committed the crime with sexual motivation. The trial court instructed the jury that to convict Goins, it must find beyond a reasonable doubt “[t]hat the assault was committed with intent to commit Indecent Liberties.” The court defined indecent liberties as occurring when a person “knowingly causes another person who is not his spouse to have sexual contact with him or another by forcible compulsion.” “Sexual contact” was defined as “any touching of the sexual or other intimate...

---

172. Id. at 726, 54 P.3d at 724–25.
173. Id. at 726, 54 P.3d at 724.
174. Id. at 726, 54 P.3d at 724–25.
175. Id. at 726, 54 P.3d at 725.
176. Id. at 726–27, 54 P.3d at 725.
177. Id. at 727, 54 P.3d at 725.
178. Id.
179. Id.
180. Id.
181. Id. at 727–28, 54 P.3d at 725.
182. Id. at 728, 54 P.3d at 725.
183. Id.
parts of a person done for the purpose of gratifying sexual desire of either party.184

The trial court judge also gave the jury a special verdict form that asked if the crime was committed with sexual motivation, pursuant to Revised Code of Washington section 9.94A.835.185 The court instructed the jury that “[s]exual motivation means that one of the purposes for which the defendant committed the crime was for the purpose of his or her sexual gratification.”186 Although the jury convicted Goins of second-degree assault with intent to commit indecent liberties, it acquitted him of committing the crime with sexual motivation.187

Division I of the Washington State Court of Appeals noted that “the verdicts [were] irreconcilably inconsistent because Goins either committed the assault for the purpose of sexual gratification or he did not.”188 Enumerating the three contexts in which inconsistent verdicts can arise involving a single defendant,189 the Goins court identified the case as one where a guilty verdict on one charge is inconsistent with a special verdict on the same charge.190 Thus, Goins argued that the Hurley and Wedner cases governed the facts of his case, mandating a mistrial.191 The court acknowledged the Hurley and Wedner cases and noted that “[u]ntil quite recently, it appeared to be settled law in Washington that, where the jury returns a general verdict of guilt accompanied by an irreconcilably inconsistent special interrogatory on the same charge, the remedy is to declare the inconsistent verdicts void and remand for a new trial.”192

The court reasoned, however, that the Washington State Supreme Court’s decisions in Ng and McNeal demonstrated an adoption of the U.S. Supreme Court’s decision in Powell, expanding judicial acceptance

184. Id.
185. Id. See also supra Part I.B.
187. Id. at 729, 54 P.3d at 726.
188. Id. at 730, 54 P.3d at 726.
189. Id. at 730–32, 54 P.3d at 726–27.
190. Id. at 732, 54 P.3d at 727.
191. Id. at 736, 54 P.3d at 729. Goins also claimed that Revised Code of Washington Section 4.44.440 mandated that the special verdict controlled and therefore that his conviction was void. Id. The court found that Revised Code of Washington section 4.44.440 does not apply to the criminal context. Id. See infra Part IV for a summary of the Goins court’s response to this argument. Finally, Goins claimed that his counsel was ineffective in not preserving his right to appeal the inconsistent verdicts. Goins, 113 Wash. App. at 743, 54 P.3d at 733. The court dismissed this argument by showing that Goins’ counsel might have strategically refrained from objecting to the inconsistent verdicts. Id. at 744, 54 P.3d at 733.
of inconsistent verdicts. The Goins court suggested that the Hurley and Wedner decisions may have been implicitly overruled by this acceptance. With Hurley and Wedner called into question, the Goins court extended the Ng and McNeal decisions, holding that in the single-defendant context “where the inconsistency is between the general verdict and a special verdict relating to that same count, the general verdict will stand if it is supported by substantial evidence.”

Although Goins argued that the McNeal decision applied only to multiple-count inconsistent verdicts, the court disagreed and reasoned that the McNeal decision extended the Ng rule to inconsistencies between a special and a general verdict in any context. The Goins court’s decision to extend the doctrine was based on several earlier cases, including Peerson. The Goins court cited the McNeal court’s reliance on Peerson for the proposition that sufficiency of the evidence review was applicable to inconsistencies between special and general verdicts. The Goins court noted that on appeal, the prosecution argued that Peerson supported the proposition that inconsistent verdicts in any context must stand when sufficient evidence supports the guilty verdict. Without passing judgment on the prosecution’s argument, the Goins court recognized that the Hurley and Wedner decisions were based on the concern that a special verdict that was inconsistent with a general verdict on a single count cast doubt on whether the state had proved its case beyond a reasonable doubt. But, ultimately, the Goins court concluded that voiding Goins’ conviction in light of the decision to uphold inconsistent verdicts in Ng, McNeal, and Peerson would be “to elevate form over substance.” Consequently, the court applied the

---

193. Id. at 734, 54 P.3d at 728.
194. Id. at 735, 54 P.3d at 729. See supra notes 165–170 and accompanying text.
195. Goins, 113 Wash. App. at 742, 54 P.3d at 733. Although the Goins court used the terms “substantial evidence” rather than “sufficient evidence,” it seems to have used “substantial” in the same way as the Ng and McNeal courts used “sufficient.” The Goins court did not distinguish between the two terms, or note the discrepancy.
196. Id. at 736, 54 P.3d at 729.
197. Id. at 741–42, 54 P.3d at 732.
200. Id.
201. Id. at 740–41, 54 P.3d at 732.
202. Id. at 742, 54 P.3d at 732.
203. Id. at 742, 54 P.3d at 732–33.
same standard of review and held that there was substantial evidence to support the guilty verdict.\textsuperscript{204}

Judge Ellington dissented to the extension of \textit{Ng}, \textit{McNeal}, and \textit{Peerson} to the context of single-count, single-defendant inconsistent verdicts.\textsuperscript{205} Citing the policy and language of \textit{Powell}, she agreed with the majority that where two general verdicts are inconsistent, considering the sufficiency of evidence is a valid and adequate remedy.\textsuperscript{206} However, she argued that “lenity offers no explanation for an inconsistent special verdict on the same count.”\textsuperscript{207} In this context, she added, “the integrity of the general verdict is called into question.”\textsuperscript{208} Thus, Judge Ellington would not apply a rule that arose from inconsistencies in different counts to cases where the inconsistency was within a single count.\textsuperscript{209} Thus, Judge Ellington recommended remanding the conviction for retrial.\textsuperscript{210}

IV. \textsc{Washington’s Criminal Court Rules and Civil Code Fail to Resolve Irreconcilably Inconsistent Criminal Verdicts}

Washington has a criminal court rule acknowledging the issue of inconsistencies between special and general verdicts in criminal trials.\textsuperscript{211} However, this rule does not provide a solution in the case of irreconcilably inconsistent verdicts because it only allows the judge to ask the jury to deliberate further.\textsuperscript{212} In addition, the Washington statute giving preference to special verdicts over the general verdicts in civil trials\textsuperscript{213} fails to resolve the issue in criminal trials.\textsuperscript{214} Consequently, Washington’s court rules and code leave the issue of irreconcilably inconsistent verdicts in criminal trials unresolved.

Washington Superior Court Criminal Rule 6.16(b), promulgated by the Washington State Supreme Court, acknowledges that inconsistencies between general and special criminal verdicts exist. The rule provides

\begin{itemize}
\item \textsuperscript{204} \textit{Id.} at 743, 54 P.3d at 733.
\item \textsuperscript{205} \textit{Id.} at 745, 54 P.3d at 734 (Ellington, J., dissenting).
\item \textsuperscript{206} \textit{Id.}
\item \textsuperscript{207} \textit{Id.}
\item \textsuperscript{208} \textit{Id.}
\item \textsuperscript{209} \textit{Id.}
\item \textsuperscript{210} \textit{Id.}
\item \textsuperscript{211} \textsc{Wash. Sup. Ct. Crim. R. 6.16(b)}.
\item \textsuperscript{212} \textit{Id.}
\item \textsuperscript{213} \textsc{Wash. Rev. Code § 4.44.440 (2002)}.
\item \textsuperscript{214} \textit{Goins}, 113 Wash. App. at 740, 54 P.3d at 732.
\end{itemize}
that “[w]hen a special finding is inconsistent with another special finding or with the general verdict, the court may order the jury to retire for further consideration. However, the rule fails to identify appropriate judicial action if the jury is unable to reconcile the two verdicts. Thus, the court rules fail to resolve the issue of irreconcilably inconsistent criminal verdicts.

The Revised Code of Washington section 4.44.440 (section 4.44.440) provides a statutory remedy for inconsistent civil verdicts in Washington. Under section 4.44.440, “[w]hen a special finding of facts shall be inconsistent with the general verdict, the former shall control the latter, and the court shall give judgment accordingly.” Consequently, in civil cases involving inconsistent verdicts, Washington judges are instructed to enforce the jury’s special verdict and not the general verdict. Although section 4.44.440 is located in the title on civil procedure, if it was applied to a criminal case, it would render the general verdict void when the general verdict directly contradicted the special verdict.

The Goins court held that section 4.44.440 does not apply in the criminal context. Until Goins, Washington courts had not determined whether section 4.44.440 applied to criminal proceedings. Although two appellate courts had been confronted with the issue, neither had ruled on the issue. The Goins court conceded that if section 4.44.440 applied the defendant’s conviction would be overturned. Then, the court reasoned that section 4.44.440 did not apply because it is in the civil procedure title of the Revised Code of Washington, and no case law had held that it applies in the criminal context. The court further noted that under prior Washington law, the remedy for single-count, single-defendant inconsistent verdicts was a mistrial, not an acquittal, as section 4.44.440 would direct. Therefore, the Goins court held that section 4.44.440 does not apply to criminal proceedings.

215. WASH. SUP. CT. CRIM. R. 6.16(b).
216. WASH. REV. CODE § 4.44.440.
220. Id. at 736, 54 P.3d at 730.
221. Id.
222. Id. at 739, 54 P.3d at 731.
223. Id. at 740, 54 P.3d at 732.
In sum, while the Washington State Legislature has addressed the inconsistent verdict issue in the civil context in section 4.44.440, indicating that the special verdict controls over the general verdict when the two are inconsistent, it has not specifically addressed the issue in the criminal context. Further, the Goins court has held that section 4.44.440 does not apply to criminal trials. The Washington criminal court rule allowing the court to send the jury back for further deliberation does not clarify the appropriate judicial action if the jury remains unable to return a consistent verdict. Thus, neither section 4.44.440 nor the Washington court rules resolve the issue of irreconcilably inconsistent criminal verdicts.

V. THE GOINS DECISION IS FLAWED BECAUSE THE COURT FAILED TO FOLLOW PRECEDENT AND EXTENDED INAPPROPRIATE CASE LAW BEYOND ITS CONSTITUTIONAL REACH

The majority decision in Goins is flawed for three reasons. First, the court did not follow the Hurley and Wedner cases, which were directly on point and binding. Second, the court extended Ng and McNeal to the single-count, single-defendant context, where these cases do not apply. Third, by extending Ng and McNeal to this context, the court reached a conclusion that violates the Due Process clause of the Fourteenth Amendment by allowing a conviction to stand when an element of the charged crime has not been proven beyond a reasonable doubt.

If the court had properly followed Hurley and Wedner, it would have reached a constitutional result. The Washington State Supreme Court should reverse the Goins decision and apply Hurley and Wedner to the single-count, single-defendant inconsistent verdict context. The court should also modify its existing criminal court rule to provide a remedy when the jury provides an irreconcilably inconsistent verdict in the single-count, single-defendant context. If the court fails to provide a remedy, the legislature should act to prevent this unconstitutional result from recurring. The legislature should adopt a criminal statute specifically applying to single-count, single-defendant inconsistent

---

224. See infra Part V.A.
225. See infra Part V.B.
226. See infra Part V.C.
Inconsistent Verdicts

verdicts mandating at least a mistrial, and at most an acquittal, if the inconsistency cannot be resolved.

A. The Goins Court Should Have Applied Hurley and Wedner, Which Were Directly on Point and Had Not Been Overruled

The Hurley and Wedner decisions are directly on point, and mandated a mistrial in Goins. All three cases involved an inconsistent special verdict and general verdict on the same count. In Hurley, the special verdict finding that the defendant was not armed negated an element of robbery.227 Similarly, in Wedner, the special verdict finding that the defendant was not armed negated an element of second-degree assault. 228 In Goins, the special verdict finding that there was no sexual motivation negated an element of second-degree assault with intent to commit indecent liberties.229 Thus, in all three cases the special verdict was irreconcilably inconsistent with the guilty general verdict because the special verdict negated one of the elements of the crime. The trial courts’ resolution in Hurley and Wedner was to grant each defendant a mistrial.230 Because the Hurley and Wedner decisions are directly on point, the Goins court should also have granted the defendant a mistrial.

At the time Division I of the Washington State Court of Appeals heard Goins, the Hurley and Wedner decisions had not been overturned. The Goins appellate court conceded that “until recently,” the Hurley and Wedner decisions controlled where the jury returned a guilty general verdict with an irreconcilably inconsistent special verdict on the same charge, directing a new trial.231 The court reasoned that the Hurley and Wedner decisions were called into doubt by recent cases, citing Ng, McNeal, State v. Barnes,232 Burke, and Holmes. 233 However, none of these cases arose in the same context as the Hurley or Wedner cases. The Ng case involved an inconsistency between an acquittal and a conviction,234 not an internal inconsistency between a general and special

227. See supra Part II.C.3.
228. See supra Part II.C.3.
verdict on the same count. The McNeal case involved an inconsistency between a conviction and an inconsistent special verdict on a different count\textsuperscript{235}—again, not an inconsistency between verdicts on the same count. The Barnes court addressed the procedural right to raise the issue of an inconsistent verdict on appeal if no objection was made at trial, not the substantive right to a new trial in the face of an internally inconsistent verdict.\textsuperscript{236} The Burke court also addressed this procedural issue, then proceeded to reconcile the verdicts while approvingly citing Wedner.\textsuperscript{237} Finally, the Holmes court was also able to reconcile the potentially inconsistent verdicts because of different ways the jury instructions could be interpreted.\textsuperscript{238} Although the Holmes court went on to question the continuing vitality of the Hurley and Wedner cases, it questioned the cases based on the Ng decision.\textsuperscript{239} But, Ng did not involve an internal inconsistency between verdicts on the same count,\textsuperscript{240} and therefore did not arise in the same context as the Hurley or Wedner cases.

In sum, the Hurley and Wedner decisions are directly on point in the single-count, single-defendant Goins context. If the appellate court had correctly applied Hurley and Wedner, it would have reversed Goins’ conviction. At the time of the Goins decision, the rule pronounced by Hurley and Wedner had not been overruled. Although the Goins court noted that some doubt existed as to the Hurley and Wedner decisions’ continuing authority, none of the cases it cited implicitly or explicitly overruled either the Hurley or the Wedner decision. Therefore, the Hurley and Wedner decisions controlled in the Goins single-count, single-defendant inconsistent verdict context, and the Goins court erred in not adhering to precedent.

B. The Goins Court Erred by Extending Inapposite Case Law

There is a fundamental difference between irreconcilable inconsistencies between a general and special verdict in the single-count, single-defendant context, such as in Goins, and inconsistencies in any other context. In the single-count, single-defendant context, a special verdict can be irreconcilably inconsistent with the general verdict only

\textsuperscript{235} State v. McNeal, 145 Wash. 2d 352, 357, 37 P.3d 280, 282 (2002).
\textsuperscript{236} Barnes, 85 Wash. App. at 668, 932 P.2d at 686.
\textsuperscript{239} Id. at 781 n.2, 24 P.3d at 1122 n.2.
\textsuperscript{240} State v. Ng, 110 Wash. 2d 32, 48, 750 P.2d 632, 640 (1988).
Inconsistent Verdicts

when the special verdict finding negates an element of the crime asserted by the general verdict. In any other inconsistent verdict context, such as inconsistencies between verdicts on different counts or different defendants, the existence of an inconsistent verdict does not per se negate an element of the crime charged.

The Goins court’s reliance on Ng and McNeal as precedent to justify acceptance of inconsistencies in the single-count, single-defendant context is misplaced because the Ng and McNeal decisions applied to different contexts. The Ng decision upheld an inconsistency between multiple convictions and acquittals on multiple counts. The McNeal court upheld an inconsistency between a special verdict on one count and a conviction on another count. In both of these cases, the inconsistencies were between verdicts on multiple counts. In contrast, the Goins case involved internally inconsistent verdicts on a single count.

Neither the Ng nor the McNeal decision involved an internal inconsistency on the same count. Because of the fundamental difference between internally inconsistent verdicts and inconsistent verdicts in any other contexts, neither the Ng nor the McNeal decision should control the outcome in the single-count, single-defendant inconsistent verdict context.

The mere existence of a logical inconsistency in all three cases does not justify upholding Goins’ conviction. The Goins court conceded that Goins logically could not have committed the assault both with and without sexual motivation, but noted that logically similar inconsistencies existed in Ng and McNeal. The court concluded that because all three cases contained logical inconsistencies, the resolution of each should be the same.

However, the rationale behind Ng and McNeal does not transfer to the Goins context because the Ng and McNeal cases, unlike the Goins case, involved inconsistent verdicts on more than one count. In Goins, an inconsistent special verdict negated

241. Id.
244. Id. at 730, 54 P.3d at 726.
245. As the Goins court noted, in McNeal, the defendant could not simultaneously have been and not been under the influence of drugs, and in Ng, the defendant could not simultaneously have been guilty of armed robbery and not guilty of felony-murder. Id. at 742, 54 P.3d at 732.
246. Id. at 742, 54 P.3d at 732–33.
247. State v. Ng, 110 Wash. 2d 32, 45, 750 P.2d 632, 639 (1988); McNeal, 145 Wash. 2d at 357, 37 P.3d at 282.
an element of the same crime on the same count on which the general verdict found the defendant guilty.248 Neither the Ng nor the McNeal court dealt with this context.

Further, concerns of jury lenity do not justify extending the Ng and McNeal holdings to the single-count, single-defendant context. In Ng, the court found that because the verdicts were a mixture of convictions and acquittals, jury lenity might have played an important role.249 In McNeal, the court found a possible showing of jury lenity in the inconsistent verdicts, noting first that a consistent special verdict would have increased the defendant’s standard sentence range,250 and also that the jury might have tried to find the defendant guilty on a basis that they thought would imply less culpability.251 In a footnote to its opinion, the Goins court commented that had the jury specially found that Goins had a sexual motivation, he would have been subject to sexual offender registration requirements.252 Thus, it is possible to argue that the jury was exercising lenity by refusing to specially find sexual motivation. However, this is unlikely because when the jury specifically asked the court about the purpose of the special verdict form, the court gave no explanation.253 Therefore, the jury was most likely unaware of the consequences of the special finding. In both Ng and McNeal, the record did not indicate whether the jury was aware of the punitive consequences of its findings. Because the record was silent, the court could reasonably assume that the Ng and McNeal juries were aware of the punitive consequences, and could therefore attribute its inconsistent findings to considerations of jury lenity. But, when a court like the one in Goins knows that the jury was probably unaware that its verdict might mitigate the defendant’s punishment, it would be disingenuous to resolve the inconsistency by assuming it was caused by jury lenity.

Finally, the Peerson case does not support the proposition that inconsistencies in the single-count context should be treated in the same manner as inconsistencies in the multiple-count context. The Goins court noted that the Peerson decision provided an example of a case upholding inconsistencies between a special and a general verdict.254 But the

249. Ng, 110 Wash. 2d at 48, 750 P.2d at 640.
250. McNeal, 145 Wash. 2d at 359 n.3, 37 P.3d at 284 n.3.
251. Id. at 361, 37 P.3d at 285.
253. Id. at 729, 54 P.3d at 726.
254. Id. at 741–42, 54 P.3d at 732 (citing McNeal, 145 Wash. 2d at 359, 37 P.3d at 283–84).
comparison to the Peerson decision is inapposite. The Peerson court actually reconciled the verdicts through liberal construction.\textsuperscript{255} Further, the alleged inconsistency in Peerson arose between a special verdict on one count and the general verdict on a second count.\textsuperscript{256} Therefore, the Peerson case neither presented irreconcilably inconsistent verdicts, as did the Goins case, nor considered the single-count context, as did the Goins case.

In sum, the Ng and McNeal cases present a rule that is inapplicable to the single-count, single-defendant context of Goins. The multiple-count context in which the Ng and McNeal decisions arose is fundamentally different from the single-count context. The mere existence of a logical inconsistency between the verdicts in all three cases does not justify extending the rationale of the Ng and McNeal cases to the single-count context. Furthermore, considerations of jury lenity underpinning the rationale in both Ng and McNeal are not present in Goins. The Goins court’s reliance on Peerson is also misplaced because that case fails to provide any support for the contention that inconsistent verdicts should be treated alike regardless of the context in which they arise. Therefore, there is no basis on which to extend the rationale and rule of Ng or McNeal to Goins.

C. The Goins Decision Unconstitutionally Allows a Conviction to Stand When All the Elements of a Crime Have Not Been Proven Beyond a Reasonable Doubt

The state failed to meet its constitutionally mandated burden of proof beyond a reasonable doubt in Goins. As the U.S. Supreme Court held in Winship, the state must prove every element of a crime charged beyond a reasonable doubt.\textsuperscript{257} In Goins, one of the elements to be proven was that the assault was committed with intent to commit indecent liberties.\textsuperscript{258} “Indecent liberties” is committed when one person forces another to have sexual contact.\textsuperscript{259} Intent to commit indecent liberties, therefore, is an intent to force such sexual contact. As the Goins majority conceded, this is a sexually motivated crime.\textsuperscript{260} Thus, sexual motivation is an element of

\begin{itemize}
  \item \textsuperscript{256} Id. at 765, 816 P.2d at 50.
  \item \textsuperscript{257} In re Winship, 397 U.S. 358, 364 (1970).
  \item \textsuperscript{258} Goins, 113 Wash. App. at 728, 54 P.3d at 725.
  \item \textsuperscript{259} Id.
  \item \textsuperscript{260} Id. at 729–30, 54 P.3d at 726.
\end{itemize}
the assault charge in the Goins case. Yet, when the jury in Goins was asked whether the defendant committed the assault with sexual motivation, it answered that he did not. Therefore, not only did the state not prove that the defendant was sexually motivated, but the jury explicitly found that the state had failed to meet its burden of proof. Despite this finding, the jury convicted Goins of the assault. Under Winship, Goins’ conviction was void, because a finding of sexual motivation was necessary to prove the crime charged.

Permitting Goins’ conviction to stand violates the constitutional requirement that the state must prove every element of the crime charged beyond a reasonable doubt. The Goins decision extends the acceptance of inconsistent verdicts beyond its constitutionally permissible reach. Before Goins, Washington courts reversed convictions when the jury in effect informed the court, through an inconsistent special verdict on the same crime on which it convicted the defendant, that it did not believe that the state had proven all elements of the crime charged beyond a reasonable doubt. In fact, several earlier cases upholding inconsistent verdicts in other contexts specifically noted that the inconsistent finding was not an element of the offense charged. In Goins, however, the court upheld a jury verdict despite the jury’s effective statement that the state had not met its burden of proof. Once it is clear, as it was in Goins, that the standard of proof has not been met, conviction is unconstitutional. Although easier to apply, sufficiency of the evidence review standard is inapposite when the state has failed to prove every element of the crime charged beyond a reasonable doubt.

The constitutional guarantee of proof beyond a reasonable doubt outweighs policy justifications for jury lenity. The U.S. Supreme Court

261. Id. at 728, 54 P.3d at 725.
262. Id. at 729, 54 P.3d at 726.
263. Id.
265. There is arguably cause to be concerned about the collective reasoning of the U.S. Supreme Court and Washington State Supreme Court on the topic of inconsistent verdicts in general. See Muller, supra note 88, at 794–820 (discussing the inadequacy of the current response to inconsistent verdicts).
Inconsistent Verdicts

In *Powell* upheld jury lenity as an exercise of the jury’s power to check the exercise of the executive power.\(^{269}\) Both the *Ng* and *McNeal* decisions relied on the reasoning of *Powell* to support the strong policy favoring the exercise of jury lenity.\(^{270}\) However, while the jury can refuse to convict a defendant *despite* the production of proof beyond a reasonable doubt that the defendant is guilty by exercising lenity, allowing the jury to convict when the state has *failed* to meet its burden of proof violates due process. Because the *Goins* jury could not agree beyond a reasonable doubt that the defendant was sexually motivated, its conviction is unconstitutional. Internally inconsistent verdicts arising in the single-count, single-defendant context are distinct from those in other contexts. The jury cannot constitutionally convict a defendant in spite of the state’s failure to prove guilt beyond a reasonable doubt—to allow such convictions in the name of jury lenity is nonsensical. No amount of respect for the jury’s power to be lenient justifies depriving the defendant of liberty by convicting the defendant on proof that is less than what is constitutionally required.

D. Either the Washington State Supreme Court or the Washington State Legislature Should Provide a Remedy to Address the Concerns Raised by the *Goins* Decision

The Washington State Supreme Court should strike down the *Goins* decision and reaffirm the *Hurley* and *Wedner* decisions by granting the defendant a mistrial. The court should do this because the *Goins* decision violated the defendant’s due process right to be convicted by proof beyond a reasonable doubt. However, it could be further argued that in cases such as *Goins*, allowing a mistrial where the jury has exposed the state’s failure to meet its burden of proof could violate the Double Jeopardy clause which prevents the state from retrying the defendant for the same crime. If this is true, then the remedy would be acquittal rather than mistrial. Prior to the *Goins* decision, Washington law directed a mistrial, not an acquittal, as the remedy for single-count, single-defendant inconsistent verdicts.\(^{271}\) The *Hurley* court addressed this issue directly: “we do not find [the] defendant entitled to a judgment

---

notwithstanding the verdict [an acquittal]. The verdict was void . . . . The posture of the case was the same as if the jury had returned no verdict at all and a mistrial had been declared. Hence, the judgment and sentence was a nullity.** In overruling the Goins decision, the court should affirm the Hurley and Wedner cases, and grant the defendant at the least a mistrial, and at the most an acquittal.

Further, the Washington State Supreme Court should modify the criminal court rules to provide at least a mistrial in the event that the jury cannot provide consistent verdicts in the single-count, single-defendant context. The court should do this in order to clarify the law and prevent recurrence of the due process violation found in the Goins decision. A modified criminal rule could read in relevant part (with additions underlined):

When a special verdict is irreconcilably inconsistent with another special verdict or with the general verdict, the court may order the jury to retire for further consideration. If the jury is unable to return a consistent verdict in the single-count, single-defendant context, the court shall declare a mistrial [or enter a judgment of acquittal].

Adopting such a court rule would provide trial courts with guidance and preserve defendants’ constitutional rights.

If the Washington State Supreme Court fails to provide a remedy for this violation, the Washington State Legislature should promulgate a statute to correct this unconstitutional result. Although Revised Code of Washington section 4.44.440 is inapplicable to the criminal context, the statute provides a good framework for a parallel criminal statute. The proposed criminal statute should have language with an effect similar to that of the proposed criminal court rule.

VI. CONCLUSION

Division I of the Washington State Court of Appeals wrongly upheld the conviction in Goins. As a result, the defendant in Goins was denied his liberty even though the state failed to meet its constitutionally required burden of proof beyond a reasonable doubt. The Goins court overturned precedent that would have granted him a mistrial and instead applied inapposite case law involving inconsistent verdicts in different contexts under the justification of jury leniency. However, jury leniency

273. See supra Part IV.
Inconsistent Verdicts

considerations are absent in the Goins case, and even if they existed they would be trumped by the defendant’s due process rights. The Washington State Supreme Court should reverse the decision in Goins. Further, the court should modify the criminal court rule to reflect a remedy of at least a mistrial and at most an acquittal in the single-count, single-defendant context. If the court fails to reverse the Goins decision and modify the criminal court rules to uphold the defendant’s constitutional rights, then the Washington State Legislature should intervene and promulgate a statute to that end.
WHO OWNS “THE LAW”? THE EFFECT ON COPYRIGHTS WHEN PRIVATELY-AUTHORED WORKS ARE ADOPTED OR ENACTED BY REFERENCE INTO LAW

Katie M. Colendich

Abstract: “The law,” including judicial opinions and statutes, is not copyrightable because neither individuals nor organizations own the law. This longstanding principle is supported by the public’s due process right to access the law. The United States Supreme Court has never determined the status of a private organization’s copyright on model codes or standards when a legislature adopts those materials into law. Federal courts have taken several different approaches to resolving this issue; however, their decisions are in direct conflict with each other. The Second and Ninth Circuits permit private authors to retain copyrights of materials subsequently enacted into law, while the Fifth Circuit does not. This Comment argues that the Fifth Circuit’s decision in Veeck v. Southern Building Code Congress International, Inc., created an unsupported exception to copyright law when it held that private organizations whose works are passed into law cannot retain their copyrights. Further, this Comment argues that the U.S. Supreme Court should resolve the current circuit split in favor of enforcing copyright to ensure that privately authored materials’ copyrights remain enforceable across the nation.

Land Management Group (LMG), a coalition of storeowners, business managers, and contractors, spent the past two years developing a comprehensive collection of sample zoning regulations for city growth and development. LMG promoted its plan in several states, leading two counties in Louisiana to adopt the zoning regulations in full, and three counties in Arizona to adopt the regulations by reference. LMG is a non-profit organization whose annual budget derives from sales of its sample regulations.

Norman Ferland, an entrepreneur in Louisiana, inquired into local zoning ordinances before drawing plans for a new restaurant. Ferland had difficulty locating the zoning regulations at the county office so he bought an electronic copy directly from LMG for $95.00. His copy included a software licensing agreement and a copyright notice stating that users may not copy or distribute the zoning regulations. Ferland copied the regulations onto his website without crediting LMG as the author. If LMG challenged Ferland’s actions as a violation of its

copyright, federal courts in Ferland’s state would likely find that LMG’s copyright became invalid when the county adopted its code into law.2

Meanwhile, Susan Rowe, the owner of a chain of spas in Arizona, wished to open two new locations. Like Ferland, she checked into the zoning ordinances of several counties but had difficulty obtaining copies, so she purchased electronic versions directly from LMG as well. Rowe also posted the zoning regulations on a website. If LMG challenged Rowe’s actions, federal courts in Rowe’s state would likely uphold LMG’s copyright and find that Rowe’s actions constitute copyright infringement.3

The unequal treatment of the same copyright in these two situations illustrates the problem that can arise when courts across the country disagree about the status of a privately held copyright once the copyrighted material is incorporated into law. The executive branch and Congress have supported government use of model codes written by private organizations.4 However, the U.S. Supreme Court has not provided guidance on how lower courts should treat copyrights on such privately authored material. The U.S. Court of Appeals for the Second and Ninth Circuits have addressed the issue and agreed that privately authored materials retain their copyright even after a government adopts them, whether in full or by reference, into the law.5 However, the Fifth Circuit took a different approach in Veeck v. Southern Building Code Congress International, Inc.,6 when it held that privately authored works automatically lose their copyrights when they are incorporated into law.7

This Comment argues that the Fifth Circuit’s Veeck decision created an unsupported exception to copyright law. The Veeck court misconstrued persuasive authority from other circuits and ignored strong support for maintaining the copyright protection of privately authored model codes and standards. This Comment argues that the U.S. Supreme Court should reject this new exception by reversing the Fifth Circuit’s

2. See Veeck, 293 F.3d at 791.
5. Practice Mgmt., 121 F.3d at 520; CCC Info. Servs., Inc. v. Maclean Hunter Mkt. Reports, Inc., 44 F.3d 61, 74 (2d Cir. 1994).
6. 293 F.3d 791 (5th Cir. 2002).
7. The merger doctrine was also part of the Veeck court’s rationale in holding that privately authored materials lose their copyright. Veeck, 293 F.3d at 800–02. This issue is beyond the scope of this Comment.
decision in *Veeck* to ensure that copyrights of privately authored laws receive uniform, nationwide protection.8

Part I of this Comment introduces the historical development of copyright law and its application to judicial opinions and statutes. Part II traces the recent development and subsequent application of the federal courts of appeals tests for assessing whether a private author’s copyright becomes invalid when a government enacts the author’s work into law. Part III argues that the Fifth Circuit’s *Veeck* decision created an unsupported exception to copyright law, contradicting decisions from the Second and Ninth Circuits as well as policies of the legislative and executive branches. Further, Part III suggests that instead of adopting a per se rule, the Fifth Circuit should have followed the Second Circuit’s lead by using the doctrine of fair use to balance the private authors’ copyright interests with the public’s due process right to access the law. Finally, Part IV concludes by recommending that the U.S. Supreme Court should resolve this circuit split by reversing the *Veeck* decision and ensure that copyrights of privately authored laws receive uniform, nationwide protection.

I. HISTORICALLY, “THE LAW” WAS TREATED AS PART OF THE PUBLIC DOMAIN, BUT MODERN DEVELOPMENTS HAVE CHALLENGED THIS APPROACH

Copyright law has origins in both the Constitution and federal statutes.9 Until recently, it was well settled that no author could copyright the text of “the law” in public domain.10 This principle, however, has become less clear as government agencies have adopted the work of private authors and incorporated it into law. It is unclear whether private authors retain their copyright in materials once they are adopted into law.11 Further, it is unclear whether state and local governments are also

---

8. The implications of holding that privately authored material is protected under copyright law are unclear. It is unclear whether a private organization such as the Southern Building Code Congress International (SBCCI) could seek royalties from a state legislative body. Therefore, it is beyond the scope of this Comment.
bound by the Copyright Act. It is possible that courts will rely on the modern “fair use” doctrine to resolve these types of copyright disputes.

A. Copyright Protections Originate in the Constitution and Title 17 of the United States Code

The current American copyright system flows directly from the Copyright Clause of the U.S. Constitution. The Copyright Clause gives Congress the authority “[t]o promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writing and Discoveries.” Since the ratification of the Constitution, Congress has protected the rights of copyright holders by enacting various statutory frameworks. Each statutory enactment has attempted to protect the core purpose of copyright law—“to secure a fair return for an author’s creative labor”—because stimulating artistic creativity ultimately benefits the public with the production of valuable work. The U.S. Supreme Court has stated “[a] copyright, like a patent, is ‘at once the equivalent given by the public for benefits bestowed by . . . the skill of individuals, and the incentive to further efforts for the same important objectives.’” Creating incentives for artists, writers, and scientists has remained a dominant theme of copyright law even though subsequent legislation has altered the scope of the protection.

Copyright is a matter of federal law governed by Congressional statutes passed under the authority of the Copyright Clause. Under the

12. See County of Suffolk v. First Am. Real Estate Solutions, 261 F.3d 179, 187 (2d Cir. 2001).
13. See, e.g., qad, Inc. v. ALN Assocs. Inc., 974 F.2d 834, 835 (7th Cir. 1992) (declaring that basis for American copyright system is Article I, section 8, clause 8 of the U.S. Constitution); accord Morely Music Co. v. Cafe Cont’l Inc., 777 F. Supp. 1579, 1582 (S.D. Fla. 1991); see also Karl Fenning, The Origin of the Patent and Copyright Clause of the Constitution, 17 GEO. L.J. 109, 109 (1929) (stating that United States copyright laws are based on Article I, section 8, clause 8 even though clause does not employ the word “copyright”).
15. Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 460 (1984) (“This Nation’s initial copyright statute was passed by the First Congress.”).
16. Twentieth Century Music Corp. v. Aiken, 422 U.S. 151, 156 (1975); accord Fox Film Corp. v. Doyal, 286 U.S. 123, 127 (1932) (The “sole interest of the United States and the primary object in conferring the monopoly lie in the general benefits derived by the public from the labors of authors”).
Copyright Act of 1976, copyright protection exists in “original works of authorship”\(^\text{20}\) and commences upon creation\(^\text{21}\)—that is, the moment the work is “fixed in tangible medium of expression.”\(^\text{22}\) Examples of copyrightable works covered by federal law include: (1) literature, (2) music, (3) drama, (4) architecture, (5) pantomimes and choreography, (6) pictorial, graphic, and sculptural works, (7) motion pictures and other audiovisual works, and (8) sound recordings.\(^\text{23}\) However, copyright protection does not extend to broad concepts such as ideas, processes, methods of operation, or discoveries.\(^\text{24}\)

The duration of copyrights for works that are now being created extends for the life of the author plus seventy years.\(^\text{25}\) If the work is pseudonymous, anonymous or made for hire, copyright extends for ninety-five years from its first publication, or 120 years from its creation, whichever expires first.\(^\text{26}\) Thereafter, the work enters the public domain and may be freely used.\(^\text{27}\)

B. Historically, Copyright Has Not Extended to Publishers of “The Law”

Certain categories of works created by government organizations, such as statutes, ordinances, regulations, and judicial opinions, have traditionally entered the public domain from the moment of their creation.\(^\text{28}\) Until recently, it has been a well-settled premise that no author can copyright the text of “the law” in the public domain. This principle developed over 150 years ago through common law decisions that prohibited authors from holding private copyrights in federal court decisions,\(^\text{29}\) state court decisions,\(^\text{30}\) and statutes.\(^\text{31}\)

\(^{20}\) Id. § 102(a).
\(^{21}\) Id. § 302(a).
\(^{22}\) Id. § 301(a).
\(^{23}\) Id. § 102(a)(1)–(8).
\(^{24}\) Id. § 102(b); see also Trotter Hardy, The Copyrightability of New Works of Authorship: “XML Schemas” as an Example, 38 HOUS. L. REV. 855, 864 (2001).
\(^{27}\) Id.
\(^{28}\) See Banks v. Manchester, 128 U.S. 244, 253 (1888).
\(^{30}\) See, e.g., Banks, 128 U.S. at 253–54; Callaghan v. Myers, 128 U.S. 617, 648–49 (1888).
The U.S. Supreme Court issued its first decision regarding copyright law in *Wheaton v. Peters*\(^{32}\) in 1834.\(^{33}\) The Court held unanimously in *Wheaton* that court reporters could not hold copyrights in the judicial opinions they compiled.\(^{34}\) Henry Wheaton, the reporter of Supreme Court cases from 1816 to 1827, claimed that the defendant had infringed his copyright in twelve volumes of Supreme Court reports when the defendant published an identical version of the decisions.\(^{35}\) The Court disagreed, holding that “no reporter has or can have any copyright in the written opinions delivered by this Court; and . . . the judges thereof cannot confer on any reporter any such right.”\(^{36}\)

The Court’s opinion in *Wheaton* marked the beginning of a policy to prevent authors from using copyright law to create monopolies on information.\(^{37}\) The Court reasoned that Wheaton obtained the opinions “not for his own sake, but for the benefit and use of the public; not for his own exclusive property, but for the free and unrestrained use of the citizens of the United States.”\(^{38}\) Accordingly, the *Wheaton* Court laid the foundation for the Court’s “bedrock policy . . . that the public should have maximum access to the law.”\(^{39}\)

The Court reaffirmed this policy in 1888, when it further defined what constitutes “the law” for copyright purposes in *Banks v. Manchester*.\(^{40}\) In *Banks*, the Court denied a copyright to a court reporter who printed the opinions of the Ohio Supreme Court.\(^{41}\) The U.S. Supreme Court held that the official reporter could not obtain a copyright for the syllabus, the statement of the case, or the opinion of the court because these elements were the work of the judges.\(^{42}\) The *Banks* court first reasoned that the

---


32. 33 U.S. (8 Pet.) 591 (1834).


35. Id. at 593–94.

36. Id. at 668.

37. See Patterson & Joyce, supra note 33, at 732.


39. See Patterson & Joyce, supra note 33, at 732.

40. 128 U.S. 244 (1888).

41. Id. at 254.

42. Id. at 253–54.
public owns judicial opinions because taxpayers pay the judges’ salaries. Relying on a public policy rationale, the court next held that “[t]he whole work done by judges constitutes the authentic exposition and interpretation of the law, which[, ] binding every citizen, is free for publication to all, whether it is a declaration of unwritten law, or an interpretation of a constitution or a statute.” The Banks Court relied on an earlier decision of the Massachusetts Supreme Judicial Court, holding that justice requires all people to have free access to judicial opinions, and that it would be against public policy to prevent access to statutes or opinions of the court. Since Banks and Wheaton, judicial opinions have been considered in the public domain and may not be copyrighted.

The same year as Banks, the U.S. Supreme Court also decided a case that expanded the scope of copyright protection to an author’s original additions to “the law.” In Callaghan v. Myers, the defendant Callaghan & Co. owned the copyright on the first thirty-one volumes of the Illinois Supreme Court Reports, while the plaintiff E.B. Myers owned the copyright on volumes thirty-two through forty-six. Callaghan attempted to buy the rights to the subsequent volumes, but refused to pay Myers’ asking price. Callaghan then reprinted Myers’ volumes without permission, including title pages, table of cases, arrangement and pagination, statements of the case, syllabi, and headnotes. The Court held that Callaghan infringed Myers’ copyright. Although the facts of Callaghan and Wheaton were similar, the Court distinguished Callaghan by recognizing Myers’ copyright in his original additions to the court opinions, including the arrangement of the cases, headnotes, tables, and pagination of the volumes. The Court noted the difference between “the law” and the additions made by the reporter using his own ingenuity.
In addition to judicial opinions, legislative acts have also historically been in the public domain and a statute’s publishers could not hold copyright in their legal works. As early as 1886, in the case of *Davidson v. Wheelock*, a federal district court for the District of Minnesota held that a compiler of state statutes “could obtain no copyright for the publication of the laws only; neither could the legislature confer any such exclusive privilege upon him.” A few years later in *Howell v. Miller*, the Sixth Circuit clarified that although an individual cannot own a copyright in statutory material produced by the government, the individual’s copyright in editorial enhancements should be protected. Although the U.S. Supreme Court has never addressed this distinction in the context of statutory materials, the *Howell* court’s rule parallels the *Callaghan* Court’s protection of editorial enhancements to judicial opinions.

C. Copyright Act of 1976

The Copyright Act of 1976 eliminated a historical dual system of copyright protection, in which unpublished works received protection under state law and published works were protected under federal law. Section 301(a) of the 1976 Act eliminated state law copyright for all works that are the subject of federal preemption. In addition, Congress codified in section 201 the general principle that a copyright in a work “vests initially in the author or authors of the work.” However, under section 105, there is no copyright protection for works of the U.S. Government. Thus, it is unclear whether the current Copyright Act would protect a private author whose copyrighted material was enacted.

---

53. 27 F. 61 (D. Minn. 1886).
55. 91 F. 129 (6th Cir. 1898).
56.  Id. at 138.
57.  See infra Part I.B.
59.  See 1 NIMMER, supra note 11, at § 2.02.
61.  See id. § 105.
Copyrights of Model Building Codes

into law. The legislative branch and executive branch have recently considered this issue and explicitly suggested that private authors whose materials are incorporated into law should retain their copyright.

Under section 201, a copyright in a work vests initially in the author of the work. Legislative history indicates that Congress enacted this section to reaffirm the basic principle that an author’s copyright is secure and cannot be taken away by involuntary transfer. Subsection 201(e) makes clear that the government cannot force authors to transfer copyright involuntarily from their owners. At least one commentator has explained that Congress did not intend to allow the government to circumvent section 201 by drafting contract provisions that allow an independent contractor to obtain a copyright in work done for the government, and then require, as a term of the contract, that the contractor transfer the copyright to the United States Government.

Although copyright generally vests in the author of the work, section 105 of the Copyright Act specifically prohibits any person from owning copyright in any work of the U.S. Government. Section 105 provides “Copyright protection under this title is not available for any work of the United States Government . . . .” All official records and documents of the U.S. Government are in the public domain and therefore cannot be copyrighted. One commentator has explained that “any work” includes all works that would be eligible for copyright if they were not “prepared by an officer or employee of the United States Government as a part of that person’s official duties.” Under this reading of the statute, government officials or employees could secure a copyright in works

62. 1 NIMMER, supra note 11, at ¶ 5.13.
64. 17 U.S.C. § 201.
65. See H.R. REP. No. 94-1476, at 60.
66. See 17 U.S.C. § 201(e). The only exception to this provision is when the government’s ownership rights are determined during Title 11 bankruptcy proceedings. 17 U.S.C. § 201(e).
67. 1 NIMMER, supra note 11, at ¶ 5.13.
68. 17 U.S.C. § 105. Two exceptions to the prohibition of government copyright exist. First, the Postmaster General may secure copyright in Post Office publications. See H.R. REP. NO. 94-1476, at 60. Second, the Secretary of Commerce may secure copyright on behalf of the United States in “standard reference data” compiled by or for the Department of Commerce. 15 U.S.C. § 290(e) (2000).
70. 1 NIMMER, supra note 11, at ¶ 5.13.
71. Id.
written outside of their official government duties.\textsuperscript{72} When passing the Act, the House specifically noted in a Committee Report that the government’s publication, or other use of privately authored work would not affect the work’s pre-existing copyright protection.\textsuperscript{73} Thus, in accordance with section 201, the legislative history of section 105 suggests that private authors retain their copyright when their material is incorporated into law.\textsuperscript{74} Nothing in the Copyright Act specifically mandates termination of a private author’s copyright protection.

Section 105 also does not explicitly prohibit private authors from holding copyright in their works when commissioned by the government or prepared under a government grant.\textsuperscript{75} In a Committee Report, the House specifically noted that private parties and institutions are not precluded from holding copyright in works produced under federal government grants and contracts.\textsuperscript{76} Thus, the legislative history of section 105 also suggests that Congress did not deem a contractor to be an “officer or employee of the United States Government.”\textsuperscript{77} Consequently, in such a situation the government agency involved could determine by contract on a case-by-case basis whether to allow an independent contractor to secure copyrights in works prepared using government funds.\textsuperscript{78}

Both the legislative branch and executive branch have recently considered copyright ownership in government works. In the National Technology Transfer and Advancement Act of 1995 (NTTA),\textsuperscript{79} Congress directed “[f]ederal agencies and departments [to] use technical standards that are developed or adopted by voluntary consensus standards bodies.”\textsuperscript{80} Congress encouraged federal agencies to use these

\begin{footnotes}
\item 72. \textit{Id}.
\item 73. H.R. REP. NO. 94-1476, at 60.
\item 74. \textit{See id}.
\item 75. 1 \textsc{Nimmer}, supra note 11, at § 5.13. Private works prepared on commission for the government are not explicitly covered by 17 U.S.C. § 105. At least one court has held that the Act allows copyright for government commissioned works. \textit{See Schnapper v. Foley}, 667 F.2d 102, 110–11 (D.C. Cir. 1981).
\item 77. Simon, supra note 31, at 426.
\item 78. 1 \textsc{Nimmer}, supra note 11, at § 5.13.
\item 80. \textit{Id}.
\end{footnotes}
organizations to support their policy objectives and activities, unless that use would be inconsistent with applicable law or impractical.  

Executive guidelines also contemplated that privately authored material would find its way into government works. The Office of Management and Budget (OMB) responded to NTTA by issuing Circular A-119.  

This circular contained instructions on the use of privately developed, consensus based standards, requiring that “if a voluntary standard is used and published in an agency document, your agency must observe and protect the rights of the copyright holder and any other similar obligations.”  

The OMB offered several reasons why government agencies should use the works of standards-setting organizations, including decreasing the cost to the government, providing incentives and opportunities to establish standards that serve national needs, and furthering the policy of reliance on the private sector to support government needs.  

Thus, under Circular A-119, federal agencies are required to observe and protect the copyrights of private authors.


The Copyright Act of 1976 does not address whether state governments can hold a copyright in “the law.” The Act prohibits the federal government from obtaining copyright protection in its work, but does not explicitly forbid state and local governments from doing the same. Although the U.S. Supreme Court has not addressed this issue, the Second Circuit and the First Circuit have interpreted the Act’s limitation on copyright ownership to extend only to the federal government. These courts have interpreted the plain language of section 105 as specifying a limitation solely on the federal government and not

---

81. Id.


83. Id. at 8555.

84. Id. at 8554–55.


86. Id. at § 105.

87. County of Suffolk v. First Am. Real Estate Solutions, 261 F.3d 179, 187 (2d Cir. 2001); Bldg. Officials Code Adm’n’s v. Code Tech., Inc., 628 F.2d 730, 735–36 (1st Cir. 1980) (“Works of state governments are therefore left available for copyright protection by the state or the individual author . . . .”).
the state government. The Second Circuit noted that because the Copyright Act specifies a limitation on ownership against the federal government, it implies that states are not similarly restricted.

A federal district court for the Northern District of Georgia, however, reached a different result in *Georgia v. Harrison Co.* The State of Georgia sued a publisher of the state code for copyright infringement, arguing that state governments can hold a copyright in state law even if private citizens cannot. The plaintiff reasoned that if Congress had intended to preclude state copyright protection, Congress would have explicitly provided for such a limitation in the Copyright Act. Despite the lack of an express prohibition by Congress, the district court followed the traditional principle that no entity may hold a copyright in judicial opinions or statutes. Further, the court held that the rationale for prohibiting copyright in judicial opinions and statutes—that the public should have open access to the law free from copyright barriers—applies equally whether the party seeking the copyright is an individual or the state itself.

Thus, historically “the law” in the form of judicial opinions and statutes has been in the public domain and has not been amenable to copyright. Congress has not only attempted to adapt to changes in technology and society by amending the Copyright Act, but also has responded to changes in the way government business is conducted, by addressing the frequent government use of documents authored by private contractors.

### E. Fair Use Defense

In 1976, Congress made major alterations to copyright law in an effort to stay current with new technological innovations. Congress codified

---

88. See Suffolk, 261 F.3d at 187; Bldg. Officials Code Adm’rs, 628 F.2d at 735–36.
89. See Suffolk, 261 F.3d at 187.
91. Id. at 114.
92. Id.
93. Id.
94. Id.
the doctrine of “fair use,” providing an affirmative defense for certain “fair uses” of copyrighted material, including reproduction. 96 Codified at 17 U.S.C. § 107, the “fair use” doctrine allows copyrighted works to be reproduced for purposes of criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, and research. 97 The doctrine “creates a limited privilege in those other than the owner of a copyright to use the copyrighted material in a reasonable manner without the owner’s consent.”98

Congress codified four factors in the Copyright Act of 1976 that courts should consider when a user of a copyrighted work raises the fair use exception as an affirmative defense. 99 First, a court should examine the purpose and character of the use, such as whether the use is for commercial or nonprofit, educational purposes. 100 Second, a court should consider the nature of the copyrighted work. 101 The more creative a work, the more it should be protected from copying; the more informational or functional the work, the broader the scope of the fair use defense should be. 102 Third, a court should assess the amount and substantiality of the portion of the copyrighted work used. 103 Finally, a court should consider the economic impact of the use, such as the effect on the potential market for the work or the value of the work. 104 In addition, courts have considered a party’s motive for using the copyrighted material. 105

The U.S. Supreme Court has noted that the fair use doctrine gives courts the flexibility to avoid rigid application of the copyright statute when it would stifle the creativity it was designed to promote. 106 The fair use doctrine balances copyright owners’ exclusive rights to what they create with the Constitution’s goal of promoting learning. 107

97. Id.
98. Fischer v. Dees, 794 F.2d 432, 435 (9th Cir. 1986).
99. 4 NIMMER, supra note 11, at § 13.05.
101. Id.
102. 4 NIMMER, supra note 11, at § 13.05.
104. Id.
105. Stewart v. Abend, 495 U.S. 207, 236 (1990); 1 NIMMER, supra note 11, at § 13.05.
106. Stewart, 495 U.S. at 236.
Commentators have stated that fair use is a well-established, equitable doctrine that fine tunes the scope of copyright on a case-by-case basis.\textsuperscript{108} It is not a blanket provision that affords or denies copyright protection to particular information.\textsuperscript{109} Instead, the doctrine is an exception to the bundle of property rights afforded to the copyright holder.\textsuperscript{110} The fair use doctrine allows courts to balance authors’ and inventors’ interests in the use of their writings and discoveries against society’s competing interest in the “free flow of ideas, information, and commerce . . . . “\textsuperscript{111}

In sum, Congress codified many fundamental copyright concepts in the Copyright Act of 1976. The Act created a comprehensive framework that incorporated the doctrine of fair use as well as prohibiting owning copyrights in government works. The well established principle that no author can copyright the text of “the law” has been challenged by recent circuit court decisions determining whether private authors retain their copyright in materials when they are incorporated into law.

II. THE FEDERAL CIRCUIT COURTS HAVE RECENTLY DIVIDED OVER THE VALIDITY OF A PRIVATELY HELD COPYRIGHT ON “THE LAW”

While it is well established that individuals and the federal government cannot claim copyright in federal judicial opinions and statutes,\textsuperscript{112} the U.S. Supreme Court has never determined whether this prohibition also applies to state statutes and regulations. One specific context in which this issue arises is when a private organization, acting in an independent capacity, authors a model code or regulation. The authoring entity owns copyright in original material it creates.\textsuperscript{113} However, courts are divided as to whether the copyright terminates when the work is subsequently adopted in whole or by reference into the official statutes and regulations of a state government.\textsuperscript{114} Circuit courts in

\begin{thebibliography}{10}
\bibitem{108} 4 \textsc{Nimmer}, \textit{supra} note 11, at § 13.05.
\bibitem{109} Maureen A. O’Rourke, \textit{Toward a Doctrine of Fair Use in Patent Law}, 100 \textsc{Colum. L. Rev.} 1177, 1188 (2000).
\bibitem{112} See Banks v. Manchester, 128 U.S. 244, 253–54 (1888); Wheaton v. Peters, 33 U.S. (8 Pet.) 591, 668 (1834).
\bibitem{113} 17 \textsc{U.S.C.} § 201(a) (2000).
\bibitem{114} See 1 \textsc{Nimmer}, \textit{supra} note 11, at § 5.12.
\end{thebibliography}
Copyrights of Model Building Codes

the Second and Ninth Circuits have held that private organizations retain their copyrights in these works.\textsuperscript{115} However the Fifth Circuit recently split from this reasoning in \textit{Veeck v. Southern Building Code International, Inc.}, and held that private code-authoring organizations automatically lose their copyright when a local government enacts their works into law.\textsuperscript{116}

\textbf{A. The Second and Ninth Circuits Have Enforced the Copyrights of Privately Authored Materials Adopted Into Law}

In recent years, two Circuit Courts have reached the conclusion that private organizations retain their copyright in model codes and regulations adopted into law. The Second Circuit first reached this decision in \textit{CCC Information Services, Inc. v. Maclean Hunter Market Reports, Inc.}\textsuperscript{117} and later extended it to a different set of facts.\textsuperscript{118} Following the Second Circuit’s example, the Ninth Circuit has also held that private authors do not lose copyright protections in their work when a government adopts the work into law.\textsuperscript{119}

\textbf{1. CCC Information Services, Inc. v. Maclean Hunter Market Reports and Subsequent Case Law}

The Second Circuit first addressed the issue of whether private authors retain copyright in materials adopted into law in \textit{CCC}. Maclean, the publisher of the “Red Book” (a compilation of used car value estimates), sued for copyright infringement when CCC Information Services, Inc. (CCC), a competitor, copied Maclean’s valuations into a computer database for sale to its customers.\textsuperscript{120} CCC combined Maclean’s Red Book valuations for each vehicle with the vehicle’s valuation in the National Automobile Dealer’s Association Official Used Car Guide (the Bluebook).\textsuperscript{121} Certain state laws incorporate the Red Book and Bluebook

\textsuperscript{115} Practice Mgmt. Info. Corp. v. Am. Med. Ass’n, 121 F.3d 516, 520 (9th Cir. 1997); CCC Info. Servs., Inc. v. Maclean Hunter Mkt. Reports, Inc., 44 F.3d 61, 74 (2d Cir. 1994).


\textsuperscript{117} 44 F.3d 61 (2d Cir. 1994).

\textsuperscript{118} County of Suffolk, N.Y. v. First Am. Real Estate Solutions, 261 F.3d 179, 195 (2d Cir. 2001).

\textsuperscript{119} \textit{Practice Mgmt.}, 121 F.3d at 520.

\textsuperscript{120} CCC, 44 F.3d at 63–64.

\textsuperscript{121} \textit{Id.} at 64.
averages as a minimum for specific insurance payments. Thus, the issue arose whether Maclean retained its copyright in the Red Book.

A federal district court granted summary judgment for CCC because it reasoned that the Red Book fell into the public domain when states incorporated it into official regulations as a factor in the average calculation. However, a three-judge panel of the Second Circuit reversed and granted summary judgment for Maclean, upholding his copyright in the Red Book. The Second Circuit stressed the need “to balance the conflicts and contradictions that pervade the law of copyright,” requiring courts to make policy determinations when conflicts exist, to determine which claim will prevail. The court considered the originality and creativity of the material in the Red Book and held that it was sufficient to warrant copyright protection.

The Second Circuit reasoned that extending copyright protection to the original elements of the work imposed little cost or disadvantage to society. The court refused to hold that a state’s reference to a copyrighted work was the same as a judicial opinion. Further, the court looked beyond the outcome of the CCC case to the possible consequences of its ruling, and expressed concern that a failure to extend copyright protection in CCC would compel undesired results in other cases. For example, if a private individual’s copyright in a work terminated upon its adoption by a state government, authors could conceivably lose their rights when a state-mandated curriculum assigned their books as school reading, because their books would be incorporated by reference into a state law. As a result, the Second Circuit held that a private individual’s copyright endures even after a state government adopts the author’s material by reference into law.

The Second Circuit recently addressed a similar issue in County of Suffolk, New York v. First American Real Estate Solutions. Suffolk

122. Id.
123. Id. at 63.
124. Id.
125. Id. at 74.
126. Id. at 68.
127. Id. at 66–68.
128. Id. at 66.
129. Id. at 74.
130. Id.
131. Id.
132. Id.
133. 261 F.3d 179 (2d Cir. 2001).
County, through its Real Property Tax Service Agency (RPTSA), created a series of original maps of tax districts and special district boundaries (tax maps) with an index system. Suffolk County registered its copyright claims for these tax maps and affixed a notice of copyright in the introduction to each tax map album. Suffolk County filed for an injunction against First American Real Estate Solutions when it published paper and CD-ROM copies of the tax maps, and placed them in the market without first obtaining a license or the county’s consent.

First American defended its actions by arguing that Suffolk County was barred from asserting its copyright under New York’s Freedom of Information Law (FOIL). FOIL requires state agencies to make all government records available for public inspection and copying. Thus, First American argued that FOIL prevented Suffolk County from receiving copyright protection for its tax maps. The Second Circuit disagreed, holding that the county could own a copyright under the Copyright Act, and that FOIL did not destroy the county’s copyright.

The Second Circuit reached this conclusion after examining the Banks decision and the U.S. Supreme Court’s rationale for holding that judges may not own copyrights in their own opinions. The Second Circuit noted two main considerations behind Banks: first, whether the work’s creator needs an economic incentive to create or has a proprietary interest in the work; and second, whether the public must access the work to have notice of the law. Applying these considerations to the facts before it, the court remanded the case for a determination of the first factor whether RPTSA required any incentive to create the maps. In regards to the second factor, the Second Circuit noted that public access existed because there was no allegation that any individual could not access the law or tax maps. Further, the court rejected First American’s argument that due process required the tax maps to be public.

134. Id. at 184.
135. Id.
136. Id.
137. Id. at 183.
138. N.Y. PUB. OFF. LAW § 84 (McKinney 1988).
139. Suffolk, 261 F.3d at 183.
140. Id. at 195.
141. Id. at 193–94.
142. Id. at 194.
143. Id. at 195.
144. Id.
domain from the moment of inception, because it found no evidence that the maps were not freely available to the public.\textsuperscript{145}

2. Practice Management Information Corp. v. American Medical Ass’n

The Ninth Circuit addressed this issue in \textit{Practice Management Information Corp. v. American Medical Ass’n},\textsuperscript{146} by applying the analysis and reasoning of the Second Circuit’s decision in \textit{CCC}. The Ninth Circuit held that the American Medical Association’s (AMA) copyright of its medical procedure code, the Physician’s Current Procedural Terminology (CPT), was still valid after a federal agency adopted the code for use on Medicaid claim forms.\textsuperscript{147} Despite Practice Management’s argument that the CPT entered the public domain when the federal Health Care Financing Agency Administration required its use, the court decided that the AMA did not lose its right to enforce its copyright simply because the government chose to use its coding system.\textsuperscript{148} The Ninth Circuit reasoned that if the code entered the public domain through governmentally mandated use, a large number of copyrights on privately authored codes, standards, and reference works would be subject to invalidation.\textsuperscript{149}

The Ninth Circuit distinguished \textit{Practice Management} from the \textit{Banks} decision in which the U.S. Supreme Court held that judicial opinions are not subject to copyright.\textsuperscript{150} According to the Ninth Circuit, the \textit{Banks} court reasoned that the public owns judicial opinions because taxpayers pay the judges’ salaries, and the due process principle of free access requires the law to be in the public domain.\textsuperscript{151} The Ninth Circuit reasoned that neither of these considerations required the invalidation of the AMA’s copyright.\textsuperscript{152} First, the court noted that the copyright system was not critical in the \textit{Banks} decision because judges had no proprietary interest over their opinions.\textsuperscript{153} In contrast, the AMA’s copyright in CPT

\begin{itemize}
\item \textsuperscript{145} \textit{Id.}
\item \textsuperscript{146} 121 F.3d 516 (9th Cir. 1997).
\item \textsuperscript{147} \textit{Id.} at 518, 520.
\item \textsuperscript{148} \textit{Id.}
\item \textsuperscript{149} \textit{Id.} at 519.
\item \textsuperscript{150} \textit{Id.} at 518.
\item \textsuperscript{151} \textit{Id.} at 518–19.
\item \textsuperscript{152} \textit{Id.}
\item \textsuperscript{153} \textit{Id.}
\end{itemize}
was critical to the AMA because the AMA authored, owned, and maintained the CPT and claimed copyright in it. Second, enforcing AMA’s copyright would not jeopardize the due process requirement to maintain free access to the law because the court found no evidence that anyone wishing to use the CPT had difficulty obtaining access to it. Thus, the Ninth Circuit joined the Second Circuit in holding that private individuals or organizations do not lose their copyright to work they have created once it is adopted into law.

B. The First and Fifth Circuits Have Declined to Enforce Copyrights of Privately Authored Materials Adopted Into Law

Not all courts confronted with this issue have enforced the copyrights of privately authored materials that are adopted or incorporated by reference into law. Although the Second and Ninth Circuits have been willing to extend copyright protection to such privately authored works, the First Circuit has strongly suggested that copyright protection does not exist. Further, the Fifth Circuit recently refused to enforce such a copyright, declaring that the copyright automatically terminated when the privately authored materials were adopted into law.

1. The BOCA Ruling Left Open Whether to Enforce Copyrights of Privately Authored Materials Adopted Into Law

In 1980, the First Circuit declined to rule on the validity of a non-profit group’s copyright in its building code when the State of Massachusetts subsequently adopted the code into law. In Building Officials & Code Administrators v. Code Technology, Inc., the plaintiff Building Officials and Code Administration (BOCA), a code writing organization, claimed copyright protection for its BOCA Basic Building Code. This private organization had developed and published the code

---

154. Id.
155. Id.
158. Veeck, 293 F.3d at 806.
159. Bldg. Officials & Code Adm’rs, 628 F.2d at 736.
160. Id. at 731.
for over twenty-eight years.\textsuperscript{161} BOCA’s procedure was to secure a copyright in the building code and then encourage government authorities to use its licensing program and adopt the code by reference, in whole, or in part.\textsuperscript{162} The defendant Code Technology, Inc. (CT), a private publisher, later published and distributed its own edition of the building code in almost identical form to BOCA’s version.\textsuperscript{163} BOCA sought an injunction to prevent CT from publishing its own edition of the building code and selling it to the public.\textsuperscript{164} CT argued that because the state adopted BOCA’s code as a set of administrative regulations having the force of law, BOCA lost its copyright protection and the code entered the public domain.\textsuperscript{165} CT relied on the \textit{Banks} and \textit{Wheaton} decisions as support and analogized the “citizen authorship” of building codes to the work of judges when writing opinions.\textsuperscript{166}

Although the First Circuit did not reach the merits of the \textit{BOCA} case, it vacated the district court’s preliminary injunction, which had prevented defendant CT from publishing and selling its own edition of the building code.\textsuperscript{167} The court reasoned that BOCA had not demonstrated a sufficient probability that it would succeed on the merits, and therefore declined to grant an injunction.\textsuperscript{168} Because the court addressed the issue only for purposes of preliminary relief, it did not specifically decide whether BOCA’s authorship of the building code entitled it to copyright protection.\textsuperscript{169} But, the court strongly suggested that copyright protection did not exist, noting that it was “far from persuaded” that BOCA could enforce its copyright in this way.\textsuperscript{170} The First Circuit reasoned that the public “owns the law,” not because the public pays the salaries of the authors of the work, as the \textit{Wheaton} and \textit{Banks} Courts had reasoned, but

\begin{itemize}
  \item \textsuperscript{161} \textit{Id.} at 732.
  \item \textsuperscript{162} \textit{Id.}
  \item \textsuperscript{163} \textit{Id.}
  \item \textsuperscript{164} \textit{Id.} at 731.
  \item \textsuperscript{165} \textit{Id.} at 733.
  \item \textsuperscript{166} \textit{Id.} at 733–34.
  \item \textsuperscript{167} \textit{Id.} at 736.
  \item \textsuperscript{168} \textit{Id.}
  \item \textsuperscript{169} \textit{Id.} at 734. In a subsequent lower court opinion in the First Circuit, a federal district court in the District of Massachusetts stated that the First Circuit has never definitively decided whether privately authored works, once adopted by government bodies into law, enter the public domain and lose their status as copyrightable works. \textit{See} John G. Danielson, Inc. v. Winchester-Conan Props., Inc., 186 F. Supp. 2d 1, 22 (D. Mass. 2002).
  \item \textsuperscript{170} \textit{Id.} at 735.
\end{itemize}
rather because “the citizens are the authors of the law.” The First Circuit deemed citizens the metaphorical “authors” of the law because the law derives its authority from the consent of the public as expressed through the democratic process. In addition, the court reasoned that the concept of due process guarantees citizens access to the law, to provide notice of their legal obligations.

At the same time, the BOCA court recognized the important public function of code drafting organizations and left open whether the copyright exception for judicial opinions created in the 1800s should be applied to technical codes. The court highlighted the federal and state trend towards adoption of privately authored works. Thus, even though the First Circuit was “far from persuaded” that BOCA could enforce its copyright, the court ultimately refused to determine whether a private organization could retain its copyright in material adopted into law.

2. **The Fifth Circuit’s Veeck Decision Terminated Copyright Protection of Privately Authored Works Enacted into Law**

Most recently, an en banc panel of the Fifth Circuit addressed private authors’ copyright interests in state laws in *Veeck v. Southern Building Code Congress International, Inc.* Peter Veeck controlled and operated RegionalWeb, a non-commercial website that posted various information of local interest to North Texas residents. Veeck posted the local building codes of two small towns in North Texas. These codes incorporated part of the Standard Building Code written by the Southern Building Code Congress International (SBCCI), a non-profit organization with approximately 14,500 members drawn from governmental bodies, the construction industry, business and trade associations, and college departments. Unlike the BOCA case, SBCCI did not execute a licensing agreement when the county adopted its

---

172. *Id.* at 734.
173. *Id.*
174. *Id.* at 736.
175. *Id.*
176. *Id.*
177. *Id.* at 735.
179. *Id.* at 735.
180. *Id.*
codes. Yet after the codes passed into law, SBCCI continued to assert its copyright. Veeck bought the model building codes directly from SBCCI for $72.00 and received a copy of the codes on a computer disk. Posting the codes onto his website, he disregarded the licensing agreement and copyright notice that prohibited copying and distribution. Consequently, SBCCI demanded that Veeck stop infringing on its copyright, so Veeck sought a declaratory judgment that he was not violating the Copyright Act.

A federal district court for the Eastern District of Texas decided in SBCCI’s favor. The court recognized that SBCCI held a copyright in the model codes, and rejected Veeck’s argument that adopting the codes into law stripped SBCCI of its copyrights. The court also rejected Veeck’s argument that protecting SBCCI’s copyright was a violation of the public’s due process right of “[f]ree access to the law,” because the court found that the codes were readily available to citizens and that Veeck himself was never denied access. The court rejected Veeck’s alleged defense of fair use, reasoning that Veeck’s use of SBCCI’s material had an adverse impact on the market for SBCCI’s copyrighted works, and that such a harm was sufficient to overcome the public’s “fair use” interests in the material. Therefore, the district court sustained SBCCI’s infringement claim, awarded statutory damages, and imposed a permanent injunction against Veeck. On appeal, a three judge panel of the Fifth Circuit affirmed the district court’s decision.

A year later, the Fifth Circuit granted rehearing en banc and reversed the district court decision in a divided nine to six opinion. The en banc panel held that Veeck’s copying was not infringement because the codes were in the public domain. First, the majority relied on the concept of “citizen authorship” of the law, which had been previously discussed by

---

181. Id. at 794.
182. Id.
183. Id. at 793.
184. See id. at 794.
185. Id.
187. Id. at 891.
188. Id. at 889.
189. Id. at 891.
190. Id. at 892–93.
193. Id. at 799–800.
the First Circuit in *BOCA*. The Fifth Circuit reasoned that the process of lawmaking demands and incorporates contributions by “the people,” that lawmakers represent the public will, and that the public is the final “author” of the law. The *Veeck* majority also relied on the *Banks* decision, which it said focused on the “metaphorical concept of citizen authorship” together with the need for citizens to have free access. By characterizing *Banks* in this way, the Fifth Circuit disagreed with the Second and Ninth Circuit’s emphasis on the “incentives” for authorship.

The Fifth Circuit also distinguished *Veeck* from *Practice Management* and *CCC* by differentiating between codes and standards. According to the majority, standards incorporated by reference, such as those at issue in *Practice Management* and *CCC*, do not become law simply because a law refers to them. In contrast, the majority found that SBCCI created its model codes for the sole purpose of having them enacted into law. However, the dissenting judges argued that the distinction between codes and standards is meaningless because they in effect operate in the same way. Judge Weiner reasoned in his dissent that the main difference in the terms is that federal law uses the term “standards” and state law uses the term “codes,” but that both terms can be used to refer to agency regulations.

Thus, the Circuit Courts are divided over whether to enforce private authors’ copyrights when a government enacts their works by reference or adoption into law. The Second and Ninth Circuits favor protection of the private copyright holder’s proprietary rights. However, the Fifth Circuit rejected the right of a private organization to retain its copyright and instead emphasized the public’s authorship of the law and its need to access the law.

---

194. Id. at 799.
195. Id.
196. Id.
197. Id.
198. See id. at 796–97.
199. Id. at 804–05.
200. Id. at 805.
201. Id.
202. Id. at 815 n.20 (Weiner, J., dissenting).
203. Id.
205. *Veeck*, 293 F.3d at 796.
III. PRIVATE AUTHORS WHOSE CODES ARE ENACTED BY REFERENCE OR ADOPTION INTO LAW SHOULD RETAIN THEIR COPYRIGHTS

The Fifth Circuit’s decision in *Veeck* created a split among the federal Circuit Courts when it held that private organizations automatically lose their copyright in works that are later enacted by reference into law. Both federal legislative and executive branch statements and public policy concerns run counter to such a rule. Moreover, the Second and Ninth Circuits’ decisions indicate that such a rule undermines copyright law. In *Veeck*, the Fifth Circuit erroneously relied on cases that are substantively distinguishable and in the process created an unwarranted exception to copyright law.

A. The Veeck Decision Conflicts with the Second and Ninth Circuits

The Fifth Circuit’s en banc decision in *Veeck* created a conflict among the federal Circuit Courts. The Fifth Circuit held that non-profit organizations that author copyrighted model codes and standards automatically lose their copyright protection when a governmental body incorporates the model codes by reference into law.206 The *Veeck* decision stands in stark contrast to the Second and Ninth Circuits holding that a private author’s copyright is not terminated when the author’s codes or standards are adopted into law.207

The division in the circuits is significant because it can lead to inconsistent application of federal copyright law around the country. Local governments in several circuits can adopt model codes drafted by organizations like SBCCI, raising the potential for inconsistent copyright protections for the same code in different jurisdictions. Until the U.S. Supreme Court, or alternatively the U.S. Congress through legislation, resolves this issue, there is the potential for a race to the courthouse in the jurisdiction most favorable to each party.208

206. *See id.* at 796.
207. *See Practice Mgmt. Info. Corp.*, 121 F.3d at 520; CCC, 44 F.3d at 74.
208. *See EUGENE F. SCOLLES ET AL., CONFLICT OF LAWS § 3.36 (3d ed. 2000).*
B. Congressional Intent and Executive Branch Directives Favor
Copyright Protection of Privately Authored Materials

The en banc Veeck court created an unsupported exception to modern copyright law by holding that adopting privately authored codes into law automatically places them in the public domain.209 This decision undermines the long-standing purpose of the Copyright Act to provide authors and inventors with incentives to produce work that will be valuable to the public.210 The plain language of the Copyright Act indicates that copyrighted model codes and standards should retain their protected status even when adopted and incorporated into the law. For example, 17 U.S.C. § 201(a) states that a copyright in a work “vests initially in the author or authors of the work.”211 Legislative history indicates that the purpose of this section was to reaffirm the basic principle that an author’s copyright is secure, and cannot be taken away by involuntary transfer.212

The Fifth Circuit misconstrued the policies behind the Copyright Act and incorrectly interpreted the policy underlying the government’s adoption of SBCCI’s codes. The purpose of the Copyright Act is to provide incentives to produce valuable work in science and the arts for the public.213 One of the incentives is granting authors copyright protection under section 201.214 The court’s ruling that copyright protection ceases when material is adopted into law could eliminate the incentive to create and effect an involuntary transfer of copyright prohibited by section 201(e).

Further, at least one commentator has argued that the text of the Copyright Act itself prohibits changing in the copyright status of privately authored codes because they have been adopted or referenced by a legislative body.215 If the code was original and expressive enough to qualify for copyright protection, then it has not lost those qualities purely because the material was enacted into law.216 In the Veeck case, the defendant never disputed that the codes were copyrighted when he

210. 1 NIMMER, supra note 11, at § 1.03.
213. See supra Part I.C.
214. See supra Part I.C.
215. Hardy, supra note 24, at 878.
216. Id.
purchased copies of the codes and a licensing agreement from SBCCI.\textsuperscript{217} The protected status of SBCCI’s works should not change simply because those works were adopted into law.

Aside from the problematic circuit split created by the \textit{Veeck} decision, it also conflicts with policies set forth by the federal legislative and executive branches. Congress has indicated that it supports the copyright protection of works commissioned by the government.\textsuperscript{218} The executive branch has reinforced this policy by instructing agencies to recognize the copyrights of private authoring organizations when using voluntary standards.\textsuperscript{219}

Although the U.S. Supreme Court has not ruled on the issue, legislative actions by the U.S. Congress indicate that it supports copyright protection for materials authored by entities like SBCCI. Congress has indicated that the federal government’s incorporation of copyrighted model codes and standards should not jeopardize private author’s copyrights.\textsuperscript{220} Section 105 prevents the federal government from holding copyrights.\textsuperscript{221} But, the House Committee Report on this section indicates it should not have a negative impact on copyrights.\textsuperscript{222} The House committee specifically noted that the government’s publication of privately authored work does not affect that work’s preexisting copyright.\textsuperscript{223} Further, in the same Committee Report, the House explicitly directed that private parties and institutions should not be precluded from holding copyright in works commissioned by the government or prepared under a government grant.\textsuperscript{224} Thus, SBCCI’s copyright in its works should not terminate because it was adopted into law.

The Fifth Circuit’s decision also conflicts with the policy embodied in the National Technology Transfer and Advancement Act of 1995 (NTTA).\textsuperscript{225} Congress directed federal agencies under NTTA to use

\begin{itemize}
  \item \textsuperscript{217} See \textit{Veeck v. S. Bldg. Code Cong. Int’l, Inc.}, 293 F.3d 791, 793 (5th Cir. 2002) (en banc).
  \item \textsuperscript{220} \textit{H.R. REP. NO. 94-1476}, at 60.
  \item \textsuperscript{221} \textit{17 U.S.C. § 105} (2000).
  \item \textsuperscript{222} \textit{See H.R. REP. NO. 94-1476}, at 60.
  \item \textsuperscript{223} \textit{Id.}
  \item \textsuperscript{224} \textit{See supra Part I.C.}
\end{itemize}
standards developed by private authoring bodies. This indicates that Congress supports federal agencies’ use of privately authored standards.

The Executive Branch also supports the use of privately authored standard and codes while protecting their copyrights. OMB revised Circular A-119 after the enactment of NTTA to guide federal agencies. In Circular A-119, the OMB interpreted this mandate as protecting the rights of the copyright holder. The OMB issued instructions on the use of privately developed, consensus based standards, requiring that “if a voluntary standard is used and published in an agency document, your agency must observe and protect the rights of the copyright holder and any other similar obligations.”

The Fifth Circuit’s decision in Veeck directly conflicts with the congressional purpose of NTTA as interpreted by the OMB. Privately authored codes and standards are important resources to the government, because they are potentially cost-saving resources. Federal copyright law does not prohibit private authors from holding copyright in their works when commissioned by the government or prepared under a government grant. To deny these contractor’s copyright protection of their work would have a negative impact on the government because it may eliminate an incentive for producing such work. Because the legislative and executive branches favor the use of privately authored works—and the private authors’ retention of their copyrights—the Fifth Circuit’s Veeck decision contradicts the policy behind NTTA.

C. Judicial Precedent Favors Copyright Protections for Privately Authored Laws

The Fifth Circuit’s en banc decision in Veeck adopted a per se rule that when a municipality adopts a privately authored model code into law by reference, the work is stripped of its copyright protection. The Fifth Circuit’s decision is flawed for several reasons. First, the Fifth Circuit reached its decision by factually distinguishing the Veeck case from the decisions of the Ninth and Second Circuits, yet it overlooked the substantive similarities between these cases. Second, the circuit court’s

---

226. Id.
228. See id. at 8555.
229. Id.
230. See supra Part I.C.
reliance on Banks and BOCA was misguided because they are inapposite and do not provide support for the court’s decision. As a result, the Fifth Circuit has created an unsupported exception to copyright law.

1. Although the Veeck Case is Factually Distinguishable from Practice Management and CCC, the Substantive Issues are Similar

There are some factual differences between the Veeck case and those decided by the Second and Ninth Circuits. The CCC case from the Second Circuit involved standards that the local government incorporated by reference into a statute,231 while Veeck involved standards directly incorporated into law.232 In addition, Practice Management involved a private company seeking to invalidate a copyright for its own purposes,233 while Veeck involved a private individual seeking to freely distribute a copyrighted public law for use by other citizens.234

However, the facts the Fifth Circuit relied upon to distinguish Veeck from CCC and Practice Management were not relevant to the legal questions presented. The Fifth Circuit distinguished Veeck from CCC and Practice Management because Veeck involved “codes” while the other two cases involved “standards.”235 But, this difference in taxonomy is meaningless. In his dissent to Veeck, Judge Wiener explained that these labels are analogous and virtually indistinguishable.236 The different labels reflect the difference between federal and local lawmaking—federal agency regulations generally address national standards, while state and local laws refer to more specific safety and building requirements typically found in state or municipal codes.237 But there is no practical difference between these terms, because both terms can refer to privately authored collections of technical data used by government agencies to regulate specialized areas of society such as

231. See CCC Info. Servs., Inc. v. Maclean Hunter Mkt. Reports, 44 F.3d 61, 63–64 (2d Cir. 1994).
234. See Veeck, 293 F.3d at 793.
235. See id. at 804.
236. Id. at 815 n.20 (Weiner, J., dissenting).
237. Id.
health, education, or construction. Thus, these factual and linguistic differences do not justify the Fifth Circuit’s departure in Veeck.

Yet, despite these minor factual differences, the three Circuit Courts all faced similar legal issues. First, the courts each addressed how far the rights of private authors should extend when the work they have created is adopted into law. Further, the courts had to answer the difficult question of how to balance the public’s rights to access to the law against the copyright holder’s proprietary interests. At the time of the Veeck decision and currently, Practice Management and CCC represent the most similar legal issues and the highest available level of authority on point, given that the Supreme Court has not yet ruled on the issue. The Fifth Circuit’s disregard for the substantive similarities of both decisions undermines the national uniformity of federal copyright law.

2. The Veeck Decision was Erroneously Based on the Banks and BOCA Cases

While the Second and Ninth Circuit’s decisions are factually distinguishable from Veeck, they are more substantively analogous to the issue that was before the Fifth Circuit than Banks. Yet, instead of relying on precedent set by its sister circuits, the Fifth Circuit based its Veeck decision on the Banks case, which is substantively distinguishable, and BOCA, a First Circuit case that was not even decided on its merits.

Although the Veeck court based its decision in part on the U.S. Supreme Court’s 1888 decision in Banks, its interpretation of Banks’ holding and application of Banks to the facts before it were incorrect. The Fifth Circuit is the only court that has relied on Banks to invalidate an author’s copyright in its privately authored “law.”

238. Id.
239. See id. at 798–800; Practice Mgmt. Info. Corp. v. Am. Med. Ass’n, 121 F.3d 516, 519 (9th Cir. 1997); CCC Info. Servs., Inc. v. Maclean Hunter Mkt. Reports, 44 F.3d 61, 67–68 (2d Cir. 1994).
240. See Veeck, 293 F.3d at 799–800; Practice Mgmt., 121 F.3d at 519; CCC, 44 F.3d at 72.
243. Veeck, 293 F.3d 791.
Supreme Court held in *Banks* that federal judges can not hold copyrights in work done in their official capacity.\(^{244}\) From this, the *Veeck* court fashioned a general rule that citizens are the metaphysical “authors” of the law, and that citizens must have free access to the law.\(^{245}\)

The Fifth Circuit should have followed the Ninth Circuit and Second Circuit’s more accurate interpretation of *Banks*’ holding and application. In *Practice Management*, the Ninth Circuit suggested two rationales for the *Banks* decision: (1) the public owns judicial opinions because taxpayers pay judges’ salaries, and (2) the due process principle of free access to the law requires that the law be in the public domain.\(^{246}\) The Ninth Circuit held that neither of these holdings required it to invalidate the AMA’s copyright.\(^{247}\) First, the Ninth Circuit reasoned that the *Practice Management* facts were different from *Banks* because private entities that write codes need the financial incentives provided by copyright law. In contrast, the judges in *Banks* were paid to write their opinions, and needed no independent incentives.\(^ {248}\) Second, the Ninth Circuit distinguished *Practice Management* because there was no evidence that copyright protection prohibited the public’s access to judicial opinions.\(^ {249}\) Moreover, the *Practice Management* court noted that even if a free access problem arose, it would be better to deal with the problem through established doctrines of copyright law, such as fair use or perhaps a due process defense for infringers, rather than simply stripping the codes of copyright protection.\(^{250}\) The Ninth Circuit also noted another possible solution to this problem—to require mandatory licensing at a reasonable royalty if the copyright holder tried to severely restrict public access to its copyrighted information.\(^{251}\)

The Second Circuit also upheld the copyright protections of privately authored materials adopted into law in *CCC*. In *CCC*, the plaintiff argued that its use of the Red Book was defensible because the Red Book had come into public domain as a result of its adoption by reference into state statutes.\(^{252}\) The Second Circuit rejected this argument and refused to hold

---

245. *Veeck*, 293 F.3d at 799.
246. *Practice Mgmt.*, 121 F.3d at 518–19.
247. *Id*.
248. *Id*.
249. *Id*.
250. *Id* at 519.
251. *Id* (citing *Abend v. MCA, Inc.*, 863 F.2d 1465, 1479 (9th Cir. 1988)).
that a state’s reference to a copyrighted work was the same as a judicial opinion. Following CCC, the Second Circuit held in County of Suffolk that the Banks decision required it to determine whether the author would have an adequate incentive to create the work absent copyright protections and whether the public had access to the law. The Fifth Circuit should have applied those factors to the Veeck case and found that copyright protection provides an incentive to SBCCI to create building codes. Moreover, as in County of Suffolk, the SBCCI’s copyright did not prevent the public from accessing the codes.

In contrast to the Second and Ninth Circuits, the Fifth Circuit stretched the facts of Banks to apply it to Veeck, and disregarded several significant facts that distinguished Veeck from Banks. First, privately authored model codes are not identical to judicial opinions and statutes because model codes are not created by publicly paid elected officials. The government does not contract with the authors of these standards to write the codes. Further, no established relationship exists between public officials and independent standards writers in which the authors are compensated by public funds.

Second, the codes and standards at issue in these cases are technical documents, narrowly focused on specific, detailed areas of law such as building codes. The detail and complexity they embody require the author to have technical expertise, and the scope of their effect is much narrower than judicial opinions and statutes that have broad application. Thus, the facts of Veeck are significantly different than those faced by the Banks court over a century ago. The Fifth Circuit should have followed the reasoning of the Second and Ninth Circuit instead of applying Banks to privately authored codes.

In addition to Banks, the Fifth Circuit incorrectly relied on the BOCA case from the First Circuit, which was never decided on the merits. Although the Veeck opinion purported to follow BOCA, the Veeck court was the first to squarely address the copyright effect when a private

253. Id. at 74.
254. County of Suffolk, N.Y. v. First Am. Real Estate Solutions, 261 F.3d 179, 194 (2d Cir. 2001).
256. See id. at 794.
257. See id. at 814 (Weiner, J., dissenting).
author’s laws are adopted verbatim. The Veeck court erroneously based its response to this novel question on an overly broad reading of the BOCA decision. The Fifth Circuit relied heavily on the First Circuit’s dicta in BOCA, which stated “[T]he public’s essential due process right of free access to the law...can be reconciled with the exclusivity afforded a private copyright holder.” From this, the Fifth Circuit jumped to the conclusion that the proper way to ensure public access to the law was to terminate SBCCI’s copyright.

The Fifth Circuit’s reliance on BOCA was misplaced because the First Circuit specifically avoided deciding whether BOCA’s model code retained its copyright after enactment. The BOCA court recognized that “the rule denying copyright protection to judicial opinions and statutes grew out of a much different set of circumstances than do these technical regulatory codes...” In fact, the BOCA court even highlighted the possible trend towards federal and state adoption of these privately authored works. The First Circuit observed that code-drafting organizations served an increasingly “important public function,” and left open whether the copyright exception for judicial opinions established in the 1880s should be adapted “to accommodate modern realities” of drafting technical codes. Thus, the First Circuit distinguished BOCA from older cases such as Banks and Wheaton, and did not reach the question presented in Veeck.

D. The U.S. Supreme Court Should Reinforce the Protections Already Existing for Privately Authored Materials in Copyright Law

Protecting the copyright of privately authored codes does not violate due process principles requiring free access to the law. The three judge panel of the Fifth Circuit noted the importance of the “free accessibility of the law,” and observed that if privately authored materials adopted into law are not accessible to the public, that lack of access would violate

259. See Veeck, 293 F.3d at 811–12 (Weiner, J., dissenting).
260. See id. at 798–800.
261. Id. at 803 (citing Bldg. Officials & Code Adm’rs, 628 F.2d at 730).
262. See id. at 799–800.
263. Bldg. Officials & Code Adm’rs, 628 F.2d at 736.
264. Id.
265. Id.
266. Id.
due process. However, in *Veeck* there was no evidence that the privately authored codes were not accessible to the public. The controversy arose in *Veeck* not because Veeck had been denied access to the law, but because he wanted to post a full copy of the code on the internet. Nothing prevented Veeck or anyone else from accessing the law at a clerk’s office.

In addition, requiring users to pay a licensing fee to get a copy of the text of a building code does not violate due process. Free accessibility of the law is different than free unlimited copying of the law. Veeck had free access to the law. He could have viewed or copied the SBCCI codes in any city clerk’s or other municipal offices in the towns that had enacted the codes by reference. Access to building codes means that a person can consult the codes. This access was present in *Veeck*, because any person could have visited the town offices, read the code, and made a personal copy. In contrast, what Veeck wanted was a copy of the full text of the model code provisions in order to place them on his website. The due process requirement of free access to the law does not guarantee everyone a personal copy of the law for free.

The Fifth Circuit’s decision in *Veeck* focuses too heavily on the public’s ability to access the law and does not adequately consider an author’s right to copyright protection. In contrast, the Second Circuit’s decision in *County of Suffolk* strikes a proper balance by providing an example of how copyright interests can coexist with the public’s interest in accessing the law. The Second Circuit determined that Suffolk County could maintain its copyrights in its official tax maps without violating New York’s Freedom of Information Law (FOIL). Further, the Second Circuit held that under FOIL, the extent of the state agency’s obligation to make the law accessible is to make its records available for public inspection and copying. Like the tax maps in *County of Suffolk,*

---


268. *Id.*


271. *See id.*

272. *Id.* at 793.

273. *See id.* at 812 (Weiner, J., dissenting) (noting that “the type of due process asserted by *Veeck* is murky at best”).


276. *Id.* at 192.
the building codes in *Veeck* also contain original material, research, compilation, and organization. Moreover, the building codes, like the tax maps, are the result of substantial work, effort, and private expense.\(^{277}\) The importance of protecting public access to the law as expressed in *County of Suffolk* is similar to the concerns expressed in *Veeck*.\(^{278}\) Unlike the Fifth Circuit, the Second Circuit resolved this concern by recognizing that the fair use doctrine protects free press and individuals seeking to use state agency records to educate other citizens or criticize the government.\(^{279}\)

The Fifth Circuit overlooked the “fair use” doctrine, a critical copyright concept that provides access to the law while still protecting the copyright holder’s proprietary rights in the material.\(^{280}\) The fair use doctrine, as codified in the Copyright Act, strikes a balance between the rights of a copyright holder and the interest of the public in accessing information.\(^{281}\) The fair use doctrine gives courts the flexibility to determine a proper outcome on a case-by-case basis, a flexibility in stark contrast to the per se rule created by the *Veeck* court.

The Fifth Circuit in *Veeck* viewed “the law” as publicly owned and therefore in the public domain for any use.\(^{282}\) The court listed acceptable uses of works in the public domain, including reproducing copies of the law for purposes such as influencing legislation, educating a neighborhood association, or simply to amuse.\(^{283}\) However, the court failed to notice that these uses are similar to the permissible uses under the fair use doctrine. What Fifth Circuit overlooked is that a legal code need not be in the public domain in order to be available for fair use.\(^{284}\) A code can be copyrighted but still accessible to citizens for educational purposes, news reporting, and criticism under the fair use affirmative defense.\(^{285}\) One of the Fifth Circuit’s justifications for why a code’s copyright must terminate when it is incorporated into law was the

---


278. *See id.* at 799–800; *Suffolk*, 261 F.3d 179 at 192–93.


280. *KELLER & CUNARD, supra* note 26, at § 8:3.1.

281. 17 U.S.C. § 107 (2000) (The “fair use of a copyrighted work . . . for purposes such as criticism, comment, news, reporting, teaching, scholarship, or research, is not an infringement of copyright.”).

282. *Veeck*, 293 F.3d at 799–800.

283. *Id.*

284. *See supra* Part I.E.

285. *See supra* Part I.E.
concern that a copyright owner might refuse to make the materials accessible on a reasonable basis.\textsuperscript{286} However, the fair use doctrine eliminates this justification by permitting courts to strike a balance between the competing interests of protecting property rights and ensuring public access.\textsuperscript{287}

The Fifth Circuit created a split among the federal Circuit Courts with its decision in \textit{Veeck} and created an unprecedented exception to copyright law.\textsuperscript{288} The Supreme Court has the opportunity to resolve this split by granting SBCCI’s petition for certiorari.\textsuperscript{289} By relying on the fair use doctrine, the U.S. Supreme Court could provide lower courts with a flexible test that balances the interests of the public and provides private authors with the copyright protection guaranteed to them by the Copyright Act.

IV. CONCLUSION

The Fifth Circuit’s decision in \textit{Veeck} created an unprecedented exception to the Copyright Act by establishing the blanket rule that once a model code or standard is enacted into law, the authoring organization’s copyright automatically terminates and the code enters the public domain. This conflicts with the Second and Ninth Circuits, which have held that a private organization retains its copyright in material later adopted into law. A private organization’s copyright in model codes and standards should be protected as long the law is still available to the public. The U.S. Supreme Court should ensure this protection by granting certiorari in \textit{Veeck} and creating a rule that relies on established principles of copyright law, such as fair use, in evaluating what is an acceptable use of copyrighted material on a case-by-case basis. This result ensures that the author’s proprietary rights are properly balanced against the public’s due process right of access to the law.

\textsuperscript{286} \textit{Veeck}, 293 F.3d at 800.
\textsuperscript{287} \textit{See} KELLER \& CUNARD, supra note 26, at § 8:3.1.
\textsuperscript{288} \textit{See supra} Part II.
CONUNDRUMS WITH PENUMBRAS: PREEMBRYOS, THE RIGHT TO PRIVACY, AND NON-GAMETE PROVIDERS WHO INTEND TO BE PARENTS

Lainie M. C. Dillon

Abstract: To date, five state high courts have resolved disputes over frozen preembryos. These disputes arose during divorce proceedings between couples who had previously used assisted reproduction and cryopreserved excess preembryos. In each case, one spouse wished to have the preembryos destroyed, while the other wanted to be able to use or donate them in the future. The parties in these cases invoked the constitutional right to privacy to argue for dispositional control over the preembryos; two of the five cases were resolved by relying on this right. The constitutional right to privacy protects intimate decisions involving procreation, marriage, and family life. However, when couples use donated sperm or ova to create preembryos, a unique circumstance arises: one spouse—the gamete-donor—is genetically related to the preembryos and the other is not. If courts resolve frozen preembryo disputes that involve non-gamete providers based on the constitutional right to privacy, they should find that the constitutional right to privacy encompasses the interests of both gamete and non-gamete providers. Individuals who create preembryos with the intent to become a parent have made an intimate decision involving procreation, marriage, and family life that falls squarely within the right to privacy. In such cases, the couple together made the decision to create a family through the use of assisted reproduction, and the preembryos would not exist but for that joint decision. Therefore, gamete and non-gamete providers should be afforded equal constitutional protection in disputes over frozen preembryos.

Leny and Eva were a married couple eager to have a baby. However, they were unable to achieve pregnancy through traditional means because Eva was born with a uterus but no ovaries. In order to have children, they decided to use Leny’s sperm, an anonymous egg donor, and in vitro fertilization (IVF). They signed up at an infertility clinic, which combined Leny’s sperm with the donated eggs, and seventeen preembryos resulted. Two preembryos were implanted into Eva’s uterus,

1. Hypothetical created by the author.
2. In vitro fertilization is a process by which ova, which are provided by either the intended mother or a donor, are placed in a medium and then fertilized by sperm, which is provided by either the intended father or a donor. See STEDMAN’S MEDICAL DICTIONARY 657 (27th ed. 2000). The preembryo that results is then implanted into a uterus of either the intended mother or a surrogate and brought to term. Id.
and the others were cryopreserved\textsuperscript{4} for possible use in the future. Eva gave birth to a healthy child, and they were happy parents. However, their marriage was troubled and the couple divorced.

During the dissolution proceedings, Leny and Eva disagreed about how to dispose of the frozen preembryos. Leny wanted to donate them to an infertile couple, but Eva wanted the opportunity to use the preembryos herself in the future. Because he had a biological connection to the preembryos and Eva did not, Leny asserted that the constitutional right to privacy granted him exclusive control over the preembryos. Eva argued that the right to privacy also sheltered her interests in the preembryos—despite the missing biological connection—and granted her an equal constitutional claim over them.

Only five frozen preembryo disputes between divorcing couples have reached state high courts.\textsuperscript{5} Parties have asserted their constitutional right to privacy in each—usually described in this circumstance as the “right to procreate” or “right not to procreate”—as a basis for determining which spouse will control the disposition of the frozen preembryos.\textsuperscript{6} In four of the five cases, both parties were gamete providers;\textsuperscript{7} they contributed their own cells, or gametes, to the preembryos.\textsuperscript{8} However, it is also common for people to make use of gamete donors and surrogates,\textsuperscript{9} as did the couple in a recent Washington State Supreme Court case, Litowitz v. Litowitz.\textsuperscript{10} In Litowitz, the Washington State Supreme Court stated in dicta that because the wife did not have a biological connection

\begin{thebibliography}{10}
\bibitem{Grobstein} Clifford Grobstein, \textit{Human Development from Fertilization to Birth}, in \textit{Encyclopedia of Bioethics} 847 (Warren Thomas Reich ed., 1995)).

\bibitem{Cryopreservation} Cryopreservation is a method of preserving the “viability of excised tissues or organs at extremely low temperatures.” \textit{Stedman’s, supra} note 2, at 432.


\bibitem{Davis} See Davis, 842 S.W.2d at 600 (stating that “[h]ere, the specific individual freedom in dispute is the right to procreate”); \textit{see also J.B.}, 783 A.2d at 711 (stating that the husband had asserted his right to procreate); Litowitz v. Litowitz, 102 Wash. App. 934, 943–44, 10 P.3d 1086, 1092–93 (2000) (holding that the husband’s right not to procreate compelled the court to give him complete control over the disposition of the preembryos), \textit{rev’d}, 146 Wash. 2d 514, 48 P.3d 261 (2002).

\bibitem{Gamete} A gamete is “[a]ny germ cell, whether ovum or spermatozoon.” \textit{Stedman’s, supra} note 2, at 725. A person is a gamete provider if his sperm or her ovum are used to create preembryos.

\bibitem{Kass} See A.Z., 725 N.E.2d at 1053; J.B., 783 A.2d at 709–10; Kass, 696 N.E.2d at 175–76; \textit{Davis, 842 S.W.2d at 589.}


\bibitem{Litowitz1} 146 Wash. 2d 514, 517, 48 P.3d 261, 262 (2002).
\end{thebibliography}
to the preembryos, the only way she could have dispositional control over them was through the contract she and her husband signed with the fertility clinic. The Litowitz court thus implied that, if the constitutional right to privacy had controlled the outcome of the dispute rather than the contract, the husband’s procreational rights would have trumped his wife’s interests merely because he had a biological connection to the preembryos and she did not.

In resolving frozen preembryo disputes, courts have a choice about what doctrine to apply: they may enforce a contract if one exists, they may resolve the dispute based on public policy, they may characterize the preembryos as marital property and dispose of them accordingly, or they may apply the right to privacy. If courts apply the constitutional right to privacy and follow the Washington State Supreme Court’s dicta regarding non-gamete providers, they will significantly, and unconstitutionally, infringe on non-gamete providers’ rights. Instead, if

---

11. See id. at 527, 48 P.3d at 267.
12. See id. (stating that because the wife did not have a biological connection to the preembryos, “[a]ny right she may have to the preembryos must be based solely upon contract”).
13. See, e.g., Kass, 696 N.E.2d at 182; Litowitz, 146 Wash. 2d at 533–34, 48 P.3d at 271.
15. No state high court has yet applied this approach, but scholars and commentators have recommended it. See, e.g., Brief of Amici Curiae Northwest Women’s Law Center, at 11, Litowitz v. Litowitz, 102 Wash. App. 934, 10 P.3d 1086 (2000) (No. 70413-9).
16. See, e.g., J.B. v. M.B., 783 A.2d 707, 715–17 (N.J. 2001); Davis v. Davis, 842 S.W.2d 588, 598–605 (Tenn. 1992). Some argue that it is both unnecessary and unwise for courts to apply constitutional rights to resolve preembryo disputes. See Brief of Amici Curiae Northwest Women’s Law Center, at 11 (No. 70413-9) (stating that “[b]ecause this case can and should be decided by resort to contract or community property law, the constitutional issue, if there is one, should not be reached”). The amicus brief argues that preembryos are marital property and, therefore, subject to contract law or equitable distribution under state law. See id. at 5–8. The brief further argues that there are “unknown and almost certainly undesirable consequences” to applying constitutional rights to preembryo disputes because doing so could enable once-anonymous gamete donors to claim that they have procreational rights in preembryos well after their gametic contribution has been made. Id. at 13–14. See also Radhika Rao, Reconceiving Privacy: Relationships and Reproductive Technology, 45 UCLA L. REV. 1077, 1122–23 (1998) (Arguing that these disputes should not be resolved based on constitutional rights because when each party possesses a personal right of privacy there is a constitutional indeterminacy in which neither one is necessarily protected. “In such areas of constitutional indeterminacy—of tragic choices between competing interests—there is no single necessarily correct constitutional resolution to a controversy.”). This Comment does not seek to argue that applying constitutional rights is the most desirable means to resolve preembryo disputes; instead, it seeks to argue that if courts do apply a constitutional rights balancing test to resolve such disputes, they should find that the right to privacy applies equally to gamete and non-gamete providers.
17. While a court may also apply the constitution of its own state, this Comment concerns only the application of the federal constitutional right to privacy.
courts apply constitutional principles in preembryo disputes where one party is a gamete provider and the other is not, they should find that the right to privacy protects both parties’ interests in the preembryos. Support for this approach can be found in the analogous area of family law where intent, not biology, determines parentage. As in parentage law, intent should determine whose interests the constitutional right to privacy protects in disputes over frozen preembryos.

This Comment argues that because the constitutional right to privacy broadly protects intimate decisions related to procreation, marriage and family life, a non-gamete provider’s intimate decision to create preembryos falls squarely within the zone of privacy protected by the federal Constitution. Part I describes the history of the constitutional right to privacy and its protection of intimate decisions related to procreation, marriage and family life. Part II discusses how courts have thus far applied the right to privacy to disputes over frozen preembryos. Part III provides an overview of parentage law, where intent is used to determine parentage for children born from the use of assisted reproduction. In Part IV, this Comment argues that in deciding whether the right to privacy encompasses the decision to create preembryos, courts should find that the Constitution protects both gamete and non-gamete providers. The creation of preembryos with the intent to become a parent is an intimate decision that is fundamental to procreation and family. Therefore, it falls squarely within the constitutional right to privacy—regardless of a biological connection.

I. THE RIGHT TO PRIVACY PROVIDES BROAD PROTECTION OVER INTIMATE DECISIONS CENTRAL TO PROCREATION, MARRIAGE, AND FAMILY LIFE

The United States Supreme Court, in its constitutional jurisprudence, has consistently applied the right to privacy to protect intimate decisions central to procreation, marriage, and family. The Court began to define

18. See infra Part III.


the nature of this right in seminal cases prior to its explicit recognition of the right to privacy. The Court has held that the U.S. Constitution protects both an individual’s basic procreative autonomy as well as the broader decisions involved with the rearing of one’s children. The Court later built on that foundation by explicitly stating that a right to privacy exists in the “penumbras” of the Bill of Rights. The Court has since applied this right to strike down laws that unduly restricted marriage, contraception, family relationships, child rearing and education.

A. Origins of the Right to Privacy

The genesis of the constitutional right to privacy can be found in *Skinner v. Oklahoma*, *Pierce v. Society of Sisters* and *Meyer v. Nebraska*. In *Skinner*, the U.S. Supreme Court recognized the right to procreate as “one of the basic civil rights of man.” In striking down a state law that sought to sterilize certain types of criminals, the Court held that the law was unconstitutional in part because “[m]arriage and procreation are fundamental to the very existence and survival of the race.”


21. See, e.g., *Skinner v. Oklahoma*, 316 U.S. 535, 541 (1942) (stating that procreation involves the basic civil rights of man); *Pierce v. Soc’y of Sisters*, 268 U.S. 510, 534–35 (1925) (stating that the right to educate is fundamental to marriage and child rearing); *Meyer v. Nebraska*, 262 U.S. 390, 399–401 (1923) (stating that the concept of liberty includes the right to learn a foreign language).

22. See *Skinner*, 316 U.S. at 541.


24. See *Griswold*, 381 U.S. at 484–85.

25. See *Loving*, 388 U.S. at 12 (striking down a law that criminalized interracial marriage).


27. See *Moore v. City of East Cleveland*, 431 U.S. 494, 495–97, 503–06 (1977) (holding that the state cannot prevent a grandmother from living with her grandchildren); see also *Prince v. Massachusetts*, 321 U.S. 158, 166 (1944) (stating that there is a “private realm of family life which the state cannot enter”).

28. See *Meyer*, 262 U.S. at 402–03 (holding that the city cannot prevent the teaching of foreign languages in elementary schools); *Pierce*, 268 U.S. at 534–35 (holding that the state cannot prevent parents from educating their children in private or parochial schools).


31. 262 U.S. 390 (1923).

32. See *Skinner*, 316 U.S. at 541.

33. Id.
Pierce and Meyer were the first Supreme Court cases to broadly characterize the general area of family life as sheltered from state intrusion.\textsuperscript{34} In Pierce, the Court struck down a law requiring children to attend public schools.\textsuperscript{35} The Court based its decision on a parent’s right to educate his or her children in a school of his or her choice, whether public, private or parochial.\textsuperscript{36} The Court held that this right to choose how to educate one’s children stems from the First and Fourteenth Amendments, which protect decisions that are fundamental to marriage and child rearing.\textsuperscript{37}

In Meyer, an elementary school teacher taught German to a ten-year-old in violation of a state statute that prohibited teaching foreign languages in elementary schools.\textsuperscript{38} The Court held that the statute violated the Constitution because the Fourteenth Amendment’s concept of liberty protects the right of a teacher to teach and of students to acquire knowledge.\textsuperscript{39} The Court noted that this protection also encompasses:

\begin{quote}
[T]he right of the individual to contract, to engage in any of the common occupations of life, to acquire useful knowledge, to marry, establish a home and bring up children, to worship God according to the dictates of his own conscience, and generally to enjoy those privileges long recognized at common law as essential to the orderly pursuit of happiness by free men.\textsuperscript{40}
\end{quote}

Forty years after Meyer, in Griswold v. Connecticut,\textsuperscript{41} the Supreme Court explicitly identified this broad shelter as the right to privacy.\textsuperscript{42}

\begin{thebibliography}{9}
\bibitem{34} See Jed Rubenfeld, The Right of Privacy, 102 Harv. L. Rev. 737, 743 (1989) (stating that Pierce and Meyer “may be seen as the true parents of the privacy doctrine” and that today these cases are frequently classified with other privacy cases); see also Rao, supra note 16, at 1093 (stating that the right to privacy’s earliest origins lie in these two cases); Moore v. City of East Cleveland, 431 U.S. 494, 499 (1977) (stating that “[a] host of cases, tracing their lineage to Meyer v. Nebraska, 262 U.S. 390, 399–401 (1923), and Pierce v. Society of Sisters, 268 U.S. 510, 534–35 (1925), have consistently acknowledged a "private realm of family life which the state cannot enter"”).
\bibitem{35} See Pierce, 268 U.S. at 534–35.
\bibitem{36} See id. at 535.
\bibitem{37} See id. at 534–35.
\bibitem{38} Meyer, 262 U.S. at 396–97.
\bibitem{39} See id. at 399–401; see also Richard C. Turkington & Anita L. Allen, Privacy Law: Cases and Materials 612 (1999) (stating that the Supreme Court determined in Meyer that under the Fourteenth Amendment individuals have certain fundamental rights which must be respected by state law, including private decisions related to education).
\bibitem{40} Meyer, 262 U.S. at 399.
\bibitem{41} 381 U.S. 479 (1965).
\end{thebibliography}
B. The Right to Privacy Protects Intimate Decisions Related to Procreation

The U.S. Constitution does not explicitly state that there is a right to privacy, but “the Court has recognized that a right of personal privacy, or a guarantee of certain areas or zones of privacy does exist under the Constitution.” This right emanates from the penumbras of the specific guarantees in the Bill of Rights, which shelter individuals from government intrusion into speech, religion, assembly, and one’s home and property. The constitutional right to privacy is also based in the Fourteenth Amendment, which provides that no state shall “deprive any person of life, liberty, or property, without due process of law; nor deny to any person within its jurisdiction the equal protection of the laws.”

In Griswold, the U.S. Supreme Court applied the right to privacy to strike down a Connecticut law making it illegal for a married couple to use “any drug, medicinal article or instrument for the purpose of preventing conception . . . .” The Court viewed the law as an undue intrusion on an intimate decision central to marriage, which the court recognized as “a relationship lying within the zone of privacy created by several fundamental constitutional guarantees.” These guarantees include those found in the First, Fourth, Fifth and Ninth Amendments.

---

42. See id. at 484-86; see also Rao, supra note 16, at 1095 (stating that Griswold was the first case in which the Supreme Court explicitly recognized the right to privacy).
43. See Roe v. Wade, 410 U.S. 113, 152 (1973); see also Vincent F. Stempel, Procreative Rights in Assisted Reproductive Technology: Why the Angst?, 62 ALB. L. REV. 1187, 1193 (1999) (stating that although the Constitution does not specifically refer to a right of privacy, over the years the Supreme Court has developed the doctrine of the right to privacy).
44. See Griswold v. Connecticut, 381 U.S. 479, 484 (1965); see also Stempel, supra note 43, at 1193 (stating that the Court has identified the penumbras of the specific guarantees in the Bill of Rights as the source of the right to privacy).
45. See Meyer, 262 U.S. at 399; Roe, 410 U.S. at 163.
46. U.S. CONST. amend. XIV.
47. See Griswold, 281 U.S. at 480 (quoting CONN. GEN. STAT. § 53-32 (1958 rev.) (repealed 1971)).
48. See id. at 484-86; see also ROBERT BLANK & JANNA C. MERRICK, HUMAN REPRODUCTION, EMERGING TECHNOLOGIES, AND CONFLICTING RIGHTS 11 (1995) (stating that in Griswold the Supreme Court forged a right of marital privacy based on the First Amendment right of association, the Third Amendment prohibition against quartering soldiers in peacetime, the Fourth Amendment prohibition against unreasonable searches and seizures, the Fifth Amendment protection against self-incrimination, and the Ninth Amendment’s vesting of rights not enumerated in the Constitution).
49. See Griswold, 381 U.S. at 484 (“The right of association contained in the penumbra of the First Amendment” is one of the rights that defines the zone of privacy.).
While the Court in *Griswold* explicitly expressed a desire to protect decisions that are fundamental to marriage, it also made clear in later cases that the right to privacy extends beyond the marital context. For example, in *Eisenstadt v. Baird* the Court struck down a Massachusetts law that made it a crime to distribute contraceptive devices to unmarried persons. The Court held that the statute was analogously unconstitutional to the Connecticut law in *Griswold* and concluded that, “whatever the rights of the individual to access . . . contraceptives may be, the rights must be the same for the unmarried and the married alike.” Procreational rights, the Court stated, are fundamental to every individual: “[i]f the right of privacy means anything, it is the right of the individual, married or single, to be free from unwarranted governmental intrusion into matters so fundamentally affecting a person as the decision whether to bear or beget a child.”

The Supreme Court further extended the right to privacy in *Roe v. Wade*, in which the Court held that pregnant women have a right to an abortion under some circumstances. Because the constitutional right to privacy protects the decision of whether and how to create a family, the Court determined that a woman has a right to have an abortion unless there is a sufficiently compelling state interest to prevent it. Therefore, in *Roe*, the Court struck down a Texas statute that prohibited abortion

---

50. See id. (“The Third Amendment in its prohibition against the quartering of soldiers ‘in any house’ in time of peace without the consent of the owner is another facet of that privacy.”).
51. See id. (“The Fourth Amendment explicitly affirms the ‘right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures.’”).
52. See id. (“The Fifth Amendment in its Self-Incrimination Clause enables the citizen to create a zone of privacy which government may not force him to surrender to his detriment.”).
53. See id (citing U.S. CONST. amend IX, which provides: “[t]he enumeration in the Constitution, of certain rights, shall not be construed to deny or disparage others retained by the people”).
55. 405 U.S. 438 (1972).
56. See id. at 454–55.
57. See id. at 445–49.
58. Id. at 453.
59. Id. (emphasis added); see also Stempel, supra note 43, at 1193–94 (stating that individual privacy rights were first set forth in *Eisenstadt*).
60. 410 U.S. 113 (1973).
61. See id. at 164–65.
62. See id. at 154, 163 (holding that a state may not regulate abortion before viability because “[w]ith respect to the State’s important and legitimate interest in potential life, the ‘compelling’ point is at viability”); see also Planned Parenthood v. Casey, 505 U.S. 833, 846 (1992) (stating that “[l]evel at viability, the State’s interests are not strong enough to support a prohibition of abortion or the imposition of a substantial obstacle to the woman’s effective right to elect the procedure”).
Preembryos and the Right to Privacy

except when necessary to save the life of the woman, holding that the right to privacy “is broad enough to encompass a woman’s decision whether or not to terminate her pregnancy.” Moreover, in a subsequent case addressing abortion rights, Planned Parenthood v. Casey, the Court reaffirmed that the right to privacy protects “the liberty relating to intimate relationships, the family, and decisions about whether or not to beget or bear a child.”

C. The Right to Privacy is a Broad Protection Encompassing the Right to Marry, the Right to Live with One’s Family, and Other Intimate Decisions Central to Highly Personal Relationships

The U.S. Supreme Court has extended the right to privacy to include situations involving marriage and living arrangements, and the Court has refused to apply it to situations that do not involve highly personal family decisions. For example, in Loving v. Virginia, the Court held that a Virginia law criminalizing interracial marriage violated the Equal Protection clause. In his opinion for the Court, Chief Justice Warren drew on preceding cases stating that the Constitution protected decisions related to marriage. He explained that “[t]he freedom to marry has long been recognized as one of the vital personal rights essential to the orderly pursuit of happiness by free men.” Justice Warren went on to specifically note that the right to privacy protects one’s decision of whom to marry because “[m]arriage is one of the ‘basic civil rights of man,’ fundamental to our very existence and survival.”

63. Roe, 410 U.S. at 153.
66. See Loving v. Virginia, 388 U.S. 1, 12 (1967) (applying the right to privacy to marriage); Moore, 431 U.S. at 503 (applying the right to privacy to living arrangements).
67. See Roberts v. U.S. Jaycees, 468 U.S. 609, 618–20 (1984) (refusing to apply the right to privacy to a men’s club); see also Paris Adult Theatre v. Slaton, 413 U.S. 49, 66 n.13 (1973) (stating that because “the constitutionally protected privacy of family, marriage, motherhood, procreation, and child rearing is not just concerned with a particular place, but with a protected intimate relationship,” it does not encompass the right to watch obscene movies in public theaters).
68. 388 U.S. 1 (1967).
69. See id. at 12.
70. See id. at 7 (citing Meyer v. Nebraska, 262 U.S. 390 (1923); Skinner v. Oklahoma, 316 U.S. 535 (1942)).
71. Id. at 12.
72. See id. (citing Skinner, 316 U.S. at 541).
Familial living arrangements may also fall within the “private realm of family life which the state cannot enter.”73 In Moore v. City of East Cleveland,74 the Supreme Court struck down a zoning ordinance that limited occupancy of dwelling units to members of a nuclear family, thereby making it a crime for a woman to have her grandchildren live with her.75 East Cleveland defended the law by arguing that the right to privacy did not extend beyond the nuclear family, because the Court had thus far only applied it to parent-child relationships.76 The Supreme Court rejected this argument, reasoning that the constitutional protection of the “sanctity of the family” includes one’s extended family.77 This protection applies broadly to families, the Court explained, because “[i]t is through the family that we inculcate and pass down many of our most cherished values, moral and cultural.”78

The Supreme Court has further delineated the scope of the right to privacy by limiting it to highly personal family decisions.79 For example, in Roberts v. United States Jaycees,80 the Court held that members of an all-male organization did not have a constitutional right to exclude women, because the organization was not the type of “highly personal relationship” the Constitution seeks to protect.81 In its holding, the Court emphasized that constitutional protection extends to those relationships that “attend the creation and sustenance of a family—marriage, childbirth, the raising and education of children, and cohabitation with one’s relatives.”82 Thus, the nature and activities of the Jaycees fell outside the zone of privacy protected by the Constitution.83

In sum, the constitutional right to privacy protects intimate decisions central to procreation, marriage and family life. As indicated by Supreme Court jurisprudence, the constitutional right to privacy is very broad. Its protection ranges from the intimate decisions related to creating a

73. See Moore v. City of East Cleveland, 431 U.S. 494, 499 (1977) (citing Prince v. Massachusetts, 321 U.S. 158, 166 (1944)).
75. See id. at 495–97.
76. See id. at 500.
77. See id. at 503–06.
78. See id. at 503–04.
81. See id. at 618–20 (finding that the Jaycees were the type of organization clearly “outside of the category of relationships worthy of this kind of constitutional protection”).
82. Id. at 619 (internal citations omitted).
83. See id. at 631 (O’Connor, J., concurring).
family—such as whom to marry and whether to have children—to decisions related to child-rearing and living arrangements.

II. FIVE STATE HIGH COURTS HAVE RESOLVED PREEMBRYO DISPUTES

Five state high courts have examined the issue of who has dispositional control over frozen preembryos upon divorce. The constitutional right to privacy was raised in each, but only two courts dealt with it directly. Those courts undertook a two-step process to resolve the disputes: first, they determined whether both spouses had constitutional rights in the preembryos; second, after concluding that they had equal constitutional rights, the courts weighed the spouses’ competing interests in the frozen preembryos. The three courts that avoided the constitutional issues nonetheless acknowledged that preembryo disputes raise constitutional privacy issues.

A. Courts Have Resolved Preembryo Disputes Based on the Constitutional Right to Privacy

In disputes over frozen preembryos, parties have consistently asserted their constitutional privacy rights in arguing for dispositional control over the preembryos. The Supreme Court of Tennessee was the first


85. See J.B., 783 A.2d at 717 (stating that the wife’s “fundamental right not to procreate” would be “irrevocably extinguished” if a surrogate were allowed to bear her preembryos); Davis, 842 S.W.2d at 602 (stating that “however far the protection of procreational autonomy extends, the existence of the right itself dictates that decisional authority rests in the gamete-providers”).

86. See Davis, 842 S.W.2d at 601, 603–05 (stating that “Mary Sue Davis and Junior Lewis Davis must be seen as entirely equivalent gamete-providers” and then proceeding to balance each party’s interests); see also J.B., 783 A.2d at 715–17 (stating that “the claims before us derive, in part, from concepts found in the Federal Constitution and the Constitution of this State” and proceeding to balance each party’s interests).

87. See Kass, 696 N.E.2d at 179 (resolved based on a contract); A.Z., 725 N.E.2d at 1059 (resolved based on public policy); Litowitz, 146 Wash. 2d at 533–34, 48 P.3d at 271 (resolved based on a contract). See also infra Part II.B.

88. See J.B., 783 A.2d at 711 (stating that the husband had asserted his right to procreate); Davis, 842 S.W.2d at 600 (determining that “[h]ere, the specific individual freedom in dispute is the right to procreate”); Litowitz v. Litowitz, 102 Wash. App. 934, 944–45, 10 P.3d 1086, 1092–93 (2002) (holding that David Litowitz, as a progenitor, had a right to procreate), rev’d, 146 Wash. 2d 514, 48 P.3d 261 (2002).
state high court to resolve such a dispute. In *Davis v. Davis* a Tennessee couple divorced after previously undergoing in vitro fertilization (IVF) and freezing preembryos for possible use in the future. Because the couple had used the wife’s ova and the husband’s sperm to create the preembryos, they were both the gamete providers and the intended parents. Upon divorce, the couple disagreed about what should be done with the preembryos. The wife, Mary Sue Davis, wished to have the preembryos donated to an infertile couple, while the husband, Junior Davis, wanted them either to remain frozen or to be discarded.

The Supreme Court of Tennessee engaged in a constitutional rights balancing test to resolve the dispute. First, the court determined that Mary Sue Davis and Junior Davis both had privacy rights in the preembryos because “the right of procreation is a vital part of an individual’s right to privacy.” To reach this conclusion, the court analogized control over preembryos to those acts the U.S. Supreme Court has explicitly stated fall under the right to use contraception and to have an abortion. Because the court determined that both spouses had equal constitutional rights in the preembryos, its second step was to compare their competing interests in them. The *Davis* court reasoned that Mary Sue’s interest in the preembryos was the “burden of knowing that the lengthy IVF procedures she underwent were futile.” The court based Junior’s interests in the preembryos on his strong desire to avoid unwanted fatherhood.

---

89. See Stempel, supra note 43, at 1192 (stating that *Davis* was the first case to deal directly with the issue of frozen preembryos).
90. 842 S.W.2d 588 (Tenn. 1992).
91. See id. at 589.
92. See id. at 591–92.
93. See id. at 589.
94. See id. at 589–90.
95. See id. at 598–605.
96. See id. at 600.
98. See *Davis*, 842 S.W.2d at 603–04.
99. See id. at 604.
100. Junior Davis had been a child of divorce and was “vehemently opposed” to becoming the father to a child that would not live with both parents. See id. at 603–04. His concerns included both
court concluded that Junior Davis’s interests were more compelling, and, therefore, the preembryos were destroyed. Notably, the Supreme Court of Tennessee also stated that Mary Sue’s interests would have been stronger if she had wanted to use the preembryos for herself and if they were the only means by which she could become a parent.

In *J.B. v. M.B.*, the Supreme Court of New Jersey echoed the Tennessee court’s determination that individuals have constitutional rights in their frozen preembryos. In *J.B.*, a New Jersey couple divorced after previously undergoing IVF and freezing preembryos. As in *Davis*, both parties were gamete providers who disagreed about what should be done with the preembryos; the husband wanted to donate them to an infertile couple, while the wife wanted them destroyed. Both parties argued that they had a constitutional right, rooted in the right to privacy, to control the disposition of the preembryos. Like the *Davis* court, the Supreme Court of New Jersey engaged in a constitutional rights balancing test to resolve the dispute. The court first considered whether the Constitution protected both spouses’ interests, and, as in *Davis*, the court held that the right to privacy protected both the husband and wife’s interests in the preembryos. Next, the court compared the parties’ competing interests and concluded that the wife’s procreational rights would be “irrevocably extinguished” if her husband was allowed to have the preembryos implanted in another woman against the wife’s wishes. Therefore, the *J.B.* court ordered the preembryos destroyed.

---

101. See id. at 604–05.
102. See id. at 604.
104. See id. at 715–17.
105. See id. at 710.
106. See id. at 708–10.
107. See id. at 710.
108. See id. at 712.
110. See id. at 715–17.
111. See id. at 717.
112. See id. at 720.
In both *Davis* and *J.B.*, the courts made it clear that the parties’ constitutional rights in the preembryos were equal.\(^{113}\) The Tennessee court explicitly stated that, “[a]s they stand on the brink of potential parenthood, Mary Sue Davis and Junior Lewis Davis must be seen as entirely equivalent gamete-providers.”\(^{114}\) Therefore, under the constitutional right to privacy, both individuals had an equal right to determine the fate of the preembryos, because neither spouse had an inherently more compelling constitutional right than the other.\(^{115}\) Consequently, the courts proceeded to the second step of balancing the parties’ competing interests.\(^{116}\)

**B. Courts That Have Resolved Preembryo Disputes Without Relying on the Constitution Have Nonetheless Indicated That Such Disputes Implicate the Constitutional Right to Privacy**

Even when courts have resolved preembryo disputes without relying on the constitutional right to privacy, they have acknowledged that the creation of preembryos implicates the right.\(^{117}\) For example, in *Kass v. Kass*,\(^{118}\) New York’s highest court enforced a dispositional agreement the divorcing couple had made at the time they began the IVF procedure.\(^{119}\) The contract stated that if the couple was ever unable to mutually agree about the disposition of the preembryos, the IVF clinic should donate them to research.\(^{120}\) Though the court decided the case based on the contract, it also noted that the dispositional control of preembryos is “a quintessentially personal, private decision.”\(^{121}\)

---


\(^{114}\) See *Davis*, 842 S.W.2d at 601.

\(^{115}\) Id. at 603.

\(^{116}\) See *id.* (stating that the court would consider both spouses’ interests by examining “the positions of the parties, the significance of their interests, and the relative burdens that w[ould] be imposed by differing resolutions”); see also *J.B.*, 783 A.2d at 716–17.


\(^{119}\) See *id.* at 182.

\(^{120}\) See *id.* at 176–77 (noting the contract specifically stated that “[i]n the event that we...are unable to make a decision regarding the disposition of our stored, frozen pre-zygotes, we now indicate our desire for the disposition of our pre-zygotes and direct [that]...our frozen pre-zygotes may be examined by the IVF Program for biological studies and be disposed of by the IVF Program for approved research investigation as determined by the IVF Program”).

\(^{121}\) *Id.* at 180.
Preembryos and the Right to Privacy

In *A.Z. v. B.Z.*, the Supreme Judicial Court of Massachusetts similarly recognized the privacy interest that individuals have in their preembryos. In *A.Z.*, the court refused to enforce a contract between a divorcing couple on the basis that it violated Massachusetts public policy. The agreement stated that in the event that the couple failed to mutually agree about the disposition of the preembryos, the wife would have full control over them. The court held that public policy prevented it from enforcing the contract, because the husband had changed his mind about becoming a parent. This decision, the Supreme Judicial Court of Massachusetts stated, “enhances the ‘freedom of personal choice in matters of marriage and family life.’”

The Washington State Supreme Court confronted this issue in *Litowitz v. Litowitz*, although in somewhat different circumstances. In *Litowitz*, a couple created preembryos during their marriage by using the sperm of the husband, David Litowitz, and ova from an anonymous egg donor. Thus, unlike the other preembryo disputes, the wife, Becky Litowitz, was a non-gamete provider. Before the couple divorced, they had a preembryo implanted into a surrogate, and they cryopreserved the excess preembryos for possible future use. Upon divorce, David Litowitz wanted the remaining preembryos destroyed or donated to an infertile couple, while Becky Litowitz wanted them implanted in a surrogate and brought to term.

While the Washington State Supreme Court resolved the dispute on contract grounds, it stated in dicta that it would not extend the right to privacy to non-gamete providers. The court reasoned that because

122. 725 N.E.2d 1051 (Mass. 2000).
123. See id. at 1059.
124. See id. at 1056.
125. See id. at 1054 (the agreement stated that if the husband and wife “should become separated, [they] both agree[d] to have the embryo(s) . . . return[ed] to [the] wife for implant”).
126. See id. at 1059.
128. 146 Wash. 2d 514, 48 P.3d 261 (2002).
129. See id. at 517, 48 P.3d at 262.
130. See id.
131. See id.
132. See id. at 517, 48 P.3d at 262–63.
133. See id. at 520, 48 P.3d at 264.
134. See id. at 533–34, 48 P.3d at 271.
135. See id. at 527, 48 P.3d at 267.
Becky Litowitz did not have a biological connection to the preembryos, “[a]ny right she may have to the preembryos must be based solely upon contract.” In other words, the Washington court suggested that when an individual decides to have a child through IVF, his or her rights and obligations to those preembryos stem solely from contractual or biological grounds. Significantly, by the time of the divorce, Becky and David Litowitz’s child had been born. Becky Litowitz’s legal status as the child’s mother was never questioned, despite her lack of a biological connection and the use of an egg donor and surrogate. Nevertheless, the Washington State Supreme Court stated that “any” right to the preembryos would be based “solely” upon contract and, thus, that Becky Litowitz did not possess a constitutionally protected privacy interest in the preembryos.

In sum, the constitutional right to privacy consistently arises in disputes over the fate of frozen preembryos. To date, two state high courts have applied constitutional rights to resolve such disputes. Significantly, even courts that resolved the disputes on non-constitutional grounds agreed that preembryo disputes implicate privacy rights. However, in one case, Litowitz, a court suggested in dicta that non-gamete providers do not have a constitutional right to control the disposition of their preembryos. If courts follow this approach in future cases, non-gamete providers will be denied privacy rights in preembryos solely because they lack a biological connection.

III. FOR CHILDREN BORN FROM ASSISTED REPRODUCTION, INTENT—NOT BIOLOGY—DEFINES LEGAL PARENTAGE

Children born from the use of assisted reproduction are often not biologically related to their parents; they may be born from processes that involve sperm donation, egg donation, surrogacy, or a combination thereof. Because biological connections do not necessarily indicate who the parents of such children are, the law increasingly looks to the

---

136. See id.
137. See id.
138. See id. at 517, 48 P.3d at 262.
140. See Robertson, supra note 9.
141. See Janet L. Dolgin, Defining the Family: Law, Technology, and Reproduction in an Uneasy Age 174 (1997) (stating that it is now recognized that biological connections do not securely anchor society’s understanding of the family).
intent of the participants in the reproductive process to determine their legal rights. The Uniform Parentage Act of 2000 (UPA) applies the intent doctrine to determine the rights of individuals who use assisted reproduction. Under this doctrine, those who intend to create a family are the legal parents to children that result, while those who do not intend to create a family—such as anonymous gamete donors—do not have parental rights. Case law also uses intent, not biology, to determine parentage. In some cases, courts have required individuals with no biological connection to a child to uphold their parental obligations because they intended to create a family. In other cases, courts have denied parentage rights to individuals with a biological connection to a child because they lacked the initial intent to create a family.

A. The Uniform Parentage Act Calls for the Use of Intent to Determine Parentage of Children Born From Assisted Reproduction

Children born from the use of egg or sperm donation are not biologically related to one or both parents; therefore, intent has emerged as the primary analysis for determining legal parenthood. The UPA,

142. See UNIF. PARENTAGE ACT §§ 702 and 703, 9B U.L.A. 355–56 (2000) (stating that a gamete donor is not a parent to a child that is conceived by means of assisted reproduction and that a husband is the father of a child resulting from assisted reproduction if he consented to the use of assisted reproduction); see also In re Marriage of Buzzanca, 72 Cal. Rptr. 2d 280, 293–94 (Cal. Ct. App. 1998) (determining that the intended parents are the legal parents to a child born from an egg donor, sperm donor, and surrogate mother); Johnson v. Calvert, 851 P.2d 776, 782 (Cal. 1993) (granting maternity rights to the gamete-provider mother over the surrogate mother because the preembryos would not exist but for the gamete-provider mother’s intent to have children); McDonald v. McDonald, 608 N.Y.S.2d 477, 480 (N.Y. App. Div. 1994) (granting maternity rights to the intended mother, who was also the gestational mother).

143. See UNIF. PARENTAGE ACT § 702, 9B U.L.A. 355 (stating that a sperm or egg donor is not the parent of a child conceived by assisted reproduction).

144. See id. at § 703, 9B U.L.A. 356 (stating that if a husband consents to assisted reproduction by his wife he is the father of a resulting child, regardless of whether he provided sperm or not).

145. See id. at § 702, 9B U.L.A. 355 (stating that a sperm or egg donor is not the parent of a child conceived by assisted reproduction).

146. See, e.g., Buzzanca, 72 Cal. Rptr. 2d at 294; Johnson, 851 P.2d at 782–87; McDonald, 608 N.Y.S.2d at 480.

147. See infra Part III.B.

148. See infra Part III.B.

149. See TURKINGTON & ALLEN, supra note 39, at 410–11. The UPA legitimizes the use of assisted reproduction by stating that it is among a handful of ways that legal parenthood may be established. See UNIF. PARENTAGE ACT, § 201, 9B U.L.A. 309 (2000). Under the UPA, the other
declares that in the context of assisted reproduction, intent, not biology, is determinative of one’s legal status as a parent. Nineteen states have fully adopted the UPA, while many others have partially adopted it. For both egg and sperm donation, the UPA states that the intended parents are the legal parents: “[i]f a husband provides sperm for, or consents to, assisted reproduction by his wife . . . he is the father of a resulting child.” Conversely, a donor is not a parent of a child conceived by means of assisted reproduction. Thus, intent determines parentage in the context of assisted reproduction.

B. Courts Also Rely on Intent to Resolve Parentage Disputes

Case law resolving parentage disputes is in line with the UPA’s approach. Courts base parentage rights for children born from assisted reproduction on intent: specifically, whether the individual acted to bring about the birth of a child with the intent to raise it as his or her own. For example, in In re Marriage of Buzzanca a child, Jaycee, was born with five potential parents: an anonymous egg donor, an anonymous sperm donor, a gestational mother, and the husband and wife who intended to be the parents. Upon divorce, which occurred shortly before the child was born, the intended father claimed that he did not...
have any legal obligation to pay child support. He based this claim on the fact that he was not biologically related to Jaycee. A California appeals court rejected the husband’s argument and held that he was obligated to pay child support because of his intent to create and parent a child. Division Four of the California Court of Appeal acknowledged that it was an undisputed fact that Jaycee would not have been born but for the intended parents’ agreement to implant a fertilized egg in a surrogate. Furthermore, the court held that the husband’s intent and affirmative act of proceeding with the surrogacy agreement were conclusive “parenthood” under the common law doctrine of estoppel.

The California court stated a clear rule: when “a child is procreated because a medical procedure was initiated and consented to by intended parents,” the intended parents are the legal parents of the child.

Similarly, in Johnson v. Calvert, the California Supreme Court relied on the intent of the parties to resolve a maternity dispute. In Johnson, a husband and wife contracted with a surrogate mother to gestate a child on their behalf. The surrogate mother asserted maternity rights based on her biological connection to the child formed through gestation and birth, despite the fact that the husband and wife provided both gametes and were therefore the genetic parents of the child. The surrogate mother claimed that her biological connection gave her a constitutional right to be the legal mother to the child. The California Supreme Court resolved the maternity dispute by relying on the intent of the parties when they undertook the procreational process.
denied maternity rights to the surrogate mother\textsuperscript{170} despite acknowledging that giving birth to the child gave her a biological connection sufficient for maternity under California law.\textsuperscript{171} The court based its decision on the reasoning that “[a] woman who enters into a gestational surrogacy arrangement is not exercising her own right to make procreative choices.”\textsuperscript{172} The court held that the wife was the legal mother because “from the outset [she had] intended to be the child’s mother”\textsuperscript{173} and because the child would not have been born \textit{but for} the couple’s desire and acts to create a family.\textsuperscript{174} The California Supreme Court focused on the fact that the couple “affirmatively intended the birth of the child, and took the steps necessary to effect in vitro fertilization.”\textsuperscript{175}

In \textit{McDonald v. McDonald},\textsuperscript{176} a New York appellate court also used intent to grant maternity rights to a woman not genetically related to her children.\textsuperscript{177} In \textit{McDonald}, a husband argued that his wife did not have maternity rights to their children because she was not their genetic mother.\textsuperscript{178} The couple had combined donor eggs with the husband’s sperm to create the preembryos, which were implanted in the wife and resulted in her giving birth to twins.\textsuperscript{179} Upon divorce, the husband argued that the children should either be found illegitimate or “genetically and legally plaintiff’s” so that his wife would be denied parentage rights.\textsuperscript{180} In rejecting the husband’s argument, the court explicitly adopted the reasoning used in \textit{Johnson v. Calvert} to hold that the wife was the legal mother of the twins because—despite her missing genetic connection—she had the necessary intent to create a family.\textsuperscript{181}

\textsuperscript{170} \textit{See} \textit{Johnson}, 851 P.2d at 787.
\textsuperscript{171} \textit{See id.} at 781 (stating that because there was “undisputed evidence that [the surrogate mother], not [the gamete provider], gave birth to the child . . . [b]oth women thus have adduced evidence of a mother and child relationship as contemplated by the Act”) (citing \textit{CAL. CIV. CODE} §§ 7003, 7004, 7015 (West 1970); \textit{CAL. EVID. CODE}, §§ 621, 892 (West 1966)).
\textsuperscript{172} \textit{Id.}
\textsuperscript{173} \textit{See id.} at 782.
\textsuperscript{174} \textit{See id.}
\textsuperscript{175} \textit{See id.}
\textsuperscript{177} \textit{See id.} at 480.
\textsuperscript{178} \textit{See id.} at 479.
\textsuperscript{179} \textit{See id.} at 478.
\textsuperscript{180} \textit{See id.}
\textsuperscript{181} \textit{See id.} at 480 (stating that “in the instant ‘egg donation’ case, the wife, who is the gestational mother, is the natural mother of the children”).

644
Preembryos and the Right to Privacy

In sum, intent, rather than biological connections, determines parentage for children born from the use of assisted reproduction. Both the Uniform Parentage Act and case law have adopted this approach. Therefore, individuals without a biological connection may nonetheless have parentage obligations, and individuals who do have a biological connection may nonetheless be denied parentage rights.

IV. THE CREATION OF PREEMBRYOS IS AN INTIMATE DECISION ABOUT WHETHER TO HAVE CHILDREN AND, AS SUCH, FALLS SQUARELY WITHIN THE ZONE OF PRIVACY PROTECTED BY THE U.S. CONSTITUTION

The constitutional right to privacy broadly protects intimate decisions central to procreation, marriage and family life.182 The decision to create preembryos in order to have children is as intimate and central to family life as procreational decisions involving contraception and abortion. As such, it falls squarely within the zone of privacy protected by the U.S. Constitution.183 Because both gamete and non-gamete providers using assisted reproduction have made the intimate decision to create a family, the right to privacy’s broad shield over decisions related to procreation, marriage and family life should protect them both. Thus, courts determining dispositional control over preembryos should consider the constitutional interests of both gamete and non-gamete providers equally. Courts have erred in this consideration to the extent that they have chained these interests to a biological connection.184

Rather than looking to biology to find that one party has a privacy interest in the preembryos and the other does not, courts should look to the intent of the parties who participated in the creation of the preembryos. Intent is a well-established method for determining parentage in family law185 and provides a useful analogy to resolving preembryo disputes. In family law, a biological connection is not required to establish parentage—what matters is the intent of the parties to become parents.186 Similarly, in the preembryo context, parties who intend to create a family have each made the intimate decision to

182. See supra Part I.
185. See supra Part III.
186. See supra Part III.
procreate and, therefore, should be equally protected by the constitutional right to privacy.

A. Courts Have Properly Applied Privacy Rights in Preembryo Disputes, Except to the Extent They Have Made the Right to Privacy Dependent Upon Biology

In each of the five cases decided by state high courts, disputes over frozen preembryos have implicated the right to privacy. In two of those cases, *Davis v. Davis* and *J.B. v. M.B.*, the courts explicitly applied the constitutional right to privacy and balanced the parties’ interests. The courts in *Davis* and *J.B.* were correct in so far as they recognized that individuals have a privacy interest in their preembryos. But, they were incorrect to the extent that they tied the right to privacy to a biological connection.

Furthermore, the Washington State Supreme Court incorrectly stated in *Litowitz v. Litowitz* that the right to privacy requires a biological connection in order to protect one’s constitutional interest in preembryos. The court failed to recognize the intimate nature of that couple’s decision—made within the highly personal relationship of their marriage—to create preembryos in order to become parents. If courts add a biological requirement to the right to privacy, they will fail to recognize that the use of assisted reproduction to create preembryos in order to have children is an intimate decision that falls squarely within the zone of privacy protected by the U.S. Constitution.

---

188. See *J.B.*, 783 A.2d at 715–17; *Davis*, 842 S.W.2d at 603–04.
189. See *J.B.*, 783 A.2d at 715–17; *Davis*, 842 S.W.2d at 603.
190. See, e.g., *Davis*, 842 S.W.2d at 602 (stating that “decisional authority rests in the gamete-providers alone, at least to the extent that their decisions have an impact upon their individual reproductive status” because “no other person or entity has an interest sufficient to permit interference with the gamete-providers’ decision to continue or terminate the IVF process, because no one else bears the consequences of these decisions in the way that the gamete-providers do”).
191. See *Litowitz*, 146 Wash. 2d at 527, 48 P.3d at 267.
B. The Right to Privacy Protects the Extremely Personal Decision of Whether to Create Preembryos in Order to Have Children

The right to privacy is sufficiently broad to encompass the intimate decision to create a family via assisted reproduction. 192 The Washington State Supreme Court’s dicta in Litowitz to the contrary utterly fails to recognize the breadth of the right to privacy. The U.S. Supreme Court acknowledged the broad scope of the right to privacy in Eisenstadt v. Baird, where the Court concluded that the right to privacy protected an individual’s decision to use contraception. 193 The Court held that “[i]f the right of privacy means anything, it is the right to the individual, married or single, to be free from unwarranted governmental intrusion into matters so fundamentally affecting a person as the decision whether to bear or beget a child.” 194 The Court has also held that the right to privacy is so broad that it shields a woman’s decision to terminate her pregnancy. 195 Yet, the right to privacy protects much more than the decision to have a child. 196 In Loving v. Virginia, the Court held that the right to privacy shielded an individual’s decision of whom to marry. 197 In Moore v. City of East Cleveland, the Court extended the right to privacy to protect a person’s decision to live with members of her extended family. 198 Despite this, the Washington State Supreme Court has suggested in dicta that the right to privacy does not encompass the fruits of a couple’s decision to combine sperm and donor eggs in an effort to create a family. 199 This flatly contradicts the recognized historic breadth of the right to privacy and, if followed, would render meaningless eighty years of U.S. Supreme Court jurisprudence.

Rather than depending on biology, courts should look to the intimate nature of the decision to create a family by using assisted reproduction and determine that the Constitution protects the interests of both gamete and non-gamete providers. As the U.S. Supreme Court determined in Roberts v. United States Jaycees, the Constitution protects highly

192. See supra Part I.
194. See id. (emphasis added and omitted).
196. See, e.g., Loving v. Virginia, 388 U.S. 1, 12 (1967) (extending the right to privacy to the decision of whom to marry); Moore v. City of East Cleveland, 431 U.S. 494, 506 (1977) (extending the right to privacy to family living arrangements).
197. See Loving, 388 U.S. at 12.
personal relationships that “attend the creation and sustenance of a family.” The creation of preembryos using assisted reproduction is a highly personal one that attends the creation and sustenance of a family and is, therefore, protected by the Constitution. Courts following the dicta in Litowitz would inappropriately make the constitutional right to privacy dependent on biology in contradiction of well-established Supreme Court jurisprudence. Instead, courts should rely on an individual’s intent to find that the constitutional right to privacy protects both gamete and non-gamete providers’ interests when they create preembryos with the intent to become a parent.

C. Courts Should Look to the Intent of the Parties at the Time They Created the Preembryos

Intent is the most practical and fair method for determining privacy rights. Courts faced with resolving preembryo disputes should analogize the issue to family law, where parentage disputes are determined by intent. In parentage disputes, biological connections are often legally meaningless. For example, individuals with absolutely no biological connection are forced to uphold their parental responsibilities if they intended to become a parent when they used assisted reproduction, while individuals who do have a biological connection are denied any parenting rights whatsoever if they lacked the initial intent to become a parent.

In re Marriage of Buzzanca, a California court forced parentage obligations on a man for a child with whom he had no biological connection, because that child would not have been born but for the man’s intent to create a family using donor sperm, donor eggs and a surrogate. In Johnson v. Calvert, the California Supreme Court denied parentage rights to a surrogate mother because she lacked the

201. See supra Part III.
202. See In re Marriage of Buzzanca, 72 Cal. Rptr. 2d 280, 293–94 (Cal. Ct. App. 1998) (applying the intent doctrine to force parental obligations on a man with no biological connection to a child); see also UNIF. PARENTAGE ACT, § 703, 9B U.L.A. 356 (2000) (stating that a husband who consented to assisted reproduction by his wife is the father of a resulting child even if he did not provide the sperm).
203. See Johnson v. Calvert, 851 P.2d 776, 787 (Cal. 1993) (using the intent doctrine to deny parentage rights to the surrogate mother and grant them to the intended mother); see also UNIF. PARENTAGE ACT, § 702, 9B U.L.A. 355 (2000) (stating that a gamete donor is not a parent to a child conceived by assisted reproduction).
204. See Buzzanca, 72 Cal. Rptr. 2d at 293–94.
Preembryos and the Right to Privacy

intent to become a parent. Despite having gestated and given birth to the child, the surrogate mother’s biological connection was insufficient to overcome the intended mother’s right to parentage because the child would not exist “[b]ut for [the] acted-on intent[]” of the intended mother to become a parent. This same approach was applied to the egg donor situation in McDonald v. McDonald, where a New York court granted parentage rights to the intended mother despite her ex-husband’s argument that because she did not have a genetic connection to the children she should not have parentage rights.

Similarly, courts should look to the intent of the parties in preembryo disputes to determine whether they made the intimate decision to create a family that falls within the zone of privacy protected by the U.S. Constitution. In Buzzanca, Johnson, and McDonald, courts looked to the intent of the parties to determine their parentage rights and obligations. Preembryo disputes likewise deal with individuals who had the intent to become parents at the time they began the process of assisted reproduction. Therefore, courts should look to the intent of these individuals when determining whether they have privacy interests in their preembryos. If both parties had the intent to become parents when they created the preembryos, then the constitutional right to privacy must apply equally to them both. A biological connection should not steal the shade provided by the umbrella of the constitutional right to privacy.

D. Deciding Preembryo Cases Based on Constitutional Privacy Interests Permits Courts to Fairly Balance Parties’ Competing Interests

If courts choose to apply the constitutional right to privacy, they should hold that it encompasses the interests of both gamete and non-gamete providers. As illustrated by the constitutional balancing test used in Davis v. Davis and J.B. v. M.B., this threshold determination enables courts to weigh the relevant interests of both parties before making a final decision about who has dispositional control over the preembryos. When courts determine that both parties have

205. See Johnson, 851 P.2d at 782.
206. See id.
207. See id.
constitutional rights, that does not determine which party ultimately prevails; it merely ensures that the interests of both parties will be considered. Significantly, when courts grant one party dispositional control over preembryos, it often enables that party to prevent his or her ex-spouse from using the preembryos against his or her wishes.\textsuperscript{210} Non-gamete providers should have this same opportunity to prevent the unfair disposition of preembryos they helped create.

The importance of granting constitutional rights to both gamete and non-gamete providers can be illustrated by looking again at the hypothetical with Leny and Eva.\textsuperscript{211} For example, assume Eva embraced the particular moral view that it would be unethical to destroy the preembryos or to donate them to scientific research.\textsuperscript{212} She and Leny discussed this issue in detail prior to engaging the IVF clinic. They agreed that once they had created their family they would donate any excess preembryos to an infertile couple (though they did not sign a contract with the IVF clinic). The preembryos would never have been created had Eva and Leny not agreed on this point, because Eva had made clear that she would not participate in the creation of the preembryos if there was a chance they would ever be destroyed. If, upon divorce, Leny wishes to have the preembryos destroyed, that desire should not be the only interest the court hears. The court should also consider Eva’s interests, as she too participated in the intimate and highly personal decision to have the preembryos created. The constitutional right to privacy must protect both Leny and Eva. For a court to hold otherwise would contradict years of U.S. Supreme Court jurisprudence establishing that the constitutional right to privacy is a broad shield protecting intimate decisions central to procreation, marriage and family life.

V. CONCLUSION

In disputes over frozen preembryos, it is inevitable that parties will invoke the constitutional right to privacy to argue that they should have dispositional control over the preembryos. The right to privacy protects

\textsuperscript{210} See \textit{supra} Part II.

\textsuperscript{211} See \textit{supra} notes 1–4 and accompanying text.

\textsuperscript{212} See, e.g., \textit{J.B.}, 783 A.2d at 710–11 for a similar example. In \textit{J.B.}, the husband argued that “as a Catholic,” he would not have agreed to participate in the creation of preembryos that would eventually be destroyed and that he and his wife had agreed “that no matter what happened the eggs would be either utilized by us or by other infertile couples.” See \textit{id.} at 710.
Preembryos and the Right to Privacy

intimate decisions involving procreation, marriage and family life. Non-gamete providers who decide to create preembryos with the intent to become a parent have made an intimate decision falling squarely within the constitutional right to privacy. Therefore, if courts choose to apply the constitutional right to privacy to resolve preembryo disputes, they should find that this right to privacy encompasses both gamete and non-gamete providers. In such cases, the couple made the decision to create a family through the use of assisted reproduction together, and the preembryos would not exist but for that joint decision. Consequently, courts should consider the interests of both gamete providers and non-gamete providers who create preembryos with the intent to become parents and refrain from granting dispositional control based solely on a biological connection.
DOES FREE EXERCISE MEAN FREE STATE FUNDING? IN DAVEY V. LOCKE, THE NINTH CIRCUIT UNDervalued WASHINGTON’S VISION OF RELIGIOUS LIBERTY

Derek D. Green

Abstract: In Davey v. Locke, a panel of the United States Court of Appeals for the Ninth Circuit ruled that Washington violated the Free Exercise Clause by refusing to allow a scholarship recipient to use state funds to pursue a theology degree. The court held that the state’s scholarship requirements facially discriminated against religion, and that the state’s interest in not violating its constitution did not serve as a compelling reason for the discrimination. In so holding, the Davey court ignored Ninth Circuit precedent and embraced a theory of the Religion Clauses at odds with United States Supreme Court jurisprudence. Furthermore, as explained in the dissent, the scholarship requirements are analogous to permissible limitations placed on other government funding programs. Based on U.S. Supreme Court precedent in other conditional funding cases, Davey should be overturned.

Washington’s Promise Scholarship program provides state funding to qualified students who plan to attend college within the state.1 As with other financial aid programs, the state placed conditions on the eligibility and use of the Promise Scholarship to ensure that the public’s money is spent as the state intended.2 These conditions include graduating from a Washington high school, achieving a certain high school class rank, and demonstrating financial need.3 The Promise Scholarship also requires that recipients not pursue a theology degree with the state funds.4 This last requirement is necessary to comply with state law5 as well as the Washington State Supreme Court’s interpretation of the state constitution.6 Since its founding, Washington’s constitution has provided that “[n]o public money or property shall be appropriated for or applied to any religious worship, exercise or instruction.”7

2. See 1999 Wash. Laws 309 § 611(6); WASH. ADMIN. CODE § 250-80-020.
3. WASH. ADMIN. CODE § 250-80-020.
4. Id. § 250-80-020(12)(f).
5. WASH REV. CODE § 28B.10.814 (2002) (“No aid shall be awarded to any student who is pursuing a degree in theology.”).
7. WASH. CONST. art I, § 11 (as amended 1993). The amendment did not alter this language.
Joshua Davey received a Promise Scholarship in 1999.\(^8\) He enrolled at Northwest College, an accredited private school in Washington State.\(^9\) Under the terms of the Promise Scholarship, Davey was allowed to use the state funds at a religiously affiliated college such as Northwest, which emphasized teaching from a “distinctly Christian” perspective.\(^10\) However, because state law prohibited funding a theology degree, Davey lost his Promise Scholarship when he decided to pursue Northwest’s “Pastoral Ministries” major, a program designed to train students to become ministers.\(^11\) Davey challenged the state’s policy,\(^12\) arguing among other claims\(^13\) that the refusal to fund his theology degree violated his First Amendment right to the free exercise of religion.\(^14\)

Davey’s challenge raised an unresolved issue involving the First Amendment’s Religion Clauses, which provide that “Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof.”\(^15\) In \textit{Davey v. Locke},\(^16\) the U.S. Court of Appeals for the Ninth Circuit had to consider whether an area exists between the Establishment and Free Exercise Clauses in which a state is free to define its own vision of religious liberty.\(^17\) Based upon the court’s holding, if such an area exists, then it does not encompass a state’s refusal to fund religious degrees.\(^18\) The Ninth Circuit held that the state’s Promise Scholarship requirement facially discriminated against a class of students on the basis of religion.\(^19\) Because the state did not have a compelling reason for the selective treatment, the court ruled that the program violated the student’s right to the free exercise of religion.\(^20\)

---

8. See Davey v. Locke, 299 F.3d 748, 751 (9th Cir. 2002).
9. See id.
10. See id.
11. See id.
12. Id.
13. Davey also argued that the program violated the Establishment Clause, as well as his rights to free speech, association, and equal protection under the federal and state constitutions. See id. at 750; Appellant’s Brief at 11–15, Davey v. Locke, 299 F.3d 748 (9th Cir. 2002) (No. 00-35962). This Note focuses on the Free Exercise Clause challenge, which was the basis of the court’s decision in Davey. See 299 F.3d at 750.
14. Davey, 299 F.3d at 750.
16. 299 F.3d 748 (9th Cir. 2002).
17. See id. at 760–61 (McKeown, J., dissenting).
18. See id. at 750.
19. Id.
20. Id.
According to Judge McKeown’s dissent, however, the majority misconstrued the issue. The state’s program did not prohibit the exercise of anything; rather, it simply reflected a decision by the state to fund certain activities and not others. Relying on U.S. Supreme Court precedent involving state funding and fundamental rights, the dissent maintained that the Promise Scholarship’s requirements were permissible.

This Note agrees with the dissent and argues that Washington’s Promise Scholarship complies with the Free Exercise Clause. In holding otherwise, the *Davey* court failed to adequately address conflicting precedent in the Ninth Circuit, as well as other courts, and contradicted Supreme Court precedent in conditional funding cases. Part I provides an overview of the Supreme Court’s historical treatment of the Religion Clauses in educational funding. Part II discusses the Court’s general framework for deciding Free Exercise Clause challenges. Part III examines the Court’s analytic approach to cases involving the conditional funding of other fundamental rights. Part IV provides a summary of the *Davey* case and an explanation of the Ninth Circuit’s evaluation of the case. Finally, Part V argues why the Ninth Circuit’s holding in *Davey* was incorrect, and Part VI concludes that Washington’s scholarship program does not violate the Free Exercise Clause. In practice, the Promise Scholarship does not suppress religion; it simply reflects a rational choice by the state to fund activities that conform with its constitution.

I. A HISTORY OF THE RELIGION CLAUSES IN EDUCATIONAL FUNDING CASES

The U.S. Supreme Court has not directly addressed the extent to which a state, consistent with both of the First Amendment’s Religion Clauses, can exclude private religious schools and students from educational funding programs available to others. However, the Court has decided a number of cases involving the Religion Clauses and

---

21. See id. at 761 (McKeown, J., dissenting).
22. Id. (McKeown, J., dissenting).
23. Id. at 761, 764–66 (McKeown, J., dissenting).
educational funding. This part summarizes the principles from these cases. The first section frames the current debate by discussing the Court’s decision nearly twenty years ago in a case very similar to the one at issue in Davey. The second section provides an overview of the Court’s past decisions involving the Religion Clauses and educational funding.

A. Setting the Stage: Witters v. Washington Department of Services for the Blind

Joshua Davey was not the first to raise the issue of educational funding and the Religion Clauses in the state of Washington. In Witters v. Washington Department of Services for the Blind, the U.S. Supreme Court reviewed the state of Washington’s refusal to allow Larry Witters to use vocational assistance monies to fund his religious training. Witters argued that the Establishment Clause did not prohibit the state from funding this training, and that the Free Exercise Clause required it. Although the U.S. Supreme Court declined to address Witters’ Free Exercise Clause challenge, it agreed with Witters’s Establishment Clause argument. Foreshadowing the reasoning of future Establishment Clause cases, the Court concluded that the government’s role was sufficiently independent from religion because it provided money to individuals instead of directly to religious organizations. As a result, the Establishment Clause did not prohibit the state from paying for Witters’ training.

However, the Court noted that a state could come to a different conclusion based on its own constitution. On remand, the Washington State Supreme Court did just that and held that the Washington

28. See id. at 483–84.
29. See id. at 489–90.
30. Id.
31. See infra Part I.B.
32. See Witters II, 474 U.S. at 488–89.
33. Id.
34. Id. at 489.
Constitution prohibited the “use of public moneys to pay for such religious instruction.”\(^35\) The state high court noted a “major difference” between the Washington and federal prohibitions on public funding of religion: the federal Establishment Clause prohibited “the *appropriation* of public money for religious instruction,” while the Washington Constitution also prohibited “the *application* of [state] funds to religious instruction.”\(^36\) Because Witters would apply the state funding towards religious education, the state was justified in denying him financial assistance.\(^37\)

The Washington State Supreme Court concluded that its reading of the state constitution did not violate the Free Exercise Clause.\(^38\) To the court, the major issue was whether a government action coerced the individual into violating a religious belief.\(^39\) The court held that it did not.\(^40\) In the court’s words, the recipient “chose to become a minister, and the Commission’s only action was to refuse to pay for his theological education. The Commission’s decision may make it financially difficult, or even impossible, for [the applicant] to become a minister, but this is beyond the scope of the free exercise clause.”\(^41\) The U.S. Supreme Court declined to review the state court decision,\(^42\) and has not directly addressed the Free Exercise Clause issue left open in *Witters II*.\(^43\) However, the Court’s multiple rulings on educational funding programs in other contexts do provide some guidance on the question.

**B. U.S. Supreme Court Jurisprudence on the Funding of Religious Education**

The U.S. Supreme Court’s decisions involving state funding of religious education reflect the Court’s “strugg[e]”\(^44\) to navigate between

\(^{36}\) Id. at 370, 771 P.2d at 1122 (emphasis in original).
\(^{37}\) Id. at 365, 771 P.2d at 1119–20.
\(^{38}\) Id. at 371, 771 P.2d at 1123.
\(^{39}\) Id.
\(^{40}\) Id.
\(^{41}\) Id. (quoting Witters I, 102 Wash. 2d 624, 631, 689 P.2d 53, 57 (1984)).
\(^{43}\) See supra note 24.
the Free Exercise and Establishment Clauses. Both clauses are written in “absolute terms, and either of which, if expanded to a logical extreme, would tend to clash with the other.” Yet despite their “internal tensions,” the Religion Clauses work together to ensure religious liberty.

Numerous cases involving state funding of religious education have focused on the Establishment Clause. In determining whether a funding program violates the Establishment Clause, the U.S. Supreme Court has considered factors that include whether a program fosters neutrality through private choice, distributes funds equally to students attending religious and non-religious schools, and uses state funds for a permissible purpose. The Court’s review of educational funding on Free Exercise Clause grounds is more limited. Although its opinions do not provide clear precedent, the Court has affirmed one ruling that the Free Exercise Clause does not require state action simply because it is permitted under the Establishment Clause.

1. The Establishment Clause in Educational Funding Cases

The U.S. Supreme Court has interpreted the Establishment Clause as reflecting a belief that “a union of government and religion tends to destroy government and degrade religion.” The Court’s jurisprudence involving religion and educational funding has historically reflected this view. For example, in the first of the modern school funding cases, Everson v. Board of Education, the Court approved the use of funds to subsidize the transportation of students, including those attending

46. Walz, 397 U.S. at 668–69.
Requiring States to Fund Religious Training

parochial schools, to and from school. However, the Court commented that there were limits to this kind of interaction, stating that the “First Amendment has erected a wall between church and state. That wall must be kept high and impregnable.” More recently, the Court has gradually allowed more interaction between state funds and private religious education, although it has not explicitly refuted Everson’s warning.

The Court’s recent Establishment Clause decisions have increasingly focused on whether a program fosters government “neutrality” through “private choice.” Disbursing public funds directly to students instead of religious institutions conveys neutrality because it leaves the decision whether to apply funds towards a religious education to “private choice” and not government action. For example, in Zelman v. Simmons-Harris, the Court ruled that Cleveland’s school voucher program was permissible because it promoted private choice by giving money directly to students for use at either religious or non-religious schools.

In addition, the Court has considered other factors as well in determining a program’s permissibility. The Court has upheld aid programs that are available to students of both public and private schools or that fund solely non-religious uses under the Establishment

55. Id. at 17.
56. Id. at 18.
57. See Chemerinsky, supra note 50. The Court devised a three-part test for Establishment Clause cases in Lemon v. Kurtzman, 403 U.S. 602 (1971). Under Lemon, the Court considers if a program (1) has an actual secular purpose, (2) has a principal effect of advancing or restricting religion, and (3) creates excessive government entanglement with religion. See id. at 612–13. The Lemon factors helped to shape the Court’s opinions in school aid cases, but the Court’s reliance on the Lemon test has wavered in recent years. See, e.g., Mitchell v. Helms, 530 U.S. 793, 807–09 (2000) (plurality opinion) (stating that under Agostini v. Felton, 521 U.S. 203 (1997), the Court relies upon only the first two Lemon factors in school aid cases).
60. See Zelman, 122 S. Ct. at 2473.
62. Id. at 2473.
63. See Mitchell, 530 U.S. at 839–40 (O’Connor, J., concurring) (stating that the Court has “never held that a government-aid program passes constitutional muster solely because of the neutral criteria it employs as a basis for distributing aid”) (emphasis in original).
Clause. For example, the Court rejected a funding program that favored students attending non-public schools over students attending public schools as an impermissible establishment of religion, but ten years later upheld a program that distributed funding evenly to students attending public and private schools.

The Court has also approved of programs with “no aid” to religion provisions, which restrict the use of public funds to solely secular purposes. For example, in *Tilton v. Richardson*, the Court reviewed a federal program that subsidized the cost of constructing new facilities at colleges and universities. The program permitted religious institutions to participate, but “expressly prohibit[ed] use of the facilities for religious purposes.” The plurality decision upheld the funding program based on the program’s explicit secular use requirement, but struck down a provision that allowed organizations to convert the facilities to religious use after twenty years, reasoning that it would violate the Establishment Clause.

However, the importance placed on these “no-aid” provisions varies in the Court’s opinions. The plurality and concurring opinions in *Mitchell*

---

66. See generally Chemerinsky, *supra* note 50. Although more recent decisions have not emphasized these factors as much, lower courts are still required to consider them when applicable. See *Agostini v. Felton*, 521 U.S. 203, 237 (1997) (“[I]f a precedent of this Court has direct application in a case, yet appears to rest on reasons rejected in some other line of decision, the Court of Appeals should follow the case which directly controls, leaving to this Court the prerogative of overruling its own decisions.”) (internal quotations and citations omitted).

67. See *Nyquist*, 413 U.S. at 782 n.38, 794.

68. See *Mueller*, 463 U.S. at 397.

69. See *Tilton v. Richardson*, 403 U.S. 672 (1971); accord *Agostini*, 521 U.S. at 211–12, 235 (allowing public school teachers to teach secular subjects at religious schools); *Mitchell*, 530 U.S. at 838 (O’Connor, J., concurring).

70. 403 U.S. 672 (1971).

71. See *id.* at 674–77.

72. *Id.* at 677.

73. *Id.* at 679–81.

74. *Id.* at 682–84. All Justices reviewing the case agreed that this provision violated the Establishment Clause. See *id.* at 692 (Douglas, J., dissenting in part); Lemon v. Kurtzman, 403 U.S. 602, 660–61 (1970) (Brennan, J., concurring in part with *Lemon* and concurring in part and dissenting in part with *Tilton*) (concluding that providing grants to “sectarian institutions” violates the Establishment Clause); *id.* at 665 n.1 (White, J., concurring in part and dissenting in part with *Lemon* and concurring in judgment with *Tilton*) (agreeing that allowing government-subsidized buildings to be used for religious purposes after twenty years is impermissible).

75. Compare *Tilton*, 403 U.S. at 682–84 (rejecting a statutory provision allowing buildings constructed with public funds to be used for religious purposes after a set amount of time), with *Zobrest v. Catalina Foothills Sch. Dist.*, 509 U.S. 1, 3 (1993) (allowing sign language interpreters at religious schools).
v. Helms\textsuperscript{76} reflect this divide. In \textit{Mitchell}, the Court upheld a program providing funds for schools to purchase equipment and supplies.\textsuperscript{77} The program required the use of “neutral, secular criteria”\textsuperscript{78} for all purchases. The plurality opinion noted that the program satisfied precedent by not providing “religious schools [with] aid that has an impermissible content,”\textsuperscript{79} but stated that diversion of funds towards religious indoctrination was otherwise permissible.\textsuperscript{80} In contrast, Justice O’Connor, joined by Justice Breyer, concurred in judgment only, and stated that she “disagree[d] with the plurality’s conclusion that actual diversion of government aid to religious indoctrination is consistent with the Establishment Clause.”\textsuperscript{81} Instead, Justice O’Connor emphasized the program’s secular requirements, including the prohibition on making payments for religious worship or instruction.\textsuperscript{82}

As the above cases reflect, the emphasis placed on “private choice,” availability of funds for both private and public school students, and “no aid” provisions varies in the Court’s Establishment Clause cases. Although recent school aid cases have focused on private choice, the Court has warned against concluding that its reasoning has “by implication” overruled earlier precedent.\textsuperscript{83} Absent clear direction to the contrary, lower courts should continue to consider all relevant factors.\textsuperscript{84}

2. \textit{The U.S. Supreme Court’s Limited Review of the Free Exercise Clause in Educational Funding Cases}

The U.S. Supreme Court has rejected the claim that excluding religious school students from receiving educational benefits violates the Free Exercise Clause.\textsuperscript{85} However, the Court did so in a memorandum

\textsuperscript{76} 530 U.S. 793 (2000).
\textsuperscript{77} Id. at 801.
\textsuperscript{78} \textit{Mitchell}, 530 U.S. at 848 (O’Connor, J., concurring).
\textsuperscript{79} Id. at 831.
\textsuperscript{80} Id. at 820.
\textsuperscript{81} Id. at 840 (O’Connor, J., concurring). Justice O’Connor also stated that these programs, even if they require “secular” use, conform with the Court’s definition of “neutrality.” \textit{See id.} at 838.
\textsuperscript{82} Id. 848–49.
\textsuperscript{83} \textit{See Agostini v. Felton}, 521 U.S. 203, 237 (1997) (“We do not acknowledge, and we do not hold, that other courts should conclude our more recent cases have, by implication, overruled an earlier precedent.”).
\textsuperscript{84} \textit{See id.}
opinion without comment. In *Luetkemeyer v. Kaufmann*, the Court affirmed a three-judge district court decision holding that providing busing services only to students attending public schools was permissible under the Free Exercise Clause. A federal district court for the Western District of Missouri ruled that the Free Exercise Clause did not require transporting students to parochial schools simply because the Establishment Clause permitted it. Instead, the district court concluded that the state’s policy fell within an area “between the Establishment Clause and the Free Exercise Clause in which action by a State will not violate the former nor inaction, the latter.”

The *Luetkemeyer* district court also held that the state’s “long established constitutional policy” of requiring a strict separation of church and state served as a “compelling state interest” justifying “any possible infringement of the Free Exercise Clause.” Seven years later, however, the U.S. Supreme Court stated that First Amendment rights can limit a state’s ability to ensure a greater separation of church and state than provided under the Establishment Clause. In *Widmar v. Vincent*, the Court held that a state university could not exclude student religious groups from using public facilities open to other organizations. The Court limited its holding to free speech and association grounds, however, and refused to consider whether “a state interest, derived from its own constitution, could ever outweigh free speech interests protected by the First Amendment.” The *Widmar* Court also did not address the

86. Id.
89. See *Luetkemeyer*, 364 F. Supp. at 386. The U.S. Supreme Court held in *Everson v. Board of Education*, 330 U.S. 1, 17 (1947) that providing transportation to school for both parochial and non-parochial school students would not violate the Establishment Clause. See also supra notes 54–56 and accompanying text (Discussing *Everson*).
91. Id.
92. *Widmar v. Vincent*, 454 U.S. 263, 276 (1981) (noting that the state’s interest in achieving a greater separation of church and state “is limited by the Free Exercise Clause and in this case by the Free Speech Clause as well”).
93. Id.
94. Id. at 265–66.
95. Id. at 273 n.13. Although the Court did not base its ruling on Free Exercise Clause grounds, the Court did comment that the Free Exercise Clause can limit a state’s ability to require a greater separation of church and state than provided under the First Amendment. See id. at 276.
96. Id. at 275–76.
Requiring States to Fund Religious Training

district court’s conclusion in Luetkemeyer that a permissible area exists between the Establishment Clause and the Free Exercise Clause in which the states are free to legislate.97

In sum, the U.S. Supreme Court has not directly ruled on the permissibility of excluding religious schools and students from educational funding under the Free Exercise Clause. However, as the Establishment Clause’s “no aid” cases reflect, the Court has used the fact that state programs exclude religious schools and students from educational funding as a justification for upholding programs against Establishment Clause challenges. These cases indicate the Court’s understanding of the narrow role the government should play in religious education, and serve as a reminder of the limitations placed on the Free Exercise Clause by the Establishment Clause.

II. THE PRINCIPLES OF FREE EXERCISE CLAUSE REVIEW

Outside of the educational funding context, the U.S. Supreme Court has decided a number of Free Exercise Clause cases.98 These cases have established some clear principles. First, the Free Exercise Clause provides absolute protection for religious beliefs, and offers some protection for religious practices.99 Second, laws lacking “neutrality” and “general applicability” toward religion are subject to closer review than those that do not.100

Other principles of Free Exercise Clause review, however, are less clear. Lower courts are split on the proper standard of review required when a state program or law lacks facial neutrality but does not have as its purpose the suppression of religion.101 Further, the U.S. Supreme Court has distinguished between programs that directly prohibit the

97. Cf. Tilton v. Richardson, 403 U.S. 672, 677 (1972) (plurality opinion) (discussing the Court’s historic attempts to find the “neutral area” between the Religion Clauses “within which the legislature may legitimately act”).
99. See, e.g., Hialeah, 508 U.S. at 533; see generally CHEMERINSKY, supra note 50, § 12.3.1.
100. See id. at 546.
101. Compare KDM v. Reedsport Sch. Dist., 196 F.3d 1046, 1050–51 (9th Cir. 1999) (refusing to apply strict scrutiny to a non-neutral program that did not reflect hostility as applied), with Peter v. Wedl, 155 F.3d 992, 996–97 (8th Cir. 1998) (applying strict scrutiny to similar program).
exercise of religion and those that condition funding on restricting a practice of religion.  

A. Established Principles of Free Exercise Clause Review

The U.S. Supreme Court’s Free Exercise Clause jurisprudence has established two clear principles. The first is that a government cannot directly regulate “religious beliefs as such.” For example, the state cannot require government officials to declare their belief in God, regardless of the state’s reasons for the requirement. As a result, most Free Exercise Clause cases do not involve direct attempts to regulate religious beliefs, but rather government actions affecting the practice of religion. Although the Free Exercise Clause does not provide absolute protection for religious conduct, it does offer some protection.  

McDaniel v. Paty provides an example. In McDaniel, a plurality held that prohibiting ministers from holding public office did not amount to a regulation based on religious beliefs, but still constituted an impermissible regulation of religious conduct. The second principle of Free Exercise Clause jurisprudence is distinguishing between laws that are “neutral” and “of general applicability” and those that are not. A government action that is neutral and generally applicable does not require strict scrutiny review. For instance, the respondents in Employment Division v. Smith lost their jobs after using peyote in a religious ceremony. The state refused

105. See, e.g., Smith, 494 U.S. at 877–80 (summarizing cases).
108. Id. at 626–27.
109. See Hialeah, 508 U.S. at 531.
110. See id.
111. Compare Hialeah, 508 U.S. at 546 (requiring strict scrutiny for a law lacking neutrality), with Smith, 494 U.S. at 890 (effectively requiring only rational basis review for a neutral, generally applicable law).
112. See Hialeah, 508 U.S. at 531.
114. Id. at 874–75.
to grant unemployment benefits to the respondents because they had violated a criminal law against the abuse of controlled substances.\textsuperscript{115} Although the Court acknowledged the effect of the law on the respondents’ exercise of religion, it stated that the Oregon law was generally applicable and neutral towards religion.\textsuperscript{116} Accordingly, the Court refused to require a “compelling interest” for the law, and instead relied upon the highly deferential rational basis standard.\textsuperscript{117} Because the state had a legitimate interest in regulating peyote, the law’s infringement upon the exercise of religious beliefs was inconsequential under rational basis review.\textsuperscript{118}

In contrast to neutral government actions, courts review laws lacking neutrality or general applicability towards religion more closely.\textsuperscript{119} If the object of the law is to suppress religion, it must be justified by a compelling state interest.\textsuperscript{120} In \textit{Church of Lukumi Babalu Aye, Inc. v. City of Hialeah},\textsuperscript{121} the Court held that a city’s attempt to prohibit the use of animal sacrifices in sacred rituals violated the Free Exercise Clause.\textsuperscript{122} The Court identified the lack of neutrality and general applicability of the law as indications of impermissible government hostility towards religion.\textsuperscript{123} As such, the Court applied strict scrutiny review and held that the city did not have a compelling reason for such discrimination.\textsuperscript{124}

\textbf{B. Circuit Split on Laws Lacking an Object of Suppressing Religion}

Two divergent theories have emerged in the federal circuit courts about the proper level of review required when a law lacks facial neutrality towards religion but does not have the object of suppressing religion.\textsuperscript{125} The first theory holds that the \textit{Hialeah} decision requires any

\begin{itemize}
\item \textsuperscript{115} \textit{Id}.
\item \textsuperscript{116} \textit{See id. at 878}.
\item \textsuperscript{117} The Court did not explicitly state that it was using the “rational basis” standard, but commentators have interpreted it as such. \textit{See, e.g., CHEMERINSKY, supra note 50, § 12.3}.
\item \textsuperscript{118} \textit{See Smith}, 494 U.S at 890.
\item \textsuperscript{119} \textit{Church of Lukumi Babalu Aye, Inc. v. City of Hialeah}, 508 U.S. 520, 531–32 (1993).
\item \textsuperscript{120} \textit{See id. at 533–34, 546}. Although the focus in \textit{Davey} is on neutrality rather than general applicability, “[n]eutrality and general applicability are interrelated . . . failure to satisfy one requirement is a likely indication the other has not been satisfied.” \textit{Id.} at 531.
\item \textsuperscript{121} 508 U.S. 520 (1993).
\item \textsuperscript{122} \textit{Id. at 524}.
\item \textsuperscript{123} \textit{See id. at 542}.
\item \textsuperscript{124} \textit{Id.} at 546–47.
\item \textsuperscript{125} \textit{Compare KDM v. Reedsport Sch. Dist.}, 196 F.3d 1046, 1050–51 (9th Cir. 1999), \textit{with} Peter \textit{v. Wedl}, 155 F.3d 992, 996–98 (8th Cir. 1998).
\end{itemize}
law that is not facially neutral to be justified by a compelling state interest. 126 In contrast, under the second theory the Hialeah decision’s strict scrutiny analysis only applies if a court determines that the object of the law is to suppress religion. 127 Lacking this purpose, heightened scrutiny is required only if the law places a “substantial burden” on the exercise of religion. 128

Under the first approach, a law that is not facially neutral towards religion is automatically subject to strict scrutiny. 129 In Peter v. Wedl, 130 the Eighth Circuit reviewed a Minnesota school district’s refusal to provide a paraprofessional assistant to an otherwise eligible disabled student because he attended a private religious school. 131 Underlying the school district’s decision was a state law that prohibited public funding of religious education. 132 The Wedl court determined that the state law “explicitly discriminated against children who attended private religious schools,” 133 and was not justified by a compelling state interest. 134 Concluding that the state law lacked facial neutrality, the court implicitly assumed it was motivated by animus towards religion. 135

In contrast, other federal circuit courts have looked beyond the face of a government action to determine if it reflected a purpose to suppress religion. 136 The Ninth Circuit adopted this approach in 1999, three years prior to Davey. In KDM v. Reedsport School District, 137 the Ninth Circuit ruled that an Oregon school district’s application of a state regulation prohibiting religious schools from receiving government-funded special education services did not violate the free exercise of religion. 138 The

126. See, e.g., Wedl, 155 F.3d at 996–98.
128. Id. (citing Hernandez v. Comm’r, 490 U.S. 680, 699 (1989)).
129. See Wedl, 155 F.3d at 996–98; Hartmann v. Stone, 68 F.3d 973, 978–79 (6th Cir. 1995).
130. 155 F.3d 992 (8th Cir. 1998).
131. Id. at 994.
132. See id. at 996–97.
133. Id. at 996.
134. Id. at 996–97 (rejecting the contention that by attempting to comply with the Establishment Clause, the law served a compelling interest). The Court also held that the law violated the student’s rights to equal protection and free speech. Id. at 997.
135. See id. at 998; cf. id. at 1001–02 (refusing to consider whether the school district’s decision to deny assistance would be permissible if it were based on a reason other than a state law “motivated by religious animus”).
136. KDM v. Reedsport Sch. Dist, 196 F.3d 1046, 1050–51 (9th Cir. 1999) (reviewing statute as applied); Strout v. Albanese, 178 F.3d 57, 65 (1st Cir. 1999).
137. 196 F.3d 1046 (9th Cir. 1999).
138. Id. at 1051.
Requiring States to Fund Religious Training

court acknowledged that the regulation was not facially neutral.\textsuperscript{139} As applied, the rule required those students attending religious schools to travel to religiously neutral settings in order to receive the same government benefits provided to public school students on-site.\textsuperscript{140} Yet the court concluded that, as applied through this program, the regulation did not “reflect a purpose to ‘suppress[] religion or religious conduct.’”\textsuperscript{141} Thus, the Ninth Circuit did not apply \textit{Hialeah}’s strict scrutiny analysis.\textsuperscript{142} The court further ruled that the regulation in practice did not impose an “impermissible burden” on the exercise of religion.\textsuperscript{143} By not subjecting the law to heightened scrutiny, the court in effect only required a rational basis for the law.

The court’s opinion in \textit{KDM} agreed with a decision by the First Circuit in a case quite similar to \textit{Davey}.\textsuperscript{144} In \textit{Strout v. Albanese},\textsuperscript{145} the First Circuit addressed the issue of whether Maine’s practice of providing funding to private secular schools—but not to religious schools—violated the Free Exercise Clause.\textsuperscript{146} The First Circuit held that the practice did not violate the Free Exercise Clause\textsuperscript{147} in part because, unlike the ordinances in \textit{Hialeah}, Maine’s program did not reflect “a substantial animus” toward religion.\textsuperscript{148}

Consequently, the First Circuit in \textit{Strout} relied on a test from \textit{Hernandez v. Commissioner},\textsuperscript{149} in which the Supreme Court ruled that heightened scrutiny was only required if a law imposed a “substantial burden” on a “central religious belief or practice.”\textsuperscript{150} Under the

\begin{itemize}
\item \textsuperscript{139} \textit{Id.} at 1050.
\item \textsuperscript{140} \textit{Id}.
\item \textsuperscript{141} \textit{Id.} at 1050–51.
\item \textsuperscript{142} \textit{Id}. Although the court left open whether the regulation could violate the Free Exercise Clause as applied to another situation, the framework for determining if a law has an impermissible purpose would remain the same. \textit{Cf. id.} at 1054 (Kleinfield, J., dissenting) (“The majority says \textit{Lukumi} does not apply here because Oregon’s regulation does not have as its object suppression of religion or religious conduct.”).
\item \textsuperscript{143} \textit{Id.} at 1050–51.
\item \textsuperscript{144} \textit{Id.} at 1051 (citing \textit{Strout} v. \textit{Albanese}, 178 F.3d 57 (1st Cir. 1999)).
\item \textsuperscript{145} 178 F.3d 57 (1st Cir. 1999).
\item \textsuperscript{146} \textit{See id.} at 59. \textit{See also} Bagley v. Raymond Sch. Dep’t, 728 A.2d 127 (Me. 1999) (holding that Maine’s tuition-funding program did not violate the Free Exercise Clause).
\item \textsuperscript{147} \textit{Strout}, 178 F.3d at 65 (listing four reasons for rejecting the Free Exercise challenge, including the lack of animus towards religion).
\item \textsuperscript{148} \textit{Id}.
\item \textsuperscript{149} 490 U.S. 680 (1989).
\item \textsuperscript{150} \textit{Id.} at 697–700. Although the Court warned against judging the “centrality of particular beliefs or practices,” it stated that it was the Court’s duty to determine how substantial the burden is
\end{itemize}
Applying the test from *Hernandez*, the *Strout* court held that Maine’s program was constitutional because the plaintiff could not demonstrate how a state’s refusal to fund attendance at a religious school resulted in a substantial burden on a central belief.\(^{152}\) In addition, the First Circuit reasoned that the government’s action did not prohibit the exercise of religion because it did not prevent the plaintiffs from attending religious schools—rather, “[a]ll it means is that the cost of religious education must be borne by the parents and not the state.”\(^{153}\)

The U.S. Supreme Court has not resolved this split in the federal circuit courts. Prior to *Davey*, both the Ninth and First Circuits first determined if a law had the object of suppressing religion before applying strict scrutiny analysis.\(^{154}\) In other circuits, any law lacking facial neutrality has required strict scrutiny review, regardless of intent.\(^{155}\) At least one Justice has recognized this conflict, and urged the Court to settle the issue.\(^{156}\)

C. **Direct Prohibition versus Conditional Funding**

A second issue which remains unclear in Free Exercise Clause jurisprudence is whether direct prohibitions on religious conduct and conditional funding of religious conduct should be subject to the same standards.\(^{157}\) Direct prohibitions, such as the ordinances prohibiting ritualistic sacrifices of animals at issue in *Hialeah*,\(^ {158}\) are different than programs that place conditions on the receipt of federal aid, such as requiring individuals receiving unemployment benefits to be available for work on particular days of the week.\(^{159}\) Both can result in a Free on the religious practice. *Id.* at 699. *But see* Employment Div. v. Smith, 494 U.S. 872, 887 n.4 (1990) (rejecting this approach as contradictory).

---


153. *Id.* The First Circuit also determined that, alternatively, the state’s interest in avoiding an Establishment Clause violation justified any potential free exercise violation. *Id.*

154. See id. at 65; *KDM v. Reedsport Sch. Dist.*, 196 F.3d 1046, 1050–51 (9th Cir. 1999).

155. See Peter v. Wedl, 155 F.3d 992, 996–97 (8th Cir. 1998).

156. See Columbia Union Coll. v. Clark, 527 U.S. 1013, 1015 (1999) (mem.) (denying cert.) (Thomas, J., dissenting) (commenting that *Wedd* and *Strout* illustrate that lower courts are “struggling to reconcile our conflicting First Amendment pronouncements”).


Requiring States to Fund Religious Training

Exercise Clause violation. However, as the U.S. Supreme Court stated in Lyng v. Northwest Indian Protective Cemetery Association, “the Free Exercise Clause is written in terms of what the government cannot do to the individual, not in terms of what the individual can exact from the government.”

A funding restriction can qualify as a violation of the Free Exercise Clause. In fact, in Sherbert v. Verner, the Court held that a state unemployment law that only provided benefits to those willing to work on Saturdays violated the Free Exercise Clause. The Court focused on the fact that the state law required some individuals to forego a central tenet of their religion in order to qualify for the state funding. Yet in more recent cases, the Court has limited the applicability of Sherbert. In Smith, for example, the Court rejected the use of the Sherbert test for challenges to generally applicable criminal laws, and noted its reluctance to expand Sherbert outside the unemployment benefits context.

Consequently, the two issues most relevant to the Davey case are unsettled. First, does a law lacking facial neutrality and general applicability automatically require strict scrutiny review, even if it does not have a purpose to suppress religion? As noted above, the Circuit Courts are split on this issue. Second, should Free Exercise Clause precedent that involves government prohibitions on religion apply to Free Exercise Clause challenges based on government funding decisions? With these uncertainties, courts must seek guidance from other fundamental rights decisions.

160. Compare id. at 403, with Hialeah, 508 U.S. at 524.
162. Id. at 451 (quoting Sherbert, 374 U.S. at 410 (Douglas, J., concurring), in upholding logging public land that is sacred to some Native Americans).
163. See, e.g., id. at 450.
165. See id. at 410.
166. See id. at 406.
168. Id.
169. Id. at 883 (“We have never invalidated any governmental action on the basis of the Sherbert test except the denial of unemployment compensation . . . . In recent years we have abstained from applying the Sherbert test [outside the unemployment compensation field] at all.”).
170. See supra Part II.B.
171. Cf. Davey v. Locke, 299 F.3d 748, 761 (McKeown, J., dissenting) (concluding that the Supreme Court’s free exercise jurisprudence did not provide “sufficient guidance” for the Davey case).
III. LOOKING BEYOND THE RELIGION CLAUSES: FUNDING VERSUS PROHIBITING THE EXERCISE OF FUNDAMENTAL RIGHTS

The distinction between a direct prohibition of conduct and the conditional funding of conduct exists in cases involving other fundamental rights as well. Supreme Court precedent indicates that the state has broad discretion to place restrictions on the use of its own funds. Because “[c]onstitutional concerns are greatest when the State attempts to impose its will by force of law,” proving that a funding program infringes upon a fundamental right is difficult. There are limits, however, on a state’s ability to selectively fund activities. Most notably, the purpose of a state’s action cannot be to “suppress dangerous ideas” by coercing individuals into refraining from expressing a disfavored viewpoint.

Although the “government may not deny a benefit to a person because he exercises a constitutional right,” it may refuse to fund the exercise of that right. is a case on point. In , a statute prohibited the use of Medicaid funding to pay for certain “medically necessary abortions” while allowing funding of other necessary procedures. The Court acknowledged that an individual has a constitutionally protected right to terminate a pregnancy, but still upheld the program against both First Amendment and Equal Protection challenges. The Court stated that although the liberty interest protected

---


174. Id. at 476.


176. Id. at 545; accord Speiser v. Randall, 357 U.S. 513, 529 (1958) (holding a state cannot deny a general tax benefit to those that speak out against the government).

177. See, e.g., Regan, 461 U.S. at 545–56 (upholding preferential tax treatment conditioned on restricting lobbying efforts); Harris v. McRae, 448 U.S. 297, 326 (1980) (prohibiting Medicaid to pay for abortions); see CHEMERINSKY, supra note 50, §11.2.4.4.


179. See id. at 301–02.

180. See id. at 312.

181. See id. at 321. The Court applied rational basis review to the Equal Protection claim. Id. at 324.
Requiring States to Fund Religious Training

“against unwarranted government interference,” it did not “confer an entitlement” to the funds needed to “realize all the advantages” of that right.  

The Court has upheld programs that restrict state funding for abortion-related services on other occasions as well. For example, in Rust v. Sullivan, the Court upheld a law that prohibited medical providers receiving government funding from even discussing abortion with their patients. The Court’s rationale was unambiguous: “[a] legislature’s decision not to subsidize the exercise of a fundamental right does not infringe the right.”

The Court has applied this distinction between selectively subsidizing a constitutionally protected right and prohibiting its exercise outside of the abortion context. For example, in Regan v. Taxation with Representation of Washington, the Court denied both free speech and equal protection challenges to a program that granted tax-exempt status to certain veteran’s organizations engaged in lobbying while refusing the status to other lobbying organizations. The Court permitted this selective funding based upon group status, and dismissed “the notion that First Amendment rights are somehow not fully realized unless they are subsidized by the State.”

The Regan Court, however, did note one limit placed upon the government’s conditional subsidizing of speech: the law cannot be “aimed at the suppression of dangerous ideas.” This limit prevents a state from using subsidies to coerce individuals into refraining from expressing a disfavored viewpoint. The Court relied on this reasoning in Rosenberger v. Rector and Visitors of the University of Virginia. In

182. Id. at 317–18.
185. See id. at 192–93.
186. Id. at 193 (internal citations omitted).
188. Id. at 545–56.
189. Id. at 546 (internal citations omitted) (quoting Cammarano v. United States, 358 U.S. 498, 515 (1959) (Douglas J., concurring)).
190. Id. at 550 (quoting Cammarano, 358 U.S. at 513). But cf. Nat’l Endowment for the Arts v. Finley, 524 U.S. 569, 596 (Scalia, J., concurring) (“It is preposterous to equate the denial of taxpayer subsidy with measures aimed at the suppression of dangerous ideas.”) (quotations omitted) (emphasis in original).
191. See Finley, 524 U.S. at 587.
Rosenberger, the Court held that a state university’s refusal to provide funds to publish religiously-oriented student newsletters, while funding the publication of non-religious student newsletters, violated the right to free speech.\footnote{193. Id. at 837.} The Court explained that the program created a “public forum” for speech, but then placed a restriction on the exercise of speech within that forum solely based on viewpoint.\footnote{194. Id. at 830–35. For a description of public forums, see, e.g., Lamb’s Chapel v. Ctr. Moriches Union Free Sch. Dist., 508 U.S. 384, 392–93 (1993).} Acknowledging that the Court had upheld other programs that declined to fund certain speech, the Court distinguished the University’s program by referencing the “suppression of dangerous ideas” exception.\footnote{195. Rosenberger, 515 U.S. at 834–35.} Although “preferential treatment of certain speakers” based on group status is permissible under Regan, placing restrictions on the “content or messages” of the speech is not.\footnote{196. Id.}

However, the Court has not consistently addressed what qualifies as the suppression of disfavored viewpoints in conditional funding cases. The Rust opinion stated that prohibiting doctors that receive federal funds from discussing abortion with their patients was not impermissible viewpoint discrimination.\footnote{197. Rust v. Sullivan, 500 U.S. 173, 193 (1991).} Instead, it simply reflected government’s choice “to fund one activity to the exclusion of the other.”\footnote{198. Id.} Yet in Legal Services Corp. v. Velazquez,\footnote{199. 531 U.S. 533 (2001).} the Court struck down a law restricting legal aid attorneys’ ability to bring suits challenging welfare programs.\footnote{200. Id. at 537.} The Court ruled that the law was designed to insulate government programs from Constitutional challenges by excluding “certain vital theories and ideas” from funding.\footnote{201. Id. at 548.} Although the programs in Rust and Legal Services both conditioned the receipt of federal funds on agreeing to comply with speech-based restrictions, they resulted in opposite outcomes in the Court.\footnote{202. See CHEMERINSKY, supra note 50, §11.2.4.4.}

Commentators have noted the difficulty in reconciling the Court’s decisions in these cases.\footnote{203. See id.} One distinction suggested in Legal Services is whether the government was acting as a speaker through its funding, or
was attempting to “facilitate private speech.” 204 For example, in *Rosenberger* the Court explained that the University’s funding was intended to encourage private speech by creating a public forum (in which the government could not engage in viewpoint discrimination), whereas the program in *Rust* was intended to convey the government’s objectives. 205 Not all justices, however, have agreed with this distinction between these cases. 206 As Justice Scalia commented, if “private doctors’ confidential advice to their patients at issue in *Rust* constituted ‘government speech,’ it is hard to imagine what subsidized speech would not be government speech.” 207

Another potentially distinguishing factor is the extent to which the government condition “distort[s]” the “usual functioning” of an “existing medium of expression.” 208 In *Legal Services*, the Court expressed concern about the impact on the court system of not allowing lawyers to present potential constitutional issues. 209 In *Rosenberger*, the Court was concerned with the restriction’s impact on what the Court had determined was an open forum. 210 In contrast, the *Rust* Court rejected the concern that the program infringed upon the physician-patient relationship. 211

In sum, U.S. Supreme Court precedent indicates that while a state is not required to subsidize the exercise of a fundamental right, it cannot use selective funding as a means of suppressing dangerous ideas. These principles are clearly established. However, as *Rust* and *Legal Services* indicate, in practice their application has proven more difficult to reconcile.

---

204. 531 U.S. at 542.
206. *See Legal Servs.*, 531 U.S. at 554 (Scalia, J., dissenting).
207. *Id.* (emphasis in original).
208. *Id.* at 543.
209. *Id.* at 545–46.
IV. IN DAVEY V. LOCKE, THE NINTH CIRCUIT HELD THAT THE PROMISE SCHOLARSHIP VIOLATES THE FREE EXERCISE CLAUSE

Although the Ninth Circuit’s decision in Davey v. Locke referenced many other constitutionally-protected rights, the court based its ruling solely on the violation of the Free Exercise Clause.212 The Ninth Circuit held that Washington’s Promise Scholarship program facially discriminated against religion, and was not justified by a compelling state interest.213 As such, the program violated the plaintiff’s right to the free exercise of religion.

The Ninth Circuit’s decision in Davey reversed a ruling by a federal district court in the Western District of Washington granting summary judgment in favor of the state. The district court had rejected Davey’s claims that the state policy violated the First Amendment’s Establishment Clause as well as his rights to the free exercise of religion, free speech, and equal protection under the state and federal constitutions.214 Reviewing the free exercise challenges, the district court concluded that Davey’s state claim was substantially the same as that rejected by the Washington State Supreme Court in Witters III.215 Because the state constitution prohibited the appropriation of state funds to subsidize religious training, the court rejected Davey’s state law claim.216 Further, the district court held that the Promise Scholarship’s requirements did not violate the First Amendment’s Free Exercise Clause,217 noting the distinction between prohibiting an activity and refusing to fund an activity.218 Finding that Davey “mistakenly presumes that he has a right to have Washington fund his religious instruction,” the district court denied his Free Exercise Clause claim.219

A panel on the Ninth Circuit reversed the district court’s decision.220 The court began its analysis by acknowledging Church of Lukumi Babalu

---

212. 299 F.3d 748, 760 (9th Cir. 2002).
213. Id. at 750.
215. Id. at *18.
216. Id.
217. Id.
218. Id. at *14.
219. Id. at *16.
220. Davey v. Locke, 299 F.3d 748, 760 (9th Cir. 2002).
Requiring States to Fund Religious Training

Aye, Inc. v. City of Hialeah as controlling. 221 The Ninth Circuit viewed the Hialeah decision as holding that a court must strictly scrutinize any law lacking facial neutrality towards religion. 222 The court stated that the Promise Scholarship was not neutral under the Hialeah standard because both the Promise Scholarship’s implementing policy and underlying statutory prohibition on state funding of religious education “refer on their face to religion.” 223

The Ninth Circuit then determined that the state’s policy of prohibiting funding for religious majors imposed an unconstitutional disability based on religious status. 224 The court relied on McDaniel v. Paty 225 for the proposition that a law cannot offer a general benefit to all, but then deny that benefit to a select group based solely on religion. 226 Comparing the state law at issue in McDaniel, which prohibited clergy from serving in the legislature, to the funding restriction in Davey, the Ninth Circuit stated that “[a] minister could not be both a minister and a delegate in Tennessee any more than Davey can be both a student pursuing a degree in theology and a Promise Scholar in Washington.” 227

The Davey majority also rejected the state’s argument that the Promise Scholarship was consistent with other conditional funding programs approved by the U.S. Supreme Court. 228 The Ninth Circuit distinguished the Promise Scholarship from these cases by relying on the Rosenberger decision, which focused on the First Amendment right to free expression in a public forum. 229 The Ninth Circuit took from Rosenberger the premise that a “restriction based on religion is aimed at ‘suppression of dangerous ideas,’” and as such was impermissible. 230 Moving from the free speech realm of the Rosenberger case to the free exercise issue in Davey, 231 the Ninth Circuit held that the Washington program was an

221. Id. at 753.
222. Id.
223. Id.
224. Id. at 754.
225. 435 U.S. 618 (1978) (plurality opinion).
226. Davey, 299 F.3d at 754.
227. Id. (emphasis in original).
228. Id. at 754–55.
230. Davey, 299 F.3d. at 755.
231. See id. at 764 (McKeown, J., dissenting) (describing the majority’s “reach[] across the First Amendment divide”).
attempt to “suppress a particular point of view,” making the program “presumptively unconstitutional.”

Having determined that the Promise Scholarship program fell under the Hialeah standard and thereby required strict scrutiny, the court held that the state lacked a compelling interest in the discrimination. The court recognized that Washington’s interest in complying with its own constitution was “indisputably strong.” However, the Ninth Circuit ruled that the Free Exercise Clause trumped this interest.

The Davey dissent, on the other hand, disagreed with the majority’s analysis of both the Free Exercise Clause and conditional funding cases. The dissent first asserted that U.S. Supreme Court precedent in Free Exercise Clause challenges was inconclusive on the issue at hand. Second, the dissent noted that the majority’s decision was inconsistent with the Court’s analysis in other fundamental rights cases, and explained that the Promise Scholarship’s rules simply reflected a permissible conditioning on the use of state funds.

As a result of Davey, a state supporting higher education through scholarships cannot exclude theology degrees without violating the Free Exercise Clause. It is irrelevant whether the object of the law is to promote a separation of church and state or simply to suppress religion. As the dissent noted, this decision curtails a state’s ability to define its own understanding of religious liberty.

V. WASHINGTON’S PROMISE SCHOLARSHIP DOES NOT VIOLATE THE FREE EXERCISE OF RELIGION

In holding that Washington’s Promise Scholarship violated the free exercise of religion, the Davey court misapplied Ninth Circuit and U.S. Supreme Court precedent. First, the court ignored Ninth Circuit

232. Id. at 755 (quoting Rosenberger, 515 U.S. at 830).
233. Id.
234. Id. at 758–60.
235. Id. at 759.
236. Id. (citing Widmar v. Vincent, 454 U.S. 263 (1981)).
237. Id. at 761–64 (McKeown, J., dissenting).
238. See id. at 761 (McKeown, J., dissenting) (concluding that “the Supreme Court’s jurisprudence in the abortion funding cases guides our decision here”).
239. Id. at 761–64 (McKeown, J., dissenting).
240. Id. at 764–68 (McKeown, J., dissenting).
241. See id. at 760.
242. See id. at 760–61, 768 (McKeown, J., dissenting).
Requiring States to Fund Religious Training

precedent by applying strict scrutiny analysis without determining if the law reflected an attempt to suppress religion.243 Had the court followed KDM and performed its analysis, the Davey court would have concluded that the Promise Scholarship did not have an object of suppressing religion, and was therefore not subject to strict scrutiny.244 Second, the Ninth Circuit failed to adequately distinguish the Promise Scholarship from permissible funding programs. As the dissent in Davey correctly explained, the state’s refusal to pay for religious training does not prohibit the exercise of religion any more than the state’s refusal to pay for an abortion prohibits the exercise of a fundamental right.245 As a result, the court’s decision contradicts U.S. Supreme Court precedent involving the funding of constitutionally-protected rights.

A. Washington’s Promise Scholarship Does Not Impermissibly Discriminate Against Religion

Washington’s refusal to fund a religious degree does not violate the Free Exercise Clause. Based on Ninth Circuit precedent, Hialeah’s strict scrutiny review is only required if the object of the law is to suppress religion.246 Because the Promise Scholarship does not have a purpose of targeting religion, the Hialeah test is inapposite.247 Further, the Davey court erred in relying on McDaniel because the Promise Scholarship does not discriminate based on religious status.248

1. The Ninth Circuit Ignored its Own Precedent by Applying Hialeah’s Strict Scrutiny Analysis in Davey

Prior to subjecting the Promise Scholarship to strict scrutiny review, the Ninth Circuit should have first determined if the program’s policies had an object of suppressing religion.249 Only three years prior to Davey, the Ninth Circuit in KDM did not apply strict scrutiny to a law lacking facial neutrality because the law as applied did not reflect an object or

243. See KDM v. Reedsport Sch. Dist., 196 F.3d 1046, 1050–51 (9th Cir. 1999).
244. See 96 F.3d 1046, 1050–51 (9th Cir. 1999).
245. See 299 F.3d at 761, 764–68 (McKeown, J., dissenting).
246. See KDM, 196 F.3d at 1050–51.
247. See id.; accord Strout v. Albanese, 178 F.3d 57, 65 (1st Cir. 1999).
248. See Davey, 299 F.3d at 753.
249. See supra notes 137–142 and accompanying text (discussing the Ninth Circuit’s approach in KDM to determining if a law lacking facial neutrality required strict scrutiny review).
purpose of suppressing religion. The court cited this lack of purposeful hostility as a distinguishing factor from Hialeah. Yet the Davey court failed to address the court’s reasoning in KDM. Although the laws at issue in Davey and KDM were different, the free exercise question was substantially the same: must a law that is not facially neutral towards religion be justified by a compelling state interest? The KDM court answered this question in the negative, instead first reviewing if the government action burdened the plaintiff and had as an object the suppression of religion. The Davey court, however, ignored this precedent and concluded that the Hialeah decision controlled because the Promise Scholarship’s policies “refer on their face to religion.”

Although both the KDM and Davey courts can draw some support from the Hialeah opinion, the KDM approach is more consistent with Supreme Court jurisprudence. As indicated by the Hialeah decision, the Court’s concern was preventing laws “designed to persecute or oppress a religion or its practices.” The Court even couched its discussion of facial neutrality in terms of deciphering if the object of a law was the suppression of religion. Summarizing why the ordinances at issue in Hialeah failed the neutrality requirement, the Court came to “one conclusion: The ordinances had as their object the suppression of religion.”

250. KDM, 196 F.3d at 1051.
251. Id. (citing Strout, 178 F.3d 57).
252. The majority opinion in Davey only refers to KDM once, and does not discuss how its analysis of the Hialeah decision is distinguishable. See Davey 299 F.3d at 759.
253. The Oregon statute did not require choosing between the pursuit of a religious practice and state funding—it only affected the manner of receiving those benefits. See KDM, 196 F.3d at 1050–51.
254. Id.
255. 299 F.3d at 753.
256. For example, the Hialeah opinion states both that “the minimum requirement of neutrality is that a law not discriminate on its face,” 508 U.S. at 533, and that “if the object of a law is to infringe upon or restrict practices because of their religious motivation, the law is not neutral.” Id.
257. Id. at 547; see also id. at 533, 542.
258. See id. at 533 (commenting that “the minimum requirement” of facial neutrality is one of “many ways” of determining the object of a law).
259. Id. at 542. Justice Kennedy, joined by Justice Stevens, also pointed to Equal Protection cases for guidance in the neutrality inquiry. 508 U.S. at 540–42 (Kennedy, J., concurring). The Court’s Equal Protection decisions support the view that the Court’s concern is with preventing animus towards particular groups. See, e.g., Romer v. Evans, 517 U.S. 620, 632 (1996); City of Cleburne v. Cleburne Living Ctr., Inc., 473 U.S. 432, 446–47 (1985); U.S. Dep’t of Agric. v. Moreno, 413 U.S. 528, 534 (1973).
Requiring States to Fund Religious Training

The Davey court’s emphasis on whether laws “refer on their face to religion”260 is also at odds with the Supreme Court’s approach in Smith. The Court noted in Smith that amending neutral laws in order to create “nondiscriminatory religious-practice exemption[s],” such as allowing peyote use in religious ceremonies, may be permissible.261 Under the Davey court’s approach,262 however, these exemptions would qualify as non-neutral, and would be subject to strict scrutiny review.

Had the Davey court followed the reasoning of KDM, it would have concluded that the Promise Scholarship did not require strict scrutiny review.263 Contrasting the facts of the Davey and Hialeah cases, it is clear that the Promise Scholarship’s requirements are not designed to suppress religion. Unlike the city ordinances at issue in Hialeah, the object of Washington’s policy does not indicate “substantial animus” against any practice of religion.264 In Hialeah, the Supreme Court emphasized that the city had specifically created ordinances in order to suppress certain religious practices.265 The Promise Scholarship has no similar purpose. In fact, the Ninth Circuit in Davey even recognized that the purpose of the Washington law was to respect the separation between church and state.266

A government’s desire to ensure the separation of church and state does not indicate an attempt to suppress religion.267 As Justice Brennan explained, courts should not “underestimate[] the role of the Establishment Clause as a coguarantor, with the Free Exercise Clause, of religious liberty. The Framers did not entrust the liberty of religious

260. Davey, 299 F.3d at 753.
262. See 299 F.3d at 753.
263. See KDM v. Reedsport Sch. Dist., 196 F.3d 1046, 1050 (9th Cir. 1999); see also Hialeah, 508 U.S. at 534, 542 (looking beyond an ordinances’ facial neutrality to determine the object of the laws).
264. See Strout v. Albanese, 178 F.3d 57, 65 (1st Cir. 1999); See also Davey, 299 F.3d at 762 (McKeown, J., dissenting) (concluding there should be “no such suspicion of animosity in this case”).
265. 508 U.S. at 542. In addition, the ordinances at issue in Hialeah also showed their discriminatory intent by carving out exceptions to certain religions, but not all. See id. at 535–36.
266. 299 F.3d at 758–59 (acknowledging Washington’s “indisputably strong interest” in not violating the state’s constitutional separation of church and state); see also id. at 761 (McKeown, J., dissenting) (“In the State of Washington’s case, it has assiduously avoided violating the first tenet of the Religion Clauses, and in doing so has not overstepped the bounds of the latter.”).
beliefs to either clause alone.”268 Neither did the Framers of the Washington Constitution.269 Article I, Section 11 of the Washington Constitution, which contains the prohibition on state funding of religion, is entitled “Religious Freedom.” It begins: “[a]bsolute freedom of conscience in all matters of religious sentiment, belief and worship, shall be guaranteed to every individual.”270 This is not the language of hostility towards religion.271 In deciding not to fund religious training, Washington is not attempting to suppress religion, but to respect the separation of church and state.

2. The Ninth Circuit Incorrectly Analogized the Promise Scholarship in Davey to the Statute in McDaniel

The majority’s reliance on the McDaniel Court’s religious status doctrine is also misplaced. The majority reasoned that the Promise Scholarship discriminated based on religious status in a manner similar to the Tennessee law in McDaniel that prohibited clergy from holding public office because of their religious status.272 Yet the two cases are dissimilar. In McDaniel, Tennessee penalized the plaintiff—took away a right to hold public office—because of his status as a minister.273 In contrast, the plaintiff in Davey demanded that the state pay for an activity

268. Id. at 256 (Brennan, J., concurring).
271. Commentators have suggested that, at the time of their passage in the late 1800’s and early 1900’s, similar provisions in state constitutions were based on hostility towards Catholic schools. See, e.g., JOSEPH P. VITERITTI, CHOOSING EQUALITY: SCHOOL CHOICE, THE CONSTITUTION, AND CIVIL SOCIETY 151–68 (1999). But cf. Noah Feldman, Non-sectarianism Reconsidered, 18 J. L. & Politics 65, 112 (2002) (acknowledging that “opposition to state funding of Catholic schools” played a role in the creation of these state provisions, but arguing that “non-sectarianism was broadly understood as a plausible and correct solution to the problem of moral education and religious heterogeneity”). Other commentators have concluded that the passage of Washington’s establishment clause, while repudiating “government-backed sectarianism,” also reflected a “lack of hostility toward religion generally.” Robert F. Utter & Edward J. Larson, Church and State on the Frontier: the History of the Establishment Clauses in the Washington State Constitution, 15 HASTINGS CONST. L.Q. 451, 472 (1988). Of note, Washington’s statutory prohibition on state funding for theology students was enacted in 1969, see 1969 WASH. LAWS EX. SESS. 222 § 15, long removed from the time period of “religious tensions that plagued the United States during the latter third of the nineteenth century.” See Toby J. Heytens, Note, School Choice and State Constitutions, 86 VA. L. REV. 117, 140 (2000).
272. Davey v. Locke, 299 F.3d 748, 754 (9th Cir. 2002).
Requiring States to Fund Religious Training

it had chosen not to subsidize—regardless of religious status.\textsuperscript{274} The plaintiff in \textit{Davey} was not being penalized for pursuing a religious degree any more than the plaintiff in \textit{Harris v. McRae} was penalized by Medicaid’s refusal to pay for an abortion.\textsuperscript{275} As the Supreme Court stated in \textit{Harris}, the funding decision simply represents “a refusal to subsidize certain protected conduct.”\textsuperscript{276}

The Court’s rationale in \textit{Harris} indicates why the Promise Scholarship must be reconciled with the Court’s review of other funding programs. As explained below, the Court has distinguished between those programs that penalize the exercise of a constitutionally protected right, and those programs that simply refuse to fund the exercise of one’s constitutionally protected right.\textsuperscript{277} The former is impermissible, the latter is not.\textsuperscript{278} \textit{Davey} falls into the latter permissible category.

\textbf{B. The Ninth Circuit Failed to Adequately Distinguish the Promise Scholarship’s Requirements from Permissible Government Funding Programs}

Comparing the Promise Scholarship to similar programs involving the conditional funding of protected activities demonstrates that the Promise Scholarship’s requirements are permissible.\textsuperscript{279} As the dissent in \textit{Davey} asserted, the requirements placed on the Promise Scholarship by Washington are similar to those approved by the U.S. Supreme Court in other conditional funding cases.\textsuperscript{280} Although the Ninth Circuit attempted to distinguish the Promise Scholarship based on \textit{Rosenberger}, analysis of that case demonstrates that the cases are dissimilar.

\textsuperscript{274} See \textit{Davey}, 299 F.3d at 751.
\textsuperscript{275} Compare \textit{Harris v. McRae}, 448 U.S. 297, 301 (1980) (reviewing the constitutionality of “denying public funding for certain medically necessary abortions”), with \textit{Davey}, 299 F.3d at 749 (reviewing the constitutionality of denying public funding of theology degrees). \textit{See also Davey}, 299 F.3d at 764–65 (McKown, J., dissenting) (noting the “indistinguishable similarity between this case and those that address the abortion funding cases”).
\textsuperscript{276} \textit{Harris}, 448 U.S. at 317.
\textsuperscript{277} See infra Part V.B.
\textsuperscript{278} See infra Part V.B.
\textsuperscript{280} \textit{Davey}, 299 F.3d at 761, 764–68 (McKown, J., dissenting).
1. The Promise Scholarship’s Requirements are Analogous to Other Permissible Limitations Placed on Government Funding Programs

As illustrated by similar funding programs upheld by the U.S. Supreme Court, the Promise Scholarship’s requirements are permissible. On a funding basis, the government action at issue in Davey is similar to the law in Regan v. Taxation with Representation of Washington. In both cases, the plaintiff claimed that the government’s decision to selectively fund certain activities violated their First Amendment rights and the Equal Protection Clause. In Regan, the First Amendment right was to free speech; in Davey the court focused on the right to the free exercise of religion. Both claimed that once the state decided to generally subsidize an activity (lobbying in Regan; education scholarships in Davey), the state could not then selectively decline to fund other activities. The Regan Court rejected this view, stating that “although government may not place obstacles in the path of a [person’s exercise of a liberty], it need not remove those not of its own creation.” The same logic applies to the scholarship program at issue in Davey. Washington cannot prohibit students from studying theology, but it can decide not to fund those studies.

As noted above, the Court’s decision in Harris also supports this conclusion. The Promise Scholarship excepted religious degrees from an otherwise general funding program; the law at issue in Harris similarly excepted specific procedures, involving constitutionally protected rights, while funding other procedures in general. The Court in Harris determined that this policy was permissible, stating that the program’s condition left a woman with “at least the same range of choice” regarding an abortion as she would have had if Congress did not subsidize health care at all.

281. Compare Regan v. Taxation with Representation of Wash., 461 U.S. 540, 542 (1983) (reviewing statute providing tax exemptions to certain nonprofit groups), with Davey, 299 F.3d at 750 (reviewing statute and policy providing scholarship to certain students); see also Appellees’ Brief at 18–20, Davey v. Locke, 299 F.3d 748 (9th Cir. 2002) (No. 00-35962).
282. Compare Regan, 461 U.S. at 542, with Davey, 299 F.3d at 750.
284. 299 F.3d at 760.
285. Compare Regan, 461 U.S. at 545, with Davey, 299 F.3d at 753.
286. 461 U.S. at 549 (quoting Harris v. McRae, 448 U.S. 297, 316 (1980)).
287. See supra Part V.A.2.
288. See 448 U.S. at 301–02.
289. Id. at 317.
The law at issue in *Davey* is indistinguishable. The state opted in general to subsidize students that met the requirements of the Promise Scholarship, but not the pursuit of a theology degree.\(^{290}\) The Promise Scholarship’s requirements leave students “with at least the same range of choice”\(^{291}\) in deciding upon a major as they would have had if the legislature had decided not to have the scholarship in the first place.\(^{292}\)

Justice White’s concurring opinion in *Harris* also supports this assertion. According to Justice White, the government’s refusal to fund abortion procedures was not intended “to interfere with or to impose any coercive restraint” on a woman’s liberty.\(^{293}\) Because there was no coercive restraint, the state’s acknowledged “legitimate interest in a potential life” was enough to rationally justify the conditional funding.\(^{294}\) This same reasoning applies equally to the Promise Scholarship. The majority in *Davey* acknowledged that the state has an “indisputably strong interest in not appropriating or applying money to religious instruction as mandated by its constitution.”\(^{295}\) Far from seeking “to interfere with or impose any coercive restraint”\(^{296}\) on Davey’s pursuit of a theology degree, the state attempted to comply with its own constitution by not getting entangled in religious training.\(^{297}\)

The *Davey* case is distinguishable from the *Legal Services* case as well. In striking down the restriction on federally-subsidized attorneys challenging the constitutionality of welfare programs, the Court in *Legal Services* emphasized the concern over manipulating the proper functioning of the courts.\(^{298}\) Attorneys could not raise potential issues to the court, suggesting a separation of powers issue by requiring judges to rule on the constitutionality of programs without the benefit of hearing all possible arguments.\(^{299}\) The Promise Scholarship raises no such separation of powers concerns. Rather, it is more analogous to the government restrictions on the advice family planning doctors receiving

---

292. *Cf. Davey*, 299 F.3d at 768 (McKeown, J., dissenting) (“Davey is just as reliant on private sources of aid for his education as he was before he applied for the scholarship funds.”).
293. *Harris*, 448 U.S. at 328 (White, J., concurring).
294. *Id.*
295. *Davey v. Locke*, 299 F.3d 748, 759 (9th Cir. 2002).
296. *Harris*, 448 U.S. at 328 (White, J., concurring).
297. *See Davey*, 299 F.3d at 761 (McKeown, J., dissenting).
299. *Id.*
federal funding could give to their patients in *Rust v. Sullivan*, which the Court upheld.300 In so holding, the Court noted that participants could reasonably be expected to know the conditional nature of the program.301 According to the Court, the program’s restrictions did not compromise the physician-patient relationship because the participants should have known the bounds of the program, and that they were free to go outside of the program to get abortion-related services.302 In a similar manner, students receiving the Promise Scholarship were free to pursue a theology degree,303 just not as a Promise Scholar. Thus, Washington’s Promise Scholarship fell within the bounds of permissible funding programs under Supreme Court precedent.

2. *Washington’s Refusal to Fund a Religious Degree Does Not Suppress Dangerous Ideas*

The majority in *Davey* also mistakenly concluded that a restriction on the funding of a religious degree is aimed “at the suppression of dangerous ideas.”304 The court’s reliance on the *Rosenberger* decision fails because the *Davey* case did not involve a state-created public forum in which the state attempted to regulate speech based on viewpoint.305 As a recipient of the Promise Scholarship, Davey could speak as freely on religious matters as he could if he were not a recipient.306 This was not the case in *Rosenberger*, where the school conditioned funds on agreeing to not use the money for religious speech within the public forum.307 As noted, the Promise Scholarship allows recipients to use state funds to

---

301. See id. at 200 (noting the program does not “justify an expectation on the part of the patient of comprehensive medical advice”).
302. Id. at 196–200; accord Legal Servs., 531 U.S. at 546.
303. Washington law prohibited the state from providing scholarships to students pursuing theology degrees; it did not prohibit students from pursuing theology degrees. See WASH. REV. CODE § 28B.10.814 (2002).
304. Davey v. Locke, 299 F.3d 748, 755 (9th Cir. 2002).
305. See Davey, 299 F.3d at 766–68 (McKeown, J., dissenting); accord Davey v. Locke, 2000 U.S. Dist. LEXIS, 22273, at *21 (W.D. Wash. 2000), rev’d, 299 F.3d 748 (9th Cir. 2002).
306. In addition, to distinguish the *Rosenberger* and *Regan* cases, the U.S. Supreme Court, in *Rosenberger*, noted that the *Regan* decision was based on the “preferential treatment” of certain groups (veterans), and not the content of the speech as in *Rosenberger*. 515 U.S. at 834. Similarly, the Promise Scholarship’s requirements are based on group status, not the content of those group’s speech. See WASH. ADMIN. CODE § 250-80-020(12)(f) (2001); WASH. REV. CODE § 28B.10.814 (2002).
307. See 515 U.S. at 825.
enroll in religion classes and to attend religiously affiliated schools; it simply declines to fund an education culminating in a theology degree. As such, it doesn’t attempt to close or limit access to a public forum based on viewpoint.\textsuperscript{308}

Further, if the \textit{Davey} court was correct and the Promise Scholarship’s conditions are aimed at the suppression of dangerous ideas, then so are all “no aid” conditions supported by a half century of Establishment Clause decisions.\textsuperscript{309} Over that time, the U.S Supreme Court has consistently approved of government funding programs that required secular or religiously-neutral use of state funds, while frequently striking down programs without these restrictions.\textsuperscript{310} Under the Ninth Circuit’s rationale in \textit{Davey}, a program could not subsidize the construction of academic facilities generally but “expressly prohibit[] use of the facilities for religious purposes.”\textsuperscript{311} Yet in \textit{Tilton}, a similar type of program survived an Establishment Clause challenge specifically because it had this requirement.\textsuperscript{312} In fact, the Court struck down the portion of this law that allowed religious schools to use these buildings for religious purposes after twenty years as a violation of the Establishment Clause.\textsuperscript{313}

\begin{footnotesize}
\begin{enumerate}
\item [308] In non-public forum cases, a funding restriction is permissible unless it threatens “to drive certain ideas or viewpoints from the marketplace.” See Nat’l Endowment for Arts v. Finley, 524 U.S. 569, 587 (1998) (internal citations omitted); see also Legal Servs. Corp. v. Velazquez, 531 U.S 533, 552 (Scalia, J., dissenting). Considering that Promise Scholars are free to express religious opinions and views openly, it is doubtful that Davey could meet this requirement. See \textit{Davey}, 299 F.3d at 766–68 (McKeown, J., dissenting).
\item [309] See supra part I.B.1 (discussing Supreme Court precedent involving “no aid” to religion provisions).
\item [310] See supra part I.B.1. In \textit{Mitchell v. Helms}, 530 U.S. 793 (2000), Justice Thomas’ plurality opinion commented that “nothing in the Establishment Clause requires the exclusion of pervasively sectarian schools from otherwise permissible aid programs, and other doctrines of this Court bar it.” \textit{Id.} at 840. Yet a majority of the Court has never accepted this view. Cf. \textit{id.} at 840 (O’Connor, J., concurring in judgment) (stating she “disagree[d] with the plurality’s conclusion that actual diversion of government aid to religious indoctrination is consistent with the Establishment Clause”).
\item [311] See \textit{Tilton v. Richardson}, 403 U.S. 672, 677 (1971) (plurality opinion); \textit{accord} Hunt v. McNair, 413 U.S. 734, 759 (1973); Roemer v. Bd. of Pub. Works, 426 U.S. 736, 759 (1976) (plurality opinion) (stating that “institutions are not so permeated by religion that the secular side cannot be separated from the sectarian”) (citation omitted).
\item [312] See 403 U.S. at 679–80; cf. Agostini v. Felton, 521 U.S. 203, 211–12, 235 (1997) (allowing public teachers to provide remedial instruction at religious schools, noting that “safeguards” ensured that no religious aid was provided through the program).
\item [313] \textit{Tilton}, 403 U.S. at 683. All justices agreed with the plurality that allowing religious use of the facilities after twenty years violated the Establishment Clause. See \textit{id.} at 692 (Douglas, J., dissenting in part); Lemon v. Kurtzman, 403 U.S. 602, 660–61 (1970) (Brennan, J., concurring in part with \textit{Lemon} and concurring and dissenting in part with \textit{Tilton}); \textit{id.} at 665 n.1 (White, J., concurring and dissenting in part with \textit{Lemon} and concurring in judgment with \textit{Tilton}).
\end{enumerate}
\end{footnotesize}
Although the Ninth Circuit borrowed from other areas of First Amendment jurisprudence in its Davey opinion, it did not address this Establishment Clause precedent. Instead, the Davey court relied heavily on the Rosenberger Court’s analysis of the Establishment Clause in public forum cases. But as Justice O’Connor stated in a concurring opinion in Rosenberger, that decision did not “signal[] the demise of the funding prohibition in Establishment Clause jurisprudence.” Had the Ninth Circuit looked at the Establishment Clause jurisprudence, it would have concluded that the Promise Scholarship does not impermissibly suppress “dangerous ideas.”

C. Absent a “Free Exercise Problem,” Strict Scrutiny is Not Required

Assuming that there is no free exercise problem with the conditions of the Promise Scholarship, Washington did not need to establish a compelling government interest. The Ninth Circuit acknowledged in Davey that “[t]he cases upon which [the state] relies . . . indicate that states may rely on their own (or the federal) establishment clause if there is no free exercise problem.” The Davey majority justified not following these cases because it found that a free exercise “problem” existed. Thus, it follows that if the Promise Scholarship’s requirements did not indicate a free exercise “problem,” then the Ninth Circuit should not have used strict scrutiny analysis. Instead, the court should

314. See supra notes 228–233 and accompanying text.
318. See, e.g., KDM v. Reedsport Sch. Dist., 196 F.3d 1046, 1050–51 (9th Cir. 1999). Strict scrutiny would not be required under the Hernandez test because Davey could not prove a “substantial burden on the observation of a central belief or practice.” See Strout v. Albanese, 178 F.3d 57, 65 (1st Cir. 1999) (citing Hernandez v. Comm’r, 490 U.S. 680, 699 (1989)) (emphasis in original). As discussed in the conditional funding cases, see supra Part V.B.1, the fact that Washington refuses to fund the pursuit of a religious degree does not place Davey in a worse position than he would have otherwise been. See Davey v. Locke, 299 F.3d 748, 768 (9th Cir. 2002) (McKeown, J., dissenting); cf. Harris v. McRae, 448 U.S. 297, 316–17 (1980) (concluding that Congress’s refusal to pay for a medically necessary abortion "leaves an indigent woman with at least the same range of choice . . . as she would have had if Congress had chosen to subsidize no health care costs at all"). In addition, as the dissent in Davey makes clear, Davey has not factually shown that the restriction substantially burdened his exercise of religion. 299 F.3d at 764 (McKeown, J., dissenting). Indeed, Davey continued to pursue his studies, “finding available after-school work to make up the difference” of the scholarship. Id. (McKeown, J., dissenting) (emphasis in original).
319. Davey, 299 F.3d at 759 (emphasis in original).
320. See supra Parts V.A.1–V.A.2.
only have applied rational basis review to determine the Promise Scholarship’s constitutionality.\textsuperscript{321}

Under a rational basis review, Washington’s funding decision is justified. The decision is rationally related to legitimate government interests, including ensuring the degree of church and state separation required under the Washington constitution.\textsuperscript{322} Although the Ninth Circuit held that Washington’s interest in not violating its own constitution could not survive strict scrutiny,\textsuperscript{323} the court nevertheless acknowledged that this interest was “indisputably strong.”\textsuperscript{324} Consequently, Washington has a legitimate interest in not violating its own constitution and satisfies rational basis review. The Ninth Circuit came to a similar conclusion in \textit{KDM}.\textsuperscript{325} Addressing the petitioner’s Equal Protection Clause claim in \textit{KDM}, the Ninth Circuit held that the state’s interest in not violating its own state constitution was a legitimate government interest.\textsuperscript{326} If strict scrutiny is not required of this law and the Promise Scholarship is subject to rational basis review, then it follows that Washington’s Promise Scholarship is constitutional.\textsuperscript{327}

\textbf{VI. CONCLUSION}

Davey has the right to pursue a theology degree, but Washington has no duty to subsidize it. Washington believes that using state funds to subsidize religious training erodes the religious liberty that its state constitution attempts to protect. Because the object of the Promise Scholarship is to support higher education within the confines of the state

\begin{itemize}
\item \textsuperscript{321} See \textit{KDM}, 196 F.3d at 1050–51.
\item \textsuperscript{322} See \textit{Davey}, 299 F.3d at 759.
\item \textsuperscript{323} \textit{Id.} at 759–60.
\item \textsuperscript{324} \textit{Id.} at 759.
\item \textsuperscript{325} 196 F.3d 1046, 1052 (9th Cir. 1999).
\item \textsuperscript{326} \textit{Id.}
\item \textsuperscript{327} The Promise Scholarship might also be able to survive strict scrutiny analysis, if required. The U.S. Supreme Court in \textit{Widmar v. Vincent} did not rule out the possibility that a state’s attempt to comply with the mandates of its own constitution could ever rise to the level of a “compelling interest” justifying an infringement on the Free Exercise Clause. See 454 U.S. 263, 275 (1981). If it is possible, the Promise Scholarship would be a good candidate. First, the state’s long history of ensuring religious liberty through the separation of church and state can be regarded as government action “protecting an interest of the highest order,” as required by \textit{Hialeah}. See Church of Lukumi Babalu Aye, Inc. v. City of Hialeah, 508 U.S. 520, 547 (internal citations omitted). Second, the Promise Scholarship’s rules are narrowly tailored to conform with the state’s Constitution without being overly broad, as evidenced by the fact that students can use the scholarship at religiously affiliated schools. The Court’s federalism jurisprudence and emphasis on state’s rights might also support this conclusion. See \textit{Davey}, 299 F.3d at 768 (McKeown, J., dissenting).
\end{itemize}
constitution and not to suppress religion, it should not be subject to strict scrutiny. The Promise Scholarship’s conditions represent a rational choice by the state legislature to comply with the state constitution. Compared to other conditional funding programs approved by the Supreme Court, Washington’s Promise Scholarship is constitutional.
THIS ISSUE OF THE WASHINGTON LAW REVIEW
IS DEDICATED TO THE MEMORY OF
LUVERN V. RIEKE
LUVERN V. RIEKE: IN MEMORIAM

William B. Stoebuck

This will be very personal. It must be, to get the measure of the man. Luvern V. Rieke was my teacher, confidante, colleague, and friend. He was a product of the State of Washington, from the small town of Cashmere—euphonious name, just east of the Cascades, between Leavenworth and Wenatchee, home of Aplets and Cotlets. His parents were German; it was their native language—and Lutheran, Lutheran in the marrow of their bones. During World War II, Captain Luvern Rieke was in the Army Air Corps, with service in China: a Flying Tiger. After the war, he studied a time, for some reason (probably because it was a Lutheran school), at Capital University, Columbus, Ohio. There he met and married Jane, his gracious life mate and now widow. Returning to Seattle, he received his B.S. in 1948, his LL.B. in 1949, from the University of Washington. He was Editor-in-Chief of the Washington Law Review and Order of the Coif.

Upon graduation from our law school, Rieke had plans to open a law practice in the little town of Quincy, east of the Cascades. But the law school had other plans for him; he was immediately hired as an assistant professor. His initial teaching assignments were Contracts and Domestic Relations, the courses that were to continue to be his main ones throughout his long teaching career. During the 1952–1953 school year, he took a leave of absence from Washington to earn an LL.M. degree at the University of Chicago, where he studied under Karl Llewellyn (whose fame probably brought Rieke to Chicago). Much later in his career, Pacific Lutheran University awarded him the honorary degree of Doctor of Laws. In 1956, he attained the rank of full professor in our law school. That was the year I entered our law school and had Rieke for Contracts in my first year, and later his course in Domestic Relations.

As a classroom teacher, he was one of the best in the law school, and his teaching style influenced my own. He taught with a good mixture of the so-called Socratic method plus lively, interesting lecture; and his level of teaching was intellectually challenging. But he was the exemplar of a true gentleman, never overbearing, never belittling. Approachable out of class as well as well as in class, he was a faculty member whom I sought for advice as a student when thinking about my career in the law.
When about to leave law practice and enter law teaching, I again sought his advice.

From the time I joined our faculty in the fall of 1967 until Rieke retired in 1986, he was my faculty colleague. He was acting dean for about two years around 1970, between the deanships of Lehan K. Tunks and Richard S. L. Roddis. That was a time of trial for the campus, the “Vietnam era,” as those of us who went through it will always call it. As acting dean, Vern Rieke was a man for that time: calm, steady, of wise judgment, and through it all a fine gentleman. After his time as acting dean, he continued teaching, as before, mainly Contracts and Domestic Relations. The latter course was, I think, closer to his heart, because it involved the things about which he most cared, human relationships and how they might be bettered. He was, if I recall correctly, a principal draftsman of Washington’s no-fault divorce statute. Always hopeful about human relations, he believed that, rather than making it difficult for couples to get out of a failed relationship, we should tighten the requirements for getting married.

After his stint as acting dean, Rieke remained a pillar of strength on the faculty, both as teacher and colleague, until (the rest of the faculty felt) his untimely retirement in 1986 at age 65 (the university’s then-mandatory age was 70). He was trim, fit, and athletic as always, and seemed a candidate, if there ever was one, to continue on till the mandatory age. But he had other things he wanted to do with his life, things he had long done, but to which he could not devote full time. A devout Christian, of Lutheran persuasion, he was already one of the leading Lutheran laymen in America. It was mainly to devote more time to his church that Vern wanted to open a new chapter in his life. Vern and Jane’s daughter, Janice Cunningham, is a graduate of our law school and practices as an estate planner with a firm in Seattle. Their son, Paul Rieke, is a Seattle lawyer, too, though not one of our graduates. Janice’s husband, John, is a lawyer, a graduate of Marquette. She had one year at Marquette when they were married, then she took her last two years at Washington and John his last year here.

After Vern retired, we missed him around the law school, where none of us recalls seeing him (though he told me he continued to do some pro bono legal work for causes in which he was interested). He and Jane attended a reception to which my wife and I invited them, less than a year and a half before his death. He was as fit, trim, and sharp as ever and looked as if he would live to be at least a hundred. Then, on the ninth of May, I heard indirectly that Vern had had a massive stroke and
Dedication to Luvern Rieke

was not expected to live. When my telephone call to his home only brought Jane’s voice on the answering machine, inquiry at Janice’s law office brought me the sad news that Vern had died that morning. I learned that he had just finished supervising the completion of a new wing onto their church—and the planting of some flowers around the new wing. In a way, that seems like a fitting close to the “new chapter” in Vern Rieke’s life—but it was still much too short a chapter. When Jane returned my call, she said it for all of us who knew him, “We have lost a mighty good man.”
FISH AS POLLUTANTS: LIMITATIONS OF AND CROSSCURRENTS IN LAW, SCIENCE, MANAGEMENT, AND POLICY

Jeremy Firestone∗
Robert Barber+

Abstract: When we think of pollutants, we either consciously or unconsciously draw a bright line between pollutants and what might be called “natural.” That which is natural cannot be a pollutant; that which is a pollutant cannot be natural. It seems odd to speak of live fish as pollutants, as odd as it would be to speak of dioxins as natural. Nevertheless, the traditional definition of fish as natural may be fading as our awareness of the adverse environmental effects of accidental or poorly planned fish introductions increases. Along these lines, a federal court recently found that non-native Atlantic salmon that escape from their pens are “pollutants” within the meaning of the Clean Water Act. Because wild Atlantic salmon is listed as an Endangered Species, Salmon mariculture provides a particularly stark example of when society might aptly consider “fish” to be pollutants. The biological, philosophical, and legal underpinning of our argument, however, transcends aquaculture into the realm of fisheries management, where we advocate that managers focus on improving water quality to the point where the native fish that historically were dominant in the habitat are once again abundant, rather than on managing aquatic ecosystems for stocked species of fish that are relatively unaffected by degraded water quality.

I. INTRODUCTION: FISH AS POLLUTANTS

When we think about pollutants, we either consciously or unconsciously construct a dichotomy of factors that affect environmental quality and draw a bright line between pollutants and what might be called “natural”—what should be in the water, land, or air, what pollutants affect or alter. In this dichotomy of thought, a given factor cannot exist on both sides of the divide. That which is natural cannot be a pollutant; that which is a pollutant cannot be natural. Fish are natural and either affect water quality positively or function as a sort of benign barometer of the effects of pollutants. In contrast, noxious chemicals,
such as dioxins, are pollutants and can only negatively affect water quality. It may seem odd to speak of live fish as pollutants, as odd as it would be to speak of dioxins as natural. Nevertheless, the traditional definition of fish as natural may be fading in both a philosophical and judicial sense as we become more aware of how accidental or poorly planned introductions of fish can adversely affect the environment.

It has been understood for a long time by fisheries managers that the introduction of exotic fish species can negatively impact aquatic ecosystems. However, it has also recently become apparent that even the introduction of native species may be undesirable. Through ever more sensitive techniques of genetic analysis, it has become possible to differentiate wild from cultured fish of the same species. This in turn has led to a great deal of speculation regarding the impact of cultured fish on wild fish populations, particularly in situations where populations of wild fish are declining. Do cultured fish compete with wild fish for resources? Can cultured fish serve as a vector by which diseases and parasites enter wild populations? Might the introduction of genetic material from hatchery or aquaculture fish to a wild population decrease genetic fitness and consequently survival in the wild population?

In this Article we examine the question of whether fish are and should be treated as pollutants in some contexts. We consider this question primarily in the context of Atlantic salmon mariculture operations—sea-based fish farming operations—that exist off the coast of Maine. Atlantic salmon mariculture presents a particularly appealing milieu in which to examine this question for the following reasons: Atlantic salmon mariculturalists employ strains or stocks of Atlantic salmon that are not native to the Gulf of Maine; there have been documented escapes of farmed Atlantic salmon; recent judicial developments suggest that as a legal matter, such escapees are pollutants; and finally, native, wild populations of Atlantic salmon are endangered. Atlantic salmon that escape from mariculture operations impact wild Atlantic salmon in numerous ways, as do other cultivated fish that are accidentally or intentionally released into waters containing wild populations of the same species. Impacts include the introduction of exotic species or

---

1. Atlantic salmon mariculture is a form of aquaculture. Aquaculture refers to the farming of aquatic organisms—not only farming of finfish, but also shellfish, crustaceans, and even aquatic plants, in either fresh or saltwater. Mariculture is saltwater aquaculture or a subset of aquaculture. See REBECCA J. GOLDBURG ET AL., PEW OCEAN COMM’N, MARINE AQUACULTURE IN THE UNITED STATES: ENVIRONMENTAL IMPACTS AND POLICY OPTIONS 3 (2001), at http://www.pewoceans.org/reports/137PEWAquacultureF.pdf.
Fish as Pollutants

varieties of fish to new bodies of water, genetic contamination of the wild genome, predation on wild fish, competition with wild fish for food and favorable space, disruptive behavior, stimulation of premature migrations, creation of unacceptably high densities of fish, mixed-stock exploitation problems, predator attraction, and disease and parasite transmission. Potentially, the most serious of these impacts, and the most difficult to document until recent advances in genetic science, is genetic contamination of the wild genome of Atlantic salmon by the genome of cultured Atlantic salmon. Parasite transmission to wild fish is already considered a potentially serious problem in other segments of the aquaculture industry, such as trout farms that provide stocked fish for sport fishing. In the case of the Atlantic salmon, concern has been expressed recently about the possibility of communication of Infectious Salmon Anemia Virus (ISAV) to wild Atlantic salmon from mariculture fish. As the Atlantic salmon mariculture industry grows, escapes of mariculture salmon will increase, and the impacts of mariculture fish on wild Atlantic salmon are likely to be correspondingly severe.

In Part II, we trace the historical abundance and decline of Atlantic salmon in the northeast United States, its life cycle, and the growth of salmon mariculture production in the United States. In Part III, we turn to the legal regime that regulates the discharge of pollutants from mariculture facilities and consider specific rulings that addressed discharges from salmon mariculture and other aquaculture facilities. In Part III, we focus on the issue of whether an escaped aquaculture


specimen is a “pollutant” within the meaning of the law. We then examine the other side of the legal equation in Part IV—the legal status of wild Atlantic salmon populations. In Part V, we consider one threat—disease transmission—that introduced fish populations pose to wild fish populations. Finally, in Part VI, we consider the policy implications of treating non-native and introduced fish as pollutants for the regulation of mariculture and perhaps, more significantly, for the stocking of fish.

II. BIOLOGY, ABUNDANCE AND DECLINE OF ATLANTIC SALMON AND THE RISE OF SALMON MARICULTURE

A. Historical Distribution, Abundance, and Decline of Atlantic Salmon

Atlantic salmon, not to be confused with the five Pacific salmon species—Chinook, Coho, Sockeye, Chum, and Pink—is a distinct species native to the North Atlantic Ocean. On the North American coast, their historic range extended from Connecticut to Labrador, while in the eastern Atlantic Ocean, they existed from northern Africa to northeast Russia. Atlantic salmon also are found in Icelandic waters and in waters of Southern Greenland.

Atlantic salmon have historically been very abundant in the northeastern United States. No more than 200 years ago, as many as 500,000 Atlantic salmon returned each year to spawn in New England rivers. Historical returns ranged from 300,000 to 500,000 fish. See U.S. Fish & Wildlife Serv., Biological Report on Atlantic Salmon Abundance, § 4.1.2, available at http://library.fws.gov/salmon/asalmon4.html (Oct. 8, 1999) [hereinafter Biological Report]. The report cites L. Stolte for the 300,000 figure and K. Beland for the 500,000 figure. L. STOLTE, DEP’T OF THE INTERIOR, THE FORGOTTEN SALMON OF THE MERRIMACK (1981); K. BELAND, ATL. SEA RUN SALMON COMM’N, STRATEGIC PLAN FOR MANAGEMENT OF ATLANTIC SALMON IN THE STATE OF MAINE (1984).
Fish as Pollutants

and possibly strayed as far south along the coast as the Hudson River. Much of this early evidence of the Atlantic salmon’s historical numbers was derived from the private journals of settlers and expeditions and constitutes a body of material that now reads like a poetic epitaph to the salmon’s former abundance. Here is one example of those early testimonials from a report compiled for the United States Fisheries Commission in 1802:

It was nothing uncommon for teams fording the rivers and creeks at night to kill salmon with their hoofs. An older settler living in the town of Hannibal told Mr. Ingersoll that one night while driving across Three-Mile Creek the salmon ran against his horses’ feet in such large numbers that the horses took fright and plunged through the water, killing one large salmon outright and injuring two others so that they were captured. The farmers living near the smaller creeks easily supplied their families with salmon caught by means of pitchforks.

By the beginning of the last century, Atlantic salmon were already in a state of serious decline in the United States, most importantly because of the construction of mainstream dams on a number of important salmon rivers that blocked passage upstream, and also because of pollution and other environmental factors. By the beginning of the Twentieth century, Atlantic salmon runs had been completely extirpated on such historically important U.S. Atlantic salmon rivers as the Connecticut and the Merrimack. Nevertheless, significant runs of salmon, at least in the sense of representing a genetic reservoir of the southernmost Atlantic salmon populations in North America if not in terms of absolute numbers, continued on a number of other salmon rivers, particularly in the northern part of the Atlantic salmon’s U.S. range. Those runs included Atlantic salmon that breed in the Pleasant,
Dennys, Machias, East Machias, Narraguagus, Sheepscot, and Ducktrap rivers in Maine, now known as the Gulf of Maine Distinct Population Segment (Gulf of Maine D.P.S.).

Over the past three decades, there has been an accelerated decline in Atlantic salmon numbers. In addition to overfishing and environmental degradation, Atlantic salmon mariculture has been cited as contributing to that decline. Taking Canadian and United States Atlantic salmon together, there has been a ten-fold drop in the number of two-sea-winter Atlantic salmon returning to spawn from the mid-1970s to the present. These fish represent almost all of the spawning potential of all wild North American Atlantic salmon. The vast majority of these fish are of Canadian origin, with only several hundred fish of United States origin.


15. See infra notes 226–39 and accompanying text.


17. For a discussion of the risks posed to wild Atlantic salmon by Atlantic salmon mariculture, see Biological Report, supra note 9.

18. Two-sea-winter Atlantic salmon are fish that spend two winters at sea between the time they leave the freshwater habitats in which they were spawned for the sea and the time they return to those same freshwater habitats to spawn. See John Kocik & Russell Brown, Nat’l Marine Fisheries Serv., Atlantic Salmon, available at http://www.nefsc.noaa.gov/sos/spsyn/af/salmon/ (Mar. 2001).


21. For the total number of salmon returning to North American rivers over the past three decades see Anderson, supra note 13, at fig. 1. For the total number of salmon returning to U.S. rivers see U.S. Atl. Salmon Assessment Comm., supra note 16, at tbl. 2.3.1. For the total number of
Fish as Pollutants

Today, Atlantic salmon are found in the United States in major river systems from the Connecticut River northward into Maine. Not all of these populations are wild. A number of populations have been restored to rivers where Atlantic salmon historically existed but were later wiped out. These populations are generally not self-sustaining, but have been maintained artificially through stocking. The most outstanding example of this is the Atlantic salmon population of the Connecticut River. A restoration program was begun on the Connecticut River in the late 1960s. Restoration efforts included the creation of fish ladders on all mainstream dams and an extensive stocking program. From 1967 to 1999 a total of 68,381,900 Atlantic salmon were stocked in the Connecticut River, primarily as fry. The total number of adult salmon returning to the river to spawn over the same time period was 4835 fish, or an average of 147 fish per year. It is highly likely that these returns would disappear completely without the support of the stocking program, which means that the Atlantic salmon has not returned to the Connecticut River in any meaningful sense.

22. For information on historic distribution and abundance of Atlantic salmon in North America see U.S. FISH & WILDLIFE SERV. & NAT’L MARINE FISHERIES SERV., supra note 14, at §§ 4.1.1–2, 4.2.1, fig. 4.1.1. Figure 4.1.1 shows the historic range of salmon in the United States. Although wild population of salmon were extirpated on the southernmost rivers, i.e., the Connecticut and Merrimack, by the end of the 20th century, minimal runs of those rivers were once again supported by stocking. For a sportsman’s view of the limited success of the Connecticut and Merrimack stocking programs, see Roger Aziz, Atlantic Salmon, the $1 Million Fish, EAGLE TRIB. (Haverhill), Dec. 15, 2002, available at http://www.eagetribune.com/news/stories/20021215/SP_014.htm. For the total number of fish returning to U.S. rivers see U.S. ATL. SALMON ASSESSMENT COMM., supra note 16, at tbl. 2.3.1.

23. See ATL. SALMON FED’N, supra note 16, tbls. 3.2.a–b. In the second half of the twentieth century, Atlantic salmon returns to the Connecticut, Merrimack, and other rivers were non-existent before the advent of the stocking program. For information on the historic occurrence of the Atlantic salmon in those rivers, see NOAA Protected Resources, supra note 12.

24. See ATL. SALMON FED’N, supra note 16, tbl. 3.2.a. Salmon fry are recently hatched fish that have fully resorbed the yolk sac from their eggs, i.e., the smallest and youngest independently-swimming salmon. Alevins are recently hatched salmon that have not resorbed the yolk sac; they typically do not leave the security of the gravel redds, or nesting sites. Fry are typically stocked despite higher mortality rates than would be the case with older fish (for example parr or smolts) because it is only when salmon are very young that they imprint with the smell of their native stream. Thus, it is thought that salmon stocked as fry will return to the streams in which they are released with a greater frequency than would fish stocked at a later life stage. For a good basic overview of the Atlantic salmon lifecycle, see ATL. SALMON FED’N, THE ATLANTIC SALMON, available at http://www.asf.ca/Overall/atlsalm.html (2002).

25. ATL. SALMON FED’N, supra note 16, tbl. 3.2.b.
Wild populations of Atlantic salmon from the Gulf of Maine D.P.S. northward are supplemented through stocking efforts; however, these populations are generally considered to be self-sustaining. The majority of the wild U.S. Atlantic salmon population is found in the Penobscot River in Maine. Over the period from 1967 to 1999, the Penobscot River accounted for 53,705 salmon, or seventy-one percent of U.S. Atlantic salmon returns. The next most important river was the Connecticut, which accounted for 4835 fish or 6.4 percent of returns over the same time period. The seven rivers that form the Gulf of Maine D.P.S. together accounted for 8027 salmon, or 10.63 percent of returns.

The total return of Atlantic salmon to U.S. waters, including both wild and hatchery-origin fish, has dropped from roughly 5000 fish a year in the mid-1980s to little more than 1000 fish a year at present. As noted earlier, the majority of these fish return to the Penobscot River in Maine, which is not part of the Gulf of Maine D.P.S., with a few hundred fish each year returning to the Connecticut and Merrimack rivers in southern New England and an even smaller number to the seven rivers of the Gulf of Maine D.P.S.

There are a number of indications that the abundance of salmon in the Gulf of Maine D.P.S. is very low. In recent years, the number of salmon from the Gulf of Maine D.P.S. returning to spawn has declined dramatically. From 1995 to 1999, 16 salmon returned to the Dennys River, 0 to the East Machias, 0 to the Machias, 1 to the Pleasant, 211 to the Narraguagus, 0 to the Ducktrap, and 32 to the Sheepscot. Also, the pre-fishery abundance index, one-sea-winter fish, also is very low in

26. The Penobscot River is adjacent to the seven rivers that form the Gulf of Maine D.P.S., but is not considered part of the Gulf of Maine D.P.S. for management purposes.
27. “Returns” are adult salmon that return to spawn; these fish are counted as they pass over fish ladders or dig redds (gravel nests) to spawn. For Penobscot River numbers, see ATL. SALMON FED’N, supra note 16, tbl. 3.2.b. The Penobscot River has been stocked heavily with Atlantic salmon over the past several decades. See ATL. SALMON FED’N, supra note 16, tbl. 3.2.a.
28. ATL. SALMON FED’N, supra note 16, tbl. 3.2.b.
29. Id.
30. Id.
31. Id.
32. The number of salmon returning to spawn can never be counted exactly in any river. In addition, there may be wide variation in the reliability of the estimates from river to river. This is because the availability of counting facilities and counters varies from river to river, as well as the visibility and accessibility of spawning redds. Therefore, the returns data must be understood as estimates and not exact counts, and are most useful in terms of trend analysis, i.e., the identification of broad trends over time such as the current long-term decline in abundance.
33. ATL. SALMON FED’N, supra note 16, tbl. 3.2.b.
Fish as Pollutants

spite of improving ocean habitat conditions. The United States Fish and Wildlife Service (FWS) and the National Marine Fisheries Service (NMFS) (collectively the “Services”) attribute the low level to depressed spawning populations of Atlantic salmon in the seven rivers comprising the Gulf of Maine D.P.S. and a consequently low index of juvenile salmon entering the sea. The production numbers of fry and parr (consecutive juvenile freshwater life stages of the Atlantic salmon) and subsequent smolt (young Atlantic salmon leaving freshwater habitats for the sea) are also very low. In part, this is due to unexpectedly high mortality of parr before reaching the smolt stage. About half of the smolt migrating do not reach the Gulf of Maine.

A study conducted on the Pleasant River found that out of a total of 707 smolt, 31 had come from a commercial hatchery upstream and 676 were wild. The fact that juvenile salmon are reared for mariculture along rivers of the Gulf of Maine D.P.S. is a cause for concern. First, escaped smolts join and potentially contribute deleterious genetic material to the wild salmon pool. And second, because salmon are thought to imprint with the odor of their home rivers as fry and parr, it increases the likelihood that mariculture escapees, imprinted with the odor of Pleasant River or other water from the Gulf of Maine D.P.S., might return to the Pleasant River to spawn.


36. Id.

37. Id.

38. Id. Most of this information comes from studies of the Narraguagus River, but because of similarities between the rivers in the Gulf of Maine D.P.S., the information is probably applicable to the entire Gulf of Maine D.P.S. See id.

39. Id.

40. In general, escaped salmon potentially contribute deleterious genetic material to the wild salmon pool. See GENETIC STATUS OF ATLANTIC SALMON, supra note 7, at 20.

A large variety of human activities have negatively impacted the abundance and distribution of Atlantic salmon, such as creating impediments to migration, water abstraction from salmon waterways, toxic pollution, acidification, deforestation, and the introduction of exotic species. Historically, the operation of commercial fisheries operations also impacted Atlantic salmon stocks. In recent years, commercial fishing pressure on the Atlantic salmon has been reduced, although this would seem to be more a function of reduced salmon stocks and subsequent regulatory actions to reduce fishing pressure on salmon than a function of the increasing Atlantic salmon mariculture industry.

B. Life Cycle of the Atlantic Salmon

Atlantic salmon are anadromous fish—that is, they are born and live as juveniles in freshwater rivers, then migrate to the sea where they mature before returning to their natal rivers to spawn. This basic life cycle may be further subdivided into the egg, alevin, parr, smolt, grilse, salmon, and kelt stages. These life stages define salmon ecology. Salmon are differentially affected by human and other impacts according to a life stage. Therefore, before assessing the effect of Atlantic salmon mariculture on wild Atlantic salmon, it is important to gain an understanding of the life stages of salmon.

Atlantic salmon spawn in late winter and early fall at the headwaters of rivers that empty into the northern Atlantic Ocean, in North America from Connecticut (formerly) northward into Canada, and in Europe at similar latitudes. The eggs are laid in redds, which are nests protected by several inches of gravel. From 2,000 to 15,000 eggs are deposited in a

42. See BIODIVERSITY REPORT, supra note 9, § 7.3.1 (including Brown Trout in a list of predators on young Atlantic salmon (all Brown trout in North America are introduced)); id. § 7.3.3 (where interspecies competition between Atlantic salmon and Brown trout is documented). The Atlantic Salmon Federation as well as the Canadian Department of Fisheries and Oceans have expressed concern over the possibility of competition between Atlantic salmon and rainbow trout. See ATL. SALMON FED’N, Alien Invasion Alert—It Continues in 2001, available at http://www.asf.ca/Research/research.html (last visited Dec. 12, 2002); NOVA SCOTIA SALMON ASS’N, N. ATL. SALMON CONSERVATION ORG., POLICY ON THE INTRODUCTION AND TRANSFER OF SALMONIDS, available at http://www.novascotiasalmon.ns.ca/newsandissues/nascopolicy.htm (last visited Dec. 12, 2002).


Fish as Pollutants

Although the number of eggs laid by salmon might seem high to the non-specialist, salmon are not particularly fecund by fish standards. The protection afforded by the gravel allows a high percentage of salmon eggs to hatch, as compared to the eggs of other fish that may circulate in the water column or otherwise be exposed to predation. Hatching takes place in late March and early April.

Newly hatched salmon are called alevins, and are dependent upon the egg yolk sac for nutrition. Alevins remain underneath the gravel until the egg sac is absorbed, at which point they emerge to feed and are known as fry. After some time in freshwater, the fry develop a coloration pattern of a series of broad vertical dark bands called parr markings; hence, at this stage of their life, Atlantic salmon are known as parr. Parr remain in freshwater for two to four years, at which point they metamorphose into smolts, the silvery-colored life stage of the salmon that goes to sea. Smolts resemble adult salmon, and once they reach the sea, they are known as salmon.

The Services cited the extremely low abundance of smolts in the Gulf of Maine discrete population segment (D.P.S.) as an important factor in their decision to list the Maine D.P.S. as an endangered species. The Services attribute this low abundance primarily to a small population of spawning adults. A study on the Narraguagus River showed that about half of the migrating smolt population does not reach the Gulf of Maine.

45. Id.
46. Compare with the fecundity of American eels, which are found with the Atlantic salmon in freshwater and may produce from 2,000,000 to 10,000,000 eggs per mature female in one spawning event. See G.P. Barbin & J.D. McCleave, Fecundity of the American Eel Anguilla Rostrata at 45°N in Maine, U.S.A., 51 J. OF FISH BIOLOGY 840, 844 (1997).
47. MILLS, supra note 44, at 9–11.
48. Id. at 11.
49. Id.
50. Id.
51. There is some disagreement about the timing of the division between the fry and parr life stages of the salmon, with some authorities saying that fry are known as parr after their first year of life in freshwater, and yet others holding that fry that have dispersed from the redd are known as parr. Id. For most fisheries workers, the existence of parr markings is enough to differentiate parr from fry and smolt.
52. Id.
55. Id. at 69,461.
Maine. Parr and smolt that escape from commercial hatcheries that supply the Atlantic salmon mariculture industry also may pose a competitive and genetic threat to wild parr and smolt. As noted earlier, a study conducted on the Pleasant River found that 31 of 707 smolt had their origins in a commercial hatchery upstream. Presumably these escapees would be among the most likely to interbreed with wild salmon because Atlantic salmon return to their natal rivers to spawn.

Salmon remain at sea from one to three years before returning to freshwater to spawn. During their time at sea, Atlantic salmon undertake long migrations; marked salmon of North American origin regularly travel as far as Greenland and some have been taken in Scottish waters on the other side of the Atlantic Ocean. Adults returning to freshwater after having spent only one year at sea are known as grilse. Adults returning to freshwater after having spent two to four years at sea are known as salmon. The spawning migration of Atlantic salmon upriver is well-known. However, Atlantic salmon differ from Pacific salmon, around which a great deal of the popular understanding of salmon life cycles in general has been formed, in that they do not necessarily die after spawning. Spent Atlantic salmon are known as kelts, and may return to the sea to pass one or more years before ascending freshwater rivers to spawn again.

C. The Value of Atlantic Salmon

Most of the exploitation of remaining wild Atlantic salmon stocks and the impetus for the creation and continued growth of the Atlantic salmon mariculture industry has been for food. The majority of farmed Atlantic salmon have historically been sold as fresh fillets, although the industry is now moving towards “value-added Atlantic salmon products aimed at meeting time-pressed consumers increasing demand for convenience,” such as frozen fillets and vacuum-marinated plastic packs. Atlantic

56. Id.
57. Id.
58. U.S. ATL. SALMON ASSESSMENT COMM., supra note 16, at tbl. 3.2.b. Salmon returning to rivers from the sea are classified as one, two, or three sea-winter fish meaning that they have spent on, two, or three years at sea.
59. MILLS, supra note 44, at 46–52.
60. Id. at 9.
61. Id. at 9–11.
62. Id.
63. See Steven Hedland, November 2003 Species Focus: Atlantic Salmon, Growth Hinges on the Farmed-Salmon Industry’s Ability to Produce and Market Winning Value-Added Products,
salmon mariculture operations help to alleviate pressure on wild salmon stocks by providing an alternative source of salmon to catch fisheries. The reduction of pressure on wild salmon for food is a positive effect of Atlantic salmon mariculture operations and must be weighed against the negative effects of salmon farming.

In 1653, Izaak Walton, the best-known early fishing author, described the Atlantic salmon as “the king of freshwater fish.” For many dedicated fly-fishermen, and others concerned with the plight of Atlantic salmon, that designation is as true today as it was then. The pursuit of Atlantic salmon is an expensive proposition and can be an important source of revenue for local communities where it can be sustained. However, in both the United States and Canada, sport fishing of Atlantic salmon has occasionally been allowed to continue on specific salmon stocks depleted beyond the point of commercial profitability.

Ted Williams, the recently deceased Boston Red Sox baseball star, was an avid Atlantic salmon fisherman for much of his life and an outspoken advocate of salmon conservation. The Atlantic salmon is one of the most cherished symbols of the wild and clean rivers. This positive image has helped to stimulate current salmon preservation efforts and will continue to play a role in the future management of the
salmon. Restoration of the Atlantic salmon to areas of its former range where it was extirpated has generated a great deal of interest and effort on the part of both the federal and state governments. At present, no sport fishing for Atlantic salmon is allowed in Maine in order to protect the endangered Gulf of Maine D.P.S.

D. The Growth of Aquaculture and Salmon Mariculture Production in the United States and Its Impact

Aquaculture production is increasing everywhere on earth, with the exception of Africa and the countries of the former Soviet Union. From 1989 to 1998, total global aquaculture production more than doubled, in terms of both weight and value. From 1987 to 1988, aquaculture production increased from 15% to 28% of the total global seafood supply. The global per capita availability of aquaculture products increased from 2.3 to 6.4 kilograms per year from 1984 to 1997.

Aquaculture production has increased steadily in the United States over the past decade and is projected to continue growing in the near future, both in terms of absolute poundage and as a percent of total U.S. seafood consumption. North American aquaculture production increased an average of 3.6% per year from 1984 to 2001, as compared to the global average of 9% per year over the same time period.


69. See, e.g., Me. Atl. Salmon Comm’n, 2003 Open Water Fishing Regulations (Apr. 18, 2003), available at http://www.state.me.us/ifw/fishing/2003openwaterlaws.html#MAINE%20ATLANTIC%20SALMON%20COMMISSION (“It is unlawful to angle, take, or possess anadromous Atlantic salmon from all Maine waters (including coastal waters). Any salmon incidentally caught, must be released immediately, alive and uninjured. At no time should Atlantic salmon be removed from the water. The number of adult anadromous Atlantic salmon returning from the sea to spawn in Maine waters is very low, and sport fisheries were suspended in December of 1999. Some areas where adult Atlantic salmon congregate have been closed to fishing for all species”). More information on the current state of regulations can be obtained at the homepage of the Maine Atlantic Salmon Commission. http://www.state.me.us/asa/.

70. See GOLDBURG ET AL., supra note 1, at iii.


72. GOLDBURG ET AL., supra note 1, at ii.

73. De Silva, supra note 71, at 434.

74. Id.

75. GOLDBURG ET AL., supra note 1, 2–5.

Fish as Pollutants

Although the increase in North American and United States production has not been as rapid as the overall increase in global production, North American consumers drive a disproportionate amount of aquaculture production in other countries. The United States is eleventh in aquaculture production but third in seafood consumption, which necessitates the importation of large amounts of fish each year.78

In North America, mariculture production79 has increased over the past two decades not only in terms of weight and value, but also as a percentage of overall aquaculture production.80 From 1988 to 1997, North American mariculture production increased from 45,000 to 209,000 metric tons, while freshwater aquaculture production increased from 233,000 to 315,000 metric tons over the same time period.81 Thus, mariculture, which in 1988 represented a little less than one quarter of total North American aquaculture production, had grown by 1997 to represent a little more than two-fifths of total North American aquaculture production.

When we narrow our focus further and consider Atlantic salmon mariculture, we find that it is increasing worldwide, especially in the United States. In both arenas, Atlantic salmon mariculture growth has been explosive. From 1984 to 1999, the value of worldwide Atlantic salmon mariculture grew from slightly more than $150 million in U.S. dollars to almost $2.5 billion, an increase of nearly 1,600%.82 From 1989 to 1999, Atlantic salmon mariculture by weight increased by 595 percent in the U.S.83 This increase in Atlantic Salmon mariculture in part reflects a shift in focus on the part of aquaculture producers from species low on the food chain to carnivorous species such as the Atlantic salmon, which tend to bring much higher prices in the market.84

The global increase in aquaculture and mariculture production has been and will continue to be driven by an increased demand for seafood worldwide in the face of dwindling stocks of wild fish. With the

77. GOLDBURG ET AL., supra note 1, at 2.
78. Id.
79. Id. at iii.
81. Id.
82. Id. The figure cited includes Atlantic salmon classified as produced either by mariculture or brackish water.
83. Id.
84. GOLDBURG ET AL., supra note 1, at 2.
exception of China, which has dramatically increased per-capita fish supply domestically since 1980, world per-capita fish supply has remained relatively stable since the mid-1980s. Over the second half of the last century, however, world landing statistics for capture fisheries grew steadily from 1950 to 1969, continued to grow through the 1970s and 1980s, albeit at a slower rate, and then leveled off during the 1990s, with total landings presently varying from 85 to 95 million metric tons per year. This leveling off of the catch is thought to represent the near complete exploitation of the most important fish stocks worldwide.

As the population of the world and consequent demand for seafood continues to grow, it unlikely that the pressure on already depleted fish stocks will lessen. Therefore, it is the opinion of many national and state governments that an ever-greater future demand for seafood can only be met through increased aquaculture, which in turn has made these governments strong advocates of increasing aquaculture. In the United States, state and federal government advocacy has played and will continue to play an important role in aquaculture development. Moreover, aquaculture is viewed not only as a solution to the seafood shortage but also as an attractive source of revenue and employment. For example, in Maine, the center of the U.S. Atlantic salmon mariculture industry, aquaculture operations generate annual revenues of nearly $70 million. Atlantic salmon mariculture alone generates annual revenues of nearly $18 million and provides nearly 700 jobs in two Maine counties. Overall, the Maine aquaculture harvest of finfish grew thirty-

86. Id.
87. Id.
89. Id. In spite of the jobs provided by the Atlantic salmon mariculture industry, both Hancock and Washington counties had unemployment rates higher in 2001 than the Maine statewide average of 4%. See ME. DEP’T OF LABOR, DIV. OF LABOR MKTG. INFO. SERV., 2001 CIVILIAN LABOR FORCE ESTIMATES FOR MAINE AND MAINE COUNTIES, BY MONTH AND ANNUAL AVERAGE, available at http://www.maine.gov/labor/lmis/data/laus/mecty01.html (last visited Dec. 12, 2002) (in cooperation with the U.S. Bureau of Labor Statistics). The unemployment rate in Hancock County was 4.5% while the unemployment rate in Washington County was more than double the statewide average at 8.1%. Id. In addition, in 1999 the per-capita income in Washington County was $19,098, or $5,122 below the statewide average of $24,220 (2000 figures not yet available). See ME. DEP’T OF LABOR, PER CAPITA PERSONAL INCOME IN THE UNITED STATES, NEW ENGLAND, MAINE, AND COUNTIES, 1981–2001, available at http://www.maine.gov/labor/lmis/data/laus/pceincome.html (Sept. 2002) (based on U.S. Dep’t of
Fish as Pollutants

six-fold from 1988 to 2000, from one million pounds in 1988 to over thirty-six million pounds in 2000. In light of these statistics from the most productive aquaculture states, together with continued predictions of shortages in world capture fisheries, it is not surprising that the United States Department of Commerce is promoting a five-fold increase in U.S. aquaculture production by the year 2025.

However, aquaculture, and in particular Atlantic salmon mariculture, will not be able to develop further in state and federal waters without successfully negotiating a variety of environmental, social, political, and technological obstacles that affect both the public perception and economic viability of aquaculture operations. Atlantic salmon mariculture releases effluents and contaminants that are difficult to monitor and control, and tends to provoke high-profile conflicts over water use. Atlantic salmon are high on the food chain and require environmentally damaging accessory fisheries. These problems will be exacerbated as Atlantic salmon mariculture becomes an increasingly important segment of overall U.S. aquaculture.

Atlantic salmon mariculture is primarily the province of large international corporations and thus has a comparatively large reserve of funding for research and development. The current regulatory framework for aquaculture is characterized by a patchwork of state and federal laws that was not conceived of with aquaculture in mind. With no agency clearly designated as the lead federal agency, no one has devised a clear blueprint detailing exactly how a prospective aquaculturalist is to navigate the maze of regulations. For example, the...
primary regulatory authority to control pests in aquaculture is given to the United States Department of Aquaculture under the 2002 Farm Bill, in which Congress defined the term “livestock” broadly to include “all farm-raised animals,” while the United States Environmental Protection Agency (EPA) regulates effluents from aquaculture facilities. Despite the potential environmental and other problems associated with expanded Atlantic salmon mariculture, the most important question currently surrounding Atlantic salmon mariculture (and other types of mariculture with the potential to harm the marine environment) is not whether these types of mariculture should be promoted and developed. Indeed, given the current global seafood shortage and the profitability and financial backing of certain segments of the mariculture industry, it seems inevitable that they will. Rather, the question currently garnering attention is: in which areas of the sea should development take place and under what regulatory framework?

The decrease in wild Atlantic salmon in U.S. waters that led to their endangered species listing coincided with a massive influx of introduced Atlantic salmon to wild Atlantic salmon habitat from two sources: Atlantic salmon stocking programs and the Atlantic salmon mariculture industry. The introduction of stocked fish was by definition intentional, with the intent of bolstering wild Atlantic salmon populations. The escape of mariculture industry fish occurred either through “leakage”—the industry term for the escape of small numbers of fish during normal operations—or, in extremely large numbers, through catastrophic events. For example, in December 2000, roughly 100,000 Atlantic salmon of non-North American origin escaped from a salmon net pen at Stone Island off the coast of Maine. In a separate incident in November 2000, 13,100 Atlantic salmon escaped from another net pen.
Fish as Pollutants

off the Maine coast.\footnote{United States Pub. Interest Research Group v. Heritage Salmon, Civ. No. 00-150-B-S, 2002 U.S. Dist. LEXIS 2706, at *10 (D. Me. Feb. 20, 2002) (recommended decision by Mag. Kravchuk).} The loss of Atlantic salmon from net pens is apparently routine in the mariculture industry. Heritage Salmon, a mariculture company, lost 90,359 fish between 1994 and 1998, although not all of those fish were escapees.\footnote{Id.} When these numbers are compared to the numbers of wild fish returning to spawn, it is easy to see why a great deal of concern has been focused on the impact of mariculture escapees on wild Atlantic salmon.

As elucidated below, Atlantic salmon mariculture operations are a point source of pollutants to navigable waters, which of course includes salmon habitat. The pollutants include copper (an ingredient in antifouling preparations used to prevent the growth of marine algae on salmon pens); feed (which includes biological wastes from the chicken industry as well as added pigments to color the salmon flesh); several diseases, viruses, and parasites including bacterial kidney disease, furunculosis, hitra, vibrios, and infectious salmon anemia; fish wastes; chemicals including antibiotics added to feed and biocides designed to kill fish lice; and Atlantic salmon escapees.\footnote{See infra notes 149–50 and accompanying text.}

The effect of escapees on wild Atlantic salmon is different than the effect of traditional pollutants. Not only may escapees directly compete with wild Atlantic salmon for food and habitat,\footnote{Introduced exotic species present a similar concern. See supra note 42.} but more importantly, escapees have the potential to dilute the genetic material of the wild stock. Although a strong wild salmon population might well support the introduction of a limited amount of foreign genetic material into the gene pool, a threatened population could be overwhelmed by the same influence.\footnote{See MILLS, supra note 44, at 270.} The number of wild Atlantic salmon in the Gulf of Maine D.P.S. returning to spawn in recent years has been extremely low, perhaps several dozens or several hundreds of fish,\footnote{Endangered and Threatened Species; Final Endangered Statute for a Distinct Population Segment of Anadromous Atlantic Salmon (Salmo salar) in the Gulf of Maine, 65 Fed. Reg. 69,459, 69,461, 69,468 (Nov. 17, 2000) (to be codified at 50 C.F.R. pt. 17).} while the number of escaped salmon from mariculture operations has been as high as 100,000 in a single incident.\footnote{United States Pub. Interest Research Group v. Atl. Salmon of Me., LLC, 215 F. Supp. 2d 239, 243 (D. Me. 2002).}
Escaped salmon from net pens also may attempt to enter rivers and spawn. In fact, Atlantic salmon escapees from net pens in the Pacific Northwest, where they have never been native, have spawned successfully in British Columbia, prompting fears that they will colonize traditional Pacific salmon waters.\footnote{105} Although Salmon mariculture in British Columbia also uses other species of salmon, about seventy percent of all salmon mariculture in British Columbia is based on Atlantic salmon stocks.\footnote{106} The government of British Columbia is currently promoting an increase in salmon mariculture and has stated that it will “begin accepting applications for new finfish aquaculture sites” starting in 2002.\footnote{107} The number of such sites currently permitted is 121.\footnote{108} In British Columbia, environmentalists and others have expressed concern about the impact of escaped Atlantic salmon on Pacific salmon stocks. Perhaps to the envy of east coast fisheries scientists, who have poured tens of millions of dollars into programs to restore the Atlantic salmon to the Connecticut, the Merrimack, and other New England rivers with little success, Atlantic salmon, with little encouragement, seem to be in the process of establishing itself in British Columbia as mariculture escapees successfully breed in British Columbia rivers.\footnote{109} While this would be considered a rousing success in Connecticut (assuming native populations were used), it has the potential to be an unmitigated disaster on the west coast, where Pacific salmon are already in severe decline due to overfishing and habitat destruction, among other causes.

E. Stocking of Native and Exotic Fish and Their Impacts

Although the idea of native fish as pollutants would be a new paradigm in fisheries management, the idea that introductions of exotic fish are potentially harmful is not new. The history of U.S. fisheries management is rife with examples of exotic fish species wreaking havoc with ecosystems, even when those exotic fish were intentionally introduced in an attempt to correct a perceived imbalance in the original
Fish as Pollutants

ecosystem or to improve fisheries. Some well-known examples include the common carp and the brown trout, both of which are now widespread throughout the United States and have altered or appropriated for themselves habitat which was previously available to native fishes; sea lampreys in the Great Lakes, which caused the complete collapse of lake trout stocks throughout the Great Lakes; and the walking catfish in Florida, which has crowded out native fishes locally and caused losses through predation in tropical fish farms.

In many cases the introduction of exotic species has been accidental, in other cases it resulted from intentional stocking. The common carp and brown trout are examples of release through intentional stocking. Both fish were widely stocked in U.S. waters over the past two centuries. Brown trout are still commonly stocked as a part of many state fishery management programs. Introductions of both common carp and brown trout to waters where they are not native are viewed as beneficial in many instances because they provide sport fishing where native species no longer exist or are reduced in numbers due to environmental factors, in the case of the brown trout, or as a food fish, in the case of the carp. However, they also may contribute to the decline of native fishes through competition and other factors as well as give a false sense of security regarding water quality because they are able to survive in waters which are unfit for native species. The carp is legendary for its ability to survive in severely degraded urban waterways, while the brown trout is the hardiest of the commonly stocked and fished trout species in the United States; it is the most resistant to the higher water temperatures, which commonly result from clearing riverbanks and subsequent loss of shade. Thus, carp and brown trout can provide fishing where no fishing exists for brook trout, cutthroat trout, and rainbow trout because of poor water quality.

Therefore, a fundamental question in fisheries management is whether aquatic ecosystems should be managed for fish that are

111. See Univ. of Fla., Center for Aquatic and Invasive Plants, at http://aquat1.ifas.ufl.edu/mcfish5d.html (last visited July 27, 2003).
113. Taylor et al., supra note 112, at 322–23.
114. Id. at 323.
relatively unaffected by degraded water quality or whether the focus should be on improving water quality to the point where native fish that historically were dominant in the habitat are once again abundant. In general, the trend in fisheries management is increasingly to exercise caution where the introduction of exotic fishes is concerned. Exotic fish introductions are no longer viewed as a panacea for all of the fisheries challenges that degraded waterways present. In fact, as the recent clamor over the accidental introduction of the snakehead into a Crofton, Maryland pond demonstrates, the introduction of exotic fish is often feared because it has the potential to disrupt and even destroy native ecosystems.\textsuperscript{115} Thus, even though the word “pollutant” is rarely used to describe such exotic fish introductions as that of the snakehead in Maryland, many fisheries managers already essentially manage exotic fish as pollutants.

The impacts of native fishes stocked for sports fishing on wild fish of the same species are essentially the same as the impacts of Atlantic salmon mariculture escapees on wild salmon—that is, the introduction of exotic species or varieties of fish to new bodies of water, genetic contamination of the wild genome, predation on wild fish, competition with wild fish for food and favorable space, disruptive behavior, stimulation of premature migrations, the creation of unacceptably high densities of fish, mixed-stock exploitation problems, predator attraction, and disease and parasite transmission.\textsuperscript{116} The American Fisheries Society categorizes the impacts of exotic species into five classes: habitat alteration, spatial alteration, trophic alteration, gene pool deterioration, and introduction of diseases.\textsuperscript{117} Of these classes, gene pool alteration is perhaps the least serious problem in the case of introduced exotic fishes because there is little chance that exotic fish will be able to breed with native fishes. However, gene pool alteration may be one of the most serious effects of the contamination of native fishes by their own genetically distinct relatives.

Exotic fish species are seen as either a boon or a boondoggle depending on the species and the evaluator. Their positive aspects include the creation of fisheries in waters where native fishes are no


Fish as Pollutants

longer abundant because of degraded water quality. This is the case with the German brown trout, which, because it is more tolerant of warm water and pollution than brook trout, provides wild, although not native, trout fishing in many eastern rivers that are no longer suitable for brook trout. However, the negative aspects of exotic species introductions include the displacement or reduction in numbers of native species from waterways that are still suitable for native species.

A number of exotic species have been introduced throughout the former range of the Atlantic salmon in North America and have established self-sustaining populations. These included the German brown trout, a European species, and the rainbow trout, from the American west. A great deal of concern has been expressed about the ability of young Atlantic salmon to successfully compete in particular with rainbow trout. All salmon and trout are closely related and the parr stage of the Atlantic salmon so closely resembles a trout and inhabits such similar habitats that for many years it was not recognized as the young of the Atlantic salmon but rather described as a species of trout. Introduced exotic trout species will undoubtedly compete with young Atlantic salmon for habitat and may replace juvenile salmon in certain habitats. Similarly, escaped parr and smolts from hatcheries that provide feedstock for Atlantic salmon mariculture operations also may compete with and displace juvenile salmon. Hatchery fish often learn or are even bred to display aggressive feeding behaviors that result in fast growth rates. These hatchery fish might out-compete wild fish in certain situations, with no hope of later completing the salmon life cycle.

F. Accidental and Intentional Releases of Fish

The two principal avenues of release into U.S. waters that are followed by native species are the same for exotic species, i.e. intentional stocking programs and accidental release. Accidental releases include the moving of fish from one waterway to another via fishermen’s bait buckets, the effects of floods, or the water in live wells in sports fishing vessels. Although undesirable, such introductions are difficult to monitor and regulate. On the other hand, a second category of accidental releases, that includes the escape of juvenile and adult fish

118. MILLS, supra note 44, at 263–65.
119. Id. at 7–9.
120. Id. at 263–65.
121. See supra notes 86–87 and accompanying text.
from hatcheries and mariculture operations, are more easily controlled through bio-security measures and escape response planning. Intentional releases include the release of stocked fish for recreational fisheries and the release of aquaculture specimens “if they are not growing rapidly enough to justify continued feeding.”122 Neither of these two forms of intentional releases are monitored or regulated, although they could be.123 While the release of intentionally stocked fish can have negative results on wild fish populations, unfortunately at present such releases are generally viewed in a positive light because they are tightly linked to sales of fishing licenses and private sector facets of the sport fishing economy. Indeed, the EPA has identified approximately 500 state fish hatcheries and cites a study of state coldwater fishery programs that indicates that those programs distributed 23.7 million pounds of trout and salmon from state hatcheries in 1996 alone.124 Yet, because the intentional release of stocked fish and the release of the second category of accidental fish release originate from “point sources,’”125 they fall within the EPA’s mandate to regulate their discharge under the Clean Water Act National Pollutant Discharge Elimination System program, if such discharges encompass the discharge of “pollutants.”

III. THE LEGAL REGIME REGULATING THE DISCHARGE OF POLLUTANTS FROM MARICULTURE FACILITIES

A. Framework

In 1972, Congress enacted the Federal Water Pollution Control Act.126 Now known as the Clean Water Act (CWA or the “Act”),127 the Act prohibits the discharge of a pollutant from a point source to the waters of the United States without a National Pollutant Discharge Elimination System permit.128

123. The EPA is seeking comment on the efficacy of banning such intentional releases. Id.
124. Id. at 57,876, 57,903. The federal government and native American tribes manage additional fish hatcheries.
125. The term “point source” means any “discernible, confined and discrete conveyance, including but not limited to any pipe, ditch, channel, tunnel, conduit, well, discrete fissure, container, rolling stock, concentrated animal feeding operation, or vessel or other floating craft, from which pollutants are or may be discharged.” 33 U.S.C. § 1362(14) (2000).
Fish as Pollutants

System (NPDES) permit. NPDES permits mandate compliance with technology-based effluent limitations and state water quality standards. To the extent that a person wishes to discharge effluents to the territorial sea, contiguous zone or ocean, section 403 of the CWA requires that the discharge also must be in compliance with any ocean discharge criteria adopted by the EPA. These criteria are primarily directed at protecting the ecological health of the marine environment. As of


131. Regulations implementing section 403 are found at Criteria and Standards for the National Pollutant Discharge Elimination System, 40 C.F.R. § 125.120–124 (2003). See also Ocean Discharge Criteria, 45 Fed. Reg. 65,942 (Oct. 3, 1980) (to be codified at 40 C.F.R. pt. 125). In the waning days of the Clinton Administration, the EPA announced a proposed revisions to Ocean Discharge Criteria and sent a proposed rule to the Office of the Federal Register. USEPA OFFICE OF WATER FACT SHEET, REVISIONS TO CLEAN WATER ACT OCEAN DISCHARGE CRITERIA REGULATIONS, EPA-842-F-01-001 (Jan. 2001). However, the proposed rule was not published in the Federal Register prior to the onset of the Bush Administration, and thus, on January 20, 2001, the Bush Administration “withdrew” the proposal to give the new EPA Administrator “an opportunity to review it.” Id.; Cooperative State Research, Education, and Extension Service Higher Education Challenge Grants Program for Fiscal Year 2001, 66 Fed. Reg. 7701, 7701–02 (Jan. 24, 2001). See National Legal Center for the Public Interest, Judicial Legislative Watch Report, 1 (Feb. 2003), available at http://www/nlci®/org/books/pdf/jlwr_march01.pdf. To date, the proposed rule as developed by the Clinton EPA (or for that matter, any modification thereof) has not been published in the federal register.

132. 40 C.F.R. § 125.122(a) (protecting against “unreasonable degradation of the marine environment”).
September 2001, 250,000 facilities nationwide were regulated under NPDES permits.\textsuperscript{133}

Although Congress designated the EPA to administer the CWA, a state may apply to manage its own program in lieu of the federal program.\textsuperscript{134} The EPA Administrator, who manages such applications, is required to approve the state application if: 1) the proposed state program is at least as stringent as the federal program; and 2) the state has adequate legal authority to carry out the program, to monitor and ensure compliance therewith, and to abate and deter violations through the imposition of civil and criminal penalties and other means of enforcement.\textsuperscript{135}

Even in the case of a state-delegated program, however, the EPA retains the right, subject to waiver, to review individual permits\textsuperscript{136} to commence an enforcement action seeking compliance with the CWA, and to require the imposition of administrative, civil judicial and criminal penalties for violations thereof.\textsuperscript{137} Additionally, in those instances when the EPA and, if delegated authority, the state having jurisdiction over the discharge, fails to undertake action to enforce compliance with the CWA, Congress authorized individual private citizens to enforce the Act through litigation.\textsuperscript{138}

\section*{B. USPIRG Cases}

In the State of Maine, the EPA had the authority to issue NPDES permits until January 2001, when that authority was assumed by the State of Maine.\textsuperscript{139} Although the EPA indicated as early as 1988 that floating net pen facilities used for salmon mariculture in state

\begin{thebibliography}{9}
\bibitem{133} U.S. ENVT. PROT. AGENCY (USEPA), AQUATIC NUISANCE SPECIES IN BALLAST WATER DISCHARGES: ISSUES AND OPTIONS, DRAFT REPORT FOR PUBLIC COMMENT 31 (2001).
\bibitem{134} 33 U.S.C. § 1342(b)–(c) (2000).
\bibitem{135} Id. § 1342(b).
\bibitem{136} Id. § 1342(d)–(e).
\bibitem{138} 33 U.S.C. § 1365; Profitt v. Rohm & Hass, 850 F.2d 1007, 1011 (3d Cir. 1988) (noting that a citizen suit operates as an "alternative enforcement mechanism" in the face of "inert" government action.) (internal citations omitted).
\end{thebibliography}

718
Fish as Pollutants

jurisdictional waters off the coast of Maine may require NPDES permits, no NPDES permit was issued for any such facility. As a result, on July 31, 2000, the United States Public Interest Research Group (USPIRG) and several individual plaintiffs filed separate lawsuits against Heritage Salmon, Inc. (Heritage Salmon), Atlantic Salmon of Maine, LLC (Atlantic Salmon), and Stolt Sea Farm, Inc. (Stolt) alleging in each case that the defendant had discharged pollutants to the U.S. waters in violation of the CWA at the defendant’s Maine salmon farm (hereinafter jointly referred to as the “USPIRG cases”).

In order to prevail on such a claim, USPIRG is required to establish that “five elements exist: ‘(1) a pollutant must be (2) added (3) to navigable waters (4) from (5) a point source.’” Because the focus of this article is on the notion of fish as pollutants, we focus on the first criterion—pollutants. Under the CWA, the term “pollutant” is defined

140. Id. at *7–11.
141. Id. at *1–2.
146. Importantly though, relying on a case involving a dam, National Wildlife Federation v. Gorsuch, 693 F.2d 156, 174–75 (D.C. Cir. 1982), and another case addressing a pumped storage facility (National Wildlife Federation v. Consumers Power Co., 862 F.2d 580, 584 (6th Cir. 1988) (stating that for there to be an “addition,” a source must “physically introduc[ ] a pollutant into water from the outside world”)), the court held that because the pollutants at issue were “put in the water by ASM as part of its operation,” and they “do not naturally occur in the bay,” they are “additions.” Atl. Salmon, 215 F. Supp. 2d at 248–49; see also Catskill Mountains Chapter of Trout Unlimited v. City of New York, 273 F.3d 481, 491 (2d Cir. 2001); Ass’n of Pac. Fisheries v. EPA, 615 F.2d 794, 806–07 n.7, 815–16 (9th Cir. 1980) (fish wastes discharged by seafood processors are pollutants). Cf. Lisa A. Brautigam, Control of Aquatic Nuisance Species Introductions Via Ballast Water in the United States: Is the Exemption of Ballast Water Discharges from Clean Water Act Regulation a Valid Exercise of Authority by the Environmental Protection Agency?, 6 OCEAN & COASTAL L.J. 33 (2001); Sandra B. Zellmer, The Virtues Of “Command and Control” Regulation: Barring Exotic Species From Aquatic Ecosystems, 2000 U. ILL. L. REV. 1233, 1285 (2000). The court also found the mariculture operations were “point sources” in that net pens were “concentrated aquatic animal production facilities” (CAAPFs) within the meaning of EPA’s regulations. EPA Administered Permit Programs: The National Pollutant Discharge Elimination System, 40 C.F.R. § 122 app. C (2003); see also id. § 122.24. More specifically, the court found that the net pens were “ponds, raceways or other similar structures” and, irrespective of the fact that the net pens did not discharge pollutants from a “discrete discharge pipe,” they released pollutants into the bay from an “identifiable, discernible, confined, and discrete emission or conveyance,” and hence were point
as dredged spoil, solid waste, incinerator residue, sewage, garbage, sewage sludge, munitions, chemical wastes, *biological materials*, radioactive materials, heat, wrecked or discarded equipment, rock, sand, cellar dirt and industrial, municipal, and agricultural waste discharged into water. 147

In *USPIRG v. Atlantic Salmon of Maine, LLC*, the court found that various materials added by the mariculture operations were “pollutants” within the meaning of the CWA.148 Specifically, the court held149 that: (1) salmon feces and urine “constitute ‘biological materials’ or ‘agricultural wastes,’” (2) the uneaten pigments, canthaxanthin, and astaxanthin, and the antibiotic, oxytetracycline which “flow out of ASM’s pens” or “falls through the net pens . . . are subsumed in the category of ‘chemical wastes,’” (3) cypermethrin, which is used to kill sea lice, and the chemicals Finquel and Parasite-S are “released into the water after their use,” are included “within the category of ‘chemical wastes,’” and (4) copper, which is a component of an antifoulant that is applied to the nets to reduce marine growth, is specifically listed by the EPA as a toxic pollutant in 40 C.F.R. § 401.15(22).150

*sources. Atl. Salmon*, 215 F. Supp. 2d at 255; see also 33 U.S.C. § 1362(14) (2000) (defining “point source” as any “discernible, confined and discrete conveyance, including but not limited to any pipe, ditch, channel, tunnel, conduit, well, discrete fissure, container, rolling stock, concentrated animal feeding operation, or vessel or other floating craft, from which pollutants are or may be discharged.”).

147. 33 U.S.C. § 1362(6). Several types of material are explicitly excluded from the definition: sewage from a vessel, discharges incidental to normal operation of an Armed Forces vessel, and certain discharges of materials associated with oil and gas production. *Id.*


150. *Atl. Salmon*, 215 F. Supp. 2d at 243–48. The court did not consider whether farmed Atlantic salmon could be viewed as pollutants because they serve as a vehicle for polychlorinated biphenyls (PCBs) and other contaminants. A recent study found that farmed salmon in British Columbia have “consistently higher levels” of total PCBs (51,216 vs. 5,302 pg/g), polybrominated diphenylethers (PBDEs) (2,688 vs. 178 pg/g), and organopesticides (other than toxaphene) (41,796 vs. 12,164 pg/g) than wild salmon. Michael D.L. Easton et al., *Preliminary Examination of Contaminant Loadings in Farmed Salmon, Wild Salmon and Commercial Salmon Feed*, 46 CHEMOSPHERE 1053, 1053 (2002). It should be noted, however, that these are preliminary findings based on a small sample size that did not allow the authors to conduct tests to determine whether the differences are statistically significant. *Id.* at 1062. See also Miriam Jacobs et al., *Investigation of Polychlorinated Dibenzo-p-dioxins, Dibenzo-p-furans and Selected Coplanar Biphenyls in Scottish Farmed Atlantic Salmon (Salmo salar)*, 47 CHEMOSPHERE 183, 191 (2002) (finding high concentrations of PCBs and other contaminants in a sample of 10 farmed-raised salmon). Much of the origin of the higher
Fish as Pollutants

Most importantly for present purposes, the Atlantic Salmon court also addressed the question of fish as pollutants. First, the court noted that “some” of the salmon grown in ASM’s net pens is of “non-North American origin.”\(^{151}\) The court also found that farm-raised salmon can be differentiated from wild salmon because they have “shortened and eroded fins, a plumper body, and a smaller head-to-body ratio . . . .”\(^{152}\) Next, the court noted that it was undisputed that some cultured salmon escape from the net pens.\(^{153}\) Finally, relying on National Wildlife Federation (NWF) v. Consumers Power Co.,\(^{154}\) the court held that fish that “do not naturally occur in the water, such as non-North American salmon,” fall within the term “biological materials” and are therefore pollutants under the Act.\(^{155}\)

contaminant levels is apparently contaminated salmon feed (which itself would be a pollutant). Michael D.L. Easton et al., Preliminary Examination of Contaminant Loadings in Farmed Salmon, Wild Salmon and Commercial Salmon Feed, 46 CHEMOSPHERE at 1053 (2002); Easton et al., supra this note, at 1071–72. According to Toxic Pollutant Effluent Standards, 40 C.F.R. § 129.105 (2003), which establishes EPA effluent guidelines for PCB discharge and which applies to such heavy industrial users of PCBs as the PCB manufacturers themselves, electrical capacity manufacturers, and electrical transformer manufacturers, PCBs are prohibited in any discharge from any existing or new source. The regulations also state that the ambient water criterion for PCBs is 0.001 µ g/l. Id. If instead of being reared in mariculture facilities, contaminated farmed salmon were sent down an outflow pipe at a PCB manufacturing facility, they could constitute an illegal discharge of PCBs. We thus might ask the question: are Atlantic salmon reared in net pens pollution in the conventional sense?

151. Atl. Salmon, 215 F. Supp. 2d 239, 244, 247 (D. Me. Feb. 20, 2002) (although it primarily raised native salmon, approximately 100,000 of the salmon were not native to North America; some farm-raised salmon is non-North American in origin).

152. See Stolt Sea Farm, 2002 U.S. Dist. LEXIS 12590, at *7; Heritage Salmon, 2002 U.S. Dist. LEXIS 2706, at *10 (stating that farm-raised salmon “can be differentiated from wild salmon by the bluntness of their fins, a deformity caused by stresses associated with crowded pens” as well as the fact that some farm-raised salmon have deformities due to physical injuries or unbalanced nutrition).


154. 862 F.2d 580, 583 (6th Cir. 1988) (stating that “live fish, dead fish and fish remains . . . are pollutants within the meaning of the CWA, since they are biological materials”). In Consumers Power, however, the Sixth Circuit held that a permit was not required because the fish, which originated in Lake Michigan adjacent to the pumped storage facility, were released back into the lake, and thus, were not “added” within the meaning of CWA. Id. That part of the Consumers Power holding is not controlling in the salmon mariculture situation because with salmon mariculture, the fish—and in particular the non-native fish—did not originate in the vicinity of the net pens.

155. It is not entirely clear what the court meant by the phrase “do not naturally occur in the water.” Atl. Salmon, 215 F. Supp. 2d at 247. First, does this mean that something must not be
Although the concern with possible genetic pollution of wild stock from mariculture escapees was well-articulated in a number of fora before the USPIRG cases, the judge in the USPIRF cases did not mention possible genetic pollution of native stocks by escapees from mariculture operations. Nor did the court rely on the endangered status of wild Atlantic salmon stocks. Rather, the court focused solely on the fact that the escapees were of different origin than the wild stock and were distinguishable from wild stock by external markers, both of which served to establish escapees as introduced to the navigable waters.

This dependence on differentiation from native stock as the sole criteria for the determination of the status of Atlantic salmon mariculture escapees as pollutants would not seem to preclude interpreting escapees from other types of aquaculture operations as pollutants under the CWA, provided the facilities in question met discharge and other criteria in

722
Fish as Pollutants

order to be regulated under the CWA. Presumably, such a criteria would label transgenic salmon and other transgenic mariculture specimens as pollutants as well.

Since the court’s orders affirming the recommended revisions in the *Atlantic Salmon* and *Stolt* cases were issued, there have been two significant developments in the USPIRG cases. First, on July 29, 2002, the court approved a detailed settlement entered into by USPIRG and Heritage to settle USPIRG’s claims against Heritage. In pertinent part, the settlement requires Heritage, prior to receiving any applicable permits from the United States or the State of Maine, to: (1) forgo the use of non-North American stocks and transgenic salmonids; (2) limit the stocking densities in its net pens; (3) fallow its salmon farms; and (4) take precautions and institute measures so that cultured fish do not escape. In pertinent part, the settlement requires Heritage, prior to receiving any applicable permits from the United States or the State of Maine, to: (1) forgo the use of non-North American stocks and transgenic salmonids; (2) limit the stocking densities in its net pens; (3) fallow its salmon farms; and (4) take precautions and institute measures so that cultured fish do not escape.159

Although the settlement with Heritage is far-reaching, it curiously fails to include three components. First, the settlement allows Heritage to continue to use Canadian stocks. This is significant given that the court ruled that non-native stocks are pollutants and, as discussed below, the National Academy of Science panel recently concluded that the evidence is “surprisingly strong” that Maine and Canadian stocks are genetically distinct. Second, although the settlement requires Heritage to obtain an NPDES permit from Maine for any new lease site, it merely requires compliance with any similar permit that may be issued for its existing sites; there is no requirement that Heritage apply for such permits for its existing sites or cease operations at some defined date at those sites if it is unable to secure the necessary permits.

Third, while the court will retain jurisdiction under 33 U.S.C. § 1365 to enforce Heritage’s commitments, the settlement does not require Heritage to pay stipulated penalties or liquidated damages in the event of a breach of its obligations. This is significant given that CWA statutory penalties do not attach to violations of the court’s decree. In contrast, if Heritage possessed an NPDES permit, Heritage would be

---

159. The settlement was lodged on June 4, 2002, two weeks prior to the court’s orders affirming the recommended decisions in *Atlantic Salmon* and *Stolt Sea Farm*. To date, no proposed settlements have been reached in the other two USPIRG cases.

160. *Heritage Salmon*, 2002 U.S. Dist. LEXIS 2706, ¶¶ 17–18, 22–23. The settlement requires Heritage to undertake measures to ensure water quality and prevent benthic impacts, *id.* ¶¶ 31–53, as well as to pay a total of $750,000 to fund wild Atlantic salmon restoration efforts and to reimburse the plaintiffs for the cost of litigation, including attorney’s fees. *Id.* ¶¶ 54, 57.


strictly liable for administrative and civil judicial fines of up to $25,000 per day for violations of the terms of the permit, and additionally, criminal felony penalties would attach for knowing violations.\textsuperscript{163} Without the more immediate effect of stipulated penalties, Heritage has much less incentive to comply with terms it has agreed to such as forgoing the use of European stocks of Atlantic salmon. This is particularly so given that, as noted earlier, Heritage is not required to obtain permits for its existing net-pen farms.

The second important development in the USPIRG cases occurred on February 13, 2003. The court entered an interlocutory order preventing ASM from introducing a “new class of fish” into its net pens until the completion of the remedial phase of the case.\textsuperscript{164} The case is thus having a potentially significant economic impact on ASM.

C. \textit{State and Federal Cases Arising out of Washington}

Interestingly, the USPIRG cases do not present the first occasion in which a court considered whether or not farm-raised Atlantic salmon constitute pollutants.\textsuperscript{165} In Washington, the State had been delegated authority to administer the NPDES permit system within its boundaries using state law. Accordingly, when a case involving the issue arose in a Washington state court, \textit{Marine Environmental Consortium v. Washington Department of Ecology},\textsuperscript{166} the court applied Washington law to the question of regulation rather than federal law, unlike the USPIRG cases.\textsuperscript{167} Moreover, rather than being the result of a citizen suit trying to enforce discharge requirements against an entity that had no permit, the Washington case arose as an appeal brought by nonprofit environmental and fishing organizations. The organizations challenged two things: (1) a permit issued by an administrative agency, the Washington Department of Ecology (WDE), and (2) a decision by a quasi-judicial body, the Washington State Pollution Control Hearings Board (WPCHB).\textsuperscript{168}

\begin{itemize}
\item \textsuperscript{163} Id.
\item \textsuperscript{166} Id.
\item \textsuperscript{167} \textit{Compare Marine Envtl. Consortium, ¶ II.I., II.VIII., II.IX with Heritage Salmon, Atl. Salmon, and Stolt Sea Farm.}
\item \textsuperscript{168} Marine Envtl. Consortium, ¶ I.I.
\end{itemize}
Fish as Pollutants

The main issue before the WPCHB was the risk posed to native Pacific salmon by farm-raised Atlantic salmon, which were not native to the Pacific Northwest—in particular, the risk that escapees might “colonize Puget Sound rivers.” The WPCHB had before it evidence of two escape incidents—105,000 and 369,000 escapees, in July 1996 and July 1997, respectively—as well as evidence of the presence of at least twelve Atlantic salmon smolts in the Tsitika River. In the face of “substantial conflict in the testimony of expert, agency and lay witnesses,” the WPCHB found inter alia that “while undesirable,” the accidental release of Atlantic salmon did “not pose a significant threat to native salmon” nor “degrade water quality.” However, the WPCHB also found substantial evidence in the record to support a finding that “regular and large releases such as those that occurred in 1996 and 1997 could constitute a threat to Pacific salmon.” On the issue of spawning, the WPCHB concluded that while there “may have been successful spawning” of escapees, there was “no evidence to support that Atlantic salmon was ‘self-sustaining.’” Here, the WPCHB took a middle ground, ordering the WDE to “take the Tsitika findings fully into account when it considers and reissues” the permits.

As an initial matter, the court found that the release of farmed Atlantic salmon from net pen facilities is regulated under the NPDES permit program. At the same time, the court, deferring to the administrative interpretation and application of the law by the WPCHB, held that the inadvertent release of Atlantic salmon is neither “pollution” within the meaning of Washington law nor, at a current level of escapement, a “nuisance.” In addition, the court held that the inadvertent release does not “render [State] waters harmful, detrimental or injurious to salmonid species” or violate water quality standards. The court also affirmed the WPCHB’s findings with regard

169. Id. ¶ I.III, I.VI.
170. Id. ¶ I.IV, I.XII.
171. Id. ¶ I.VIII–IX.
172. Id. ¶ I.LX.
173. Id. ¶ I.XIII.
174. Id.
175. Id. ¶ II.VII.
176. Id. ¶ II.IV–V.
177. This leads to the anomalous result that the federal standard of “pollution” is more stringent than the Washington standard since any escape of non-Native farm-raised salmon would constitute the release of a “pollutant” under the CWA. See infra note 332 and accompanying text.
Another case arising out of the State of Washington, a Ninth Circuit opinion, addresses the CWA regulation of mussel-harvesting facilities and bears on the question of whether or not salmon mariculture escapees are pollutants, as well as the broader issue of fish as pollutants. In Ass’n to Protect Hammersley, Eld, & Totten Inlets v. Taylor Resources, Inc. (APHETI), the defendant-appellee, Taylor Resources, Inc. (Taylor) operated two mussel-harvesting facilities in Puget Sound’s Totten Inlet and produced more than 20,000 pounds of mussels annually. Taylor harvested a species of mussels known as Gallo that had been present in the Puget Sound for approximately twenty-five years and that “reproduce naturally in Puget Sound, albeit in limited numbers.”

There was some dispute over whether Gallo mussels were introduced into Puget Sound solely as the result of the actions of mussel harvesters, or whether, in addition, they independently found their way to the Sound. A mesh net surrounded the farm-raised mussels. However, unlike Atlantic salmon farming operations that add fish food and other chemicals to the water, the Gallo mussels are “nurtured exclusively by the nutrients found naturally in the waters of Puget Sound, with nothing added.” Nevertheless, as a byproduct of their metabolism, the mussels “produce and release” particulates, feces and pseudofeces and generate ammonium and inorganic phosphate; mussel shells are released from the nets as well.

When Taylor applied for an NPDES permit, the WDE informed Taylor it would “neither accept nor process” the application. According to the WDE, the mussel facilities did not require discharge permits because fish food (nutrients) is not used to promote shellfish growth. Subsequently, APHETI filed a citizen suit under the CWA.
alleging that Taylor was violating the Act by discharging mussel feces, mussels shells and ammonia into Puget Sound without a permit. 188

Two policy considerations complicated the legal question of whether the mussel operations resulted in the “discharge of a pollutant” from a “point source.” 189 First, mussels filter excess nutrients that otherwise can harm marine ecosystems, thus enhancing water quality. 190 Second, regulation of these types of operations could divert regulators’ finite financial and personnel resources away from other more environmentally-significant activities such as pollution prevention. 191

Taking the issue of whether these facilities constitute point sources first, the court deferred to the EPA’s determination of which mariculture facilities should be considered point sources under the CWA. The court concluded that the mussel operations were not “point sources” within the meaning of the Act because the EPA had excluded facilities that feed less than approximately 5,000 pounds of food per month in a given calendar year. 192 The implication for Atlantic salmon mariculture operations of this portion of the Ninth Circuit opinion is that those operations that exceed the feeding and production criteria set forth in EPA regulations are “point sources” within the meaning of the CWA. 193

The more difficult substantive issue the APHETI court faced was what constitutes “biological materials” for the purposes of the CWA. The Ninth Circuit noted the term “biological materials” was ambiguous because it was unclear whether it included all “biological matter regardless of quantum and nature and regardless of whether generated by living creatures, or whether the term is limited to biological materials that are a waste product of some human [or industrial] process.” 194 In resolving the ambiguity in favor of the latter interpretation, the court

188. Id.
189. Id. at 1010–11. Although the United States filed an amicus brief in this matter, it took no position on the issue of “biological materials” as “pollutants.” Personal communication with David Mann, counsel for APHETI (Oct. 1, 2002).
190. Ass’n to Protect Hammersley, Eld, & Totten Inlets v. Taylor Res., Inc., 299 F.3d 1007, 1010 n.2 (9th Cir. 2002). This argument was advanced by several Native American tribes that participated as amici curiae. Id.
191. Id. at 1011 n.3. These concerns were raised by amici Jamestown S’Klallam Tribe and People for Puget Sound. Id.
192. Id. at 1018–19 (noting that the mussel operations were not entitled to the other regulatory exclusion—minimum production levels). The APHETI court also might have been influenced by the fact that the statutory definition of point source includes a “concentrated animal feeding operation.” 33 U.S.C. § 1362(14) (2000) (emphasis added).
194. APHETI, 299 F.3d at 1016.
advanced five reasons. First, the court observed that under the doctrine of *ejusdem generis*, the more specific illustrative lists of pollutants—“radioactive materials,” “wrecked or discarded equipment,” “garbage,” “sewage sludge,” “solid waste,” and “incinerator residue,” set forth in 33 U.S.C. § 1362(6), suggest that the more general term “biological materials” refers to a “waste material of a human or industrial process.” Second, when it enacted the CWA, Congress specifically listed the “propagation” of “shellfish” as a goal. Third, there was no evidence in the record that the release of mussel shells, mussel feces, or other byproducts results in any harm. Fourth, viewing biological materials as requiring transformation by a human or industrial process was in accord with other courts that had considered the question. Finally, the CWA defines a closely related term, “pollution,” to mean “man-made or man-induced alteration of the chemical, physical, biological, and radiological integrity of water.” In sum, the Ninth Circuit held that mussel shells, mussel feces, and other byproducts, although released into the environment, “come from the natural growth and development of mussels” rather than from the “waste product of a transformative human process,” and, as such, are not regulated under the CWA.

195. “[W]hen a statute contains a list of specific items and a general item, we usually deem the general item to be of the same category or class as the more specifically enumerated items.” *Id.* (quoting *Sutton v. Providence St. Joseph Med. Ctr.*, 192 F.3d 826, 834 (9th Cir. 1999)) (internal citations omitted).

196. We question why the court felt it necessary to rely on this rule of statutory construction given the breadth of so-called “specific” terms such as “garbage,” “sewage sludge,” and “solid waste” and the fact that it had other bases for its holding.

197. *Id.* (quoting 33 U.S.C. § 1251(a)(2)); see also 33 U.S.C. §§ 1312(a), 1314(a)(2).

198. *APHETI*, 299 F.3d at 1016.

199. *Id.* at 1017 (citing *Concerned Area Residents for Env’t. v. Southview Farm*, 34 F.3d 114, 117 (2d Cir. 1994) (liquid manure that had been spread on farm fields); *United States v. Plaza Health Labs.*, Inc., 3 F.3d 643, 645 (2d Cir. 1993) (holding that glass vials containing human blood that were placed into a river were biological materials); *Nat’l Wildlife Fed’n v. Consumers Power Co.*, 862 F.2d 580, 583 (6th Cir. 1988) (holding that “live fish, dead fish and fish remains” released through a dam’s turbines were biological materials); *United States v. Frezzo Bros.*, 461 F. Supp. 266, 269–70 (E.D. Pa. 1978), aff’d, 602 F.2d 1123 (3d Cir. 1979) (holding that “mushroom compost” pile runoff was a biological material). Although those courts found that each of the above materials are “biological materials,” the judicial opinions in those matters neither examined the meaning of “biological materials” in depth nor limited that term’s meaning in the manner suggested by the *APHETI* court.

200. *APHETI*, 299 F.3d at 1017 (quoting 33 U.S.C. § 1362(19)).

201. *Id.* at 1017–19. The Ninth Circuit is not entirely persuasive on the distinction it is attempting to draw. For example, we question in what sense these releases were not “human-induced” given that, but for the placement of the mussels by humans in the Sound, their feces and shells would not have been released.
Fish as Pollutants

Nonetheless, the APHETI court was careful to note that it was not suggesting that materials found naturally in the water can never be “biological materials” within the meaning of the CWA. As an example, the court indicated that discarded fish parts and shells, “although naturally occurring, are altered by a human or industrial process, and as waste materials in significant amounts, might affect the biological composition of the water.” More importantly, by citing the critical holding in NWF v. Consumers Power Co. with approval, the Ninth Circuit also implied that, although its view of what constituted “biological materials” under the CWA was narrower than that advocated by the plaintiffs, even a more limited interpretation of biological materials would include the discharge of “live fish, dead fish and fish remains” through a dam turbine. Thus, under the APHETI reasoning, to the extent that farm-raised Atlantic salmon escape as a result of a transforming human process, they would be considered to be “biological materials” and hence “pollutants” within the meaning of the CWA.

The Washington cases are important not only because they raise the issue of escaped farmed fish as pollutants, but also because in conjunction with the recent Maine cases, they illustrate the need for a coherent and cohesive national policy regulating the intentional and unintentional release of farmed aquatic life into the nation’s waters. Without such a policy, courts will continue to arrive at potentially conflicting interpretations of the law in cases such as the Washington mussel case and the Washington and Maine salmon mariculture cases.

202. Id. at 1016–17.
203. Id. at 1017 (citing Ass’n of Pac. Fisheries v. Envtl. . . Prot. Agency, 615 F.2d 794, 802 (9th Cir. 1980)). It should be noted that such a finding may run head long into the Consumers Power case discussed earlier that found that, while live fish and fish parts were “pollutants” within the meaning of the CWA, no “addition” occurs when they are released back into the water. See NWF v. Consumers Power Co., 862 F.2d 580, 584 (6th Cir. 1988).
204. 862 F.2d 580 (6th Cir. 1988); see also supra note 150 and accompanying text.
205. APHETI, 299 F.3d at 1017 (quoting Consumers Power, 862 F.2d. at 583) (dealing with a pump storage facility rather than a dam).
206. See id. As previously noted, the court in the USPIRG cases did not rely on the endangered species status of wild Atlantic salmon as a justification. Rather, the court rested its decision on the literal meaning of “biological materials,” one of the items Congress defined as a “pollutant” under the CWA, and the fact that the stocks in question were not native to the Gulf of Maine. Id.
D. Proposed EPA Action to Regulate Escapes: A Step Forward or Backward?

Recently, the EPA has taken a cautious step toward distinguishing which farmed fish should be regulated as pollutants and which ones should not. Specifically, in proposed aquaculture effluent guidelines, the EPA states that persons operating certain net pen systems must “develop and implement [best management] practices [‘BMPs’] to minimize the potential [unintended] escape of non-native species.”207 These practices—such as installing double netting in a net pen operation208—would be embodied in a non-native species escapement plan.209

The EPA’s proposal is important in several respects. First, given the EPA’s intention to regulate the escape of net-pen farm-raised fish, the EPA has implicitly determined that their escape constitutes the “discharge” of a “pollutant” from a point source.210 Second, in the proposed rule, the EPA defines a non-native aquatic animal species as an “individual, group or population of a species . . . [t]hat is introduced into an area or ecosystem outside its historic or native geographic range” and that “has been determined and identified by the appropriate State or Federal authority to threaten native aquatic biota.”211 While this definition rightly recognizes that non-nativeness should be examined at the subspecies level, it may have limited applicability given that it only applies in those instances where a state or federal entity has made a formal determination that the non-native strain constitutes a threat. It would nonetheless appear to encompass the rearing of non-native strains of Atlantic salmon off the coast of Maine. Third, as discussed next, the EPA’s proposed rule is limited in a number of respects, some of which

209. Id.
significantly undercut the otherwise far-reaching holding of the USPIRG cases.

To begin with, the proposed effluent guidelines explicitly exempt “species raised for stocking by public agencies” from the definition of “non-native aquatic animal species.” As discussed in more detail below, while a large number of federal and state fish hatcheries raise non-native species and have NPDES permits, those permits do not authorize the intentional or unintentional release of fish. This proposal would appear to authorize such releases without regard to the biological consequences. Second, the EPA’s proposal is limited to the escape of non-native fish rather than any farmed fish. As noted elsewhere in this article, there are likely to be subtle genetic distinctions between wild and captive-bred populations. Third, although the EPA regulates concentrated aquatic animal production facilities (CAAPFs) that produce as little as 20,000 pounds annually of any cold water species, the proposed effluent limitations do not apply to facilities that produce less than 100,000 pounds annually. The EPA established the 20,000-pound threshold based on economic modeling that led it to conclude that the proposed limitations would have adverse economic effects on trout producers that produce less than 94,000 pounds annually.

Fourth, the guidelines do not take into account the potential biological impact of an escape. For example, the stringency of the requirements is not based on whether or not wild populations potentially affected are endangered or threatened. Fifth, the proposed guidelines

212. Id.
213. Intentional releases include stocking for sport fishing.
214. See, e.g., Telephone Interview with Gary Whelan, Fish Production Manager, Michigan Department of Natural Resources (MDNR) (Apr. 15, 2002) (while MDNR managed hatcheries have discharged permits, they do not have permits authorizing releases for stocking purposes).
216. The EPA does not explain why it chose 100,000 rather than 94,000 pounds as the cutoff. See Effluent Limitation Guidelines, 67 Fed. Reg. at 57,925.
217. See id. The EPA found that “small facilities would experience compliance costs that exceed 5% of their revenues which is higher than for large facilities.” See id. Moreover, a facility that produces between 20,000 and 100,000 pounds annually would not be exempted from the NPDES permit requirement. Rather, it would be subject to conditions in an NPDES permit based on the “best professional judgment” of the permit writer. See id.
218. Id. at 57,928.
are directed at minimizing rather than preventing escapes.\textsuperscript{219} Moreover, they do not even require that escapes actually be minimized. Rather, they only establish the principle that operators should implement best management practices to minimize escapes.\textsuperscript{220} In other words, rather than an escape leading to a strict liability permit violation, any CWA liability for escapes under the proposed guidelines is premised solely on the failure of a facility operator to use best management practices. Finally, the EPA underscored its ambivalence to address this “potential area of concern”\textsuperscript{221} with its statement that it was “considering whether it should establish national requirements for net pens systems at all.”\textsuperscript{222} In sum, given the limitations in the EPA’s proposed effluent guidelines as outlined here, the EPA’s proposal would appear to cut the legs out from under the Maine district court’s rulings on escapees.\textsuperscript{223}

Although the EPA did not specifically propose any rule related to the intentional release of farmed fish—for example, in the event that net penned farmed fish are not “growing rapidly enough to justify continued feeding”—the EPA indicated that it was considering banning such practices.\textsuperscript{224} This statement highlights another peculiarity of the proposed rule. It is not clear how the EPA can reconcile a requirement that facility operators minimize escapes while otherwise providing those same facility operators with carte blanche to intentionally release farmed fish.

From the standpoint of the Atlantic salmon mariculture industry, the money saved in building “leaky” Atlantic salmon net pens from which a considerable number of fish can escape is more than the money represented by lost fish.\textsuperscript{225} Nevertheless, the EPA adopts as a final rule even the rather limited if regulatory framework proposed and mandates that Atlantic salmon operators use best management practices to minimize escapes as part of any Atlantic salmon mariculture NPDES permit; the industry will either have to comply with that directive or go out of business.

\textsuperscript{219} Id.
\textsuperscript{220} Id.
\textsuperscript{221} Id. at 57,912.
\textsuperscript{222} Id. at 57,901.
\textsuperscript{224} Id. An intentional release of farmed fish would appear to fit easily within the \textit{APHETI} court’s standard of a waste product generated by a human-induced process. \textit{APHETI}, 299 F.3d at 1017–18.
\textsuperscript{225} Volpe, \textit{supra} note 105, at 14.
Fish as Pollutants

E. A Draft Code of Conduct

Finally, the National Maritime Fisheries Service (NOAA) has recently taken steps to establish guidelines for aquaculture with the publication of a draft code of conduct for mariculture operations in federal waters, which may assist Atlantic salmon recovery efforts.226 While potentially far-reaching in some respects, unlike the EPA effluent guidelines mentioned above, the proposed code of conduct is “soft” law—that is, if it is adopted, compliance will be voluntary.227 In pertinent part, the code of conduct expresses many of the concerns voiced here and calls for the adoption of the “precautionary approach” combined with “adaptive management” to be the “guiding principles” while recognizing the imperative of preventing escapes, combined with the remedial action to address significant escape incidents.228

IV. LEGAL STATUS OF WILD ATLANTIC SALMON

In this section we trace the legal status of wild Atlantic salmon over the past several decades. Most of our focus is on the consideration of Atlantic salmon under the Endangered Species Act, which culminated in a recent decision by the federal government to list the Gulf of Maine Atlantic salmon D.P.S. as endangered. However, the controversy surrounding the legal status of wild Atlantic salmon did not end with that listing. Indeed, the controversy led Congress to ask the National Research Council (NRC) for scientific advice on “understanding and reversing the declines in Maine’s salmon population,” the issuance of an interim report by the NRC Committee on Atlantic Salmon in Maine on


227. See CODE OF CONDUCT, supra note 226, § 3.

228. See id. §§ 6.5.1, 6.6.3. The Code also calls for best management practices; conservation of genetic diversity and the maintenance of the functional integrity of ecosystems; regulation of non-indigenous aquatic organisms and genetically-altered species; adoption of necessary measures to minimize the potential for the incidence and transmission of diseases and parasites; and protection of critical habitats, protected areas, endangered species, etc. through sitting criteria, monitoring, assessment, and enforcement. Id. §§ 6.3.2–3, 6.5.2–3, 6.5.4–5, 6.6.3.
the genetic makeup of Atlantic salmon populations in Maine, and a legal challenge to the listing.

A. Legal Framework

In 1976, Congress enacted the Fishery Conservation and Management Act and, among other things, established eight regional fishery management councils that are each tasked to prepare and submit to the Secretary of Commerce a fishery management plan for fisheries within their respective regions. The New England Fishery Management Council (NEFMC) has jurisdiction over fish in federal waters bordering Maine’s and other New England states’ coastal waters. In 1988, the NEFMC adopted a Fishery Management Plan (FMP) for Atlantic salmon. The FMP prohibits the possession of wild Atlantic salmon harvested in federal waters. The Magnuson Act authority complemented existing federal regulatory authority over Atlantic salmon on the high seas—the area beyond nations’ exclusive economic zone that was conferred to the United States by the Convention for the Conservation of Salmon in the North Atlantic Ocean and the North Atlantic Salmon Conservation Organization—as well as state authority. A 1999 amendment to the Atlantic salmon FMP created an administrative process for persons desiring to operate an Atlantic salmon mariculture project in federal waters.

229. NRC GENETIC STATUS INTERIM REPORT, supra note 156.
233. Id.
239. 50 C.F.R. § 648.41. There is some question whether, as a matter of law, the Magnuson Act actually confers on NOAA, NMFS and the regional Fisher Management Councils (FMCs) the power to regulate aquaculture. There is no explicit authorization in the Magnuson Act to regulate
The Endangered Species Act (ESA) sets forth a far-reaching regulatory regime that provides protection for any species threatened or endangered with extinction.\(^{240}\) The ESA defines “species” as “any subspecies of fish or wildlife or plants, and any distinct population segment of any species of vertebrate fish or wildlife, which interbreeds when mature.”\(^{241}\) The ESA directs the federal government to determine whether any species has become endangered or threatened due to habitat destruction, over-utilization, disease or predation, inadequacy of other regulatory mechanisms, or other natural or manmade factors.\(^{242}\) Because salmonids are anadromous and spend a portion of their life cycle in both freshwater and saltwater, both NMFS and FWS have a role in determining the status of Atlantic salmon under the ESA.\(^{243}\) That determination must be rendered “solely on the basis of the best scientific and commercial data available . . . .”\(^{244}\)

The listing of a species as threatened or endangered has a whole host of implications. To begin with, the Services are required to designate critical habitat for the species,\(^{245}\) and develop and implement a recovery plan.
plan that provides for the “conservation and survival” of the species.\(^{246}\)

In addition, the ESA prohibits the taking\(^{247}\) of any endangered fish or wildlife species in most circumstances. In contrast, takings of threatened species are only prohibited upon the adoption of a rule pursuant to section 4(d) of the ESA, and with respect to any resident species in a state that has entered into a cooperative agreement with the federal government, only to the extent that that state also has adopted the rule.\(^{248}\)

Finally, section 7 of the ESA, by mandating interagency consultation and biological analyses, attempts to “insure that any action authorized, funded, or carried out by” any federal government agency “is not likely to jeopardize the continued existence of any endangered species or threatened species or result in the destruction or adverse modification of habitat.”\(^{249}\)

---

\(^{246}\) 16 U.S.C. § 1533(f). Recovery plans are required, “to the maximum extent practicable” to include “site-specific management actions,” “objective, measurable criteria” for removing the species from the ESA list, and time and cost estimates. Id.; Fund for Animals, 903 F. Supp. at 115 n.8.

\(^{247}\) 16 U.S.C. § 1538. “Take” is defined as to “harass, harm, pursue, hunt, shoot, wound, kill, trap, capture, or collect, or to attempt” the same. Id. § 1532(19) (emphasis added). By regulation, both the FWS and NMFS have separately defined “harm” to include “significant habitat modification or degradation where it actually kills or injures wildlife by significantly impairing essential behavioral patterns, including breeding, feeding or sheltering.” Endangered and Threatened Wildlife and Plants, 50 C.F.R § 17.3, § 222.102 (2003). The U.S. Supreme Court upheld the validity of the FWS regulation in Babbitt v. Sweet Home Chapter. 515 U.S. 687, 708 (1995). The regulatory definition of “harass” is arguably broader than “harm” encompassing not only acts but also “omission[s]” and requiring only the “likelihood of injury” rather than actual injury. 50 C.F.R. § 17.3. Moreover, the terms “wound” and “kill” suggest an emphasis on consequences in addition to intent (e.g., hunt, shoot, trap, etc.). See 16 U.S.C. § 1531(18). Exceptions to the broad taking prohibition include takings for scientific purposes and to enhance species survival or if such taking is “incidental to, and not the purpose of, the carrying out of an otherwise lawful activity.” Id. § 1539(a)(1).

\(^{248}\) Id. § 1533(d).

\(^{249}\) 16 U.S.C. § 1536(a)(2). The Services have jointly defined “jeopardize the continued existence of” as engaging in an action that “reasonably would be expected, directly or indirectly, to reduce appreciably the likelihood of both the survival and recovery of a listed species in the wild by reducing the reproduction, numbers, or distribution of that species.” Interagency Cooperation, Endangered Species Act of 1973, 50 C.F.R. § 402.02 (2003). Such action includes continuing action. Tenn. Valley Auth. v. Hill, 437 U.S. 153, 173–74, 184 (1978). Each federal agency proposing to undertake action shall “request of the Secretary [of the Interior or Commerce, as appropriate] information whether any species which is listed or proposed to be listed may be present in the area of such proposed action. If the Secretary advises . . . that such species may be present, such agency shall conduct a biological assessment for the purpose of identifying any endangered
Fish as Pollutants

B. ESA Listing of Atlantic Salmon

The Services have followed a tortured track in their regulation of wild Atlantic salmon, reversing their position on more than one occasion. This story begins in 1991, when the FWS designated Atlantic salmon in five Maine rivers as candidate species under the ESA. In 1994, in response to petitions to list Atlantic salmon under the ESA, the Services found sufficient scientific information to suggest that listing may be warranted. In 1995, the Services conducted a joint status review of the species. That review led the Services to conclude that D.P.S. of Atlantic salmon on seven Maine rivers—the Dennys, East Machias, Machias, Pleasant, Narraguagus, Ducktrap and Sheepscot Rivers—should be listed, and as a result, the Services proposed to list the seven-river Atlantic salmon D.P.S. as a threatened species.

By proposing to list the Atlantic salmon D.P.S. as a threatened rather than an endangered species, the Services were able to include within the proposal a special rule promulgated pursuant to section 4(d) of the ESA that would permit the State of Maine to adopt a plan, subject to the approval of the Services, that would define the measures to be undertaken to conserve the species. After assessing the adequacy of species or threatened species which is likely to be affected by such action.” 16 U.S.C. § 1536(c)(1) (emphasis added). If the biological assessment identifies any such species that is “likely to be affected,” formal consultation under § 1536(a)(2) is triggered. Id. At the conclusion of consultation, the Secretary is required to issue a biological opinion “detailing how the agency action affects the species or its critical habitat.” Id. § 1536(b)(3). And, if “jeopardy or adverse modification is found,” the ESA requires the Secretary to suggest “reasonable and prudent alternatives” to the action proposed by the federal agency. Id.


252. In general, “any distinct population segment of any species of vertebrate fish or wildlife that interbreeds when mature” may be formally defined under the ESA as a distinct population segment (D.P.S.). 16 U.S.C. § 1532(15).


255. See 16 U.S.C. § 1533(c)(1) and text accompanying supra note 249.

the State of Maine’s conservation plan, and reviewing public submittals
and other current information, the Services concluded that ongoing
actions, including those identified in Maine’s conservation plan, “have
substantially reduced threats to the species [and] . . . will facilitate the
rehabilitation of the seven rivers D.P.S.” As a result, in December
1997, the Services withdrew the proposed rule to list the Atlantic salmon
D.P.S. as a threatened species.258

Defenders of Wildlife, a nonprofit organization, and other
organizations concerned with status of Atlantic salmon, sued the
Services over their decision to withdraw the proposed listing. After
conducting an updated status review,259 and presumably with insight
gained on the likelihood that it would prevail on the merits of the legal
challenge to its withdrawal decision given that the legal case had been
fully briefed,260 the Services reversed course again. On November 17,
1999, the Services proposed to once again add Atlantic salmon to the
ESA list261—this time proposing that it be listed not as threatened, but as
derangered, and designating the D.P.S. as the Gulf of Maine in
acknowledgement of the fact that other Atlantic salmon populations
could be added to the D.P.S. “if they were found to be naturally
reproducing and to have wild stock characteristics.”262 On November 17,
2000, the Services adopted a final rule that found that the Atlantic
salmon D.P.S. was distinct in that it met both discreteness conditions. It
was (1) “markedly separated from other populations” due to “physical,
physiological, ecological, or behavioral factors,” and (2) delimited by
international boundaries such that significant differences in the “control
of exploitation, management of habitat, conservation status, regulatory
mechanisms exist between the United States and Canada.”263 Further, it

257. Endangered and Threatened Wildlife and Plants; Withdrawal of Proposed Rule to List a
258. Id.
259. The 1999 status review was made publicly available on Oct. 19, 1999. Availability of a
260. In anticipation of a proposed listing, the parties to the case entered into a stay. Telephone
Interview with Howard Crystal, Counsel for Defenders of Wildlife (Apr. 2002).
261. Endangered and Threatened Species; Proposed Endangered Status for a Distinct Population
Segment of Anadromous Atlantic Salmon (Salmon salar) in the Gulf of Maine, 64 Fed. Reg. 62,627
262. Endangered and Threatened Species; Final Endangered Status for a Distinct Population
was a significant population segment of the species and in danger of extinction. That decision is presently being challenged by the State of

264 Id. More specifically, the Services have identified three criteria that establish status as a D.P.S.: (a) the discreteness of a given population segment, (b) the significance of a given population segment to the species or subspecies, and (c) the conservation status of a given population segment in relation to the ESA listing standards. The Services D.P.S. policy is found at Policy Regarding the Recognition of Distinct Vertebrate Population Segments Under Endangered Species Act. 61 Fed. Reg. 4722 (Feb. 7, 1996). A summary of the Services’ evaluation taken almost verbatim from Endangered and Threatened Species; Final Endangered Status for a Distinct Population Segment of Anadromous Atlantic Salmon (Salmon salar) in the Gulf of Maine, 65 Fed. Reg. at 69,459–69,462, follows (parenthetical comments added by authors):

To be discrete, a population must be: 1) “markedly separated from other populations of the same taxon as a consequence of physical, physiological, ecological, or behavioral factors” or 2) “delimited by international governmental boundaries within which differences in control of exploitation, management of habitat, conservation status, or regulatory mechanisms exist that are significant in light of section 4(a)(1)(D) of the ESA.” The Services used biogeographical and other information to define three historically distinct population segments for the Atlantic salmon: the Long Island Sound D.P.S., the Central New England D.P.S., and the Gulf of Maine D.P.S. The two southernmost D.P.S.s are now essentially extinct. The boundaries of the Gulf of Maine population segment are described as running from the Kennebec River north and including eight rivers. River-specific hatchery fish are included with wild fish as part of the D.P.S. The Services conclude from data on the straying of spawning fish, genetic information, geographic segregation, and limited stocking from outside the population segment that the Gulf of Maine D.P.S. is separate from other salmon populations. The Services found that probably no salmon are genetically pure, but that the remaining salmon are the only genetic legacy of the salmon that ranged as far south as the Housatonic River in Connecticut. The Services conclude that both criteria for discreteness were satisfied by the Gulf of Maine D.P.S., but note that either one would suffice.

Significance means: 1) “persistence of the [D.P.S.] in an ecological setting unusual or unique for the taxon.” 2) “evidence that loss of the [D.P.S.] would result in a significant gap in the range of the taxon,” 3) evidence that the [D.P.S.] are abundant elsewhere (introduced), or 4) that the [D.P.S.] is different (markedly) from the rest of the species. The Services determined that D.P.S. is the southernmost population of salmon still existing. The Services note that the genome is no longer pure, but point out that important genetic resources remain within the D.P.S., because stocking was primarily from D.P.S.-derived fish. (This implies that non-D.P.S. fish could damage the genome.). The Services point out that D.P.S. genetic resources are “vital to the species’ future survival.” (This may be a reference to the potential effects of global warming, which presumably would warm river waters throughout much of the Atlantic salmon’s historical habitat and thus favor the survival of salmon adapted to warmer waters.). The Services conclude that Gulf of Maine salmon are both discrete and significant, i.e. a D.P.S.

The conservation status of a given population segment refers to the abundance of that population segment relative to the ESA listing standards. The Services conclude that abundance of the Gulf of Maine D.P.S. is extremely low. Fish counts annually are in no more than tens of fish for the whole system, although the Services caution that this is not a full count. The pre-fishery abundance index (one-sea-winter or 1SW fish means fish that have not yet returned from the sea to ascend rivers and spawn) is low in spite of improving ocean habitat conditions. The Services attribute this to depressed spawning populations in the rivers and consequently few juvenile salmon entering the sea. 0+ (fry) and 1+ and 2+ (parr) numbers are also low. Smolt production is also very low. There is unexpectedly high mortality from 1+ and older parr to the smolt stage. Also, about half of the smolt that do migrate don’t reach the Gulf. Most of this information comes from the Narraguagus River. A similar study was conducted on the Pleasant River and found 31 smolt that had come from a commercial hatchery upstream and 676 that were evidently wild. (This implies genetic pollution of native fish by escapees from the hatchery stage of the commercial mariculture operations—that is, a release of juvenile
Maine, two of the aquaculture defendants in the USPIRG cases (Stolt and Atlantic Salmon), and other parties. In their listing of the Gulf of Maine Atlantic salmon D.P.S. as endangered, the Services made a number of important findings that are relevant to the issue of cultured salmon as pollutants:

It is “unlikely that any Atlantic salmon populations in the United States exist in a genetically pure native form . . . .” However, “present populations are descendants of these aboriginal stocks, and their continued presence in indigenous habitat indicates that important heritable local adaptations still exist.” Indeed, despite 128 years of stocking, “hatchery fish have not substantially introgressed with the remnant populations and genomes” of the Gulf of Maine population segment.

Three of the factors cited by the Services in support of their decision to list the Gulf of Maine D.P.S. as endangered are: (1) a “large number of aquaculture hatchery origin juveniles” in the Pleasant River; (2) the growing threat of Infectious Salmon Anemia (ISA), a fatal viral disease, and (3) the “increasing use of European strain Salmon by the Maine Aquaculture industry.” Studies of Northwest Ireland rivers have “clearly demonstrated” that escaped juvenile salmon have “completed their entire life cycles in the wild, including accurate homing to natal rivers and interbreeding with wild salmon.” Rivers in Canada and Norway that bear a resemblance to Gulf of Maine rivers “provide substantial evidence that negative impacts [from escaped farm salmon] to the D.P.S. such as disruption of redds, competition for food and habitat, disease or parasite transfer, and interbreeding “can be reasonably anticipated to occur in Maine.”

The Services conclude that the abundance of salmon in the Gulf of Maine D.P.S. is extremely low.

267. Concerns regarding the documented presence of ISA on the Canadian-side of Cobscook Bay led the Services to forgo stocking rivers with rivers-specific fish raised on the U.S.-side of that Bay.
269. Id. (citing S.L. Clifford et al., Genetic Changes in an Atlantic Salmon Population Resulting from Escaped Juvenile Farm Salmon, 52(1) J. FISH BIOLOGY 118 (1998)).
270. Id. at 69,477.
271. Id. at 69,478 (citations omitted).
Fish as Pollutants

Escaped farmed salmon have been found in the St. Croix, Penobscot, Dennys, East Machias and Narraguagus Rivers.272

C. Review by the National Academy of Science

The demarcation of wild Atlantic salmon found in the seven Maine rivers as a distinct population segment led to a call for the National Academy of Science (NAS) to review the conclusions of the Services.273 The Services’ findings were generally seconded by the important report issued by the NAS in the spring of 2002 on the Gulf of Maine D.P.S.274 The National Resource Council’s Committee on Atlantic Salmon in Maine found that North American and European stocks were “clearly genetically distinct,” and there was “surprisingly strong” evidence of genetic diversity between Maine and Canadian stocks, and even “considerable genetic divergence” among populations on the eight Maine rivers.275 Moreover, the NRC panel concluded that farmed fish differ in genetic makeup from the Gulf of Maine D.P.S. due to non-native strains, selection by breeders (growth rate, fat content, disease resistance, and delayed maturity), and “inadvertent selection by the novel environment (e.g., reduced fright response, disease resistance, and altered aggressive behaviors).”276 Particularly important for present purposes, the NRC noted that researchers estimate a 3% escape rate in British Columbia.277 For Maine, a 3% escape rate of farmed salmon translates to 180,000 escapees per year from net pens.278 Even if one assumes an “escape rate as low as 0.17 percent, which would be impressive,” there still would be “10,000 escapees per year, 100 times the number of adults that returned to spawn” in the Gulf of Maine rivers in 2000.279

272. Id.
273. NRC GENETIC STATUS INTERIM REPORT, supra note 156, at 1.
274. Id.
275. Id. at 20 (“Those same traits may not be adaptive in the wild.”).
276. Id. at 21 (citing M. Gross, Net Risk: Assessing Potential Impact of Fish Farming on BC’s Wild Salmon, in GHOST RUNS: THE FUTURE OF WILD SALMON ON THE NORTH AND CENTRAL COASTS OF BRITISH COLUMBIA (B. Harvey et al. eds., 2002)).
277. Id.
278. Id. While wild salmon likely have an adaptive advantage over farmed salmon, even a “10:1 adaptive advantage” may not be sufficient to overcome a 100:1 numerical disadvantage. Id. In Marine Envtl. Consortium v. Wash. Dep’t of Ecology, the court affirmed the Washington State Pollution Control Hearings Board’s findings on the legal and biological effect of the escape of farmed Atlantic salmon on the Pacific coast. No. 99-2-00797-0, slip op. at ¶ II.VII. (Wash. Super. Ct. Dec. 1, 2000). See also supra notes 153–65 and accompanying text. The Board found, and the
The genetic impact of mariculture escapees on wild salmon thus may include the introduction of foreign genetic material to the wild genome. This foreign material may be from other parts of the Atlantic salmon’s range (the use of Norwegian sperm stock for example in U.S. mariculture operations) or it may be altered genetic material and include parts of the genome of fish other than the Atlantic salmon. The first genetically engineered animal in the U.S. awaiting U.S. Food and Drug Administration approval, so that it may be farmed commercially, is a Chinook salmon with Ocean pout genes that cause it to grow seven times faster than normal. It is not inconceivable that genetically engineered Atlantic salmon may soon exist and be awaiting approval for mariculture operations in the United States.

D. Conservation Measures in Support of Wild Atlantic Salmon

In addition to undertaking regulatory measures, such as listings that have legal implications, the Services have also undertaken conservation measures in relation to wild Atlantic salmon. Specifically, the Services maintain a hatchery program for the Gulf of Maine D.P.S. at Craig Brook National Fish Hatchery in Orland, Maine. This program is river-specific, meaning that fish from a given river are used as broodstock for that same river. Currently, fish from five rivers in the Gulf of Maine D.P.S. are used in the hatchery program. Fish from a sixth river also were captured as broodstock, but were destroyed due to the presence of Salmon Swimbladder Sarcoma Virus. A second attempt is being made to establish a broodstock population for this court affirmed, that the release of farmed salmon “while undesirable, does not pose a significant threat to native salmon” in terms of competition, predation, disease transmission or hybridization. See supra notes 165–79 and accompanying text. Yet, the Board found, and the court affirmed that “regular and large releases such as occurred in 1996 [105,000 Atlantic salmon] and 1997 [369,000 Atlantic salmon] could constitute a significant threat to Pacific salmon.”

---

280. NRC GENETIC STATUS INTERIM REPORT, supra note 156, at 20.
283. NRC GENETIC STATUS INTERIM REPORT, supra note 156, at 18.
285. Id.
Fish as Pollutants

Although fry were first stocked from this hatchery program into some rivers in 1996, and because a minimum of four years are needed to evaluate the success of the stocking, as of 2000, the effect of the river-specific stocking program had yet to be evaluated.

Within this hatchery program, the Services have established breeding protocols to help “ensure that genetic integrity is maintained.” Nevertheless, the genetic makeup of stocked fish is rarely the same as wild fish, even if the stocked fish are part of a river-specific breeding program. There are two reasons for this. One is that it is nearly impossible to avoid selection of any kind in a hatchery operation. That is, hatchery fish by definition survive according to their fitness in relation to selection pressures within the hatchery, which may be different from selection pressures in the wild. For example, aggressive feeding without regard to the presence of other fish may be a beneficial behavior in a hatchery, but in the wild, it may unnecessarily expose young salmon to predators. The second reason is that, even in the absence of differing selection pressures between hatcheries and the wild, the abundance of the specific genome of the hatchery brood-stock is amplified with respect to other segments of the overall genome in the wild—that is, in the hatchery, broodstock will on average contribute more offspring to the overall gene pool than will wild salmon, and thus the diversity of the overall genome is in effect narrowed.

V. THE POTENTIAL FOR DISEASE TRANSMISSION FROM CULTURED TO WILD STOCKS

The parallel harms that result from the intentional release of stocked fish and the accidental release of cultured fish suggest that a narrow policy intervention focused solely on aquaculture would be incomplete. Although the first cases that have considered fish as pollutants have arisen as a result of accidental releases in the context of aquaculture and power production, even fish that are intentionally released into the environment, often at the behest of recreational fishers, can negatively

286. Id.
287. Id.
288. The fry should spend at least two years in freshwater and two years at sea before returning to spawn or four years total from the time of stocking to the time of returning to spawn.
289. 65 Fed. Reg. at 69,467.
290. YURI P. ALTUKHOV ET AL., supra note 41, at 252–60.
291. Id.
292. Id.
impact wild populations. For example, native species reared in fish hatcheries have the potential to genetically pollute the genome of wild fish stocks, while both native and exotic species raised in fish hatcheries, like aquaculture stocks, may transmit pathogens and viruses to wild stocks. One such disease that scientists have studied extensively in freshwater salmonid populations is whirling disease.

Even when stocking policies do consider ecosystem-level effects, the nature of fish is fundamentally different than the nature of other pollutants. Most pollutants are passive, and their dynamic of spread can be more or less accurately predicted. However, live fish and parasites actively move through water systems, across oceans, and upstream. It is likely that anadromous rainbow and steelhead trout, as well as bull trout and Chinook salmon, are susceptible to infection by *M. cerebralis* and may help to spread the parasite through their migrations. This in turn implies the spread of the parasites across stretches of ocean by the straying of adult fish from their natal rivers. When fish move across sovereign boundaries they pose additional management challenges. Indeed, in any state fish stocking program, once the fish and their associated parasites are in the water their spread is largely out of the hands of fisheries managers.

In many cases, live fish are already treated as de facto pollutants by fish and game agencies. For example, the fear of the spread of whirling disease has prompted states to take expensive and sometimes extreme measures to prevent its spread. In 1968 in Michigan, trout which had been distributed to one hundred and fifty-nine waters were destroyed, because they could be traced back to three commercial facilities which tested positive for whirling disease. This was typical of measures which were taken at one time in response to reports of whirling disease in the United States. Recently however, such extreme measures have been modified in favor of policies designed to limit the spread of

---

293. The *Salmonidae* family is composed of salmon, trout, and char species.


297. Id. at 3.
whirling disease without destroying fish.\textsuperscript{298} Such policies include cleaning up hatcheries, limiting the stocking of infected fish, and managing affected rivers so that parasite numbers are reduced or even eliminated.\textsuperscript{299}

Moreover, researchers and fish managers already study and employ, respectively, elaborate and expensive technologies to prevent, detect, and treat parasites. For example, DNA-based techniques are being tested so that fisheries managers will have a rapid and accurate tool with which to detect \textit{M. cerebralis} infection.\textsuperscript{300} Traditional tools of parasite detection may be more expensive and time-consuming, and include spore-staining techniques, mechanical and enzymatic isolation of spores, histological assessment, and immunologic methods employing the use of labeled antibodies.\textsuperscript{301} Once detected, treatment of \textit{M. cerebralis} infections range from treatment of the worm host\textsuperscript{302} through drying of ponds, application of the lampricide agent TFM, and raising of water temperatures, to disinfection of water through chlorine, ozonation, filtration, or the application of ultraviolet light in order to attack the myxospore and actinospore stages of the \textit{M. cerebralis} life cycle\textsuperscript{303}, to drug treatment, habitat alteration (particularly the reduction of high-sediment areas of high \textit{M. cerebralis} infection), reduced or targeted stocking of infected fish, and as has already been discussed, the sacrifice of infected fish, in order to treat infected fish.\textsuperscript{304} These measures are analogous to measures which are taken in order to detect and limit the

\footnotesize
298. Id. at 20.
299. Id.
301. See id. at 205 for the expense and time needed for traditional methods. For the methods themselves, please see the appropriate sections in the article. Sections are titled according to the method they describe.
302. The lifecycle of \textit{M. cerebralis} is complex and may include several different strategies. A typical lifecycle includes the utilization of oligochaete hosts (typically Tubifex tubifex worms) and an infectious \textit{triaclinomyxon} stage that is free in the water column, in addition to the stages of the lifecycle spent within salmonid hosts. For a review of the \textit{M. cerebralis} lifecycle see Ronald P. Hecrick & Mansour El-Matbouli, Recent Advances with Taxonomy, Life Cycle, and Development of Myxobolus cerebralis in the Fish and Oligochaete Hosts, in Whirling Disease: Reviews and Current Topics, 29 AM. FISHERIES SOC’Y SYMP. 45–53 (2002).
303. See supra note 9 for a summary of the \textit{M. cerebralis} life cycle.
304. For information on methods used to treat whirling disease-positive fish, see Eric J. Wagner, Whirling Disease Prevention, Control, and Management: A Review, in Whirling Disease: Reviews and Current Topics, 29 AM. FISHERIES SOC’Y SYMP. 197–212 (2002). Sections are titled appropriately according to the treatment they describe.
dissemination of ordinary pollutants, which raises the question: If fish are already treated as pollutants by the agencies that work with them on a daily basis, shouldn’t this treatment be codified in the law?

VI. POLICY IMPLICATIONS AND CONCLUSIONS

The recent descriptions of escaped Atlantic salmon as pollutants are an indication that we have reached a turning point in our understanding of fisheries management. The concept of fish as pollutants, especially if it is solidified and extended through further court cases, has the potential to restructure not only our thoughts about regulation of the increasingly important U.S. aquaculture and mariculture industries, but also our management of fishes in inland and coastal waters in general. This is especially true for the type of expensive sport fish stocking programs that have been criticized by some fisheries managers and scientists for more than a century as ineffective at best and destructive at worst to natural assemblages of game and other fish. Critics of such programs have historically lacked the legal and scientific tools with which to argue their case. The recent decisions in the USPIRG cases, along with the endangered species listing of Atlantic salmon, the National Academy of Sciences report, and the increasing evidence of disease transmission between hatchery and wild stocks, may provide the philosophical, scientific and legal basis with which to challenge the status quo of fisheries management.

The purpose of the CWA is expressed in section 101(a): “The objective of the Act is to restore and maintain the chemical, physical, and biological integrity of the Nation’s waters.”305 Section 301(a) of the CWA states that “[e]xcept as in compliance with this section and sections 302, 306, 307, 318, 402, and 404 of this Act, the discharge of any pollutant by any person shall be unlawful.”306 As noted earlier, the term “pollutant” includes “biological materials.”307 The only exception relevant to fish hatcheries are allowable discharges under the NPDES permit system. The purpose of the NPDES permit system, defined in 40 C.F.R. § 131.2, provides “water quality standards should, wherever attainable, provide water quality for the protection and propagation of fish, shellfish and wildlife and for recreation in and on the water and take into consideration their use and value of public water supplies.

306. Id. § 1311(a).
307. See id. § 1362(6); see also supra notes 178–90 and accompanying text.
Fish as Pollutants

"propagation of fish, shellfish, and wildlife, recreation in and on the water, and agricultural, industrial, and other purposes including navigation . . . ."\textsuperscript{308}

Clearly, the phrase “biological materials,” as it is normally understood, would include fish.\textsuperscript{309} From this, it ought to be inescapable that the discharge of fish from mariculture facilities, fish hatcheries and fish hatchery trucks that are already defined as point sources for the purposes of the CWA should be regulated under the NPDES permit system. Water quality standards that do not address the discharge of live fish as pollutants from NPDES-regulated facilities are incomplete in light of the expressed purpose of both the CWA and NPDES permit system and evolving science. A large number of federal and state fish hatcheries are NPDES-regulated facilities, and yet, do not have provisions in their permits authorizing the unintentional or intentional release of fish into waters of the United States.\textsuperscript{310} To the extent our interpretation of the law is correct, this is equivalent to discharging dioxins or any other type of pollutant from a point source into waterways without a NPDES permit. The CWA and NPDES statements of purpose are perhaps the strongest indication that the discharge of fish should be regulated under the NPDES permit system. It is difficult to conceive of a more radical alteration of the “biological integrity” of a body of water as expressed in the CWA statement of purpose than the introduction of fish. How can an introduction of fish that negatively impacts the native biofauna “take into consideration . . . [the] propagation of fish, shellfish, and wildlife” as expressed in the CWA and NPDES statements of purpose?

Turning more specifically to the USPIRG cases, they are for the most part well-reasoned. Although there is no legislative history to shed any light on the congressional intent in defining the word “pollutant” to include the term “biological materials,” mariculture escapees—being living organisms—fit neatly within the common meaning of that term,\textsuperscript{311} and hence, are “pollutants” within the meaning of the CWA.\textsuperscript{312} However, by considering only non-North American Atlantic salmon to


\textsuperscript{309}. “Biological” is defined as “of, relating to, caused by, or affecting life or living organisms,” \textit{THE AMERICAN HERITAGE COLLEGE DICTIONARY} 139 (3d ed. 1993). “Material” is defined as the “substance or substances out of which a thing is or can be made.” \textit{Id.} at 837.

\textsuperscript{310}. See, e.g., Gary Whelan, supra note 214.

\textsuperscript{311}. \textit{THE AMERICAN HERITAGE COLLEGE DICTIONARY} 139, 837 (3d ed. 1993).

\textsuperscript{312}. 33 U.S.C. § 1362(6).
be “non-native” and by confining its holdings to “non-native” farm-raised Atlantic salmon, the court made two errors.

First, as a factual matter, farm-raised Atlantic salmon of Canadian origin are not native to the Gulf of Maine. As noted in Part III, the National Academy of Science as well as federal fish agencies found subtle, yet important, genetic distinctions between wild Atlantic salmon originating in the Gulf of Maine and those that spawn in Canadian waters. Moreover, even farm-reared native fish differ from their wild cousins because of breeder selection and the novel environment posed by net pens. Second, the court appears to have placed emphasis on the word “pollutant,” in narrowly tailoring its holding, when the operative phrase is “biological materials.” This latter term does not in itself have the negative connotation of the word “pollutant.” Thus, the addition of any living organism, including fish from mariculture operations, fish hatcheries or fish trucks, to the waters of the United States from a point source falls neatly within the ambit that the CWA seeks to regulate.

It is true that in an unpublished order a Washington superior court affirmed a state administrative board order’s that held that the “inadvertent release of Atlantic salmon” at the then existing escape level did not cause “pollution,” thereby supporting the issuance of discharge permits at Atlantic salmon aquaculture facilities. Yet, for a number of reasons that case is of limited value to the question of escapes under federal law. First, the case was decided on the basis of state, rather than federal, law. Moreover, the state law in question defines “pollution” to mean, inter alia, “contamination … or alteration of the physical, chemical or biological properties” of, or the discharge of a “substance” into, any waters of the state “as will or is likely to create a nuisance or render such waters harmful, detrimental or injurious to the public health, safety or welfare, or to domestic, commercial, industrial, agricultural, recreational, or other legitimate beneficial uses, or to livestock, wild animals, birds, fish or other aquatic life.” In other words, to

314. See supra note 277and accompanying text.
317. “Whenever the word ‘pollution’ is used in this chapter, it shall be construed to mean such
Fish as Pollutants

demonstrate that escaped Atlantic salmon are “pollution” under Washington state law, a complainant must establish not only a biological alteration of or discharge of a substance into state waters, but that such alteration or discharge causes a “nuisance” or otherwise renders the waters “harmful, detrimental or injurious” to native salmonoid species. In comparison, under the CWA, the EPA or a citizen plaintiff as appropriate, must only demonstrate the discharge of material that is “biological” in nature. In other words, no showing of harm is required under federal law. Lastly, as an unpublished lower court state order, the finding and conclusions are without value as precedent.

Regardless of the legal merits of a finding that the escape of net raised Atlantic salmon constitutes the discharge of a pollutant from a point source, we are left with the question of whether regulating such escapes is good policy from a social and scientific vantage point. One might reasonably ask: do we really want to think of a few Atlantic salmon escaping from a fish farm as pollutants?

The short answer is yes. There is sufficient evidence that escaped Atlantic salmon negatively affect wild salmon to consider the escapees pollutants. The long answer is that in order to fully appreciate the effect of escaped Atlantic salmon on wild Atlantic salmon, it is important to


319. In Public Utilities District v. Department of Ecology, the Washington Supreme Court stated that the definition of “pollution” under state law “is, if anything, broader than the definition of ‘pollution’ in the Clean Water Act” in the context of a case involving the establishment of minimum instream flows. 146 Wash. 2d 778, 820, 51 P.3d 744, 765 (2002). This statement, however, addressed the fact that Washington explicitly regulates “physical” changes to waters while the CWA does not. The court did not address the issue of harm. The conclusion that in some respects Washington’s discharge permit program is less encompassing than the federal program is a peculiar result given that under the CWA, state programs are supposed to be at least as stringent as the federal program. See supra note 135 and accompanying text.

320. Dahl-Smyth, Inc. v. City of Walla Walla, 148 Wash. 2d 835, 839 n.4, 64 P.3d 15, 17 n.4 (2003) (en banc) (“We do not suggest that this or any other unpublished opinion should be relied on as precedent.”); cf. WASH. R. APP. P. 10.4(h) (prohibiting the citation of unpublished opinions as authority).
consider this question within the context of the life cycle and historical abundance of wild Atlantic salmon including their current legal status, their value as food and sports fish, their cultural significance, the impact of anthropogenic factors on salmon abundance, and the development of Atlantic salmon and other mariculture operations in the United States. Finally, because the decision in the USPIRG cases (and the APHETI case as well) may affect the regulation of other aquaculture ventures and other fish management practices, we might want to consider the history of aquaculture in general in the United States, including food production, the aquarium industry, and trout and other game fish stocking programs, and ask ourselves if it also makes sense to think of other species of fish as pollutants.

As an alternative to present Atlantic salmon farming practices in Maine, we recommend that: (1) Atlantic salmon hatchery operations be relocated to rivers outside of the Gulf of Maine D.P.S. so that maricultured salmon are not biologically imprinted with the same biological markers as endangered wild populations; and (2) offshore mariculture facilities be moved out of state jurisdictional waters and into that portion of the territorial sea that is exclusively federal (three to twelve nautical miles from shore) and the U.S. Exclusive Economic Zone (EEZ) that extends to 200 nautical miles from shore in order to minimize potential ecological harm and prevent water use conflicts. Unfortunately, inasmuch as it is an impediment to the movement of mariculture further from the shore, the current regulatory framework for U.S. mariculture is incompletely developed.

State regulation of mariculture is inconsistent, with such diverse rules as a blanket prohibition of finfish farming in Alaska, a submerged lands lease from the Department of Agriculture and approval by the State Cabinet required in Florida, and aquaculture sitting on a local and county level in Texas and Washington respectively. 321 Federal statutes and their administering agencies with potential regulatory authority over mariculture operations include: the CWA; sections 102 and 304 of the Marine Protection, Research and Sanctuaries Act (MPRSA) (EPA and NOA); 322 section 10 of the Rivers and Harbors Act of 1899 (RHA) (U.S. Army Corps of Engineers); 323 the ESA, (the Services); 324 the Marine

Fish as Pollutants

Mammal Protection Act (MMPA), (FWS and NMFS);\textsuperscript{325} the Coastal Zone Management Act, (NOAA and affected states);\textsuperscript{326} the Magnuson Act (NMFS and Regional Fish Management Councils), the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) (the EPA);\textsuperscript{327} the Federal Food, Drug, and Cosmetic Act, (Food and Drug Administration);\textsuperscript{328} and Outer Continental Shelf Lands Act, (Marine Minerals Service, Department of the Interior).\textsuperscript{329} The scope of future interaction between these statutes, their administering agencies, and mariculture operators is not yet clear. Recommendations, however, by the United States Ocean Commission will be forthcoming in 2003,\textsuperscript{330} on how to structure U.S. Ocean policy institutions and more specific work by an interdisciplinary research team that is presently developing an operational framework—administration, planning, site assessment, leasing, permitting, environmental reviews, monitoring, compliance and enforcement—for offshore aquaculture in federal waters.\textsuperscript{331}

From the mariculture industry’s standpoint, moving into federal waters would have several benefits. To begin with, whatever regulatory framework is eventually adopted for federal waters, it will provide the

\textsuperscript{325} 16 U.S.C. § 1371.
\textsuperscript{326} 16 U.S.C. § 1456.
\textsuperscript{327} 7 U.S.C. § 136(a)–(y).
\textsuperscript{328} 21 U.S.C. §§ 301–397.
\textsuperscript{331} The website of the research project is: http://darc.cms.udel.edu/sgeez (Sept. 16, 2002). A final report is anticipated in late summer or early fall, 2003. An earlier report can be found at http://darc.cms.udel.edu/sgeex1final.pdf (last accessed July 28, 2003). The research team includes one of the co-authors. The life cycle of the Atlantic salmon presents policymakers and fisheries managers with an additional chore—the necessity of managing the species cooperatively across state, national and international jurisdictional boundaries. For example, however the Gulf of Maine D.P.S. might be protected in Maine’s rivers as a result of its endangered species status, its protection under U.S. regulations from commercial or recreational fishing impacts outside the U.S. EEZ is more problematic. For example, the ESA prohibition against takes on the high seas applies only to persons "subject to the jurisdiction of the United States.” 16 U.S.C. § 1538(a)(1) (2000). Likewise, the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) is directed toward prohibiting international trade in wild animals and plants that threatens their survival rather than prohibiting the taking of endangered species. In addition, because there is some gene flow between salmon populations on both sides of the Atlantic, a change in the genome in one part of the salmon’s range might manifest itself in a separate stock some distance from the first-affected population. Available at http://www.cites.org/eng/disc/text.shtml and implementing U.S. law; 16 U.S.C. § 1538(a)(1)(c)–(d).
industry with a more consistent regulatory environment than presently exists as it moves from one state jurisdiction to another, up and down the Atlantic, Pacific, and Gulf Coasts. Moreover, federal waters are unique because state law does not apply to them, and thus, for the most part, mariculturalists operating in federal waters can avoid potentially conflicting and duplicative federal and state regulations. In the future, both the State of Maine and the federal government will have input into the terms and conditions of NPDES permits issued for Atlantic salmon mariculture operations in Maine coastal waters, while only the federal government would have input into NPDES permits granted for operations in federal waters. Moreover, the farther mariculture operations move offshore, the more likely they are to able avoid conflicts with other uses (e.g., the conflicts between wild and cultured salmon stocks and among mariculture, commercial fishing, oil and gas, and preservation), other users, and interested parties. Moving farther offshore also should ameliorate some of mariculture’s ecological impacts. For example, the capacity of the ocean to assimilate nutrient loadings from mariculture operations should be greatly enhanced as one moves from near-shore to the wide expanse of the EEZ. Ultimately, Atlantic salmon mariculture firms and other mariculture operators will have to weigh the benefits of avoiding state regulation against the costs of moving operations to the EEZ, which include higher operating costs and the risk of storm damage.

A consideration of aquaculture also points toward a broader issue and direction—that is, whether the EPA should formulate national policy and

332. Section 401 of the CWA, 33 U.S.C. § 1341, requires a federal permit applicant to receive certification from a state that its discharge is in compliance with state water quality standards. Section 401, however, is not applicable to discharge that occur in federal offshore waters because the jurisdictional reach of the states under Section 401 only extends three miles offshore. See id. § 1362(8); Natural Resources Defense Council v. EPA, 863 F.2d 1420 (9th Cir. 1988). It is true, however, that even if mariculture operations occur in federal waters, some state laws and programs may still have bite. For example, section 307 of the Coastal Zone Management Act (CZMA) requires federal permit applicants to obtain state certification that any permitted action that will affect land uses, water uses or natural resources of a state’s coastal zone are consistent with that state’s coastal zone management plan. 16 U.S.C. § 1456. States also, for example, could require a permit to transport live fish through state jurisdictional waters. See, e.g., ALASKA ADMIN. CODE tit. 5, § 41.005 (1988); CONN. GEN. STAT. ANN. § 26-57 (West 1991). Finally, states may participate in permit and lease development through the environmental evaluation and public participation process mandated by National Environmental Policy Act (NEPA). 42 U.S.C. § 4332 (2000); Natural Res. Def. Council v. U.S. Dep’t of the Navy, CV-01-07781 CAS (RZx) (C.D. Cal. Sept. 17, 2002) (opinion by Christina A. Snyder) (despite the presumption against extraterritoriality, NEPA applies in EEZ).

333. See GOLDBURG ET AL., supra note 1, at 5.
Fish as Pollutants

pass comprehensive rules under the CWA regulating the intentional and unintentional release of all food, sport, and all other categories of live fish from all aquaculture, mariculture, tropical fish farm, and fish hatchery facilities in the country. The EPA has thus far shown a reluctance to consider live fish as pollutants even though they appear to fit within the CWA definition of pollutants and for the most part have been recognized as such by courts.334 This reluctance to associate stocking or releasing fish with pollution is probably shared by most fisheries managers, even those who in effect already treat live fish as pollutants through their stocking policies. To overcome the reluctance to consider fish as CWA pollutants, Congress should consider passing separate legislation regulating the intentional and unintentional introduction of all fish to waters of the United States. We emphasize the word “regulating” because we do not envision legislation that would prohibit all fish stocking and aquaculture operations. Indeed, it may make little sense at this late date to prohibit, for example, the stocking of brown trout in a body of water that has been stocked annually for a century. At the same time, any stocking proposal must be sensitive to the ecosystem into which stock are proposed to be introduced and should not be used as an excuse to avoid mitigating damage to a degraded habitat.

The spread of such parasites as whirling disease, the introduction of exotic fish to aquatic ecosystems, and the escape of cultured Atlantic salmon cannot be regulated adequately through a patchwork of differing state agencies, laws, and policies. The current architects of the vast majority of fish-stocking policy are state fish and game agencies, whose perceived interests are not always in agreement with the thoughtful stewardship of aquatic ecosystems. These agencies are often pressured to provide a given number of fish in given waterways without much thought to the possible ecological effects of their stocking policies, which can include genetic damage to the native populations of fish or the spread of disease and parasites.335

334. The EPA may be forced to show its hand, however, in response to comments on its proposed aquaculture effluent guidelines or in response to its draft Aquatic Nuisance in Ballast Water Policy. U.S. ENVTL. PROT. AGENCY (USEPA), supra note 133.

Whirling disease is a syndrome of spinal deformities and erratic behaviors, such as tail-chasing, caused by the presence of a parasite, *Myxobolus cerebralis*, in trout. Whirling disease can be fatal to juvenile trout. The introduction of whirling disease has potential to decimate wild trout populations. For example, in a series of experiments on the upper Colorado River in Colorado, young brook trout and cutthroat trout exposed to the parasite in the field suffered 85 percent or higher mortality within four months of exposure. A reduction of a similar magnitude, 90 percent, was evidenced on Montana’s Madison River, with the population stabilizing at a level that represents 10 percent of its historical abundance.

A review of the relevant literature by Jerri L. Bartholemew and Paul W. Reno shows that whirling disease has now been documented in 22 of the 50 states. In addition to being widespread the presence of whirling disease has now been confirmed in some of the nation's most important fisheries, including Pyramid Lake in Nevada and the Ausable River in Michigan. In 1998, whirling disease infection was documented in a population of Yellowstone Lake cutthroat trout. Those fish represented the world's largest native cutthroat trout population, residing in the largest intact zone of lake cutthroat trout habitat. In western states, the spread of whirling disease has been particularly notable. Whirling disease is present in all coldwater drainages in Colorado, with the exception of the Animas and North Republican rivers. In Montana, whirling disease has been confirmed in 87 waters in the Beaverhead,

---

337. Id. at 1.
338. Id.
342. Id. at 10–12.
343. Id.
344. YELLOWSTONE NATIONAL PARK NEWS RELEASE, Anglers Need to Catch Lake Trout in Yellowstone Park (Sept. 5, 2002), available at http://www.nps.gov/yell/press/02106.htm. The subject of this press release, the threat posed to Yellowstone Lake Cutthroat trout by introduced Lake Trout, is itself illustrative of the problems posed to native populations of fish by introduced species.
345. See Bartholemew, supra note 341, at 10–12.
Fish as Pollutants

Gallatin, Madison, Bighole, Bitterroot, Blackfoot, Jefferson, Swan, Clark Fork, Missouri, Sun, Flathead, and Yellowstone river drainages. 346 Whirling disease is not native to North America. 347 Whirling disease arrived in North America through the importation of frozen rainbow trout; it was first documented in Pennsylvania from where it rapidly spread to other states. 348 Vectors for the spread of M. cerebralis include import of fish and fish products as well as oligochaete worms imported as food for ornamental fish. 349 The shipment of fish from one location to another is the most likely vector for the spread of whirling disease in the United States. 350 Fish hatcheries are perhaps the most important vector for the spread of whirling disease.

The private and public hatchery system throughout the U.S. is an important component of the shipment of fish from one location to another. Of the 22 states in which whirling disease has been documented, 20 have fish culture facilities which are either currently positive for whirling disease or which have been historically positive. 351 In 1998, whirling disease was found in 12 of 15 of Colorado’s trout hatcheries and, through unwitting introduction to previously uninfected waters, had killed off 90 percent of the trout in six of Colorado’s most important trout rivers. 352 Whirling disease is thought to have arrived in Colorado through a shipment of diseased fish from a private hatchery in Idaho. 353 In some cases, the detection of whirling-disease positive fish in fish hatcheries has led to extreme management measures. In Utah, the sacrifice of all potentially affected fish followed the discovery of whirling-disease in a state hatchery. 354 The response of the Utah Division of Wildlife Resources to the presence of whirling disease in a state fish hatchery is one example of de facto treatment of live fish as pollutants by fisheries managers.

346. Id. at 10–12.
347. Id. at 8.
348. Id. at 13.
351. See Bartholomew, supra note 341, at 10–12.
353. Id.
354. See Bartholomew, supra note 341, at 19.
The transmission of whirling disease from hatchery fish to wild stocks suggests that the ecological impacts that stem from the introduction of non-wild fish populations neither are limited to salmon net pen operations nor the accidental release of fish from aquaculture operations, but rather, transcend the line between accidental and intentional fish releases.\textsuperscript{355}

\textsuperscript{355} The issue of disease transmission (in this case, infectious salmon anemia (ISA), infectious hepatopoietic necrosis (IHN), and infectious pancreatic necrosis (IPN) rather than the whirling disease) from cultured stocks of Atlantic salmon, whether they are native or non-native, to the endangered Gulf of Maine Atlantic salmon DPS is an additional cause of concern.
**MIRANDA’S POISONED FRUIT TREE:**
THE ADMISSION OF PHYSICAL EVIDENCE DERIVED FROM AN UNWARNED STATEMENT

Kirsten Lela Ambach

Abstract: *Miranda v. Arizona* created an exclusionary rule that prohibits using, as part of the prosecution’s case in chief, evidence that is obtained as the result of unwarned custodial interrogation. In *Michigan v. Tucker* and *Oregon v. Elstad*, the United States Supreme Court narrowed the scope of this rule in relation to the “fruit of the poisonous tree” doctrine that excludes all evidence derived from constitutional violations. The *Tucker* Court held that the testimony of a witness identified from an unwarned statement should be admitted, and the *Elstad* Court held that a warned statement following an unwarned statement should also be admitted. In both cases, the primary rationale for the exceptions was the prophylactic status of the *Miranda* rule. This status distinguished *Miranda* violations from constitutional violations to which the fruits doctrine was applicable. In 2000, the U.S. Supreme Court reversed its earlier characterization of *Miranda* as a prophylactic rule, and instead reaffirmed the rule’s constitutional status in *Dickerson v. United States*. After *Dickerson*, a circuit split developed regarding the admissibility of physical evidence derived from unwarned statements excluded at trial. This Comment argues that the U.S. Supreme Court should hold that both the unwarned statement and its derivative physical evidence should be excluded from the prosecution’s case in chief. Fifth Amendment precedent mandates applying the fruits doctrine to physical evidence derived from an unwarned statement in order to effectively deter violations of *Miranda* and to ensure the trustworthiness of evidence obtained in the interrogation setting.

In 1966, *Miranda v. Arizona* dramatically changed the face of custodial interrogation. The United States Supreme Court held that incommunicado custodial surroundings and contemporary interrogation techniques are inherently compelling and in violation of the Fifth Amendment. See *Joshua Dressler, Understanding Criminal Procedure* § 22.01 (3d ed. 2002). “Incommunicado” interrogations are done in a setting that isolates the suspect from anyone and anything familiar in order to deprive the suspect of every psychological advantage and bestow such advantages on the interrogators. See *Miranda*, 384 U.S. at 449–50.

1. See *Joshua Dressler, Understanding Criminal Procedure* § 22.01 (3d ed. 2002).
2. “Incommunicado” interrogations are done in a setting that isolates the suspect from anyone and anything familiar in order to deprive the suspect of every psychological advantage and bestow such advantages on the interrogators. See *Miranda*, 384 U.S. at 449–50.
3. See *id.* at 467. By the 1960s, police conducting interrogations had largely adopted and used psychological tactics rather than physical abuse. See *id.* at 448–49. The *Miranda* Court focused on the psychological methods described in two criminal interrogation manuals. *Id.* at 449–54 (citing *Inbau & Reid, Criminal Interrogation and Confessions* (1962); *O’Hara, Fundamentals of Criminal Investigation* (1956)). Psychological pressures were more difficult to prove than physical abuse when the voluntariness of the defendant’s confession was in question. The Court concluded that police were exploiting the incommunicado nature of custodial interrogation in order to destabilize the suspect by focusing on his insecurities about himself or his surroundings. Once the suspect was off balance, the police used patience, persistence, and trickery to inhibit the defendant...
Amendment.\textsuperscript{5} This inherent compulsion can be dispelled only by the police explicitly informing the suspect of his or her rights,\textsuperscript{6} known as the \textit{Miranda} warnings,\textsuperscript{7} and the suspect’s voluntary and knowing waiver of those rights.\textsuperscript{8} The \textit{Miranda} decision established the broad, presumptive exclusionary rule\textsuperscript{9} that “no evidence obtained as a result of interrogation can be used against” the defendant unless the prosecution demonstrates at trial that the warnings were given and the defendant’s waiver was voluntary and knowing.\textsuperscript{10} However, the Court did not clarify whether “no evidence” was limited to the defendant’s statement, or extended to evidence derived from that statement as well. The Court has since carved out three exceptions to \textit{Miranda}’s sweep.\textsuperscript{11}

A separate but related doctrine in the context of exclusion of evidence is the “fruit of the poisonous tree”\textsuperscript{12} doctrine (fruits doctrine). The fruits doctrine generally forbids the prosecution from using either from exercising his constitutional rights. \textit{See id.} at 455.

\begin{itemize}
\item \textsuperscript{5} \textit{See id.} at 467.
\item \textsuperscript{6} \textit{See id.} at 444–45.
\item \textsuperscript{7} An in-custody suspect:
\begin{quote}
must be warned prior to any questioning that he has the right to remain silent, that anything he says can be used against him in a court of law, that he has the right to the presence of an attorney, and that if he cannot afford an attorney one will be appointed for him prior to any questioning if he so desires.
\end{quote}
\textit{Id.} at 479.
\item \textsuperscript{8} \textit{See id.} at 444–45.
\item \textsuperscript{9} \textit{See id.} at 476.
\item \textsuperscript{10} \textit{See id.} at 479.
\item \textsuperscript{11} \textit{See Oregon v. Elstad, 470 U.S. 298, 314 (1985)} (creating the cured statement exception); \textit{New York v. Quarles, 467 U.S. 649, 655 (1984)} (creating the public safety exception); \textit{Michigan v. Tucker, 417 U.S. 433, 447–48 (1974)} (creating the limited retroactivity exception). The cured statement exception requires courts to exclude an unwarned statement but admit a subsequent statement if \textit{Miranda} warnings were administered and the suspect knowingly waived those rights before the later statement was made. \textit{See Elstad, 470 U.S.} at 314. Under the public safety exception, courts may admit both the defendant’s statement and any evidence derived from that statement when \textit{Miranda} is violated for public safety concerns. \textit{See Quarles, 467 U.S.} at 655. Finally, in the limited retroactivity exception, courts must exclude an unwarned statement but may admit the testimony of a witness identified through that statement if the statement was obtained prior to the \textit{Miranda} decision and the trial occurred afterwards. \textit{See Tucker, 417 U.S.} at 447–48. The Court has also established an impeachment exception whereby an otherwise inadmissible unwarned statement is admissible to attack the credibility of the defendant’s trial testimony. \textit{See Harris v. New York, 401 U.S. 222, 226 (1971)}. However, this exception is beyond the scope of this Comment because it still does not allow the evidence into the prosecution’s case in chief. Another possible exception to \textit{Miranda} is the independent source doctrine. So far, however, this doctrine has only been applied within the Fifth Amendment context to immunity, not interrogation, and therefore it is also beyond the scope of this Comment. \textit{See Kastigar v. United States, 406 U.S. 441, 460 (1972)}.
\item \textsuperscript{12} \textit{See Nardone v. United States, 308 U.S. 338, 341 (1939)}.
\end{itemize}
unconstitutionally obtained primary evidence or its derivative evidence in its case in chief. Primary evidence is the direct product of police action violative of the constitution, such as an illegally seized bag or an unwarned statement. Derivative evidence is the indirect product of police action. The police obtain or become aware of derivative evidence through primary evidence. Examples of derivative evidence include testimony by a non-party witness whose identity was revealed in the unwarned statement, a defendant’s subsequent statement, or drugs located through an unwarned statement. Generally, an exclusionary rule presumptively applies to the primary evidence. Unless the prosecution proves by a preponderance of the evidence that an exception to the fruits doctrine applies, the derivative evidence is excluded.

The U.S. Supreme Court first extended the general exclusionary rule to derivative evidence, thereby creating the fruits doctrine in the Fourth Amendment context. In Wong Sun v. United States, the Court held

14. See Wong Sun, 371 U.S. at 484.
15. See id.
16. See id.
17. See United States v. Faulkingham, 295 F.3d 85, 91 (1st Cir. 2002).
18. See Miranda v. Arizona, 384 U.S. 436, 476 (1966); Wong Sun, 371 U.S. at 484. The general exclusionary rule for a Fifth Amendment violation is not merely a remedy for a constitutional right; it is inherent in the right itself. See DRESSLER, supra note 2, § 23.02[E][1]. The Fifth Amendment provides that no person “shall be compelled in any criminal case to be a witness against himself.” U.S. CONST. amend. V. In other words, if the evidence is in any way linked to self-incrimination, it is inadmissible in the prosecution’s case in chief. When Miranda found custodial interrogation inherently compulsive, it extended the Fifth Amendment’s exclusionary rule to custodial, unwarned statements. See Miranda, 384 U.S. at 458.
19. See Nix v. Williams, 467 U.S. 431, 443 (1984); see also Wong Sun, 371 U.S. at 486.
20. See Nix, 467 U.S. at 444.
21. See Silverthorne Lumber Co. v. United States, 251 U.S. 385, 392 (1920) (finding that “[t]he essence of a provision forbidding the acquisition of evidence in a certain way is that not merely evidence so acquired shall not be used before the Court but that it shall not be used at all”) (emphasis added).
that evidence “come at by exploitation”\textsuperscript{23} of an unconstitutional act (“the poisonous tree”\textsuperscript{24}) is inadmissible derivative evidence (“the tainted fruit”\textsuperscript{25}) of the tree). The fruits doctrine applies unless the connection between the Fourth Amendment violation and the derivative evidence is too attenuated.\textsuperscript{27} The passage of time and a break in events are two factors relevant to determining whether the fruits doctrine applies in the Fourth Amendment.\textsuperscript{28} While the fruits doctrine was first developed under the Fourth Amendment context,\textsuperscript{29} the U.S. Supreme Court has also applied it in the context of the Fifth\textsuperscript{30} and Sixth Amendments.\textsuperscript{31} For example, in the Fifth Amendment context, the Court has held that courts can compel a state witness to offer incriminating testimony only if both the compelled testimony and the evidence derived from that testimony are inadmissible in a criminal prosecution against the witness.\textsuperscript{32}

In \textit{Dickerson v. United States},\textsuperscript{33} the U.S. Supreme Court clarified the status of \textit{Miranda}’s rule, which requires the administration of warnings and an effective waiver prior to admitting the statement in the prosecution’s case in chief, as a constitutional rule under the Fifth Amendment’s privilege against self-incrimination.\textsuperscript{34} The \textit{Dickerson} Court held that \textit{Miranda}’s strictures on the admissibility of statements made during custodial interrogation are constitutional, not prophylactic.\textsuperscript{35} Notably, the exceptions to \textit{Miranda} survived \textit{Dickerson},\textsuperscript{36} despite their primary reliance on \textit{Miranda}’s former prophylactic status.\textsuperscript{37} The Court distinguished them on the ground that

\begin{itemize}
\item \textsuperscript{23} See \textit{id.} at 488 (internal quotes omitted).
\item \textsuperscript{24} See \textit{Nardone v. United States}, 308 U.S. 338, 341 (1939).
\item \textsuperscript{25} See \textit{Nix}, 467 U.S. at 441.
\item \textsuperscript{26} See \textit{Wong Sun}, 371 U.S. at 488.
\item \textsuperscript{27} See \textit{Nardone}, 308 U.S. at 341.
\item \textsuperscript{28} See \textit{id.}
\item \textsuperscript{29} See \textit{Silverthorne Lumber Co. v. United States}, 251 U.S. 385, 392 (1920).
\item \textsuperscript{30} See \textit{Murphy v. Waterfront Comm’n}, 378 U.S. 52, 79 (1964).
\item \textsuperscript{31} See \textit{United States v. Wade}, 388 U.S. 218, 235–36 (1967).
\item \textsuperscript{32} See \textit{Murphy}, 378 U.S. at 79.
\item \textsuperscript{33} See \textit{United States v. Dickerson}, 530 U.S. 428 (2000).
\item \textsuperscript{34} See \textit{id.} at 437–38.
\item \textsuperscript{35} See \textit{id.} at 444. A prophylactic rule is formulated to prevent something. \textsc{Black’s Law Dictionary} 1234 (7th ed. 1999).
\item \textsuperscript{36} See \textit{Dickerson}, 530 U.S. at 441.
\end{itemize}
"no constitutional rule is immutable." The Court recognized that constitutional rules can be modified to have exceptions because courts are unable to foresee all applications of a rule at the time of the original decision.

Since the Dickerson decision, a circuit split has developed regarding the admissibility of physical evidence derived from unwarned statements excluded at trial. The crux of the split lies in the circuit courts’ interpretations of the impact of Miranda’s warnings and waiver as a constitutional right, and not merely a prophylactic safeguard, on two of Miranda’s progeny involving derivative evidence. Two circuits have held that the fruits doctrine does not apply to derivative physical evidence. Conversely, two other circuits have held that the fruits doctrine does apply to Miranda violations. The U.S. Supreme Court recently granted certiorari to United States v. Patane, and therefore this issue will soon be resolved.

This Comment argues that the U.S. Supreme Court should hold that, absent an exception, both the unwarned statement and its derivative physical evidence should be excluded when Miranda rights are violated. Part I explores the scope of Miranda’s exclusionary rule. Part II discusses the impact of Dickerson’s reaffirmation of Miranda’s constitutional status on the scope of Miranda’s exclusionary rule. Part III reviews the post-Dickerson lower federal court split on the application of the fruits doctrine to derivative physical evidence. Part IV argues that the U.S. Supreme Court should hold that the fruits doctrine

38. See Dickerson, 530 U.S. at 441.
39. See id.
41. See Patane, 304 F.3d at 1025; Faulkingham, 295 F.3d at 93; Sterling, 283 F.3d at 219; DeSumma, 272 F.3d at 180.
42. See Sterling, 283 F.3d at 219; DeSumma, 272 F.3d at 180.
43. See Patane, 304 F.3d at 1029; Faulkingham, 295 F.3d at 90.
45. This Comment is limited in scope to the admissibility of physical fruits of unwarned statements. The current state of other forms of derivative evidence, namely witness identification and defendants’ statements, is beyond this Comment’s scope—except as they pertain to an established exception to Miranda’s exclusionary rule.
should be uniformly applied to physical evidence derived from unwarned statements.

I. MIRANDA’S EXCLUSIONARY RULE

In *Miranda v. Arizona*, the U.S. Supreme Court created a default exclusionary rule: courts cannot admit evidence obtained prior to the suspect being given *Miranda* warnings and making a voluntary and knowing waiver of those rights to prove the prosecution’s case in chief at trial. The *Miranda* Court did not address the applicability of the fruits doctrine, which mandates exclusion of all evidence derived from a constitutional violation, absent an established exception. Yet in subsequent cases, the Court restricted the scope of *Miranda*’s exclusionary rule by carving out three exceptions to its presumption of inadmissibility. First, the public safety exception allows the prosecution to use both the unwarned statement and its fruits in the prosecution’s case in chief. The second and third exceptions are derivative evidence exceptions, requiring courts to exclude the unwarned statement but admit its fruits. The Court has not explicitly ruled on whether an exception exists for physical evidence derived from an unwarned statement excluded at trial, although some Justices have expressed opinions on the issue.

---


50. See *Elstad*, 470 U.S. at 314; *Quarles*, 467 U.S. at 655; *Tucker*, 417 U.S. at 448. This Comment is limited to scope to exceptions to the exclusion of evidence in the prosecution’s case in chief. See supra note 11.

51. See *Quarles*, 467 U.S. at 655.

52. See *Elstad*, 470 U.S. at 318; *Tucker*, 417 U.S. at 448.


54. See, e.g., *Quarles*, 467 U.S. at 666 (O’Connor J., concurring in part and dissenting in part); *id.* at 688 (Marshall, J., dissenting).
Miranda’s Poisoned Fruit Tree

A. The Scope of Miranda’s Exclusionary Rule

_Miranda v. Arizona_ governs the admissibility of evidence obtained through custodial interrogation.\(^{55}\) _Miranda_ was one of four consolidated cases in which the defendants had provided the police with incriminating statements while in custody and subject to interrogation.\(^{56}\) While the cases were from different jurisdictions, they had four significant facts in common: (1) each suspect was in custody; (2) each suspect was questioned in the police-dominated atmosphere of an interrogation room; (3) each suspect was run through police interrogation procedures; and (4) none of the suspects was informed of the privilege against self-incrimination.\(^{57}\) The question before the Court was whether custodial interrogation rendered the defendants’ statements inadmissible as violations of the Fifth Amendment.\(^{58}\) Prior to the _Miranda_ decision, the admissibility of a confession was determined by a voluntariness test inquiring whether a defendant’s will was overborne.\(^{59}\) This voluntariness test was derived from two constitutional bases: the Fifth Amendment privilege against self-incrimination\(^{60}\) and the Fourteenth Amendment’s Due Process Clause.\(^{61}\) After reviewing police practices and interrogation manuals, however, the _Miranda_ Court concluded that there was an inherent compulsion about custodial interrogation that the voluntariness test could not safeguard.\(^{62}\)

The Court determined that custodial surroundings and interrogation procedures blurred the line between voluntary and involuntary statements, thus heightening the risk that an individual’s Fifth Amendment privilege against self-incrimination would not be protected.\(^{63}\) To alleviate this risk, the Court created a concrete set of

---


\(^{56}\) See _Miranda_, 384 U.S. at 456–57.

\(^{57}\) See id. at 445, 456–57; DRESSLER, supra note 2, § 24.04[A].

\(^{58}\) See _Miranda_, 384 U.S. at 445.

\(^{59}\) See Malloy v. Hogan, 378 U.S. 1, 7 (1964).

\(^{60}\) U.S. CONST. amend. V.

\(^{61}\) U.S. CONST. amend. XIV, § 1. See Malloy, 378 U.S. at 6–7; Brown v. Mississippi, 297 U.S. 278, 286 (1936) (holding that state courts must suppress involuntary confessions under the Fourteenth Amendment’s Due Process Clause); Bram v. United States, 168 U.S. 532, 542–43 (1897) (holding that federal courts must suppress compelled confessions under the Fifth Amendment).

\(^{62}\) See _Miranda_, 384 U.S. at 457–58.

procedural safeguards for law enforcement agencies and courts to follow and grounded the admissibility of any statement made during custodial interrogation upon the administration of the four Miranda warnings and waiver of those rights. The prosecution has the burden of demonstrating that the Miranda warnings were administered and that the defendant knowingly and intelligently waived those rights before the court can admit the defendant’s statement. Thus, the Miranda Court announced a broad exclusionary rule: “unless and until [the Miranda] warnings and waiver are demonstrated by the prosecution at trial, no evidence obtained as a result of interrogation can be used against [the defendant].”

The Court directed that “the prosecution may not use statements . . . stemming from custodial interrogation of the defendant unless it demonstrates the use of procedural safeguards effective to secure the privilege against self-incrimination.” The Court did not address whether this use was limited to the defendant’s statements, or whether it was more broadly meant to include the fruits doctrine, thus prohibiting the prosecution from using derivative evidence obtained from the unwarned statement as well. The fruits doctrine arises in the Miranda context when an unwarned statement leads to: (1) the identification of a witness whom the prosecution seeks to use in its case in chief, (2) self-incriminating statements from the defendant, or (3) information regarding the location of physical evidence. Thus far, the

64. See Miranda, 384 U.S. at 442.
65. See id. at 444–45. The four Miranda warnings require that an in-custody suspect “be warned prior to any questioning”: (1) “that he has the right to remain silent,” (2) “that anything he says can be used against him in a court of law,” (3) “that he has the right to the presence of an attorney,” and (4) “that if he cannot afford an attorney one will be appointed for him prior to any questioning if he so desires.” Id. at 479.
66. See id. at 444, 479.
67. See id. at 479 (emphasis added).
68. Id. at 444 (emphasis added).
70. See id. at 436–37.
Miranda’s Poisoned Fruit Tree

U.S. Supreme Court has neither explicitly adopted nor rejected the fruits doctrine in the Miranda context.73

B. Exceptions to Miranda’s Exclusionary Rule

The U.S. Supreme Court has carved out three exceptions to Miranda’s exclusionary rule.74 First, courts can admit an unwarned statement and its fruits when the defendant was providing information regarding pressing public safety concerns.75 In New York v. Quarles,76 the U.S. Supreme Court determined that an unwarned statement and its fruits are admissible when public safety would be immediately jeopardized by any delay in obtaining the evidence.77 The Court held that an unwarned statement revealing the location of a gun and the gun itself should be admitted, even without Miranda warnings, because the defendant had discarded the gun in a populated supermarket.78 The Court characterized the situation as one in which “the need for answers to questions . . . outweighs the need for the prophylactic rule” protecting the defendant’s Fifth Amendment rights.79 Therefore, where public safety is at issue, the Miranda rule does not apply and both the primary and derivative evidence are admissible.80

The second exception to Miranda’s exclusionary rule stems from Michigan v. Tucker.81 In Tucker, the police gave the defendant the warnings constitutionally required before Miranda,82 but at the time, the

73. See Patterson v. United States, 485 U.S. 922, 922–23 (1988) (White, J., dissenting from denial of certiorari); Patane, 304 F.3d at 1023; Faulkingham, 295 F.3d at 90–91.
74. See Elstad, 470 U.S. at 314; New York v. Quarles, 467 U.S. 649, 655 (1984); Tucker, 417 U.S. at 447–48. These are exceptions to the exclusion of evidence from the prosecution’s case in chief. For other exceptions outside the scope of this Comment, see supra note 11.
75. See Quarles, 467 U.S. at 657.
77. See id. at 655.
78. See id. at 652.
79. See id. at 657.
80. See id. at 659–60.
82. See id. at 447 (finding that “at the time respondent was questioned these police officers were guided, quite rightly, by the principles established in Escobedo v. Illinois”) (citation omitted). The Escobedo decision set the stage for Miranda by holding that the custodial interrogation of a suspect without honoring the suspect’s request for an attorney and not effectively warning him of his right to remain silent or his right to consult with his attorney renders any statement elicited from him inadmissible under the Sixth Amendment. Escobedo v. Illinois, 378 U.S. 478, 490–91 (1964).
warnings did not include the right to free counsel. After waiving his rights, the defendant identified Robert Henderson as an alibi witness. Mr. Henderson, however, discredited the defendant’s statement, and the prosecution used him in its case in chief. After the interrogation, but before the trial date, the U.S. Supreme Court announced the Miranda rule, which included the indigent defendant’s right to state-provided counsel. At trial, the judge retroactively applied Miranda and excluded the defendant’s statement. The U.S. Supreme Court, however, held that the derivative witness identification evidence should be admitted.

Under Tucker, the Court created a limited retroactivity exception for witnesses identified by a defendant’s statement made before Miranda was decided, but after the defendant received the warnings required at the time. The Tucker Court rationalized admitting the derivative evidence by characterizing Miranda as a prophylactic rule. The Court reasoned that the Miranda warnings were not a constitutionally protected right. Instead, the warnings were prophylactic measures designed to protect the privilege against self-incrimination, and the failure to give them did not abridge the Fifth Amendment. The Tucker Court also determined that the traditional fruits doctrine does not apply when the police do not actually infringe upon the defendant’s constitutional rights. The Court distinguished Tucker from Wong Sun, where the fruits doctrine applied because a constitutional violation had occurred. Consequently, the traditional fruits doctrine requiring

83. See Tucker, 417 U.S. at 436.
84. Id.
85. Id. at 436–37.
86. See id. at 437.
88. See Tucker, 417 U.S. at 447–48; see also Johnson v. New Jersey, 384 U.S. 719, 733 (1966) (holding that Miranda would not be applied retroactively to cases in which the trial had begun before the decision was announced, but that it would apply retroactively to cases in which the trial began after the date of decision).
90. See id. at 448.
91. See id. at 446.
92. See id. at 444.
93. See id. at 446.
94. See id.
96. See Tucker, 417 U.S. at 445–46.
exclusion of derivative evidence was not controlling for prophylactic Miranda violations.97

Finally, the U.S. Supreme Court created an additional derivative evidence exception to Miranda’s exclusionary rule in Oregon v. Elstad.98 In Elstad, the defendant implicated himself in an answer to a police question after being arrested, but before receiving his Miranda warnings.99 An hour later at the stationhouse, the defendant received the Miranda warnings, waived them, and gave the police a fully incriminating statement.100 The Court decided that a suspect who responds to unwarned but non-coercive questioning does not thereby lose the ability to waive his rights and confess after being given the requisite Miranda warnings.101 Thus, the Court held that the defendant’s warned statement, although derived from an earlier unwarned statement, should be admitted.102 Unlike a Fourth Amendment violation,103 sufficient passage of time or a break in events is not required to cure the second, derivative statement.104 The administration of the Miranda warnings and the defendant’s waiver of those rights were sufficient to validate the derivative statement.105

The Elstad Court identified the factors necessary to cure a warned confession derived from an earlier unwarned statement.106 The Court held that “a careful and thorough administration of Miranda warnings serves to cure the condition that rendered the unwarned statement inadmissible.”107 Comparatively, a Fourth Amendment violation triggers the fruits doctrine and taints all derivative evidence until a sufficient break in events neutralizes it.108 Yet, the Court concluded that this

97. See id. at 448.
99. See id. at 301.
100. See id.
101. See id. at 318.
102. See id.
103. See Nardone v. United States, 308 U.S. 338, 341 (1939) (holding that derivative evidence of a Fourth Amendment violation may be admitted if it has become so attenuated from the original violation as to sufficiently dissipate the taint).
104. See Elstad, 470 U.S. at 306.
105. See id. at 318.
106. See id. at 309–10.
107. See id. at 310–11.
108. See Nardone, 308 U.S. at 341.
“break in events” test is unnecessary to cure a *Miranda* violation\(^\text{109}\) because in any custodial interrogation setting, arming the defendant with his rights dispels the inherent compulsion.\(^\text{110}\) If the defendant subsequently knowingly and voluntarily waives his or her rights, courts must admit the statement under this cured statement exception.\(^\text{111}\)

The *Elstad* Court reaffirmed the prophylactic rationale behind the *Miranda* warnings and determined that *Miranda* violations warranted less severe consequences than constitutional violations.\(^\text{112}\) The Court’s leniency toward *Miranda* violations relied in part on the differences between the Fourth and Fifth Amendments.\(^\text{113}\) Although the Court has not clearly articulated these differences, it has noted that “[w]here a Fourth Amendment violation ‘taints’ the confession, a finding of voluntariness for the purposes of the Fifth Amendment is merely a threshold requirement in determining whether the confession may be admitted in evidence.”\(^\text{114}\) *Miranda* created the additional requirement that an otherwise voluntary confession may still be inadmissible if *Miranda* warnings were not administered.\(^\text{115}\) Consequently, the *Elstad* Court determined that *Miranda* provides a remedy for the defendant who has not suffered a constitutional harm.\(^\text{116}\) In this respect, “[t]he *Miranda* exclusionary rule . . . serves the Fifth Amendment and sweeps more broadly than the Fifth Amendment itself.”\(^\text{117}\)

Although the failure to administer *Miranda* warnings constitutes a violation of the defendant’s Fifth Amendment right against self-incrimination,\(^\text{118}\) the U.S. Supreme Court has refused to apply the traditional fruits doctrine to *Miranda* violations.\(^\text{119}\) Having characterized *Miranda* warnings as prophylactic, not constitutional,\(^\text{120}\) the Court reasoned that the fruits doctrine, a constitutional violation remedy, did

---

111. See *Elstad*, 470 U.S. at 318.
112. See id. at 307–08.
115. See id. at 307.
116. See id.
117. Id. at 306.
119. See *Dickerson*, 530 U.S. at 441; *Elstad*, 470 U.S. at 308.
Miranda’s Poisoned Fruit Tree

not apply. Instead, the Court carved out three exceptions to Miranda’s exclusionary rule, enabling the prosecution to use derivative evidence previously obtained illegally in its case in chief.

C. The Undecided Question of Derivative Physical Evidence

The U.S. Supreme Court has yet to address the applicability of the fruits doctrine to physical evidence derived from unwarned statements excluded at trial. Although the Court determined that derivative physical evidence should be admitted in New York v. Quarles, the evidence at issue in Quarles involved a public safety concern and the defendant’s statement was admitted as well. The Court has not determined whether courts should admit derivative physical evidence of an excluded unwarned statement. But, several Justices have expressed views on the subject. In Quarles, four of the Justices opined on the admissibility of derivative physical evidence in the Miranda context after rejecting the majority’s “public safety” exception.

In his dissenting opinion in Quarles, Justice Marshall, with whom Justices Brennan and Stevens joined, determined that both the unwarned statement and the derivative physical evidence must be presumed inadmissible. Justice Marshall recognized the U.S. Supreme Court’s
prior extension of the fruits doctrine to the Fifth Amendment context and reasoned that *Miranda* was in fact a constitutional rule. Therefore, he suggested that precedent required suppressing physical evidence discovered as a direct result of an unwarned statement.

Justice O’Connor, on the other hand, concurred in admitting the derivative physical evidence on the ground that *Miranda*’s exclusionary rule was limited to testimonial evidence. She reasoned that the Fifth Amendment only prohibits compelling a suspect to disclose evidence of a testimonial or communicative nature. Nontestimonial or physical evidence is outside the scope of the Fifth Amendment. Justice O’Connor based this conclusion on a series of cases admitting evidence obtained from the defendant’s body. Specifically, the Court has upheld police action compelling defendants to provide blood samples, voice exemplars, handwriting samples, and speak during a lineup, as not violating the Fifth Amendment. The Court reasoned that procuring evidence from the defendant’s body did not in any way compel self-incrimination. In these cases, the Court limited testimonial evidence invoking the Fifth Amendment to an accused’s communication that explicitly or implicitly relates to a factual assertion or discloses information. Consequently, Justice O’Connor concluded that in light of the nontestimonial evidence precedent, the non-

reconsideration in light of *Nix*. See *Quarles*, 467 U.S. at 689–90 (Marshall, J., dissenting).


132. See *Quarles*, 467 U.S. at 683 (Marshall, J., dissenting).

133. See *id.* at 688–89 (Marshall, J., dissenting).

134. See *id.* at 660 (O’Connor, J., concurring in part and dissenting in part).

135. See *id.* at 666 (O’Connor, J., concurring in part and dissenting in part).

136. See *id.*


142. See *Schmerber*, 384 U.S. at 764.

Miranda’s Poisoned Fruit Tree

communicative nature of the derivative physical evidence itself foreclosed excluding such evidence on Fifth Amendment grounds. 144 Ultimately, the majority in Miranda did not draw such distinctions; instead, it created a broad exclusionary rule that evidence obtained from a suspect without the full administration of the Miranda warnings and a knowing and voluntary waiver should not be admitted in the prosecution’s case in chief. 145 The U.S. Supreme Court has limited the application of Miranda’s exclusionary rule in three situations. 146 In carving out these exceptions, the Court characterized the Miranda warnings as a prophylactic safeguard, as distinguished from a constitutional rule. 147 The Court has not ruled on the admissibility of derivative physical evidence when the unwarned statement has been excluded. 148

II. DICKERSON V. UNITED STATES: MIRANDA IS A CONSTITUTIONAL RULE

The Dickerson decision represents the U.S. Supreme Court’s break from characterizing Miranda’s requirements as prophylactic. 149 In Tucker, Quarles, and Elstad, the Court characterized Miranda as a prophylactic rule and used that characterization as a rationale for carving out the three exceptions. 150 In Dickerson, however, the Court held that Miranda was a constitutional rule. 151 The issue before the Court was whether Congress had the constitutional authority to supersede Miranda when it enacted 18 U.S.C. §§ 3501. 152 Under §§ 3501, admissibility of

146. See Oregon v. Elstad, 470 U.S. 298, 314 (1985); Quarles, 467 U.S. at 655; Michigan v. Tucker, 417 U.S. 433, 447–48 (1974). These are exceptions to the exclusion of evidence in the prosecution’s case in chief. For other exceptions outside the scope of this Comment, see supra note 11.
147. See Elstad, 470 U.S. at 307–08; Quarles, 467 U.S. at 653; Tucker, 417 U.S. at 446.
150. See Elstad, 470 U.S. at 308–09; Quarles, 467 U.S. at 657; Tucker, 417 U.S. at 445–46.
151. See Dickerson, 530 U.S. at 444.
statements made during custodial interrogation was based solely on their voluntariness.\textsuperscript{153} Thus, § 3501 represented a return to the pre-
\textit{Miranda} voluntariness test, rejecting \textit{Miranda} per se.\textsuperscript{154} If \textit{Miranda} warnings were a prophylactic safeguard, then Congress had the constitutional authority to enact § 3501.\textsuperscript{155} If \textit{Miranda} warnings were required under the Constitution, on the other hand, then Congress had overstepped its authority.\textsuperscript{156}

This issue surfaced when the United States Court of Appeals for the Fourth Circuit determined the admissibility of an unwarned custodial statement under § 3501.\textsuperscript{157} The Fourth Circuit held that an unwarned but voluntary confession was admissible under § 3501.\textsuperscript{158} In \textit{Dickerson}, the police arrested and interrogated the defendant in connection with a bank robbery without administering \textit{Miranda} warnings.\textsuperscript{159} At trial, the defendant argued that his incriminating statements should be suppressed under \textit{Miranda}.\textsuperscript{160} The trial court granted his motion to suppress.\textsuperscript{161} On appeal, the Fourth Circuit ruled that while the defendant had not received \textit{Miranda} warnings, the confession was voluntary, thus meeting § 3501’s admissibility requirements.\textsuperscript{162}

The U.S. Supreme Court reversed and held that \textit{Miranda}’s warning-based rule governing the admissibility of statements made during custodial interrogation was a constitutional decision, which a legislative act could not overrule.\textsuperscript{163} The Court returned to its rationale for the warnings in \textit{Miranda}: “the coercion inherent in custodial interrogation blurs the line between voluntary and involuntary statements, and thus heightens the risk that an individual will not be ‘accorded his privilege under the Fifth Amendment.’”\textsuperscript{164} The Court reasoned that the warnings were constitutionally required to protect a suspect’s Fifth Amendment

\begin{footnotesize}
\begin{footnotes}{153.}\textit{See Dickerson}, 530 U.S. at 435–36.\footnotesend
\begin{footnotes}{154.}\textit{See id. at 436.}\footnotesend
\begin{footnotes}{155.}\textit{See id. at 437.}\footnotesend
\begin{footnotes}{156.}\textit{See id.}\footnotesend
\begin{footnotes}{157.}\textit{See id. at 432.}\footnotesend
\begin{footnotes}{158.}\textit{See id.}\footnotesend
\begin{footnotes}{159.}\textit{See id.}\footnotesend
\begin{footnotes}{160.}\textit{See id.}\footnotesend
\begin{footnotes}{161.}\textit{See id.}\footnotesend
\begin{footnotes}{162.}\textit{See id.}\footnotesend
\begin{footnotes}{163.}\textit{See id. at 444.}\footnotesend
\begin{footnotes}{164.}\textit{Id. at 435} (quoting \textit{Miranda v. Arizona}, 384 U.S. 436, 439 (1966)).\footnotesend
\end{footnotes}
\end{footnotesize}
Miranda’s Poisoned Fruit Tree

rights. Thus, while Miranda left the door open for the legislature to design safeguards that exceed Miranda’s constitutional requirements, it also represents the minimum constitutional standard that the legislature cannot go below.

Although the Dickerson Court rejected the prophylactic rationale used to create the three exceptions in earlier jurisprudence, it did not overrule any of these cases. In fact, it incorporated the prophylactic Miranda opinions directly into its holding. Yet, the Court did not explain how it reconciled the prophylactic Miranda holdings with Miranda’s “reaffirmed” constitutional status. Instead, the Court held that Miranda’s exceptions merely illustrated “that no constitutional rule is immutable.” In other words, because courts cannot foresee every circumstance in which counsel will seek to apply a general rule, it is necessary for courts to be able to modify such rules accordingly. The Court noted that these modifications are as much a part of constitutional law as the original decision. Therefore, while the Dickerson opinion greatly undermined the principal reasoning underlying both Tucker and Elstad, their exceptions are still good law. The Court’s decision in Dickerson, however, threw into question whether the Court would adopt another rationale to further limit the scope of Miranda’s exclusionary rule, or whether it would use the fruits doctrine to exclude physical evidence derived from a now constitutional Miranda violation.

III. POST-DICKERSON APPLICATION OF MIRANDA TO DERIVATIVE PHYSICAL EVIDENCE

Following Dickerson, a circuit split developed regarding the admissibility of physical evidence derived from unwarned statements.

165. Id. at 439.
166. Id. at 440.
167. See id. at 437–38.
168. See id. at 432.
169. See id. (“We therefore hold that Miranda and its progeny in this Court govern the admissibility of statements made during custodial interrogation in both state and federal courts.”) (emphasis added).
170. See id. at 437–38, 441.
171. Id. at 441.
172. See id.
173. See id.
174. See id. at 432.
175. See United States v. Patane, 304 F.3d 1013 (10th Cir. 2002); United States v. Faulkingham,
The United States Courts of Appeals for the Third and Fourth Circuits rejected the fruits doctrine in the \textit{Miranda} context,\footnote{295 F.3d 85 (1st Cir. 2002); United States v. Sterling, 283 F.3d 216 (4th Cir. 2002); United States v. DeSumma, 272 F.3d 176 (3d Cir. 2001).} and reaffirmed their pre-\textit{Dickerson} holdings that physical fruits are admissible after excluding the unwarned statement.\footnote{176. See \textit{Sterling}, 283 F.3d at 219; \textit{DeSumma}, 272 F.3d at 180.} In contrast, the United States Court of Appeals for the First Circuit reaffirmed its pre-\textit{Dickerson} holding that the fruits doctrine applies to intentional \textit{Miranda} violations.\footnote{177. See \textit{Sterling}, 283 F.3d at 219; \textit{DeSumma}, 272 F.3d at 180.} The United States Court of Appeals for the Tenth Circuit went a step further and held that the fruits doctrine applies to negligent \textit{Miranda} violations.\footnote{178. See Faulkingham, 295 F.3d at 93 (citing United States v. Byram, 145 F.3d 405, 410 (1st Cir. 1998)).} 

\textbf{A. Physical Fruits Are Admissible}

Since \textit{Dickerson}, two circuits have held that the fruits doctrine does not apply to physical evidence derived from an unwarned statement.\footnote{179. See \textit{Sterling}, 283 F.3d at 219; \textit{DeSumma}, 272 F.3d at 180.} In \textit{United States v. DeSumma},\footnote{180. See \textit{id}, at 1019.} the Third Circuit refused to apply the fruits doctrine to the physical fruits of a \textit{Miranda} violation.\footnote{181. 272 F.3d 176 (3d Cir. 2001).} An FBI agent asked an arrested, unwarned defendant if he had any weapons in his possession.\footnote{182. See \textit{id}, at 180.} The defendant replied that he had a gun in his car, which the agent subsequently found.\footnote{183. See \textit{id}, at 180–81.} The Third Circuit held that the gun could be admitted without violating the Fifth Amendment, even though it was obtained through a non-Mirandized statement.\footnote{184. See \textit{id}.} 

Relying on the \textit{Dickerson} Court’s affirmation of the distinction between Fourth and Fifth Amendment violations in \textit{Elstad},\footnote{185. See \textit{Patane}, 304 F.3d at 1019.} the Third Circuit determined that \textit{Dickerson}’s characterization of \textit{Miranda} as a constitutional rule did not fatally undermine \textit{Elstad}’s rationale for not applying the fruits doctrine.\footnote{186. See \textit{id}, at 180.} The court focused on \textit{Elstad}’s...
determination of Miranda’s exclusionary rule as sweeping more broadly than the Fifth Amendment itself. The court reasoned that while voluntary unwarned statements were excluded from evidence, “the Miranda presumption . . . does not require that the statements and their fruits be discarded as inherently tainted.” Consequently, the court held that the fruits doctrine does not apply to Miranda violations because the police would be no more deterred after the trial court excludes the unwarned statement.

Similarly, in United States v. Sterling, the Fourth Circuit refused to apply the fruits doctrine to physical evidence derived from an unwarned statement. The arresting officer asked the arrested, unwarned defendant whether he had any weapons. The defendant stated that he had a shotgun in his truck, which police subsequently found. The Fourth Circuit held that the trial court properly admitted the shotgun because the fruits doctrine did not extend to Fifth Amendment violations.

The Sterling court determined that Dickerson did not impact the Fourth Circuit’s earlier holding that the fruits doctrine does not apply in a Miranda context. The court reasoned that Dickerson’s incorporation of Miranda’s progeny into its holding was evidence that “the established exceptions, like those in Tucker and Elstad, survive.” Further, the Sterling court suggested that Miranda’s exclusionary rule did not extend beyond the unwarned statements made in violation of the Miranda rule, regardless of whether Miranda was constitutional or not. Therefore, while acknowledging Miranda’s constitutional status, the court held that Tucker and Elstad continued to support its rejection of the application of the fruits doctrine to derivative physical evidence.

188. See id. at 179 (citing Oregon v. Elstad, 470 U.S. 298, 306 (1985)).
189. Id. (quoting Elstad, 470 U.S. at 307).
190. See id. at 180.
191. 283 F.3d 216 (4th Cir. 2002).
192. See id. at 219.
193. See id. at 218.
194. See id. at 219.
195. See id. (citing United States v. Elie, 111 F.3d 1135 (4th Cir. 1997)).
196. Id.
197. See id.
198. See id.
B. Exclusion of Derivative Physical Evidence Limited to Willful Miranda Violations

Before *Dickerson*, the United States Court of Appeals for the First Circuit held that *Elstad* did not prevent the application of the fruits doctrine in the *Miranda* context.199 After *Dickerson*, the court reaffirmed this position in *United States v. Faulkingham*,200 but limited the application of the fruits doctrine to exclude only derivative evidence obtained by willful violations of *Miranda*, not merely negligent acts.201 In *Faulkingham*, the police failed to administer *Miranda* warnings to the defendant before a series of events unfolded where the police assisted the defendant in connecting with his drug supplier.202 However, to achieve that end, the arrested and unwarned defendant made incriminating statements, which led the officers to the drug supplier, who in turn led the officers to heroin.203 Because there was no evidence that the police deliberately failed to give the warnings or did so to get more possibly incriminating evidence, the court determined that the *Miranda* violation at issue was the result of negligent police conduct.204 The court held that while the fruits doctrine applied to the *Miranda* context,205 courts can still admit derivative evidence where the unwarned statement was suppressed and the law enforcement agents merely acted negligently, not willfully or maliciously.206

The *Faulkingham* court reasoned that *Elstad* did not bar the application of the fruits doctrine because in *Elstad* the defendant’s warned statement, not a derivative witness or physical evidence, was at issue.207 Nevertheless, the court noted the *Elstad* Court’s broad language discouraging the use of the fruits doctrine in the *Miranda* context, regardless of the nature of the derivative evidence.208 Consequently, the *Faulkingham* court framed a limited application of the fruits doctrine.209

199. See *United States v. Byram*, 145 F.3d 405, 410 (1st Cir. 1998).
200. 295 F.3d 85, 90–91 (1st Cir. 2002).
201. See id. at 93–94.
202. See id. at 87–88.
203. See id.
204. See id. at 93–94.
205. See id. at 90.
206. See id. at 94.
207. See id. at 90–91.
208. See id. at 91 (citing Oregon v. *Elstad*, 470 U.S. 298, 307 (1985)).
209. See id. at 93–94.
Miranda’s Poisoned Fruit Tree

The court held that in the Fifth Amendment context, once the unwarned statements are suppressed, the rationale of deterring police misconduct becomes less primary.210 The First Circuit balanced the reduced importance of the deterrence goal against the importance of the evidence to the truth-seeking process.211 Applying this balancing analysis, the court concluded that derivative physical evidence is admissible where police negligently fail to administer Miranda warnings.212

C. Exclusion of Derivative Physical Evidence in Negligent Miranda Violations

Contradicting the First Circuit, the Tenth Circuit in United States v. Patane213 applied the fruits doctrine to physical evidence derived from a negligent Miranda violation.214 Without administering the full Miranda warnings, the officer in Patane asked the arrested defendant about his Glock .40 caliber pistol.215 The defendant replied, “The Glock is in my bedroom . . . .”216 The court determined that the police officer’s conduct was negligent and not willful,217 although the court did not elaborate on this conclusion. Ultimately the Tenth Circuit held that the firearm, discovered as a result of the defendant’s statement following the police officer’s negligent administration of Miranda warnings, must be suppressed under Miranda’s exclusionary rule.218

The Patane court held that Elstad and Tucker were not controlling.219 The court reasoned that while these cases declined to apply the traditional fruits doctrine of Wong Sun, their holdings were grounded on the rationale that suppression only applies to the fruits of unconstitutional conduct, and “the violation of a prophylactic rule [does]

210. See id.
211. See id.
212. See id. at 94. For an example of the application of the fruits doctrine to a willful violation, see United States v. Byram, 145 F.3d 405, 410 (1st Cir. 1998).
213. 304 F.3d 1013 (10th Cir. 2002).
214. See id. at 1023. This case represents a reversal of the Tenth Circuit rule that the physical fruits of Miranda violations need not be suppressed. See id. (citing United States v. McCurdy, 40 F.3d 1111, 1117 (10th Cir. 1994)).
215. See id. at 1015.
216. Id.
217. See id. at 1027.
218. See id. at 1019.
219. See id. at 1023.
not require the same remedy.\textsuperscript{220} Because \textit{Dickerson} fatally undermined such logic in \textit{Tucker} and \textit{Elstad},\textsuperscript{221} the Tenth Circuit concluded that \textit{Tucker} and \textit{Elstad} did not foreclose the application of the fruits doctrine in the \textit{Miranda} context.\textsuperscript{222}

The \textit{Patane} court also suggested that the Third and Fourth Circuits erred in their refusal to exclude physical fruit.\textsuperscript{223} The court criticized the blanket bar of the fruits doctrine as having two fundamental problems.\textsuperscript{224} First, the \textit{Patane} court suggested that \textit{Dickerson}'s referral to “\textit{Miranda} and its progeny”\textsuperscript{225} did not foreclose suppression of physical evidence for the simple reason that \textit{Elstad} and \textit{Tucker} did not involve derivative physical evidence.\textsuperscript{226} Second, the court suggested that \textit{Dickerson}'s reaffirmation of \textit{Elstad}'s distinction between Fourth and Fifth Amendment violations did not bar the application of the fruits doctrine to \textit{Miranda} violations.\textsuperscript{227} The \textit{Patane} court reasoned that the difference between the violations was in the breadth of the fruits doctrine, not in its application.\textsuperscript{228} In other words, the court determined that the \textit{Miranda} exceptions created a narrowed fruits doctrine, not an absolute bar on the doctrine itself.\textsuperscript{229}

Finally, the \textit{Patane} court criticized the First Circuit's selective use of the fruits doctrine in the \textit{Miranda} context.\textsuperscript{230} Specifically, it disagreed with the \textit{Faulkingham} court's differentiation between negligent and willful violations of a defendant's \textit{Miranda} rights.\textsuperscript{231} The Tenth Circuit asserted that the Fifth Amendment privilege against self-incrimination was equally violated whether the police were negligent or willful in their failure to administer the \textit{Miranda} warnings.\textsuperscript{232} Moreover, limiting the fruits doctrine to willful violations would not vindicate the deterrence

\textsuperscript{220}. \textit{Id.} at 1019.
\textsuperscript{221}. \textit{See id.}
\textsuperscript{222}. \textit{See id.} at 1022–23.
\textsuperscript{223}. \textit{See id.} at 1027.
\textsuperscript{224}. \textit{See id.} at 1024.
\textsuperscript{225}. \textit{See Dickerson} v. United States, 530 U.S. 428, 432 (2000).
\textsuperscript{226}. \textit{See Patane}, 304 F.3d at 1024.
\textsuperscript{227}. \textit{See id.} at 1025.
\textsuperscript{228}. \textit{See id.}
\textsuperscript{229}. \textit{See id.}
\textsuperscript{230}. \textit{See id.} at 1027 (citing United States v. Faulkingham, 295 F.3d 85 (1st Cir. 2002)).
\textsuperscript{231}. \textit{See id.}
\textsuperscript{232}. \textit{See id.} at 1028.
Miranda’s Poisoned Fruit Tree

goal of the exclusionary rule.\textsuperscript{233} Relying on the subjective view of the police officer would also make it difficult for courts to apply the Miranda presumption consistently.\textsuperscript{234} Instead, the Patane court held that applying the fruits doctrine to physical fruits, regardless of the officer’s subjective intent, “provides certainty in application and clarity for the officers charged with operating under it [and therefore] better serves the interests of citizens, officers, and judicial efficiency.”\textsuperscript{235}

The U.S. Supreme Court granted certiorari to Patane,\textsuperscript{236} and therefore, this circuit split will soon be resolved. The circuit split outlines three possible outcomes. If the Court agrees with the reasoning of the Third and Fourth Circuits, derivative physical evidence will be admitted into the prosecution’s case in chief.\textsuperscript{237} On the other hand, the Court could agree with the First Circuit and conclude that the fruits doctrine is applicable in the Miranda context,\textsuperscript{238} but limited to willful violations of Miranda.\textsuperscript{239} This holding would admit physical evidence derived from negligent Miranda violations.\textsuperscript{240} Finally, the Court could affirm the Tenth Circuit’s ruling and hold that derivative physical evidence should be excluded regardless of officer intent.\textsuperscript{241}

IV. PHYSICAL EVIDENCE DERIVED FROM UNWARNED STATEMENTS SHOULD BE EXCLUDED FROM THE PROSECUTION’S CASE IN CHIEF

Having recently granted certiorari to United States v. Patane,\textsuperscript{242} the U.S. Supreme Court should hold that physical evidence derived from statements obtained in violation of Miranda should be excluded from the prosecution’s case in chief for two reasons. First, no U.S. Supreme Court precedent forecloses exclusion of physical evidence derived from an

\textsuperscript{233}. Id. at 1029.
\textsuperscript{234}. Id.
\textsuperscript{235}. Id.
\textsuperscript{236}. 71 U.S.L.W. 3530 (U.S. Apr. 21, 2003) (No. 02-1183).
\textsuperscript{237}. See United States v. Sterling, 283 F.3d 216, 219 (4th Cir. 2002); United States v. DeSumma, 272 F.3d 176, 180 (3d Cir. 2001).
\textsuperscript{238}. See United States v. Faulkingham, 295 F.3d 85, 90–91 (1st Cir. 2002).
\textsuperscript{239}. See id. at 93; see also United States v. Byram, 145 F.3d 405, 409–10 (1st Cir. 1998).
\textsuperscript{240}. See Faulkingham, 295 F.3d at 94.
\textsuperscript{241}. See United States v. Patane, 304 F.3d 1013, 1029 (10th Cir. 2002), cert. granted, 71 U.S.L.W. 3530 (U.S. Apr. 21, 2003) (No. 02-1183).
excluded unwarned statement. Tucker and Elstad are distinguishable on their facts. Furthermore, Dickerson undermined extending Tucker and Elstad by clarifying Miranda’s constitutional status. Consequently, the Third and Fourth Circuits erred in extending Tucker and Elstad to admit the physical fruits of excluded unwarned statements. Second, U.S. Supreme Court precedent directs courts to apply the fruits doctrine to derivative physical evidence of an unwarned statement. This application should not be conditioned on officer intent. Anything less than an objective application of the fruits doctrine to physical evidence would undermine the effect of Miranda’s exclusionary rule. Although the First and Tenth Circuits have correctly applied the fruits doctrine to derivative physical evidence, the First Circuit incorrectly limited its application to willful violations of Miranda. The U.S. Supreme Court has not conditioned Miranda’s exclusionary rule or its exceptions on officer intent. The Court has not created an exception for derivative physical evidence, and therefore, courts should exclude such evidence along with the unwarned statement.

A. There Is No U.S. Supreme Court Precedent Governing the Application of the Fruits Doctrine to Derivative Physical Evidence

The U.S. Supreme Court has not decided whether the fruits doctrine applies to physical evidence derived from unwarned statements excluded

243. See Patane, 304 F.3d at 1023; Faulkingham, 295 F.3d at 90.
244. See Patane, 304 F.3d at 1024.
245. See id. at 1023–25.
246. See id. at 1024–25.
248. See Patane, 304 F.3d. at 1028–29. But see United States v. Faulkingham, 295 F.3d 85, 93–94 (1st Cir. 2002).
249. See Patane, 304 F.3d at 1029; cf. Miranda v. Arizona, 384 U.S. 436, 479 (1966) (finding that the protection of the Fifth Amendment outweighs any hardships the Miranda rule imposes upon law enforcement).
250. See Patane, 304 F.3d at 1021–27; Faulkingham, 295 F.3d at 90–91.
251. See Patane, 304 F.3d at 1027–28. But see Faulkingham, 295 F.3d at 93–94.
252. See, e.g., Quarles, 467 U.S. at 656; Miranda, 384 U.S. at 467, 475.
254. See id. at 1029.
Miranda’s Poisoned Fruit Tree

at trial.255 In New York v. Quarles, the Court created a public safety exception to Miranda’s exclusionary rule allowing the admission of derivative physical evidence when public safety would be immediately jeopardized by any delay in obtaining evidence.256 This public safety exception, however, does not broadly apply to admit all derivative physical evidence. The Court has recognized that Quarles is a “narrow exception”257 for situations in which public safety is threatened.258 Further, the Court determined that where public safety is at issue, the evidence is outside the scope of Miranda altogether.259 Therefore, the issue of whether derivative physical evidence is admissible when Miranda does apply and the unwarned statement has been excluded remains open.

Neither Michigan v. Tucker nor Oregon v. Elstad resolve this issue because they are grounded on Miranda’s former prophylactic status.260 The Tucker and Elstad decisions provide two exceptions to Miranda’s exclusionary rule that allow derivative evidence to be admitted after the court has excluded the unwarned statement.261 The Dickerson case fundamentally undermined these holdings by characterizing Miranda as a constitutional decision.262 Once stripped of its prophylactic rationale, the Tucker case exists as a unique case limited to the timing of the interrogation and the Miranda decision.263 The Elstad rule becomes a rule for curing derivative evidence from being excluded under the fruits doctrine.264 The Third and Fourth Circuits erred by expanding these exceptions to admit the physical fruits of unwarned statements.265

255. See Patterson, 485 U.S. at 922–23 (White, J., dissenting from denial of certiorari); Patane, 304 F.3d at 1023; Faulkingham, 295 F.3d at 90–91.
256. See Quarles, 467 U.S. at 655.
257. See id. at 658.
258. See id. at 657–58.
259. See id. at 657.
263. See Tucker, 417 U.S. at 447.
264. See Elstad, 470 U.S. at 306.
1. Tucker Is Limited to Its Facts

The Tucker decision does not control the issue of whether derivative physical evidence should be admitted. The Tucker Court explicitly declined “to resolve the broad question of whether evidence derived from statements taken in violation of the Miranda rules must be excluded regardless of when the interrogation took place.” Instead, the Court rooted its holding “on a narrower ground” because of the unique facts of the case. The Miranda violation at issue in Tucker occurred prior to the Court’s decision in Miranda, but the trial occurred after Miranda’s announcement. The Tucker Court held that because the police did not actually violate the law while they were interrogating the defendant, a special limited retroactivity exception applied to admit the derivative evidence.

Because the Dickerson Court characterized Miranda as a constitutional rule, police failure to either administer Miranda warnings or obtain an effective waiver should result in the suppression of both the unwarned statement and its fruit, unless the public safety exception in Quarles applies. The Tucker Court concluded that because Miranda warnings were merely prophylactic, violating Miranda should not reap consequences as severe as those for a constitutional violation. After Dickerson, this rationale has been undermined. Moreover, once the prophylactic language and conclusions are withdrawn from the Tucker opinion, Wong Sun is controlling. The Tucker Court distinguished Wong Sun on the ground that it involved a

266. See Tucker, 417 U.S. at 447.
267. Id.
268. See id.
269. See id. at 447–48.
270. See id. at 437.
271. See id. at 447–48.
273. Cf. Tucker, 417 U.S. at 446–47 (distinguishing the consequences of police conduct abridging a constitutional right from police conduct departing only from prophylactic standards).
276. See Dickerson, 530 U.S. at 432.
277. Cf. Tucker, 417 U.S. at 445–46 (distinguishing actual abridgement of a defendant’s constitutional rights, in which case suppression of derivative evidence is warranted, from departure from a prophylactic standard); see also Wong Sun v. United States, 371 U.S. 471, 485, 488 (1963) (excluding verbal and physical evidence derived from a constitutional violation).
Miranda’s Poisoned Fruit Tree

constitutional violation whereas Miranda violations depart only from prophylactic standards. The Court therefore acknowledged that actual constitutional violations require suppression of the fruits of that conduct under Wong Sun. Thus, after a police failure to either administer the Miranda warnings or obtain an effective waiver, both the unwarned statement and its physical fruit should be suppressed at trial unless the public safety exception applies.

2. Elstad Established the Rule for Curing a Miranda Violation

As an exception limited to admitting warned statements derived from unwarned statements, Elstad does not control whether derivative physical evidence should be admitted. The Elstad Court held that the defendant’s derivative statement should be admitted based on the rationale that Miranda warnings are merely prophylactic safeguards requiring a narrower exclusionary rule than the fruits doctrine applied to constitutional violations. The Court reasoned in Elstad that Miranda’s warnings and waiver requirements swept more broadly than the Fifth Amendment itself. The Court suggested that while Miranda’s prophylactic safeguards barred otherwise constitutional confessions, the severe consequences of the fruits doctrine were inapplicable for such non-constitutional violations. However, this reasoning does not survive after Dickerson’s reaffirmation of Miranda as a constitutional rule. The logical consequence of Dickerson’s characterization of Miranda as a constitutional decision is that the scope of Miranda’s exclusionary rule is the same as the Fifth Amendment. Therefore, in direct contrast to the Court’s holding in Elstad, if law enforcement officers fail to properly administer the Miranda warnings and obtain a knowing and voluntary waiver from the defendant, these violations

279. See id.; see also Wong Sun, 371 U.S. at 484 (determining that “[t]he exclusionary prohibition extends as well to the indirect as the direct products of [unlawful] invasions”).
282. See Elstad, 470 U.S. at 305–06.
283. See id. at 306.
284. Id. at 308–09.
should “breed the same irreremediable consequences as police infringement of the Fifth Amendment itself.”

Once shed of its prophylactic rationale, Elstad survives only as a curing rule for Miranda violations, distinct from the Fourth Amendment attenuation doctrine. The Elstad Court based its holding in part on the distinction between unwarned interrogation under the Fifth Amendment and unreasonable searches under the Fourth Amendment. Because the evidence at issue in Elstad was in fact a Mirandized statement, impliedly the difference between unwarned interrogation and unreasonable searches is that the warnings and a waiver can cure a Miranda violation. In Elstad, the Court held that “a careful and thorough administration of Miranda warnings serves to cure the condition that rendered the unwarned statement inadmissible.” In contrast, courts apply the attenuation doctrine to Fourth Amendment violations to determine when derivative evidence is admissible. Under this doctrine, unreasonable searches and seizures are not “cured” until the connection between the illegality and the contested evidence has “become so attenuated as to dissipate the taint.”

Distinguishing the consequences of Fourth and Fifth Amendment violations based on how or when those consequences end is reasonable. Once police undertake a search and seizure, it is arguably impossible to cure the initial invasion of privacy. Instead, a court must review the proximity of the Fourth Amendment violation and the discovery of the derivative evidence to determine whether there has been a sufficient break to untaint the evidence. Curing is easier under the Fifth Amendment because informing a suspect of his or her Miranda rights at any point in the interrogation process dispels the inherent compulsion of interrogation. Consequently, the potential taint from the earlier

286. See Elstad, 470 U.S. at 309.
287. See id. at 311.
288. See Nardone v. United States, 308 U.S. 338, 341 (1939) (concluding that derivative evidence may be admissible when the connection between the illegality and the evidence becomes “so attenuated as to dissipate the taint”).
289. See Elstad, 470 U.S. at 306.
290. See id. at 306, 310–11.
291. Id. at 310–11 (emphasis added).
292. See Nardone, 308 U.S. at 341.
293. Id.
294. See id.
Miranda’s Poisoned Fruit Tree

violation to any statement made subsequent to a waiver dissipates.296
Under this interpretation, the difference between the Fourth and Fifth
Amendments relates to the limits of the fruits doctrine, not its
application.297

Dickerson reaffirmed Elstad’s distinction between unreasonable
searches and unwarned interrogations.298 However, contrary to the Third
and Fourth Circuits’ interpretations,299 this reaffirmation did not provide
a blanket admission of all derivative evidence of unwarned
statements.300 Instead, this distinction highlighted Elstad’s enunciation
of a curing rule,301 separate from the attenuation doctrine in the Fourth
Amendment.302

Furthermore, Dickerson’s characterization of Elstad as “refusing to
apply the traditional fruits doctrine developed in the Fourth
Amendment”303 does not foreclose application of the fruits doctrine in the
Miranda context. The “traditional” fruits doctrine incorporates the
attenuation doctrine.304 Elstad held that this was not necessary in the
Fifth Amendment context because the administration and waiver of the
Miranda warnings prior to the statement in question sufficiently cured
the statement of illegality.305

On review in Patane, the Court should recognize that Miranda’s
exclusionary rule would not require a curing test if the fruits doctrine
was not applicable in the Miranda context. Curing inherently recognizes
that a violation has occurred with ongoing consequences until the taint
of the violation appropriately dissipates. In the Miranda context,
dissipation occurs when warnings are administered and a waiver is
given.306 In the case of physical evidence derived from an unwarned
statement, however, there is no possibility of intervening warnings
between the unwarned statement (i.e., the poisonous tree) and the

296. See Elstad, 470 U.S. at 311; Miranda, 384 U.S. at 479.
297. See United States v. Patane, 304 F.3d 1013, 1025 (10th Cir. 2002).
299. See United States v. Sterling, 283 F.3d 216, 219 (4th Cir. 2002); United States v. DeSumma,
272 F.3d 176, 180 (3d Cir. 2001).
300. See Patane, 304 F.3d at 1026.
301. See Elstad, 470 U.S. at 311.
303. Dickerson, 530 U.S. at 441 (internal quotations omitted).
304. See Nardone, 308 U.S. at 341.
305. See Elstad, 470 U.S. at 311.
306. See id.
physical evidence (i.e., the fruit) to cure its taint. Therefore, the Court should hold that both the unwarned statement and its physical fruit should be excluded from the prosecution’s case in chief, absent a public safety exception.

3. The Third and Fourth Circuits Erred in Extending Tucker and Elstad to Exclude Derivative Physical Evidence from Miranda’s Exclusionary Rule

The Third and Fourth Circuits improperly extended Elstad and Tucker’s derivative evidence exceptions to admit physical evidence derived from an excluded unwarned statement.307 In upholding their pre-Dickerson interpretations of Tucker and Elstad, the Third and Fourth Circuits relied on the Dickerson decision’s recognition of Miranda’s progeny and its distinction between the Fourth and Fifth Amendments to find that the Elstad and Tucker exceptions survived Dickerson.308 The courts interpreted the survival of these exceptions to mean that Tucker and Elstad should be extended to bar the application of the fruits doctrine to derivative physical evidence of Miranda violations.309 Yet, the Tucker and Elstad cases are sufficiently distinguishable on their facts and should be limited to their holdings.310

Further, the circuit courts failed to recognize the implications of Miranda’s constitutional status.311 Although the Fourth Circuit noted that its own pre-Dickerson characterization of Miranda as prophylactic was no longer good law,312 it failed to consider how this limited the Tucker and Elstad holdings. In addition, the Third Circuit incorrectly relied on the Dickerson decision’s reaffirmation of Elstad as ground to dismiss the constitutional argument.313 This dismissive treatment of Miranda’s constitutional status does not survive the logical consequences of Dickerson’s impact on Tucker and Elstad.314

308. See Sterling, 283 F.3d at 219; DeSumma, 272 F.3d at 180.
309. See Sterling, 283 F.3d at 219; DeSumma, 272 F.3d at 180.
310. See supra Part IV.A.1–2.
311. See Sterling, 283 F.3d at 219; DeSumma, 272 F.3d at 180.
312. See Sterling, 283 F.3d at 219.
313. See DeSumma, 272 F.3d at 180.
314. See supra Part IV.A.1–2.
Miranda’s Poisoned Fruit Tree

B. The U.S. Supreme Court Should Hold that Both the Unwarned Statement and Its Physical Fruits Should Be Excluded, Regardless of Officer Intent, to Ensure Full Protection of the Defendant’s Fifth Amendment Rights

Fifth Amendment precedent requires courts to exclude unwarned statements and their physical fruits from the prosecution’s case in chief in order to enforce Miranda’s substantive protections. Because the fruits doctrine applies to Fifth Amendment violations, it should also apply to physical evidence derived from violations of Miranda’s Fifth Amendment safeguards. Furthermore, Miranda’s exclusionary rule is most effective at deterring police violations of suspects’ Fifth Amendment rights when courts exclude both the unwarned statement and its physical fruits at trial. The First and Tenth Circuits followed Fifth Amendment precedent and applied the fruits doctrine to physical fruit. But, contrary to the First Circuit’s approach, the fruits doctrine should not be limited to willful violations of Miranda because officer intent is irrelevant to the admissibility of evidence obtained in violation of Miranda.

1. The Fruits Doctrine’s Application to the Fifth Amendment Encompasses Physical Evidence in the Post-Dickerson Miranda Context

The Dickerson decision clarified that Miranda’s warnings are constitutionally required to protect a suspect’s Fifth Amendment rights. The U.S. Supreme Court has held that the fruits doctrine applies to the Fifth Amendment. A court cannot compel a witness to make self-incriminating statements “unless the compelled testimony and

316. See Murphy, 378 U.S. at 79.
317. See Quarles, 467 U.S. at 688 (Marshall, J., dissenting).
318. See Patane, 304 F.3d at 1026–27.
319. See Patane, 304 F.3d at 1021, 1026; United States v. Faulkingham, 295 F.3d 85, 90–91, 93–94 (1st Cir. 2002).
its fruits cannot be used” against the witness in a criminal trial. When the Court extended the scope of the Fifth Amendment to encompass pre-trial custodial interrogation in *Miranda*, it should have clearly incorporated the fruits doctrine as part of the Fifth Amendment precedent.

Moreover, the Court’s language in both *Tucker* and *Elstad* suggests that the doctrine expressed in *Wong Sun*, “that fruits of a constitutional violation must be suppressed,” controls and bars the admissibility of physical fruits. In *Tucker*, the Court decided that *Wong Sun* was not controlling specifically because “the police conduct at issue here did not abridge [the defendant’s] constitutional privilege against compulsory self-incrimination.” Once the *Dickerson* Court clarified that a violation of *Miranda* did in fact abridge the defendant’s Fifth Amendment rights, the *Wong Sun* decision should control. Similarly, in *Elstad*, the Court determined that the defendant’s contention that the fruits doctrine applied to his case “assumes the existence of a constitutional violation.” Now that such an assumption is correct, all evidence discovered as a result of that violation should be excluded as tainted fruit.

The U.S. Supreme Court has not carved out an exception for the physical fruits of excluded unwarned statements, although some of its opinions have used language that could reasonably lead courts to conclude that such an exception exists. In *Quarles*, Justice O’Connor

323. *See id.* (emphasis added).
324. *See Miranda*, 384 U.S. at 444; *see also Dickerson*, 530 U.S. at 440 n.5 (highlighting *Miranda* as a Fifth Amendment rule).
325. *Cf. Quarles*, 467 U.S. at 688–89 (Marshall, J. dissenting) (discussing the inadmissibility of derivative physical evidence under Supreme Court precedent).
330. *See id.* at 305–06.
331. *See United States v. Patane*, 304 F.3d 1013, 1023 (10th Cir. 2002) (concluding that *Elstad* and *Tucker* do not bar exclusion of derivative physical evidence); *United States v. Faulkingham*, 295 F.3d 85, 90–91 (1st Cir. 2002) (finding that *Elstad* does not “wholly bar the door” to excluding derivative evidence).
332. *See, e.g., Elstad*, 470 U.S. at 304 (“The Fifth Amendment, of course, is not concerned with nontestimonial evidence.”); *see also New York v. Quarles*, 467 U.S. 649, 666 (1984) (O’Connor, J., concurring in part and dissenting in part) (stating that the Fifth Amendment’s mandate “does not protect an accused from being compelled to surrender nontestimonial evidence against himself”).
Miranda’s Poisoned Fruit Tree

wrote a concurring opinion in which she justified admission of derivative physical evidence on the ground that Miranda’s exclusionary rule was limited to testimonial evidence.\(^\text{333}\) Because nontestimonial evidence falls outside the scope of the Fifth Amendment, Justice O’Connor reasoned that the fruits doctrine could not apply to derivative physical evidence.\(^\text{334}\) Writing for the majority in Elstad, Justice O’Connor again noted in dicta that nontestimonial evidence is outside the scope of the Fifth Amendment.\(^\text{335}\)

Yet, the principal flaw in equating derivative physical evidence with nontestimonial evidence is that derivative evidence is never inadmissible because of a defect in the manner in which it was obtained; it is excluded because it is the indirect product of an earlier constitutional violation.\(^\text{336}\) In other words, the constitutional violation that results in an unwarned statement taints the derivative physical evidence—not the act of locating the evidence itself. The inadmissibility of the derivative evidence is grounded in the rule that the prosecution cannot profit from a constitutional violation.\(^\text{337}\) The nontestimonial evidence line of cases involved physical evidence unconnected to any constitutional violation.\(^\text{338}\) Thus, nontestimonial evidence is outside the scope of the Fifth Amendment strictly because the act of obtaining the evidence does not itself compel the defendant to self-incriminate.\(^\text{339}\)

The Court should recognize on review in Patane that Fifth Amendment precedent excludes both the unwarned statement and its physical fruits obtained from a Miranda violation.\(^\text{340}\) Miranda created a

\(^{333}\) See Quarles, 467 U.S. at 660 (O’Connor, J., concurring in part, and dissenting in part).

\(^{334}\) See id. at 666.

\(^{335}\) See Elstad, 470 U.S. at 304.


\(^{339}\) See Doe v. United States, 487 U.S. 201, 210 (1988) (defining testimonial evidence to mean that an accused’s communication explicitly or implicitly relates a factual assertion or discloses information); see also Schmerber, 384 U.S. at 765 (finding that “[n]ot even a shadow of testimonial compulsion upon or enforced communication by the accused was involved either in the extraction or in the chemical analysis” of the blood sample).

presumption of inadmissibility for evidence obtained without warnings and a waiver. Although the Court has created three narrow exceptions to this rule, it has not created such an exception for derivative physical evidence. Fifth Amendment precedent should apply to Miranda now that the Court has recognized Miranda as a Fifth Amendment rule. Fifth Amendment precedent suggests that derivative physical evidence is an unwarned statement’s tainted fruit. Therefore, the U.S. Supreme Court should hold that both the unwarned statement and its physical fruits should be excluded, absent a public safety concern or an interim warning and waiver.

2. Applying the Fruits Doctrine to Derivative Physical Evidence Is Necessary to Give Miranda’s Exclusionary Rule Its Intended Weight

The exclusionary rule is a harsh penalty for law enforcement officers and prosecutors, but it is the deterrent the Court has chosen for Miranda violations. The Miranda Court concluded that society’s need for interrogation did not outweigh the constitutional privilege against self-incrimination. Preventing the government from profiting from its illegal conduct is necessary to give effect to the fundamental right against self-incrimination. Admitting the physical fruits of an excluded statement allows the prosecution to make the same if not physical fruits of a Miranda violation would mark a dramatic departure from Supreme Court precedent”).

343. See Patane, 304 F.3d at 1023.
345. See Murphy, 378 U.S. at 79.
346. See Elstad, 470 U.S. at 314; Quarles, 467 U.S. at 655.
347. See Miranda v. Arizona, 384 U.S. 436, 481 (1966). The U.S. Supreme Court has suggested that the principal reason for extending the exclusionary rule to fruit of the unlawful police conduct is that “this admittedly drastic and socially costly course” is necessary to deter police from violating suspects’ constitutional and statutory rights. See Nix v. Williams, 467 U.S. 431, 442–43 (1984).
“greater use of the confession than merely introducing the words themselves.”

Excluding the defendant’s statement but admitting its physical fruits creates negative incentives. It encourages the police to interrogate the defendant so that crucial physical evidence can be discovered before Miranda warnings are administered. Admitting derivative physical evidence, but not the unwarned statement, is a loophole exacting the same “heavy toll on individual liberty” as custodial interrogation did before Miranda. Instead, the Court should exclude both the unwarned statement and its physical fruits because this is the most effective means for achieving the twin purposes behind Miranda’s exclusionary rule: trustworthiness and deterrence. First, admitting the physical fruits of inherently compelling custodial interrogations jeopardizes the integrity of the fact-finding process in court. Second, excluding the physical fruits of an unwarned statement best deters compelled interrogations and safeguards the defendant’s Fifth Amendment rights. It forewarns police officers that the prosecution bears the heavy burden of demonstrating the suspect was warned and sufficiently waived his or her rights before a court can admit an unwarned statement and its physical fruits, thereby encouraging compliance with Miranda’s safeguards.

3. The First and Tenth Circuits Correctly Applied the Fruits Doctrine to the Miranda Context, Although the First Circuit Erred in Restricting Its Application to Intentional Violations of Miranda

The First and Tenth Circuits determined that in order to achieve the goals of Miranda, derivative physical evidence must be suppressed in certain situations. These circuit courts limited Elstad’s application to

350. See United States v. Patane, 304 F.3d 1013, 1027 (10th Cir. 2002) (emphasis omitted); cf. Nix, 467 U.S. at 443 (reasoning that “the prosecution is not to be put in a better position than it would have been in if no illegality had transpired”).

351. See Patane, 304 F.3d at 1026.

352. Miranda, 384 U.S. at 455.


354. See Miranda, 384 U.S. at 466.

355. See id. at 457–58.

356. See id. at 479.

357. See United States v. Patane, 304 F.3d 1013, 1029 (10th Cir. 2002); United States v.
only admit warned statements following unwarned statements.\textsuperscript{358} The Tenth Circuit limited the holding in \textit{Tucker} to apply only to derivative evidence stemming from violations preceding the \textit{Miranda} opinion itself.\textsuperscript{359}

While both the First and Tenth Circuits applied the fruits doctrine to derivative physical evidence,\textsuperscript{360} the First Circuit narrowed its application to intentional violations only.\textsuperscript{361} However, the First Circuit’s reliance on the officer’s subjective intent clouds \textit{Miranda}’s bright-line rule presuming exclusion unless the prosecution can demonstrate the police properly administered \textit{Miranda} warnings and the suspect waived those rights.\textsuperscript{362} It also places courts in the position of weighing subjective belief.\textsuperscript{363} Not only is it difficult for courts to uncover and evaluate the subjective motivation of a police officer, but such attempts can lead to inconsistent results.

\textit{Miranda}’s presumption of exclusion reflects the reality that negligent conduct violates a defendant’s personal right against self-incrimination to the same extent as willful conduct.\textsuperscript{364} The \textit{Miranda} Court did not condition the compulsion of the interrogation environment on officer intent; instead, it reasoned that only warned statements could be the product of a defendant’s free choice.\textsuperscript{365} The \textit{Miranda} Court created a severe penalty for obtaining unwarned statements to deter officers from foregoing or undermining the defendant’s rights and to ensure the trustworthiness of admitted evidence.\textsuperscript{366} The Court recognized the severity of \textit{Miranda}’s exclusionary rule, but held that the Fifth Amendment required such extreme measures.\textsuperscript{367} There is language in the

\begin{quote}
\textsuperscript{358} See \textit{Patane}, 304 F.3d at 1021 (discussing the exclusion of derivative evidence outside of the \textit{Elstad} facts); \textit{Faulkingham}, 295 F.3d at 90–91 (same).
\textsuperscript{359} See \textit{Patane}, 304 F.3d at 1020 (stating that the U.S. Supreme Court’s ruling rested largely on pre-\textit{Miranda} facts).
\textsuperscript{360} See id. at 1019; \textit{Faulkingham}, 295 F.3d at 90–91.
\textsuperscript{361} See \textit{Faulkingham}, 295 F.3d at 93–94.
\textsuperscript{362} See \textit{Miranda} v. Arizona, 384 U.S. 436, 479 (1966); \textit{Patane}, 304 F.3d at 1029.
\textsuperscript{363} \textit{Patane}, 304 F.3d at 1029.
\textsuperscript{364} Id. at 1028.
\textsuperscript{365} See \textit{Miranda}, 384 U.S. at 458.
\textsuperscript{366} See id. at 479; see also id. at 467.
\textsuperscript{367} See id.
\end{quote}
Miranda’s Poisoned Fruit Tree

Court’s opinion to suggest that deterrence is only achieved by administering Miranda warnings and adequately waiving those rights.368 The U.S. Supreme Court has also created a separate presumption that absent an exception, evidence derived from a constitutional violation is tainted fruit that should be excluded from court.369 The Court has placed the burden on the prosecution to prove by a preponderance of the evidence that an exception applies.370 The Court has not created an exception for the negligent failure to administer the Miranda warnings. Indeed, the Court has never suggested that officer intent is relevant to the exceptions carved out of Miranda’s exclusionary rule.371 Thus, the Court should hold that both the unwarned statement and derivative physical evidence should be excluded from the prosecution’s case in chief, regardless of officer intent.

V. CONCLUSION

Currently, there is a split among circuit courts regarding the admissibility of physical evidence derived from non-Mirandized statements. In addressing this issue, the U.S. Supreme Court should provide a definitive answer in Patane on the admissibility of physical evidence directly derived from unwarned statements. After reviewing Miranda and its progeny, the Court should apply the fruits doctrine to physical evidence obtained by negligent or willful Miranda violations. It is also the most effective measure for ensuring that Miranda warnings are properly “employed to dispel the compulsion inherent in custodial surroundings.”372 Consequently, absent an established exception, the

368. See id. at 457–58 (“The current practice of incommunicado interrogation is at odds with one of our Nation’s most cherished principles—that the individual may not be compelled to incriminate himself. Unless adequate protective devices are employed to dispel the compulsion inherent in custodial surroundings, no statement obtained from the defendant can truly be the product of his free choice.”); id. at 467 (“In order to combat [custodial interrogation’s inherently compelling] pressures and to permit a full opportunity to exercise the privilege against self-incrimination, the accused must be adequately and effectively apprised of his rights and the exercise of those rights must be fully honored.”).


370. See Nix, 467 U.S. at 444.


372. See Miranda, 384 U.S. at 458.
Court should hold that physical evidence derived from an unwarned statement should be excluded from the prosecution’s case in chief.
DR. JEKYLL’S WAIVER OF MR. HYDE’S RIGHT TO REFUSE MEDICAL TREATMENT: WASHINGTON’S NEW LAW AUTHORIZING MENTAL HEALTH CARE ADVANCE DIRECTIVES NEEDS ADDITIONAL PROTECTIONS

Nick Anderson

Abstract: Mental health care advance directives are gaining popularity nationwide. Following a growing trend, the Washington State Legislature has recently passed a law allowing patients to draft mental health care advance directives that could be irrevocable. Patients who sign an irrevocable directive essentially waive their fundamental right to refuse treatment in the future. The United States Supreme Court has held that waivers of fundamental rights must be made knowingly, voluntarily, and intelligently. However, as passed, Washington’s new law contains insufficient safeguards to guarantee such a waiver. This Comment proposes that the Washington State Legislature amend this law to require two additional protections: a “rights advocate” to explain the potential waiver of rights, and a written warning in the advance directive form. These safeguards will help ensure that patients make knowing and intelligent waivers of their fundamental right to refuse treatment.

Dr. Jekyll,¹ a respected physician, lives in Seattle with his wife. Dr. Jekyll has dissociative identity disorder.² About once a year his other personality, Mr. Hyde, takes over, rendering him unable to perform as a physician. Mr. Hyde always refuses to take the psychotropic medication that could control the symptoms and speed up Dr. Jekyll’s recovery. Yet, because Mr. Hyde never poses a danger to himself or to others, the State of Washington will not commit him to a mental hospital against his will or force him to take medication. Nevertheless, Dr. Jekyll and his wife agree that taking medication is important. Dr. Jekyll’s wife persuades him to sign an “advance directive,” a document authorizing her to commit him to a mental health facility where he can receive treatment and medication when he slips into the role of Mr. Hyde. This document sounds like a good idea to Dr. Jekyll; it will allow him to receive the treatment he needs so he can recover and return to work. However, when Dr. Jekyll signs the document, he does not realize he is effectively waiving his fundamental right to refuse medical treatment later on. Once

¹. Character names are the invention of Robert Louis Stevenson. Robert Louis Stevenson, The Strange Case of Dr. Jekyll and Mr. Hyde (1886).
². Hypothetical created by the author for illustrative purposes.

the document is signed, Mr. Hyde no longer has the right to refuse medication, no matter how strongly he protests.

This scenario could soon take place in Washington. The Washington State Legislature has passed a new law, Mental Health—Advance Directives for Health Care, which allows a competent individual to sign a binding, and in some cases irrevocable, mental health advance directive.3 This directive is a document in which a patient gives advance consent for mental health care treatments.4 A competent Dr. Jekyll could therefore instruct health care providers how to treat him when he becomes incompetent and slips into the role of Mr. Hyde. While little dispute exists over whether there is a constitutional right to refuse medical treatment in advance,5 questions remain about a person’s right to irrevocably consent to such treatments in advance.6 Washington’s new law could allow the state to force an unwilling, mentally ill patient to take medication if the patient previously gave irrevocable consent to such treatment.7 The patient’s present unwillingness would conflict with the previously given consent, and this conflict raises a serious concern about whether irrevocable advance directives might unconstitutionally infringe upon a patient’s right to refuse medical treatment.

The United States Supreme Court has recognized that a patient’s ability to refuse medical treatment is a fundamental right.8 State legislation cannot permit a person to waive a fundamental right in advance unless the waiver meets the constitutionally-required “knowing, voluntary, and intelligent” test.9 This Comment argues that Washington’s new law fails to ensure that patients make such a waiver. Specifically, the law does not guarantee that patients’ waivers of their fundamental right to refuse treatment are made knowingly and intelligently.

4. Id. § 5.
5. See Cruzan v. Dir., Mo. Dep’t of Health, 497 U.S. 261, 278 (1990); see also infra note 123 and accompanying text.
7. See Mental Health Advance Directives, supra note 3, § 5.
8. Cruzan, 497 U.S. at 278.
Mental Health Advance Directives

Part I of this Comment describes advance directives and Washington’s new law. Part II examines a patient’s right to refuse medical treatment and the three requirements for valid waivers of that right. Part III explores how existing statutes, rules of procedure, and court-mandated processes provide protections to guarantee that waivers of fundamental rights are made knowingly, voluntarily, and intelligently. Finally, Part IV argues that the Washington State Legislature should amend the mental health advance directive law to include similar protections. These protections are necessary to ensure that patients make knowing and intelligent waivers, and to prevent the new law from being struck down as unconstitutional.

I. FOLLOWING A GROWING TREND, THE WASHINGTON STATE LEGISLATURE HAS PASSED A LAW AUTHORIZING MENTAL HEALTH ADVANCE DIRECTIVES

Advance directives allow patients to give informed consent for certain treatments in advance. Such directives were first used for end of life situations, but many patients in the State of Washington have pressed for statutory authorization to draft advance directives for mental health care. In response, Washington legislators have passed a law that will allow patients to create mental health advance directives.

A. Advance Directives Allow Patients to Make Decisions About Their Future Health Care

An advance directive is a legal document that declares a patient’s wishes about medical treatment to be provided should the patient become incompetent or unable to communicate. Patients sign advance directives while they are competent, and the documents remain in effect even if the patients later become incompetent. There are two different

11. See infra note 37 and accompanying text.
12. Mental Health Advance Directives, supra note 3, § 5.
types of advance medical directives that allow patients to control their future health care: proxy directives and instructional directives.\textsuperscript{15}

A proxy directive simply allows a patient to authorize someone else to make decisions on that patient’s behalf.\textsuperscript{16} The principal—the patient who signs the directive—allows the proxy—an agent or personal representative—to make decisions on the principal’s behalf if and when the principal becomes incompetent.\textsuperscript{17} Such directives are similar to a durable power of attorney.\textsuperscript{18} The agency relationship is “durable” because it lasts even when the principal becomes incompetent.\textsuperscript{19} The benefit of this arrangement is that it enables the agent to respond flexibly to changing circumstances when the patient’s current wishes are unknown.\textsuperscript{20}

In contrast, an instructional directive is a legal document in which a person defines what kinds of treatment may be performed in the future if he or she becomes incompetent and cannot make decisions.\textsuperscript{21} One popular type of instructional advance directive is called a living will.\textsuperscript{22} A living will specifies a person’s desire to be taken off life support when there is little chance of recovery by stating that the signer’s life shall not be artificially prolonged when an extreme physical or mental disability makes it highly unlikely that the signer will recover.\textsuperscript{23} Many states have statutes authorizing living wills.\textsuperscript{24} In Washington, the Natural Death Act recognizes “the right of an adult person to make a written directive” to explain when the person would like life-sustaining treatment removed.\textsuperscript{25}

The legal basis for advance directives is rooted in the tort theory of informed consent. Advance directives attempt to allow the competent

\textsuperscript{15} Elizabeth M. Gallagher, \textit{Advance Directives for Psychiatric Care: A Theoretical and Practical Overview for Legal Professionals}, 4 PSYCHOL. PUB. POL’Y & L. 746, 749 (1998).
\textsuperscript{16} Id. at 751.
\textsuperscript{17} JESSE DUKEMINIER & STANLEY M. JOHANSON, WILLS, TRUSTS, AND ESTATES 404 (6th ed. 2000).
\textsuperscript{18} Id.
\textsuperscript{20} DUKEMINIER & JOHANSON, supra note 17, at 404.
\textsuperscript{22} DUKEMINIER & JOHANSON, supra note 17, at 403–04.
\textsuperscript{23} Id.
\textsuperscript{25} WASH. REV. CODE § 70.122.010 (2003).
Mental Health Advance Directives

“present self” to give informed consent on behalf of the “future self” who may become incapacitated. All patients must give informed consent before medical treatment can be administered. As a general rule, a patient’s consent should be based on adequate information about the benefits of medical treatment, the available alternatives, and the risks involved. When treatment is unauthorized or performed without any consent at all, the health care provider administering the treatment may be liable for having committed a battery.

The rationale behind allowing advance directives is that a person is in a better position to give true informed consent to mental health care treatments when he or she is competent than when incompetent. Typically, a competent person can carefully weigh the benefits, alternatives, and risks of treatment and make an informed decision. By comparison, an incompetent person’s ability to evaluate information is impaired to such an extent that the person lacks the capacity to make treatment decisions. Consent is essential to the provision of medical treatment, and advance directives allow informed consent to be given ahead of time.

B. Many Washington Patients Wanted Statutory Authorization to Write Mental Health Advance Directives

In recent years there has been a groundswell of support for advance directives by mental health consumers, providers, and advocacy

---

26. See Dresser, supra note 6, at 814 n.134.
29. Id. at 557.
30. See Dresser, supra note 6, at 785 n.31.
31. See Mental Health Advance Directives, supra note 3, § 2(8).
33. See Winick, supra note 10, at 70–71 (arguing that advance directives refusing treatment are more likely to be valid than those requesting treatment, but that the latter does constitute a measure of proof of informed consent).
groups.34 Mental health patients—especially those who alternate between competence and incompetence—have pushed for legislation allowing mental health advance directives.35 Sixteen states have responded by enacting statutes specifically authorizing mental health advance directives.36 Many mental health care patients in Washington state pressured the Washington State Legislature to pass a similar law.37

One of the major reasons some families of patients wanted statutory authorization to execute mental health advance directives was to clarify uncertainty about Washington law.38 Mental health treatment can always be obtained voluntarily if patients are willing to check themselves into a treatment facility.39 However, patients who suffer from mental illness are often “treatment resistant” and either refuse treatment or do not realize they need it.40 Even those who do take medication often stop taking it prematurely as they begin to feel better.41 When a mentally ill person refuses treatment, friends and family must go through a complex public legal process to have their family member committed involuntarily.42 Moreover, under Washington State’s Involuntary Treatment Act, courts only authorize involuntary commitment when patients are “gravely disabled” or pose a risk of serious harm to themselves or to others.43

Because of the difficult legal standard that existed under Washington law, families felt frustrated by their inability to help mentally ill

---

34. See Gallagher, supra note 15, at 748.
35. Email from Erin Speck, Legislative Assistant to Washington State Senator Karen Keiser, to author (Jan. 27, 2003) (on file with author) [hereinafter Speck Email].
36. DEBRA SRENNIK & LISA BRODOFF, IMPLEMENTING PSYCHIATRIC ADVANCE DIRECTIVES: SERVICE PROVIDER ISSUES AND ANSWERS 28 (Working Paper, 2003); see, e.g., ALASKA STAT. §§ 47.30.950–.980 (Michie 2002); HAW. REV. STAT. ANN. § 327F (Michie 2002); IDAHO CODE §§ 66-601 to -613 (Michie 2002); ILL. COMP. STAT. 43/1-115 (2002); MINN. STAT. ANN. § 253B.03(6)(d) (West 2002); N.C. GEN. STAT. §§ 122C-71 to -77 (2002); OKLA. STAT. ANN. tit. 43A, §§ 11-101 to -113 (West 2002); OR. REV. STAT. §§ 127.700–.737 (2002); TEX. CIV. PRAC. & REM. CODE ANN. §§ 137.001–.011 (Vernon 2002); UTAH CODE ANN. §§ 62A-15-1001 to -1004 (2002).
37. See REPORT FROM THE SENATE COMMITTEE ON CHILDREN & FAMILY SERVICES ON S.B. 5223, 58th Sess., at 3 (Wash. 2003).
38. Speck Email, supra note 35.
39. WASH. REV. CODE § 71.05.050 (2003).
40. See Gallagher, supra note 15, at 747.
42. WASH. REV. CODE § 71.05.370.
43. Id. § 71.05.030; see also In re Anderson, 17 Wash. App. 690, 692, 564 P.2d 1190, 1192 (1977).
Mental Health Advance Directives

relatives. A family’s hands were effectively tied—it could not send a mentally ill family member to a treatment facility if the person did not meet the strict definition of “gravely disabled,” even if the person could benefit from treatment. In the vast majority of involuntary and voluntary commitment cases, it is the family that seeks commitment or treatment. A legislative assistant paraphrased one patient’s passionate testimony regarding this problem to the Washington State Legislature: “[w]hen someone is allowed to ‘decompose’ so severely before they can get help under the Involuntary Treatment Act, they never come back quite the same.” Many families believe that a mental health advance directive would allow patients to get the treatment they need before the patients’ mental health deteriorates to the point where it fits the definition of “gravely disabled.”

Further, proponents of the new law argued that advance directive legislation would help cure another inadequacy in Washington law: situations where incompetent patients cannot communicate the treatment they want. When incompetent patients are unable to let others know what type of treatment they prefer, Washington law allows courts to “make a substituted judgment for the patient as if he or she were competent to make such a determination.” In In re Ingram, the Washington State Supreme Court held that when a court is asked to make a substituted medical decision for an incompetent person, the court must attempt to decide as the individual would if he or she were competent. An advance directive allows patients to describe recurring mental health issues treatments that have worked in the past.

45. Id.
47. Speck Email, supra note 35.
49. WASH. REV. CODE § 71.05.370(7)(b) (2003).
51. Id. at 838, 1369.
Therefore, a mental health advance directive could be strong evidence of a patient’s wishes on which a court could rely when making such a substituted judgment.53

Another reason that some Washington citizens wanted statutory authorization to execute mental health advance directives is that mental health care providers and patients had concerns about the legal validity of advance directives based on existing state law.54 Prior to the bill’s passage, it was unclear whether patients who wanted to execute irrevocable advance directives could do so.55 Therefore, patients and attorneys believed that legislation explicitly authorizing mental health advance directives would be more useful than the general common law and vaguely analogous laws allowing advance directives for end of life situations.56

Attorneys in Washington had attempted to apply general principles of state law on living wills and powers of attorney to the unique circumstances of mental health care.57 Although these laws did not prohibit mental health care advance directives, the actual enforceability of mental health directives drafted according to such laws was somewhat uncertain.58 Mental health care providers were unsure about how to deal with mental health advance directives in the absence of express statutory authorization, and were concerned about potential liability for following, or refusing to follow, mental health care advance directives.59 Finally, many attorneys prefer to rely on a specific grant of statutory authority when representing a client who wants to draft a mental health advance directive.60

C. A New Washington Statute Specifically Authorizes Patients to Write Mental Health Advance Directives

The Washington State Legislature recently passed a law that specifically authorizes patients to write mental health advance

54. See REPORT FROM THE SENATE COMMITTEE ON CHILDREN & FAMILY SERVICES ON S.B. 5223, 58th Sess., at 3 (Wash. 2003).
55. See Speck Email, supra note 35.
56. See Gallagher, supra note 15, at 763.
57. See id. (specifically considering Washington State).
58. See id. at 772.
59. See generally SREBNIK & BRODOFF, supra note 36, at 11.
60. Gallagher, supra note 15, at 771.
Mental Health Advance Directives

directives. The law is called “Mental Health—Advance Directives For Health Care.” In December of 2002, Representative Patricia Lantz and Senator Karen Keiser proposed House Bill 1041 and Senate Bill 5223, respectively, to the Washington State Legislature. On February 25, 2003, the Judiciary Committee combined the House and Senate bills into a Substitute Bill that passed both houses. On May 14, 2003, Governor Gary Locke signed the bill into law. The law became effective on July 27, 2003. This law is the result of a two-year collaborative drafting process that involved patients, mental health care providers, attorneys, and legislators.

Senator Keiser first sponsored this bill when a family with a schizophrenic son asked her to propose advance directive legislation. Their son was able to function when he was taking his medication, but periodically he would stop taking it. When this happened, he often ended up on the streets of downtown Seattle in a state of confusion, where he would threaten passers-by. Sometimes the young man would find himself in jail or in the emergency room. He and his family wished there were some way to break this cycle and get him help earlier when he first presented symptoms. The family wanted to be able to

---

61. Mental Health Advance Directives, supra note 3, § 5.
62. Id.
66. See Mental Health Advance Directives, supra note 3.
67. Id.
68. Telephone Interview with Debra Srebnik, Professor, University of Washington Department of Psychiatry and Behavioral Sciences (Oct. 15, 2002) (Professor Srebnik was a frequent participant in advance directive stakeholder meetings).
70. Speck Email, supra note 35.
71. Id.
72. Id.
73. Id.
secure the treatment their son needed earlier, and they felt that this legislation would allow them to do so.\footnote{Id.}

The new law gives patients a statutory right to execute mental health care advance directives that will be valid if certain formalities are met.\footnote{Mental Health Advance Directives, supra note 3, \S 5(2).} The pertinent section of the law provides:

A directive shall:

(a) Be in writing;

(b) Contain language that clearly indicates that the principal intends to create a directive;

(c) Be dated and signed by the principal or at the principal’s direction in the principal’s presence if the principal is unable to sign;

(d) Designate whether the principal wishes to be able to revoke the directive during any period of incapacity or wishes to be unable to revoke the directive during any period of incapacity; and

(e) Be witnessed in writing by at least two adults, each of whom shall declare that he or she personally knows the principal, was present when the principal dated and signed the directive, and that the principal did not appear to be incapacitated or acting under fraud, undue influence, or duress.\footnote{Id. \S 6.}

This new Washington law allows principals to make many decisions for themselves. For example, the law permits patients to choose when their directive will become effective.\footnote{Id. \S 6(3).} It also includes a sample form that patients can use to create their own directives.\footnote{Id. \S 26.} This form includes a place for patients to indicate, at the time the directive is executed, whether the directive will be revocable or irrevocable during subsequent incapacity.\footnote{Id. \S 26 (Part IV).}

Patients or their families can challenge the creation of a mental health care advance directive in court.\footnote{Id. \S 20 (stating that “[a]ny person with reasonable cause to believe that a directive has been created or revoked under circumstances amounting to fraud, duress, or undue influence may petition the court for appointment of a guardian for the person or to review the actions of the agent or person alleged to be involved in improper conduct under RCW 11.94.090 or 74.34.110”).} The legislation specifically authorizes

\begin{itemize}
\item \footnote{Id.}
\item \footnote{Mental Health Advance Directives, supra note 3, \S 5(2).}
\item \footnote{Id. \S 6.}
\item \footnote{Id. \S 6(3).}
\item \footnote{Id. \S 26.}
\item \footnote{Id. \S 26 (Part IV).}
\item \footnote{Id. \S 20 (stating that “[a]ny person with reasonable cause to believe that a directive has been created or revoked under circumstances amounting to fraud, duress, or undue influence may petition the court for appointment of a guardian for the person or to review the actions of the agent or person alleged to be involved in improper conduct under RCW 11.94.090 or 74.34.110”).}
\end{itemize}
Mental Health Advance Directives

a principal—the person who signs an advance directive—to bring an action to contest the validity of his or her directive. For example, patients could argue that their advance directives are not valid because they were signed under duress, fraud, or undue influence. Patients could also argue that they were not competent at the time they signed the directive. Further, families could potentially challenge the directive’s validity by arguing that it was signed unknowingly or unintelligently.

An irrevocable mental health advance directive signed under the new Washington law is, in effect, a waiver of the patient’s right to make future health care decisions because it involves more than merely giving advance consent for certain medical treatments. There is a distinction between giving consent for treatment and giving up the right to refuse treatment. With a traditional consent form, a patient can revoke consent at any time. However, with an irrevocable advance directive, the patient gives up his or her future right to withdraw consent, effectively waiving the future right to refuse medical treatment during subsequent incapacity. At the time the directive is signed, the “present self” waives a right that the “future self” would otherwise possess—the right to refuse treatment.

In sum, in response to public demand, the Washington State Legislature has passed a new law to ensure that mental health advance directives are enforceable and to clarify the requirements for executing such a directive. The law permits patients to execute mental health advance directives consenting to certain treatments in advance. Patients have the right to challenge their advance directives in court on the

81. Id. § 12.
82. See id. § 20.
83. See id. § 12.
84. Id. § 20.
85. Dresser, supra note 6, at 819 (arguing that signing a voluntary commitment contract involves a waiver of rights).
86. See Winick, supra note , at 70–71.
88. See Mental Health Advance Directives, supra note 3, § 6.
89. Dresser, supra note 6, at 819.
90. Mental Health Advance Directives, supra note 3, § 5.
91. Id.
II. PATIENTS HAVE A FUNDAMENTAL RIGHT TO REFUSE MEDICAL TREATMENT THAT CAN ONLY BE WAIVED VOLUNTARILY, KNOWINGLY, AND INTELLIGENTLY

Fundamental rights are rights that are constitutionally protected because they are vital to a democratic society. One of the rights the U.S. Supreme Court has deemed fundamental is the right to refuse medical treatment. Fundamental rights cannot be waived unless a person’s waiver is voluntary, knowing, and intelligent. Therefore, patients can only waive the fundamental right to refuse medical treatment if the waiver is made knowingly, voluntarily, and intelligently.

A. Fundamental Rights Are Protected by the Constitution

The U.S. Supreme Court considers some rights to be fundamental. As Justice Cardozo explained, a fundamental right represents the “very essence of a scheme of ordered liberty . . . a ‘principle of justice so rooted in the traditions and conscience of our people as to be ranked as fundamental.’” Some fundamental rights are expressly stated in the Bill of Rights, while others have been held to originate in the concept

92. Id. §§ 12, 20.
93. See Dresser, supra note 6, at 819.
98. Palko, 302 U.S. at 325–26 (alteration in original); see also Moore v. City of E. Cleveland, 431 U.S. 494, 503 (1977) (characterizing fundamental rights as those liberties that are “deeply rooted in this Nation’s history and tradition”).
99. U.S. CONST. amends. I–X. For example, the Court has held that the right to free speech and the right to free exercise of religion are fundamental rights stemming from the Bill of Rights. See O’Lone v. Estate of Shabazz, 482 U.S. 342, 348 (1987).
Mental Health Advance Directives

of “liberty” under the Due Process Clause of the Fourteenth Amendment. Certain fundamental rights protect individuals from overreaching by the government, like the right to be free from government intrusion or to have assistance of counsel at a criminal trial. Other rights protect people’s bodily integrity, such as the rights to have children, to marital privacy, to use contraception, and to terminate one’s pregnancy.

The U.S. Supreme Court has held that fundamental rights and liberties deserve more protection and require a greater degree of vigilance than do other rights. The Due Process Clause of the Fourteenth Amendment provides “heightened protection against government interference with certain fundamental rights and liberty interests.” When a state statute implicating a fundamental right is challenged, the Court applies the “strict scrutiny” standard of review. Under this standard, the Court will strike down the law infringing on the fundamental right unless the state can demonstrate that the law is justified by a compelling state interest and is narrowly tailored.

B. The Right to Refuse Medical Treatment Is a Fundamental Right

The U.S. Supreme Court has recognized that the fundamental liberty interest under the Due Process Clause includes the right to refuse unwanted medical treatment. As early as 1891, the Court noted that

100. See Palko, 302 U.S. at 325; see also Charfauros v. Bd. of Elections, 249 F.3d 941, 951 (9th Cir. 2001) (stating that the Due Process Clause is the basis for the fundamental rights to get married and to vote); Konvitz, supra note 94, at 13.


105. Glucksberg, 521 U.S. at 720.


“no right is held more sacred” than the right of every individual to the possession and control of his or her own person. \(^{110}\) This “sacred” right includes the right to refuse medical treatment. \(^{111}\)

The U.S. Supreme Court first recognized the right to refuse medical treatment in *Cruzan v. Director, Missouri Department of Health*. \(^{112}\) In *Cruzan*, a young woman was in a terrible car accident that left her on permanent life support. \(^{113}\) The young woman’s parents wanted to remove her from the artificial life support, arguing that she had virtually no chance of recovering her cognitive faculties. \(^{114}\) The State of Missouri refused to allow her family to take her off life support. \(^{115}\) The State wanted to require extra evidence beyond the family’s testimony to ensure that the patient herself, if competent, would have wanted to be removed from life support. \(^{116}\) The Court agreed and concluded that a state could require more evidence of a patient’s desire to be removed from life support than her parents’ belief. \(^{117}\)

In *Cruzan*, the Court suggested that the right to refuse treatment is closely related to the requirement of informed consent. \(^{118}\) Patients must give informed consent before receiving medical treatment, because the forcible treatment of a non-consenting person represents a “substantial interference with that person’s liberty.” \(^{119}\) As the Court reasoned, if patients must give informed consent, then, by a “logical corollary,” a patient necessarily has the right to refuse to give consent. \(^{120}\)

Recognizing the young woman’s right to refuse treatment, the *Cruzan* Court noted that even prisoners possess a significant liberty interest in avoiding unwanted administration of antipsychotic drugs under the Due Process Clause of the Fourteenth Amendment. \(^{121}\) Justice O’Connor,


\(^{113}\) Id. at 265.

\(^{114}\) Id.

\(^{115}\) Id.

\(^{116}\) Id.

\(^{117}\) Id. at 280.

\(^{118}\) Id. at 269 (reasoning that informed consent is needed to protect the liberty interests inherent in the Fourteenth Amendment).


\(^{120}\) *Cruzan*, 497 U.S. at 270.

\(^{121}\) *Harper*, 494 U.S. at 221–22.
Mental Health Advance Directives

concurring in the judgment, reasoned that the liberty guaranteed by the Due Process Clause must protect an individual’s deeply personal decision to reject medical treatment.122 The Court recognized the “principle that a competent person has a constitutionally protected liberty interest in refusing unwanted medical treatment.”123 Applying this principle to the case at hand, the Court held that Missouri could require stronger evidence of what the patient would have wanted, other than her parents’ belief, had she been competent.124

Individual states have also statutorily recognized the fundamental right to refuse treatment.125 For example, the Washington State Legislature recognized the fundamental right to refuse treatment in the Natural Death Act, which states: “[t]he Legislature finds that adult persons have the fundamental right to control the decisions relating to the rendering of their own health care.”126

People with mental illnesses have a limited right to refuse medical treatment, although the actual scope of an incompetent person’s right to refuse treatment is unclear.127 Still, the U.S. Supreme Court has recognized that, at a minimum, incompetent persons retain a basic right to refuse treatment.128 Some lower courts have held that incompetent persons have the right to refuse medical treatment absent an emergency situation129—such as when they pose a danger to themselves or others.130

122.   

123.   

124.   


126. Id.  

127. See Fleischner, supra note 14, at 790.  

128. The U.S. Supreme Court has focused on the right of an incompetent person to refuse treatment in the criminal context. See, e.g., Riggins v. Nevada, 504 U.S. 127, 135 (1992) (stating that forced medication of a pretrial detainee during trial deprived him of a fair trial); Washington v. Harper, 494 U.S. 210, 229 (1990) (noting that an incompetent prisoner had a significant constitutional due process interest in avoiding unwanted administration of mental drugs).  

129. See Rogers v. Okin, 634 F.2d 650, 653 (1st Cir. 1980), vacated and remanded sub nom. Mills v. Rogers, 457 U.S. 291 (1982), on remand, 738 F.2d 1 (1st Cir. 1984); see also Guardianship of Roe, 421 N.E.2d 40, 42 (Mass. 1981) (holding that an incompetent mentally ill patient may have a right to refuse medication based on the constitutional right to privacy and on the common law).  

If patients are incompetent and cannot make any treatment decision, the patients’ guardians can still refuse treatment on their behalf.131 Washington courts also adhere to this principle.132 The Washington State Supreme Court has recognized that an “incompetent’s right to refuse treatment should be equal to a competent’s right to do so.”133 As noted previously, in Washington a person retains the right to refuse treatment until he or she meets the strict definition of “gravely disabled.”134 In addition, the Washington State Supreme Court has held that an incompetent person’s expressed wish to refuse treatment must be given substantial weight, even if made while the person is incompetent.135

C. Fundamental Rights Cannot Be Waived Unless the Waiver Is Voluntary, Knowing, and Intelligent

The U.S. Supreme Court has been reluctant to recognize waivers of fundamental rights.136 The Court has instructed lower courts to indulge all reasonable presumptions against the waiver of fundamental constitutional rights.137 The United States Court of Appeals for the Ninth Circuit has held that the waiver of a constitutional right is “not to be implied and is not lightly to be found.”138 Therefore, courts will presume that a person has not waived a fundamental right absent clear evidence of the person’s intent, and courts are extremely hesitant to recognize an implied or inferred waiver.139

The U.S. Supreme Court applies a three-part test to determine whether a person has made a valid waiver of a fundamental right: a

135. Guardianship of Ingram, 102 Wash. 2d at 840, 689 P.2d at 1370.
138. Ostland v. Bobb, 825 F.2d 1371, 1373 (9th Cir. 1987) (quoting United States v. Provencio, 554 F.2d 361, 363 (9th Cir. 1977)).
Mental Health Advance Directives

waiver is only valid if it was made voluntarily, knowingly, and intelligently.\textsuperscript{140} This three-part test has developed gradually over time. In 1938, the Court established the “knowing” requirement in \textit{Johnson v. Zerbst}.\textsuperscript{141} The Court later required valid waivers to be made knowingly and voluntarily in \textit{Brady v. United States}.\textsuperscript{142} Finally, the Court added the “intelligent” requirement in both \textit{Brady}\textsuperscript{143} and \textit{Miranda v. Arizona}.\textsuperscript{144} This three-part test for waiver of fundamental constitutional rights applies equally to criminal and civil cases.\textsuperscript{145}

The U.S. Supreme Court began developing the three-part waiver test in \textit{Johnson}, where the Court held that criminal defendants must make knowing waivers of their fundamental rights to counsel\textsuperscript{146}. In \textit{Johnson}, two U.S. Marines were convicted of passing counterfeit twenty-dollar bills.\textsuperscript{147} The soldiers, who were not represented by counsel during their trial, argued on appeal that they were not even aware they had a right to counsel.\textsuperscript{148} Therefore, they claimed their Sixth Amendment rights were violated.\textsuperscript{149} The government argued that the two soldiers waived their constitutional right to counsel by not asking the trial judge to appoint lawyers for them.\textsuperscript{150}

Ruling in favor of the soldiers, the \textit{Johnson} Court noted that the trial judge has a responsibility to determine on the record whether a defendant has made a competent waiver of the right to counsel.\textsuperscript{151} The Court stated that that waiver of a fundamental constitutional right should

\begin{thebibliography}{12}
\bibliographystyle{chicago}
\bibitem{141} 304 U.S. 458 (1938).
\bibitem{142} 397 U.S. 742, 748 (1970).
\bibitem{143} \textit{Id}.
\bibitem{144} 384 U.S. 436, 444 (1966).
\bibitem{146} \textit{Johnson}, 304 U.S. at 464.
\bibitem{147} \textit{Id} at 460.
\bibitem{148} \textit{Id} at 467.
\bibitem{149} \textit{Id} at 467–68.
\bibitem{150} See \textit{id} at 460, 464.
\bibitem{151} \textit{Id} at 465.
\end{thebibliography}
be “an intentional relinquishment or abandonment of a known right.”

Therefore, if the soldiers did not even know they had a right to request counsel, they could not have validly waived that right.

The Court added another requirement to ensure valid waivers in *Brady* by holding that waivers of fundamental rights must also be made voluntarily. *Brady*, fearing that a co-defendant who had already plead guilty would testify against him, plead guilty to kidnapping, thus avoiding the death penalty. On appeal, Brady argued that his right to plead “not guilty” was waived involuntarily because he was coerced under the pressure of the possible application of the death penalty. The Court rejected this argument, relying on the fact that Brady was twice questioned by the trial court judge as to the voluntariness of his plea, as well as the fact that he was represented by competent counsel.

Therefore, a person’s waiver of a fundamental right is not valid if it is coerced under force or threat of force. For example, waivers of fundamental rights cannot be obtained by the threat of continued incarceration. The Court has noted several factors that may indicate when a waiver of a fundamental right is not voluntary, including “coercion, terror, inducements, [and] subtle or blatant threats.”

The *Brady* Court also reaffirmed that waivers of fundamental rights must be made intelligently, and held that “[w]aivers of constitutional rights not only must be voluntary but must be knowing, intelligent acts.” The Court relied on *Miranda* in requiring intelligent waivers.

---

152. *Id.* at 464.
154. *Brady v. United States*, 397 U.S. 742, 748 (1970); *see also Miranda v. Arizona*, 384 U.S. 436, 444 (1966) (noting that “[t]he defendant may waive effectuation of these rights, provided the waiver is made voluntarily”).
156. *Id.* at 744.
157. *Id.* at 743.
158. *Id.* at 749.
161. *Boykin*, 395 U.S. at 243; *see also DANIEL PENOFSKY, GUIDELINES FOR INTERROGATIONS: WAIVER OF RIGHTS UNDER MIRANDA* 92, 114 (1967) (noting that other factors indicative of an involuntary waiver include physical abuse, psychological coercion, trickery, or cajolery).
162. *See Brady*, 397 U.S. at 748.
163. *Id.* at 748.
Mental Health Advance Directives

An intelligent waiver of rights requires that the party waiving the right understands the consequences of the waiver. Factors that courts consider when deciding if a waiver is intelligently made include the age, educational background, and mental state of the person waiving the right. Depression, for example, may be enough to make a waiver unintelligent because “[d]epression can impair a patient’s ability to understand information, to weigh alternatives, and to make a judgment that is stable.”

D. One Federal Court Has Applied the Knowing, Voluntary, and Intelligent Test to Waivers of the Fundamental Right to Refuse Medical Treatment

Although one federal court has applied the three-part waiver test to a waiver of the right to refuse medical treatment, no other federal courts have specifically addressed the issue. In Rogers v. Okin, a federal district court for the District of Massachusetts held that a patient’s signature on a form, by itself, was not sufficient to meet the knowing, voluntary, and intelligent waiver test. In Rogers, all patients were required to sign a form before being admitted to a mental hospital in Massachusetts, which stated: “I understand that during my hospitalization and after any care, I will be given care and treatment which may include the injection of medicines.” Patients who had

165. See Hardling v. Lewis, 834 F.2d 853, 857 (9th Cir. 1987).
166. PENOFSKY, supra note 161, at 115.
170. Id. at 1368. Although other parts of the district court’s opinion were overturned on appeal, on subsequent remand neither the U.S. Court of Appeals for the First Circuit nor the U.S. Supreme Court disturbed the district court’s holding that there was not a valid waiver of the right to refuse medical treatment under the facts of the case. See Rogers, 634 F.2d 650; Mills v. Rogers, 457 U.S. 291 (1982).
172. Id. at 1367.
signed this form argued that the form’s language was so vague that it could not constitute a “knowing[,] voluntary waiver of a constitutional right to refuse treatment.”\textsuperscript{173} The district court agreed that merely signing this form did not constitute a knowing waiver by the patients, and held that “[i]n order for a court to find a waiver of a right to refuse [treatment], the evidence must be clear that the patient understood such a right existed and then elected knowingly and voluntarily to waive such a right."\textsuperscript{174} The court concluded that the patients had no real choice about whether to sign the consent form—it was similar to a “take it or leave it” adhesion contract—and that the written warning the patients signed was inadequate to notify them that they were waiving fundamental rights.\textsuperscript{175} Under these facts, the court held the waiver was invalid because the knowing and voluntary prongs of the waiver test were not met.\textsuperscript{176} The opinion in Rogers did not specifically address the intelligent prong.\textsuperscript{177}

Thus, the U.S. Supreme Court has adopted a three-part waiver test in order to protect fundamental constitutional rights. For a person to validly waive a fundamental right, the waiver must be made knowingly, voluntarily, and intelligently. Consequently, it is likely that such a waiver is required for a person to give up the fundamental right to refuse medical treatment.

III. RULES, COURT-MANDATED PROCESSES, AND STATUTES PROVIDE SAFEGUARDS TO GUARANTEE VOLUNTARY, KNOWING, AND INTELLIGENT WAIVERS OF FUNDAMENTAL RIGHTS

There are many contexts in which special procedural safeguards are in place to ensure that knowing, voluntary, and intelligent waivers are made. For example, the Federal Rules of Criminal Procedure allow defendants to waive the fundamental rights that accompany a criminal trial, but include a process to ensure that such waivers meet the “knowing, voluntary, and intelligent” test.\textsuperscript{178} Similarly, the U.S. Supreme Court requires police officers to warn arrestees of their

\textsuperscript{173} \textit{Id.} at 1368.
\textsuperscript{174} \textit{Id.}
\textsuperscript{175} \textit{Id.}
\textsuperscript{176} \textit{Id.}
\textsuperscript{177} \textit{Id.}
\textsuperscript{178} \textsc{Fed. R. Crim. P.} 11(b)(2).
Mental Health Advance Directives

fundamental rights to assistance of counsel and to remain silent before they can validly waive any of those rights. 179 Finally, statutes like Oregon’s Death with Dignity Act, which allow a person to waive the fundamental right to life, contain procedural safeguards in order to ensure knowing, voluntary, and intelligent waivers of that right. 180

A. The Federal Rules of Criminal Procedure Ensure That a Defendant’s Waiver of Fundamental Trial Rights Is Made Knowingly, Voluntarily, and Intelligently

An accused individual has a fundamental right to a trial by jury, a privilege against compulsory self-incrimination, and a right to confront his or her accusers. 181 Entering a plea of guilty or no-liam contedere is, in effect, a waiver of all of these rights that accompany a trial. 182 Therefore, a defendant’s unknowing and involuntary guilty plea will be void. 183 Thus, the Federal Rules of Criminal Procedure require federal courts to ensure that defendants’ waivers of trial rights are made knowingly, voluntarily, and intelligently. 184

First, Federal Rule of Criminal Procedure 11 requires procedural safeguards to ensure that federal defendants make knowing waivers of their rights. 185 The judge must address the defendant in open court and explain specific facts concerning the charge to which the defendant has pleaded guilty and the rights he or she has waived. 186 This requirement guarantees that defendants know precisely what rights are at stake. 187

181. U.S. CONST. amend. VI (guaranteeing defendants the right to trial by an impartial jury and to be confronted with witnesses against them); see also City of Mobile v. Bolden, 446 U.S. 55, 77 n.24 (1980) (concluding that defendant in a criminal case has the “fundamental” right to trial by a jury of his peers).
182. See Dresser, supra note 6, at 816 (arguing that entering a plea of guilty under FED. R. CRIM. P. 11(c)(4) before trial is “a waiver of all the safeguards that accompany a trial”).
184. Id. at 466.
185. See FED. R. CRIM. P. 11(b)(1)(H)–(I) (requiring a judge to tell the defendant the maximum and minimum penalties the defendant will be subject to should the defendant choose to plead guilty).
187. See FED. R. CRIM. P. 11(b)(1)(B)–(E) (requiring judges to tell defendants what rights are at stake).
Second, in addition to informing the defendant of his or her rights, courts also must ensure that a defendant’s waiver is made voluntarily. Under Rule 11, the court must determine whether a defendant’s guilty plea is “voluntary and did not result from force, threats or promises” other than a plea agreement. Courts may only admit a defendant’s confession into evidence when they are satisfied that the confession was made voluntarily. Many states require that the judge’s determination of whether there has been an “effective waiver of the right to plead not guilty” be made on the record.

Finally, Rule 11 includes requirements to ensure intelligent waivers. When a defendant has chosen to plead guilty as part of a plea bargain agreement reached with the prosecution, the judge must “advise the defendant that the defendant has no right to withdraw the plea if the court does not follow [the prosecution’s] recommendation or request.” This procedural safeguard helps to ensure that the defendant waives the right to trial only after being fully informed of the consequences and risks of the waiver. The scope of Rule 11 was considerably expanded in 1973 to reflect the intent of the original rule which stated simply that a court “shall not accept . . . [a plea of guilty] without first addressing the defendant personally and determining that the plea is made . . . with understanding of the nature of the charge and the consequences of the plea.” Therefore, Federal Rule of Criminal Procedure 11 ensures that a defendant’s waiver of the fundamental right to trial is made knowingly, voluntarily, and intelligently.

B. The Miranda Warnings in Criminal Interrogations Assure That a Defendant’s Waiver of Fundamental Civil Rights Is Made Knowingly, Voluntarily, and Intelligently

The U.S. Supreme Court adopted “prophylactic” rules in *Miranda v. Arizona* that protect suspects from invalidly waiving fundamental rights. Under the Sixth Amendment, criminal suspects have the

190. *See Boykin*, 395 U.S. at 242 (citing Jackson v. Denno, 378 U.S. 368, 387 (1964)).
191. *Id. at 244 n.6.*
Mental Health Advance Directives

fundamental right to assistance of counsel, and under the Fifth Amendment, suspects have the right to remain silent. When a criminal suspect is taken into custody, police officers must give five warnings before interrogating the suspect. At trial, the prosecution bears the burden of proving that a defendant’s waivers were valid.

Miranda warnings protect criminal suspects from unknowingly waiving fundamental rights. Police officers must tell suspects specifically what their rights are. The Miranda warnings require officers to tell suspects “[y]ou . . . have the right to remain silent” and “[y]ou are entitled to consult with an attorney before interrogation.” Courts will not presume that criminal defendants knew that they had a right to remain silent in the absence of a Miranda warning. These warnings ensure that suspects and defendants are waiving known rights.

In addition, Miranda warnings protect criminal suspects from involuntarily waiving fundamental rights. Once a suspect is prosecuted, courts will examine whether a suspect’s decision to waive a right was voluntary in light of the totality of the circumstances. Miranda warnings are one factor that courts will take into consideration because “[o]nce warned, the suspect is free to exercise his own volition in deciding whether or not to make a statement to the authorities.”

196. See U.S. Const. amend. V.
197. PENOFSKY, supra note 161, at 14 (noting that the five required warnings are: (1) that the accused has the right to remain silent; (2) that any statement made can be used against the accused; (3) that the accused has the right to consult with an attorney; (4) that the accused has the right to have an attorney present during any interrogation; and (5) if the accused cannot afford an attorney, one will be appointed for him).
198. See id. at 7.
200. See id.
202. See Miranda, 384 U.S. at 468 (explaining that “the Fifth Amendment privilege is so fundamental to our system of constitutional rule and the expedient of giving an adequate warning as to the availability of the privilege so simple, we will not pause to inquire in individual cases whether the defendant was aware of his rights [to remain silent] without a warning being given”).
203. See PENOFSKY, supra note 161, at 15.
204. Miranda, 384 U.S. at 475.
Miranda warnings were designed to protect fundamental rights from government “compulsion, subtle or otherwise” that “operates on the individual to overcome free choice in producing a statement.”207 Later judicial review ensures that a suspect’s waivers were made voluntarily.208

Finally, Miranda warnings protect criminal suspects from unintelligently waiving fundamental rights.209 Police officers inform suspects of the potential consequences of a decision to waive the right to silence when they state that “[a]nything you say can and will be used against you in a court of law.”210 Thus, the Miranda warnings protect criminal suspects from invalidly waiving their fundamental rights to assistance of counsel and to remain silent.211

C. Oregon’s “Death with Dignity” Statute Ensures That a Patient’s Waiver of the Fundamental Right to Life Is Made Knowingly, Voluntarily, and Intelligently

Statutes allowing citizens to waive fundamental rights can include procedural protections to guarantee that a waiver is knowingly, voluntarily, and intelligently made.212 For example, Oregon’s Death with Dignity Act allows patients with terminal conditions to waive their right to life by taking prescribed medication that will cause their own death.213 The right to life is a fundamental right.214 Therefore, Oregon has included special protections in its Death with Dignity Act to ensure that any waiver of the fundamental right to life meets the U.S. Supreme Court’s three-part waiver test.215

Oregon’s legislation—originally an initiative that voters passed—includes safeguards to protect vulnerable individuals.216 Similar ballot

207. Miranda, 384 U.S. at 474.
208. Schneckloth, 412 U.S. at 227.
209. Miranda, 384 U.S. at 475.
210. GRISSO, supra note 201, at 11.
211. Miranda, 384 U.S. at 475.
212. See, e.g., OR. REV. STAT. § 127.815(1)(c) (2003) (requiring physicians to make sure patient is “making an informed decision” before physician can prescribe life-ending medication to that patient).
213. Id.
215. See OR. REV. STAT. § 127.815.
measures failed in California and Washington because opponents successfully convinced voters that the bills did not contain adequate safeguards.\textsuperscript{217} The drafters of the subsequent Oregon measure took this into consideration.\textsuperscript{218} They knew that Oregon voters would be more likely to support the bill if it contained protections to ensure that people would seriously consider their decision to end their own lives, and to guarantee that this decision would be made in light of all the relevant information.\textsuperscript{219}

As a result, Oregon’s Death with Dignity Act provides in relevant part that before a physician can prescribe life-ending medication:

(1) The attending physician shall:

(a) Make the initial determination of whether a patient has a terminal disease, is capable, and has made the request voluntarily;

(c) To ensure that the patient is making an informed decision, inform the patient of:

(A) His or her medical diagnosis;
(B) His or her prognosis;
(C) The potential risks associated with taking the medication to be prescribed;
(D) The probable result of taking the medication to be prescribed; and
(E) The feasible alternatives, including, but not limited to, comfort care, hospice care and pain control;
(d) Refer the patient to a consulting physician for medical confirmation of the diagnosis, and for a determination that the patient is capable and acting voluntarily;
(e) Refer the patient for counseling if appropriate . . . ;

(h) Inform the patient that he or she has an opportunity to rescind the request at any time and in any manner.\textsuperscript{220}

\textsuperscript{217} Id.
\textsuperscript{218} Id.
\textsuperscript{219} Id.
\textsuperscript{220} OR. REV. STAT. § 127.815.
Although the Oregon statute specifically addresses the voluntary and intelligent prongs, it does not contain provisions specifically aimed at ensuring “knowing” waivers, such as a provision requiring physicians to tell patients that they have a fundamental right to life.\textsuperscript{221} Perhaps the drafters assumed that it is common knowledge that the right to continue living is fundamental.\textsuperscript{222}

Oregon’s Death with Dignity Act does include protections to guarantee that waivers of the fundamental right to life are made voluntarily.\textsuperscript{223} The physician must make a good faith effort to determine whether the patient is voluntarily asking to give up his or her right to life.\textsuperscript{224} Moreover, Oregon’s statute addresses the intelligent waiver prong.\textsuperscript{225} Oregon requires health care providers to sit down with a patient who is considering physician-assisted suicide and explain the potential risks of the decision.\textsuperscript{226} When patients choose to be removed from life support or to take a prescribed medication to end their own lives, a physician is required to explain the options, alternatives, and risks.\textsuperscript{227} These requirements ensure that any waiver is made intelligently.

Oregon requires a face-to-face verbal warning from the doctor before patients are allowed to waive their rights.\textsuperscript{228} By requiring the attending physician to have this conversation with a person who is considering waiving his or her right to life, Oregon’s Death with Dignity Act protects its citizens from unknowingly, involuntarily, or unintelligently waiving their rights.

Thus, Federal Rule of Criminal Procedure 11, \textit{Miranda} warnings, and Oregon’s Death with Dignity Act all include provisions to ensure knowing, voluntary, and intelligent waivers of fundamental rights. Under Rule 11, a federal judge must tell criminal defendants what rights they forfeit by pleading guilty.\textsuperscript{229} Likewise, \textit{Miranda} warnings require

\begin{itemize}
\item \textsuperscript{221} \textit{Id.} § 127.815(1)(a)–(c).
\item \textsuperscript{222} \textit{See} Woods v. Niertheimer, 328 U.S. 211, 216 (1946) (noting that the “fundamental rights to life and liberty are guaranteed by the United States Constitution”).
\item \textsuperscript{223} \textit{OR. REV. STAT.} § 127.815(1)(a).
\item \textsuperscript{224} \textit{See Grant & Linton, supra note 167, at 528.}
\item \textsuperscript{225} \textit{OR. REV. STAT.} § 127.815(1)(c)(A)–(E).
\item \textsuperscript{226} \textit{See id.} § 127.815(1)(c)(C).
\item \textsuperscript{227} \textit{See id.} § 127.815(1)(a)–(c).
\item \textsuperscript{228} \textit{See id.} While the Oregon statute does not explicitly require that the communication be face-to-face, it is reasonable to assume that most doctors talk with their patients in person before prescribing medication to the patient allowing the patient to end his or her own life.
\item \textsuperscript{229} \textit{FED. R. CRIM. P.} 11(a)(1)(B)–(E).
\end{itemize}
Mental Health Advance Directives

police officers to tell suspects what their rights are and explain the consequences of waiving those rights. Finally, Oregon’s Death with Dignity Act requires physicians to meet with patients to discuss the risks and alternatives before the physician can prescribe lethal medication for patients wanting to waive their right to life. These procedures help ensure that waivers of fundamental rights meet the U.S. Supreme Court’s three-part test.

IV. WASHINGTON’S ADVANCE DIRECTIVE LAW SHOULD BE AMENDED TO INCLUDE PROTECTIONS AT THE EXECUTION STAGE THAT WILL GUARANTEE KNOWING AND INTELLIGENT WAIVERS

Although Washington’s new law authorizing mental health advance directives provides many benefits, it may be unconstitutional as passed because it does not ensure that waivers of the fundamental right to refuse treatment will be made knowingly and intelligently. With minor changes to the new law, the Washington Legislature could fix this shortcoming and guarantee that patients’ waivers of fundamental rights will meet the U.S. Supreme Court’s three-part test.

A. Washington’s Mental Health Advance Directive Law Contains Insufficient Safeguards in the Execution Stage to Meet the “Knowing, Voluntary, and Intelligent” Test

Washington’s mental health advance directive law fails to guarantee that patients will make knowing and intelligent waivers of their fundamental right to refuse medical treatment. Under the law, patients could sign an irrevocable advance directive that effectively constitutes a waiver of their right to refuse treatment in the future. Although the Washington law includes sufficient safeguards to ensure that such a waiver is voluntary, it fails to guarantee that a waiver is made knowingly

231. OR. REV. STAT. § 127.815(1)(a)-(c).
232. Interview with Lisa Brodoff, Professor, Seattle University School of Law, in Seattle, Wash. (Nov. 14, 2002) (explaining that the drafters of the proposed legislation tried to ensure a great deal of choice to the mentally ill); see, e.g., Mental Health Advance Directives, supra note 3, § 6(3), (allowing patients the freedom to choose the point in time when their directive will become effective).
and intelligently in accordance with the Johnson test.\textsuperscript{233} In fact, the bill provides even less protection against unknowing waivers than the protections held insufficient in Rogers.\textsuperscript{234}

Signing an irrevocable advance directive constitutes a waiver of a patient’s fundamental rights.\textsuperscript{235} Even incompetent persons retain the fundamental right to refuse treatment.\textsuperscript{236} In Washington, a patient retains the right to refuse treatment until the person meets the strict definition of “gravely disabled.”\textsuperscript{237} Signing an irrevocable advance directive can cause a patient to effectively give up his or her right to refuse treatment in the future.\textsuperscript{238} Therefore, signing an irrevocable mental health advance directive constitutes a waiver of a fundamental right. Because Washington’s mental health advance directive law allows patients to waive their future right to refuse treatment, it must include safeguards to ensure that such waivers meet the knowing, voluntary, and intelligent test.\textsuperscript{239}

The drafters of Washington’s mental health advance directive law adequately addressed one part of the U.S. Supreme Court’s three-part “knowing, voluntary, and intelligent” test.\textsuperscript{240} The law ensures that patients will make voluntary waivers by requiring two adult witnesses to attest to the fact that the patient signed voluntarily at the time the advance directive was executed.\textsuperscript{241} The witnesses must certify that the

\begin{footnotesize}
\begin{enumerate}
    \item See supra note 89 and accompanying text.
    \item Wash. Rev. Code § 71.05.030 (2003); see also In re Anderson, 17 Wash. App. 690, 692, 564 P.2d 1190, 1192 (1977).
    \item Dresser, supra note 6, at 819 (arguing that signing a voluntary commitment contract, which is similar to an irrevocable advance directive, involves a waiver of rights).
    \item See Whitmore v. Arkansas, 495 U.S. 149, 165 (1990); Miranda v. Arizona, 384 U.S. 436, 444 (1966); Brookhart v. Janis, 384 U.S. 1, 4 (1966); Patton v. United States, 281 U.S. 276, 312 (1930); see also D.H. Overmyer Co. v. Friek Co., 405 U.S. 174, 185 (1972); Schell v. Witek, 218 F.3d 1017, 1023 (9th Cir. 2000); Gete v. INS, 121 F.3d 1285, 1293 (9th Cir. 1997); Leonard v. Clark, 12 F.3d 885, 889 (9th Cir. 1993); Davies v. Grossmont Union High Sch. Dist., 930 F.2d 1390, 1394 (9th Cir. 1991).
    \item Mental Health Advance Directives, supra note 3, § 6(1)(e).
    \item Id.
\end{enumerate}
\end{footnotesize}
Mental Health Advance Directives

patient did not appear incapacitated at the time of signing. Although the legislation protects against involuntary waivers, no provisions are included to guarantee that waivers of the right to refuse treatment are made knowingly or intelligently. As a result, Washington’s law contains insufficient safeguards to ensure that the advance directives will withstand constitutional scrutiny.

The Washington statute fails to meet the Johnson test that requires waivers to be made knowingly. Much like the two Marines in Johnson who did not know they had a fundamental right to counsel, some patients who execute mental health advance directives may not know they have a fundamental right to refuse treatment. Moreover, the new Washington law includes no requirement that those executing an advance directive be informed that they are giving up a fundamental right to refuse treatment. This is in striking contrast to Johnson where the U.S. Supreme Court required trial judges to inform defendants of their right to request counsel.

Further, the Washington law provides even less protection against an unknowing waiver of the fundamental right to refuse treatment than the mental hospital did in Rogers v. Okin. In Rogers, patients had to sign a form acknowledging that they would be given treatment once they entered the hospital. Yet, the court held that this form was not enough to constitute a knowing waiver of the right to refuse treatment. Thus, a court may require more than a mere written acknowledgment of rights to ensure that patients are knowingly waiving a right. The form included in the Washington law refers to potential treatment or medication in terms of “preferences,” but never mentions the waiver of a right to refuse treatment. The Washington law thus provides even less protection

242. Id. (requiring witnesses to certify that the person signing an advance directive does not appear to be acting under fraud, undue influence, or duress).
243. See id.
245. Id. at 467.
246. See Mental Health Advance Directives, supra note 3.
249. Id. at 1367.
250. Id.
251. See id.
252. Mental Health Advance Directives, supra note 3, § 26 (Part V).
against an unknowing waiver than the protection that the court held insufficient in *Rogers*. Therefore, the Washington law does not meet the knowing requirement set forth by the U.S. Supreme Court.

Additionally, Washington’s statute does not guarantee that patients will make *intelligent* waivers of their fundamental rights, because it does not ensure that patients will know the consequences, risks, and alternatives to signing an irrevocable advance directive. In contrast to Federal Rule of Criminal Procedure 11, which requires judges to warn defendants of the possible consequences of waiving the right to a trial, the Washington statute does not require someone to explain to patients the consequences of signing an irrevocable advance directive. Also, unlike Oregon’s Death with Dignity Act, which requires a physician to explain alternatives to a patient, Washington’s statute does not require anyone to explain to patients the alternatives to signing irrevocable advance directives. Therefore, the Washington law does not ensure that patients will intelligently waive the fundamental right to refuse treatment.

Thus, Washington’s mental health advance directive law fails to ensure that patients will make knowing and intelligent waivers of their fundamental right to refuse medical treatment. Under the new law, it would be possible for a fully competent person to sign an irrevocable advance directive without realizing that the document waived his or her fundamental right to refuse medical treatment at a later time. Under such circumstances, the directive would be open to legal challenges. For example, imagine that two witnesses certify that a patient is “competent” to sign a directive, but that patient is clinically depressed. It is certainly possible for two witnesses to be unaware of a patient’s clinical depression. Though the patient signs the directive voluntarily, the depression may be enough to make the waiver unknowing or unintelligent because “depression can impair a patient’s ability to understand information, to weigh alternatives, and to make a judgment that is stable.” Without more, Washington’s statute permits such a patient to waive the fundamental right to refuse medical treatment without even realizing it. This unfortunate possibility stands in stark

253. *See id.*
Mental Health Advance Directives

contrast to the expressed desires of those citizens who initially advocated for mental health advance directives in Washington.257

B. A “Rights Advocate” and a Written Warning Would Provide the Necessary Protections Against Unknowing Waiver

With two minor changes, the Washington State Legislature could guarantee that the advance directive law would ensure knowing and intelligent waivers. First, the statute should be amended to require a “rights advocate” to explain to the patient that an irrevocable directive is a waiver of the fundamental right to refuse treatment, and discuss the possible risks of and alternatives to using an advance directive. Second, the advance directive form should include a written warning that signing an irrevocable advance directive could constitute a waiver of the fundamental right to refuse treatment.

The Washington State Legislature should require a “rights advocate” to explain that the patient is waiving a fundamental right, and to discuss the potential consequences of that waiver before the patient signs an irrevocable advance directive. This rights advocate should be a disinterested person with no other relation to the principal who has a basic understanding of Washington law and of the fundamental right to refuse treatment. The rights advocate could be an attorney, a social worker, or a health care professional. Similar to Miranda warnings258 and Oregon’s Death with Dignity Act,259 this requirement should be mandatory when the patient selects an irrevocable advance directive in Washington.260 If the patient retains the right to revoke the directive at any time,261 the fundamental right to refuse medical treatment is not waived and a “rights advocate” is not needed. However, when a patient decides that the directive should be irrevocable during subsequent incapacity,262 a “rights advocate” can ensure that the patient understands that doing so waives the fundamental right to refuse medical treatment.

257. SREBINIK & BRODOFF, supra note 36, at 10 (noting that “a prominent concern of [potential advance directive users in Washington] was whether [they] would have sufficient information and competency to execute” advance directives).
258. See Miranda v. Arizona, 384 U.S 436, 475 (1966); see also supra note 197.
259. See OR. REV. STAT. § 127.815.
260. Mental Health Advance Directives, supra note 3, § 6(1)(d).
261. Id.
262. Id.
Essentially, the rights advocate should act like the physician in Oregon’s Death with Dignity Act, explaining the risks, consequences, and alternatives to signing an irrevocable advance directive. Thus, a rights advocate would bring an advance directive in line with the theory of informed consent. This will help guarantee that the patient’s waiver is intelligent and made in light of all the relevant information.

Once the rights advocate has informed the patient, the advocate should sign a certification stating that the patient has been advised of his or her rights. A court could refer to the certification if the directive is subsequently challenged. To accomplish this, the Washington Legislature should amend its mental health advance directive law to include the following:

If the principal has elected to be unable to revoke the directive during any period of incapacity, the directive shall include a certification by a rights advocate that states in substance as follows: “I am a disinterested attorney, social worker, or health care professional. I certify that I am familiar with the provisions of the Involuntary Treatment Act, the mental health advance directive authorized under this statute and the constitutionally protected fundamental right to refuse medical treatment. I met with the principal prior to the execution of this directive and advised him/her concerning his/her rights. In connection with this directive, I advised him/her about the risks involved, the possible legal and medical consequences of signing or not signing this directive, and available alternatives to signing the directive. In particular, I advised the principal that by signing this directive he/she was effectively waiving the fundamental right to refuse treatment in the future.”

This language will ensure that the patient is advised—by an impartial rights advocate—of the legal consequences of signing the directive, including the fact that it constitutes a waiver of a fundamental right.

A second change to the advance directive statute, adding a warning in the advance directive form, would also help prevent patients from signing advance directives without knowing they are waiving fundamental rights. The new statute includes a sample form directive

---

263. See OR. REV. STAT. § 127.815.
264. See Mental Health Advance Directives, supra note 3, §§ 12, 20.
265. Similar language was first proposed by Karen Boxx, Professor, University of Washington School of Law, at a Mental Health Advance Directives for Health Care stakeholders meeting in Olympia, Washington on December 3, 2002.
Mental Health Advance Directives

that patients can use.\textsuperscript{266} This form should be amended to include a
warning that signing an irrevocable directive may constitute a waiver of
a fundamental right. At present, the form requires those who want an
irrevocable directive to sign under a statement that says:
“I . . . understand that if I choose this option and become incapacitated
while this directive is in effect, I may receive treatment that I specify in
this directive, even if I object at the time.”\textsuperscript{267} This statement should be
expanded and clarified to include a second sentence that reads: “I further
understand that I have a fundamental constitutional right to refuse
medical treatment. I understand that by choosing to execute an
irrevocable directive I may be waiving my fundamental right to refuse
medical treatment.” Much like the explanations provided by rights
advocate, such a written warning would ensure that patients make
knowing waivers of the fundamental right to refuse treatment. Although
the rights advocate may be a more personal and effective means of
guaranteeing a knowing waiver, requiring both oral and written warnings
would create two levels of procedural protection.

C. By Making These Changes, the Washington State Legislature Can
Ensure That Mental Health Advance Directives Will Be
Constitutional and Enforceable

If the Washington advance directive statute is amended to include
these two changes, courts will most likely uphold irrevocable advance
directives as knowing, voluntary, and intelligent waivers of the
fundamental right to refuse treatment. As passed, the law satisfies the
voluntary requirement,\textsuperscript{268} but fails to meet the knowing and intelligent
requirements. The proposed changes will rectify this failure in multiple
ways.

First, the amended statute would ensure knowing waivers if it requires
a rights advocate to explain that patients are giving up a fundamental
right when they sign an irrevocable directive. This explanation of rights
is similar to police officers telling suspects that they “have the right to
remain silent” as part of a \textit{Miranda} warning.\textsuperscript{269} It is also akin to a judge

\textsuperscript{266} Mental Health Advance Directives, \textit{supra} note 3, § 26.
\textsuperscript{267} Id. § 26 (Part IV).
\textsuperscript{268} Id. § 6(1)(e); see also \textit{supra} note 241 and accompanying text.
\textsuperscript{269} See \textit{Miranda v. Arizona}, 384 U.S 436, 475 (1966); see also \textit{supra} note 197.
informing defendants that they have the right to a trial and do not have to plead guilty. The rights advocate would be able to guarantee that the patient knows about his or her rights. Thus, requiring a rights advocate would ensure that the Washington statute meets the Johnson test. Further, if the advance directive form states that the person is waiving a fundamental right, it will be more clear and specific than the form used in Rogers. Therefore, with these two changes, the law would ensure that patients make knowing waivers of their fundamental right to refuse treatment.

Additionally, the amended statute would ensure intelligent waivers if it requires a rights advocate to tell a patient about the risks, consequences, and alternatives of signing an irrevocable directive. Such information is analogous to the explanations that must be given by a physician under Oregon’s Death with Dignity Act, and by a judge acting under Rule 11. Adding a requirement for a rights advocate would allow patients to make an informed decision before signing an irrevocable directive. Therefore, the rights advocate would ensure that the patient intelligently waives his or her fundamental right to refuse treatment, as required by both Brady and Miranda.

Finally, these protections would make irrevocable directives more enforceable. If a patient becomes incompetent and subsequently tries to challenge his or her directive in court, the person’s signature on the form and the certification of the rights advocate would provide the court with evidence that the patient validly waived his or her right to refuse treatment.

V. CONCLUSION

The new Washington statute authorizing mental health advance directives stems from good intentions. However, there is a risk that the

276. Mental Health Advance Directives, supra note 3, §§ 12, 20.
Mental Health Advance Directives

statute could be successfully challenged as a violation of the U.S. Supreme Court’s requirement that fundamental rights can only be waived knowingly, voluntarily, and intelligently. By adding the simple procedural safeguards of a “rights advocate” at the signing stage, and a clear warning in the form, an amended version of the advance directive law should withstand a constitutional challenge. Consequently, Dr. Jekyll would be protected from inadvertently waiving Mr. Hyde’s fundamental right to refuse treatment.
THE BONDS OF JOINT TAX LIABILITY SHOULD NOT BE STRONGER THAN MARRIAGE: CONGRESSIONAL INTENT BEHIND § 6015(c) SEPARATION OF LIABILITY RELIEF

Svetlana G. Attestatova

Abstract: Spouses who file joint tax returns are jointly and severally liable for any resulting tax deficiency. In the past, only innocent spouses—those with no knowledge of the tax understatement—could qualify for relief from such liability. In 1998, Congress expanded existing innocent spouse relief and added two new forms of relief—the separation of liability and discretionary relief provisions. Codified at 26 U.S.C. § 6015(c), separation of liability relief allocates items that give rise to a deficiency to each spouse as if they had filed separate returns, and is only available to spouses who are divorced, separated, or living apart. However, a claimant spouse cannot obtain separation of liability relief if the Secretary of the Treasury can prove the claimant had actual knowledge of “any item giving rise to a deficiency.” Courts and the Department of the Treasury have interpreted this exception to refer to the transaction underlying a deficiency. In contrast, taxpayers and members of the Tax Section of the American Bar Association argue that this statutory language means an item on a joint return. This Comment examines the text, structure, and legislative history of § 6015(c), and concludes that there is strong support for the latter reading. Therefore, this Comment urges Congress and the Treasury to amend § 6015(c) and its implementing regulations to clarify Congress’ intended meaning.

Filing a joint tax return generally subjects spouses to joint and several liability.¹ Until 1971, this rule was absolute and resulted in cases of “grave injustice.”² The typical case of inequity arose when one spouse failed to report embezzled funds on a joint return, secretly squandered them, and then deserted the other spouse.³ Often the other spouse had derived no benefit from the funds and had no knowledge of the embezzler’s activities or the resulting omission on the tax return, but was nevertheless held jointly and severally liable for the resulting tax deficiency.

¹ 26 U.S.C. § 6013(d)(3) (2000). Joint and several liability is “[l]iability that may be apportioned either among two or more parties or to only one or a few select members of the group, at the adversary’s discretion.” BLACK’S LAW DICTIONARY 926 (7th ed. 1999). See also Richard C.E. Beck, The Innocent Spouse Problem: Joint and Several Liability for Income Taxes Should Be Repealed, 43 VAND. L. REV. 317, 319 n.2 (1990) (“This rule has been in the statute since 1938, but before that date it was controversial and the subject of litigation.”).
² See S. REP. NO. 91-1537, at 2 (1970) (citing Louise M. Scudder, 48 T.C. 36, 41 (1967) (recognizing the inflexibility and harshness of the statute when a wife was held liable for tax liability created by her husband’s failure to report embezzled funds)).
³ See id.
deficiency. In 1971, Congress recognized the injustice of holding such a spouse liable, and lessened the harshness of the joint and several liability rule. Congress enacted a relief provision that became known as innocent spouse relief, which alleviated liability for spouses who had no knowledge of and did not significantly benefit from the omitted income. This provision, codified at 26 U.S.C. § 6013(e), was intended “to bring government tax collection practices into accord with basic principles of equity and fairness.” Since the enactment of § 6013(e) in 1971, the scope of joint and several liability relief has been expanded twice—first in 1984, and again in 1998.

While inequity is apparent in the case of the embezzler spouse, it is not as clear in a case where one spouse receives retirement distributions, erroneously claims them as non-taxable on a joint return, and uses the funds to pay off the mortgage on the family home. If the spouses later divorce, the second spouse, who did not receive the retirement funds but knew about their receipt by the first spouse, may be awarded the family home pursuant to a divorce decree. Should that second spouse, who significantly benefited from the retirement funds, nonetheless be relieved from joint and several liability for the deficiency caused by the funds’ omission from the joint return? Under § 6013(e), the second spouse remained liable because he or she knew or had reason to know of

4. Id.
7. Although former § 6013(e) did not actually use that term, the courts did. See Ewing v. Comm’r, 118 T.C. 494, 518 n.3 (2002) (Laro, J., dissenting) (noting that the term innocent spouse was apparently spawned in Spanos v. United States, 212 F. Supp. 861, 864 (D. Md. 1963), where the court described a taxpayer who had filed a joint return with her husband as an innocent spouse after noting that the taxpayer had no income of her own and was innocent of her husband’s fraudulent failure to file a federal income tax return when it was due).
13. See id. at 330.
Joint Tax Liability Relief

the tax understatement, which has been interpreted by courts to mean knowledge of the underlying transaction.14

Congress enacted extensive changes to § 6013(e) in 1998, replacing it with § 6015.15 Section 6015 provides three different forms of relief: 

\textit{innocent spouse} relief under § 6015(b), \textit{separation of liability} relief under § 6015(c), and \textit{discretionary} relief under § 6015(f).16 Separation of liability relief is only available to spouses who are divorced, separated, or living apart.17 However, that form of relief is not available if a claimant spouse has actual knowledge of the “item giving rise to a deficiency” on a joint tax return.18 The language \textit{item giving rise to a deficiency} can have two meanings: (1) the spouse knows about the underlying transaction that gave rise to the omitted income (or the erroneously claimed deduction or credit, as the case may be), or (2) the spouse knows that the item on the return was incorrect when it was signed.19 Under the \textit{underlying transaction} interpretation, separation of liability relief would not be available to the second spouse in the example above because that spouse had actual knowledge of the retirement funds.20 However, under the \textit{item on a return} interpretation, the second spouse could obtain separation of liability relief despite knowing about the retirement funds, as long as the spouse did not know that the funds’ tax treatment on a return was incorrect.21

15. See infra Part I.A.
17. Id. § 6015(c)(3)(A).
18. Id. § 6015(c)(3)(C).
20. See \textit{Mitchell v. Comm’r}, 292 F.3d 800, 805 (D.C. Cir. 2002); \textit{Cheshire v. Comm’r}, 282 F.3d 326, 335 (5th Cir. 2002). This Comment will use the term \textit{underlying transaction} to encompass a variety of phrases used by courts. See \textit{Mitchell}, 292 F.3d at 806 (holding that the Service does not have to show “actual knowledge of the improper tax treatment of an item”); \textit{Cheshire}, 282 F.3d at 337 (using the phrase “an item of income, deduction or credit”); \textit{Cheshire}, 115 T.C. at 195 (using the phrase “an actual and clear awareness of the omitted income” in cases of omitted income). \textbf{But cf. \textit{Cheshire}, 115 T.C. at 200 (Thornton, J., concurring) (concluding that the majority’s standard requiring knowledge of omitted income “inherently rejects” the Internal Revenue Service’s argument that “actual knowledge of an ‘item’ means actual knowledge merely of the event or transaction giving rise to the deficiency”).}
21. See \textit{Cheshire}, 115 T.C. at 203–07 (Colvin, J., dissenting). This Comment will use the term \textit{item on a return} to cover a variety of phrases used to describe knowledge of incorrect tax reporting. See, e.g., \textit{Cheshire}, 282 F.3d at 335–36 (rejecting taxpayer’s argument that knowledge refers to “incorrect tax reporting of an item of income, deduction or credit” and refusing to construe the statute to bar relief only for spouses who know that “an entry on the joint tax return is incorrect”).
This Comment argues that Congress intended separation of liability relief to apply to spouses who did not have actual knowledge that a particular item on a joint return was incorrectly claimed, even if they did have knowledge of the underlying transaction giving rise to that item. Part I describes the legislative history of § 6015, and its three types of relief, focusing on the operation of the separation of liability relief provision. Part II examines how the actual knowledge exception to separation of liability relief has been interpreted. Part III looks at the case law interpreting the knowledge requirement of innocent spouse relief under § 6015(b) and former § 6013(e). Finally, Part IV analyzes the text, structure, and legislative history of the actual knowledge exception, and argues that spouses who do not have actual knowledge of an incorrect item on a joint return, but who know about the underlying transaction, should still be allowed to separate their tax liability. This Comment further suggests that an overbroad judicial interpretation of the actual knowledge exception should be remedied by a legislative or administrative amendment.

I. SECTION 6015 PROVIDES SPOUSES WITH THREE AVENUES OF RELIEF FROM JOINT AND SEVERAL LIABILITY

When Congress broadened joint and several liability relief in 1998, it was responding to the inadequacy of taxpayer protection under the pre-1998 innocent spouse provision.22 With that remedial purpose in mind, Congress added two new forms of relief and expanded the existing innocent spouse relief.23 One of the new forms of relief is the separation of liability provision, which is premised on the principle of individual responsibility and is granted only to spouses who are divorced, separated, or living apart.24 The second new form of relief is the discretionary relief provision, which permits the Secretary of the Treasury (the Secretary) to extend relief in cases where a taxpayer...

---

Cheshire, 115 T.C. at 206–07 (Colvin, J., dissenting) (concluding that the knowledge requirement means “knowledge that the return is incorrect, not knowledge that there was an income-producing activity or transaction”).


23. See id.

Joint Tax Liability Relief

cannot qualify for either innocent spouse or separation of liability relief, but still merits protection.25

A. Congress Enacted § 6015 to Remedy the Inadequate Protection of Spouses Under Former § 6013(e)

Congress first expanded the original innocent spouse relief provision in 1984.26 This amendment to § 6013(e) extended relief not only to cases of income omitted from a return,27 but also to cases involving an erroneously claimed deduction, credit, or basis.28 To obtain relief under the 1984 amendment, the claimant spouse had to prove five elements: (1) a joint return had been filed; (2) the tax understatement was “substantial” (over $500 and a specified percentage of a spouse’s income); (3) this understatement was attributable to “grossly erroneous” items (either omitted from gross income or having no basis in fact or law in case of a deduction, credit, or basis); (4) the claimant had no actual or constructive knowledge of such understatement; and (5) it would be inequitable to hold the claimant liable for the tax deficiency.29

Despite this expansion, § 6013(e) proved inadequate because its many stringent requirements were extremely difficult for a claimant spouse to prove.30 Out of the five statutory relief requirements under former § 6013(e),31 the only requirement that was easily satisfied was showing that a joint return had been filed. The substantial understatement requirement32 caused hardships to some taxpayers if their liability was $500 or less.33 In addition, courts differed widely in their interpretations

25. See 26 U.S.C. § 6015(f); see also infra text accompanying notes 74–78.
29. See id. § 6013(e)(1)–(4).
32. Id. § 6013(e)(1)(B), (e)(3); see also supra text accompanying note 29.
of what deductions met the standard of “having no basis in fact or law” under the “grossly erroneous” requirement. The requirement that a claimant spouse must have no actual or constructive knowledge of the tax understatement—the “innocence” requirement—resulted in a large body of conflicting case law with unpredictable outcomes. Commentators harshly criticized this innocence requirement for having “degenerated into a subjective [inquiry] best characterized as whether the spouse seeking relief can move the judge to sympathy.” Similarly, the requirement that holding a claimant spouse liable for a tax deficiency must be inequitable—the “equity” requirement—was criticized for being “vague and unpredictable.”

Proposals to modify § 6013(e) ranged from urging the complete repeal of joint and several liability to abolishing the filing of joint returns. In 1998, Congress replaced § 6013(e) with § 6015 as part of the Internal Revenue Service Restructuring and Reform Act of 1998. Congress was concerned with the inadequacy of the innocent spouse provisions and enacted § 6015 to make relief from joint and several liability more accessible. Like former § 6013(e), § 6015 had a broad remedial purpose.
Joint Tax Liability Relief

Section 6015 was the result of a compromise between different House and Senate versions of the bill. The innocent spouse provision of the House bill preserved the basic framework of former § 6013(e) but lessened the requirements a spouse must meet to qualify for relief. In particular, the House bill provided that the tax understatement no longer had to be substantial and required the understatement to be attributable to only erroneous (rather than grossly erroneous) items.51

In contrast, the Senate amendment liberalized the House bill. The Senate version rejected the constraints of the § 6013(e) framework and approached the relief provision on a different theoretical level, endorsing a separation of liability method based on a proposal by the American Bar Association. The separation of liability approach allowed a claimant spouse to sever his or her tax liability from that of the other spouse by allocating income, deductions, or credits as if they had filed separate returns. Thus, liability for a tax deficiency caused by omitted income, an erroneous credit, or an erroneous deduction would follow only the person responsible for that error. When Senator Graham presented the Senate amendment, he explained that the Senate

---

47. See Price v. Comm’r, 887 F.2d 959, 963 n.9 (9th Cir. 1989) (stating that giving § 6013(e) an unduly narrow and restrictive reading would hinder Congress’ broader purpose in seeking to remedy an injustice).


49. See Angney, supra note 30, at 614–15.


51. See id. at 61.


53. See S. REP. NO. 105-174, at 56. This concept has been also referred to as (1) the “item” approach, ABA Proposal, supra note 37, at 7; IRS Restructuring (Innocent Spouse Tax Rules): Hearing on H.R. 2676 Before the Senate Comm. on Finance, 105th Cong., 2d Sess. 182 (1998) [hereinafter IRS Restructuring Hearing] (statement of Marjorie O’Connell, tax attorney, O’Connell & Associates); (2) the “division of liability” method, Brief of Amici Curiae in Favor of Reversal at 2, Cheshire v. Comm’r, 282 F.3d 326 (5th Cir. 2002) (No. 00-60855) [hereinafter Cheshire Amici Brief], available at http://www.abanet.org/tax/groups/domrel/cheshirebrief.pdf (last visited June 22, 2003); and (3) the “accounting” approach, 144 CONG. REC. S4474 (daily ed. May 7, 1998) (statement of Sen. Graham). This Comment uses the phrase separation of liability approach.

54. See ABA Proposal, supra note 37, at 6–7.

55. See S. REP. NO. 105-174, at 56.

56. See id. at 55–56.

57. Senators Graham, D’Amato, and Feinstein introduced an amendment to the actual knowledge exception, which became part of the final bill, to make actual knowledge relevant at the time an individual signed the return, and to remove cases of duress from the exception’s coverage. See 144 CONG. REC. S4473–74 (daily ed. May 7, 1998).
Committee on Finance had adopted the American Bar Association separation of liability approach to replace joint and several responsibility with “individual responsibility.”

A Conference Committee was appointed to resolve the disagreement between the House and Senate versions of the bill; the Committee successfully reached a compromise that was approved by both houses. The resulting 1998 amendment repealed § 6013(e) and replaced it with § 6015, which combined the House and Senate versions. The Senate’s liberal separation of liability relief was included, although it was limited to taxpayers who were divorced, separated, or had been living apart for at least one year. Further, the 1998 amendment included the House provision that modified the innocent spouse relief and was available to taxpayers who did not qualify for separation of liability relief. The Conference Committee added the third form of relief to cover “appropriate situations to avoid the inequitable treatment of spouses” in the Secretary’s discretion. The price of this compromise was the mixture of three forms of relief, each operating under different policies: prerequisite innocence and equity embedded in innocent spouse relief, individual responsibility for a spouse’s own taxes underlying separation of liability relief, and inequity avoidance under discretionary relief.

B. Section 6015 Transformed the Innocent Spouse Relief Under Former § 6013(e) into Three Forms of Relief

As a result of the Conference Committee’s compromise, § 6015 provides for three distinct types of relief that are codified at 26 U.S.C. § 6015(b), (c), and (f). The first avenue of relief, § 6015(b), is a reformed version of the former innocent spouse provision, § 6013(e),

58. See id. at 54474.
63. 26 U.S.C. § 6015(b); see H.R. CONF. REP. NO. 105-599, at 251.
65. See supra text accompanying notes 36–42.
66. See supra text accompanying notes 53–58.
67. See supra text accompanying note 64.
Joint Tax Liability Relief

while the other two provisions, § 6015(c) and (f), are completely new additions to the statute.68

1. Innocent Spouse Relief Under § 6015(b)

Section 6015(b) provides innocent spouses with complete relief from liability if (1) a joint return has been made; (2) the tax understatement is attributable to the non-claimant spouse’s erroneous items; (3) the claimant spouse proves that in signing the return he or she had no actual or constructive knowledge of the understatement; (4) it is inequitable to hold the claimant spouse liable for the tax deficiency; and (5) the claimant spouse elects the benefits of § 6015(b) within two years after the date the Secretary has begun collection activities.69 In addition, partial relief is available if the claimant did not know and had no reason to know of the extent of the understatement.70

Although it is based on former § 6013(e), § 6015(b) broadened the scope of its predecessor.71 The tax understatement is no longer required to be substantial, and the items to which it is attributable do not have to be grossly erroneous.72 However, the 1998 amendment preserved the innocence and equity requirements of former § 6013(e).73

2. Discretionary Relief Under § 6015(f)

Section 6015(f) authorizes the Secretary to provide equitable relief at his or her discretion if it is unfair to hold the taxpayer liable.74 Discretionary relief is authorized only if a requesting taxpayer cannot obtain either innocent spouse or separation of liability relief under § 6015(b) or (c).75 The Secretary’s determination must be made “taking into account all the facts and circumstances.”76 On review to the courts, the Secretary’s determination is subject to the “abuse of discretion”

68. See Cheshire v. Comm’r, 282 F.3d 326, 331 n.9 (5th Cir. 2002).
70. Id. § 6015(b)(2). Before the 1998 amendment, it was unclear whether a court could grant partial innocent spouse relief. 1998 BACKGROUND REPORT, supra note 34, at 6 (citing Wiksell v. Comm’r, 90 F.3d 1459 (9th Cir. 1997)).
73. Id. § 6015(b)(1)(C)-(D); see also supra notes 36–37, 40–41 and accompanying text.
75. Id. § 6015(f)(2).
76. Id. § 6015(f)(1).
In order to prevail, taxpayers appealing the denial of discretionary relief must demonstrate that the Secretary exercised his or her discretion arbitrarily, capriciously, or without sound basis in fact or law.

3. Separation of Liability Relief Under § 6015(c)

Unlike innocent spouse relief under § 6015(b) and discretionary relief under § 6015(f), which are available to any spouse who meets the prescribed statutory requirements, separation of liability relief is only available to taxpayers who are divorced, legally separated, or living apart for at least one year. Congress believed that an elective system based on separate liability would provide better protection for this class of taxpayers because the inequities experienced by taxpayers facing collection attempts by the Internal Revenue Service (the Service) are most appalling in cases of divorce or separation.

Josephine Berman’s story, related to the Senate Committee on Finance, is such a case. Although her husband left her in 1970, the Service tried for years to recover a joint tax deficiency that was created when the Service disallowed her husband’s deductions for legal...
Joint Tax Liability Relief

expenses incurred during litigation with his partner in an S corporation and claimed on the Bermans’ 1968, 1969, and 1970 joint returns. At the time of Ms. Berman’s testimony in 1998, the Service had placed a lien on her home, destroyed her credit rating, and seized her retirement savings. In part because of stories of inequities suffered by taxpayers like Ms. Berman, the Senate Committee adopted the separation of liability relief provision. Although this provision allows taxpayers to separate their joint tax liability as if they had filed separate returns, there are procedural and substantive limitations on its availability to taxpayers.

a) Operation of § 6015(c)

Section 6015(c) is based on the separation of liability approach proposed by the ABA, “in which liability for the tax follows responsibility for the item, and represents a departure from strictly proportional liability.” Separation of liability relief allows spouses to divide their joint tax deficiency through the allocation of items of income, deduction, and credit. When a taxpayer requests relief under § 6015(c), each item giving rise to a tax deficiency must first be allocated to each spouse as if the spouses had filed separate returns. Next, the claimant’s tax liability for those items is calculated. For example, if the Service assesses a deficiency attributable to $70,000 of the husband’s unreported income and $30,000 of the wife’s disallowed deductions, and the husband is qualified to elect separation of liability

84. Id. at 153–54. Apparently, the Service concluded the claimed deductions were disallowed because they were incurred to protect Mr. Berman’s equity investment rather than his income interest. Id. at 153.
85. Id. at 154–55.
86. See S. REP. NO. 105-174, at 55.
89. Id. § 6015(d)(3)(A) (“Any item giving rise to a deficiency on a joint return shall be allocated to individuals filing the return in the same manner as it would have been allocated if the individuals had filed separate returns for the taxable year.”) (emphasis added). There are two exceptions to this general allocation rule: (1) a tax benefit to the claimant under § 6015(d)(3)(B), see infra note 92; and (2) fraud of one or both individuals authorizing the Secretary to provide for a different allocation under § 6015(d)(3)(C).
90. Id. § 6015(d)(1) (“The portion of any deficiency on a joint return allocated to an individual shall be the amount which bears the same ratio to such deficiency as the net amount of items taken into account in computing the deficiency and allocable to the individual under paragraph (3) bears to the net amount of all items taken into account in computing the deficiency.”).
relief under § 6015(c), then his liability would be limited to seventy percent of the total tax deficiency.

b) Limitations on § 6015(c)’s Availability

Congress has set several limitations on the availability of the separation of liability relief provision. Section 6015 imposes two technical limitations: (1) a taxpayer’s request for relief must be made within two years from the beginning of the Service’s collection activities, and (2) no credit or refund is available as a result of the § 6015(c) election. Further, there are three additional limitations, so-called “special rules,” that are intended to prevent inappropriate use of separation of liability relief. These limitations originated in the Senate amendment and were incorporated into the final bill by the Conference agreement. First, neither spouse can elect to separate liability if assets were transferred between the spouses as part of a fraudulent scheme joined in by both spouses. Second, if the spouses transferred disqualified assets for the principal purpose of tax avoidance, the liability of the electing spouse is increased by the value of any assets so
Joint Tax Liability Relief

Third, if an electing spouse had actual knowledge of “any item giving rise to a deficiency” at the time he or she signed the return, separation of liability relief is not available for any tax deficiency that is attributed to that item (the actual knowledge exception). The Service has the burden of showing that a claimant has actual knowledge of an item giving rise to a deficiency.

In sum, § 6015(c)’s allocation rules allow a claimant spouse to separate his or her liability for a jointly owed tax deficiency. Separation of liability relief is only available to a limited class of taxpayers, and is subject to many procedural and substantive limitations. Notably, if taxpayers had actual knowledge of any item giving rise to a deficiency, they cannot separate liability for that deficiency.

II. TWO INTERPRETATIONS OF THE ACTUAL KNOWLEDGE EXCEPTION TO § 6015(c): UNDERLYING TRANSACTION AND ITEM ON A RETURN

Spouses with actual knowledge of any “item giving rise to a deficiency” cannot obtain separation of liability relief for that deficiency. Congress’ use of that phrase has sparked disagreement over what “item” the taxpayer can know about to qualify for relief. The United States Courts of Appeals for the Fifth Circuit and District of Columbia Circuit, as well as the United States Tax Court, have all agreed with the Service and adopted the underlying transaction interpretation, thus barring taxpayers from obtaining relief if they knew about the transaction that gave rise to a tax deficiency. However, Judge Colvin agreed with the taxpayers in his dissent from the Tax Court opinion and determined that this statutory phrase means an “item on the return” so that the actual knowledge exception would bar relief only if taxpayers knew that they were signing a joint return containing an incorrectly claimed item.

99. See id. § 6015(c)(4).
100. See id. § 6015(c)(3)(C).
101. See id.
102. Id. ("If the Secretary demonstrates that an individual making an election under this subsection had actual knowledge, at the time such individual signed the return, of any item giving rise to a deficiency (or portion thereof) which is not allocable to such individual under subsection (d), such election shall not apply to such deficiency (or portion).”).
103. See Mitchell v. Comm’r, 292 F.3d 800, 806 (D.C. Cir. 2002); Cheshire v. Comm’r, 282 F.3d 326, 337 (5th Cir. 2002); Cheshire v. Comm’r, 115 T.C. 183, 195 (2000); see also supra note 20.
104. See Cheshire, 115 T.C. at 206–07 (Colvin, J., dissenting); see also supra note 21.
A. The Underlying Transaction Interpretation

The first case to address this issue arose in the Fifth Circuit. In *Cheshire v. Commissioner*, Kathryn Cheshire’s husband received retirement distributions, more than half of which were used to pay off the Cheshires’ home mortgage and buy a new family car. On the joint return for that year, the Cheshires incorrectly claimed that only a small portion of the retirement distributions was taxable. Before signing the return, Mrs. Cheshire questioned her husband about the tax consequences of the retirement distributions, and he replied that an accountant advised him that the proceeds used to pay off the mortgage were nontaxable. In fact, he had never consulted an accountant, and the proceeds were taxable. The Cheshires subsequently divorced, and Mrs. Cheshire received the house and family car under a divorce decree. Consequently, under the joint and several liability that arises from filing a joint return, the Service assessed a deficiency against Mrs. Cheshire for understating the taxable amount of the retirement distributions and other income.

Mrs. Cheshire argued that § 6015(c) should relieve her of joint and several liability because the phrase “item giving rise to a deficiency” means “incorrect tax reporting of an item.” In other words, she contended that § 6015(c) relief should be available to her because she did not have actual knowledge that the joint tax return incorrectly omitted the retirement funds that should have been taxed. affirms the Tax Court’s decision, the Fifth Circuit disagreed with Mrs. Cheshire and held that “item” refers to “an item of income, deduction, or credit.” Thus, the Fifth Circuit adopted the underlying transaction interpretation of “item” under § 6015(c) and held that Mrs. Cheshire could not separate her liability because she had “actual and clear awareness” of Mr. Cheshire’s retirement distributions and knew how the

---

105. 282 F.3d 326 (5th Cir. 2002).
106. Id. at 329–30.
107. Id. at 330.
108. Id.
110. Cheshire, 282 F.3d at 330.
111. Id. at 329–30.
112. Id. at 335; see also supra note 21.
113. See Cheshire, 282 F.3d at 335.
114. Id. at 337; see also supra note 20.
Joint Tax Liability Relief

distributions were spent. In reaching its decision, the Fifth Circuit relied on the plain meaning of the term “item” in § 6015 and other sections of the Internal Revenue Code (the Code) to support the meaning adopted by the court. Additionally, the court suggested that Mrs. Cheshire’s interpretation “runs afoul of the general rule that ignorance of the tax laws is not a defense to a tax deficiency.” Because the court found that the statute was not ambiguous on its face, it declined to give any deference to § 6015’s “inconclusive” and “ambiguous” legislative history.

Both the Fifth Circuit and the Tax Court relied on the principle that ignorance of the law is not a defense in rejecting Mrs. Cheshire’s argument. However, unlike the Tax Court, the Fifth Circuit did not recognize that a conflict existed between its interpretation of “item” and numerous references in the legislative history of § 6015(c). The majority of the Tax Court resolved this conflict in favor of a broader reading of the actual knowledge exception, relying in part on a self-employment example in the report issued by the Joint Committee on Taxation for the 1998 amendment. The example posits a husband with $20,000 of unreported self-employment income, $5,000 of which his wife knew about, and concludes that after the divorce, the wife could not use § 6015(c) to separate joint and several liability for the $5,000, of which she knew, although she could do so for the remaining $15,000, of which she did not know. The Tax Court concluded that in omitted income situations, a taxpayer who actually knows about “the item of
income that should have been reported on the return” has actual knowledge of an item giving rise to the deficiency; such taxpayer is therefore ineligible for separation of liability relief.\textsuperscript{124}

The Court of Appeals for the District of Columbia Circuit reached a similar conclusion in \textit{Mitchell v. Commissioner}.\textsuperscript{125} In \textit{Mitchell}, the taxpayer Mrs. Mitchell lost her husband who, prior to his sudden death, received a retirement distribution and used the proceeds to purchase treasury securities.\textsuperscript{126} Mrs. Mitchell knew about the distribution and what her husband did with it, although she did not know the correct tax treatment of the distribution on a return.\textsuperscript{127} The Service determined a tax deficiency against Mrs. Mitchell, but she argued that she should receive separation of liability relief because the actual knowledge exception should only apply if she knew that the retirement distributions were not properly reported.\textsuperscript{128}

The court rejected Mrs. Mitchell’s argument and held that the Service was not required to prove that she had actual knowledge of the improper tax treatment of an item.\textsuperscript{129} The \textit{Mitchell} court relied on the literal meaning of the phrase “item giving rise to a deficiency” and rejected Mrs. Mitchell’s interpretation as “semantically awkward.”\textsuperscript{130} Like the \textit{Cheshire} court, the \textit{Mitchell} court considered the principle that ignorance of the law is not a defense, and held that it was “unlikely that Congress would have employed such subtle and ambiguous phrasing” to overrule this well-established concept.\textsuperscript{131} The court also noted that the remedial purposes of \textsection 6015 were satisfied by shifting the burden of proving Mrs. Mitchell’s actual knowledge of the item to the Service and by requiring the Service to prove actual, and not merely constructive, knowledge.\textsuperscript{132} In sum, in adopting the underlying transaction interpretation, courts have primarily relied on the plain meaning of the word “item” and the principle that ignorance of the law is not a defense.

\begin{itemize}
\item \textsuperscript{124} \textit{Id.} at 195.
\item \textsuperscript{125} 292 F.3d 800 (D.C. Cir. 2002).
\item \textsuperscript{126} \textit{Id.} at 801.
\item \textsuperscript{127} \textit{Id.} at 801–02.
\item \textsuperscript{128} \textit{See id.} at 805.
\item \textsuperscript{129} \textit{Id.} at 806.
\item \textsuperscript{130} \textit{Id.} at 805.
\item \textsuperscript{131} \textit{Id.}
\item \textsuperscript{132} \textit{See id.} at 805–06; \textit{see also} 26 U.S.C. \textsection 6015(c)(3)(C) (2000).
\end{itemize}
Joint Tax Liability Relief

B. The Item on a Return Interpretation

Dissenting from the Tax Court’s decision in *Cheshire*, Judge Colvin agreed with the taxpayer’s interpretation of “item” and concluded that the history and context of § 6015(c) required the Service to prove a taxpayer’s actual knowledge that an item was incorrectly reported on a return. Judge Colvin reasoned that the legislative history consistently supported this interpretation, pointing to the Senate Committee on Finance report, which stated that: “if the IRS proves that the electing spouse had actual knowledge that an item on a return is incorrect, the election will not apply to the extent any deficiency is attributable to such item.” Further, another portion of the report stated:

“The Committee intends that this election [to separate liability] be available to limit the liability of spouses for tax attributable to items of which they had no knowledge. The Committee is concerned that taxpayers not be allowed to abuse these rules by knowingly signing false returns, or by transferring assets for the purpose of avoiding the payment of tax by the use of this election.”

Thus, Judge Colvin determined that the Senate Committee intended the actual knowledge exception to apply to taxpayers who knew that a return was incorrect—in other words, to taxpayers who were “knowingly signing false returns.”

In addition, Judge Colvin considered the floor remarks by Senators Graham and D’Amato persuasive. Both senators were members of the Senate Committee on Finance, and they introduced the amendments to the actual knowledge exception. Senator Graham commented that the Secretary had the burden of demonstrating that a spouse making a § 6015(c) election “had actual knowledge of the conditions within that return which led to this deficiency” in order to be “100 percent responsible.” Similarly, Senator D’Amato explained the policy

---

134. *Id.* at 203–07.
135. *Id.* at 204–05 (quoting S. REP. NO. 105-174, at 59 (1998)).
136. *Id.* at 205 (quoting S. REP. NO. 105-174, at 55–56).
137. *Id.* at 205–06.
138. *Id.* at 205.
139. See *supra* note 57.
140. *Cheshire*, 115 T.C. at 205 (Colvin, J., dissenting) (quoting 144 CONG. REC. S4474 (daily ed. May 7, 1998)).
behind the actual knowledge exception as alleviating lawmakers’ concerns “that some taxpayers may try to abuse the innocent spouse rules by knowingly signing false returns, or transferring assets for the purpose of avoiding the payment of tax, and then claim to be innocent.” Because “no one would want to open the door to that type of fraud,” the actual knowledge exception was included in the bill.

Finally, Judge Colvin noted that the language used by the Conference Committee was identical to that used by the Senate Committee, and determined that this legislative history “unequivocally show[ed]” Congress’ intent to require the Commissioner to prove that the spouse knew his or her tax return was incorrect. Accordingly, Judge Colvin concluded that “Congress intended the knowledge requirement to mean knowledge that the return is incorrect, not knowledge that there was an income-producing activity or transaction.”

C. Treasury Regulations

Having won in Cheshire and Mitchell, the Service, acting through the Treasury, promptly fortified its foothold with the Final Regulations, published on July 18, 2002. The Regulations interpret the actual knowledge exception differently depending on whether a case involves omitted income, an erroneous deduction, or fictitious or inflated deductions or credits. For cases where a spouse omits income, the Service adopted the holding of the Tax Court in Cheshire. In such cases, the standard is whether the requesting spouse actually knew about the item, including the receipt of the income, rather than simply knowing about the tax consequences of the item. On the other hand, in cases where a spouse has claimed an erroneous deduction, the Service

---

141. Id. (quoting 144 CONG. REC. S4474).
142. Id. (quoting 144 CONG. REC. S4474).
144. See supra text accompanying note 135.
146. Id. at 206–07 (citations omitted).
Joint Tax Liability Relief

adopted a standard developed by the Tax Court in *King v. Commissioner*,\(^{152}\) where the relevant inquiry is whether the requesting spouse actually knew about the factual circumstances that made the item unallowable as a deduction, rather than the proper tax consequences of the item.\(^{153}\) In cases involving fictitious or inflated deductions or credits, the standard is whether the spouse actually knew that the expenditure was not incurred, or not incurred to that extent.\(^{154}\)

Before these Regulations became final, commentators criticized the Service’s interpretation of the actual knowledge exception, claiming that it was contrary to the provision’s legislative history.\(^{155}\) However, the Service reasoned that narrowing these standards would give the actual knowledge exception a superfluous meaning.\(^{156}\) The Service also rejected these criticisms based on the self-employment income example contained in the legislative history.\(^{157}\) The Service read this example to refer to both the income tax and the self-employment tax deficiency attributable to the $5,000 portion of the self-employment income about which the wife had actual knowledge.\(^{158}\) While the wife’s knowledge of her husband’s self-employment income might also imply that she knew that the omitted $5,000 is subject to income tax, it does not mean that she is aware of an additional self-employment tax on the omitted income, because many taxpayers are unaware of that tax.\(^{159}\) Because this example did not indicate that the Service must prove that the wife actually knew that self-employment income was subject to income tax

\(^{152}\) 116 T.C. 198, 204 (2001).


\(^{154}\) Treas. Reg. § 1.6015-3(c)(2)(i)(B)(2).

\(^{155}\) See, e.g., Unofficial Transcript of IRS Hearing on Innocent Spouse Regs., 2001 TAX NOTES TODAY 112-100 (June 11, 2001) [hereinafter June 2001 IRS Hearing] (statement of David L. Keating, Senior Counsel, National Taxpayers Union); AICPA Suggests Changes to Innocent Spouse Regs., 2001 TAX NOTES TODAY 133-26 (July 11, 2001).

\(^{156}\) See June 2001 IRS Hearing, supra note 155, ¶ 102, 107 (remarks of Judy Wall, Branch Chief, IRS Procedure and Administration) (“In that legislative history where they talk about knowingly signing false returns, . . . there are several limitations in the statute. One of which is fraud . . . . I’m having trouble understanding why we would have needed the separate, actual knowledge limitation if we already had the fraud limitation . . . .”).

\(^{157}\) See supra text accompanying notes 121–23.


and self-employment tax, the Service rejected the standard that actual knowledge of the item means actual knowledge of the proper tax treatment of the item. 160

Contrary to the Regulations’ approach, the National Taxpayer Advocate, the head of the Service’s Taxpayer Advocate Service, recommended eliminating the actual knowledge requirement from § 6015(c) because it “frustrates [c]ongressional intent” to have separation of liability relief as “a largely mechanical application of law that permits divorced or separated taxpayers to end their joint financial obligation to the IRS.” 161 According to the National Taxpayer Advocate’s statistics, the denial rate of all claims under the actual knowledge standard is only slightly lower than the denial rates under the constructive knowledge standard, even though the constructive knowledge standard, which is more difficult for a claimant spouse to prove, should logically lead to significantly higher denial rates. 162

Courts and the Treasury have interpreted the actual knowledge exception under § 6015(c) to disallow separation of liability relief to claimants with actual knowledge of an underlying transaction. 163 In contrast, Judge Colvin’s dissent in Cheshire urged a pro-taxpayer interpretation of this exception, under which only claimants who knew they had signed an incorrect return would be barred from § 6015(c) relief. 164 Additionally, the Regulations’ approach, interpreting the actual knowledge exception as three different standards, has been met with substantial criticism from both commentators and the National Taxpayer Advocate. 165

162. 2000 NATIONAL TAXPAYER ADVOCATE’S ANNUAL REPORT TO CONGRESS, supra note 161, at 91.
163. See supra Part II.A, C.
164. See supra Part II.B.
165. See supra notes 155, 161–62 and accompanying text.
III. THE CONSTRUCTIVE OR ACTUAL KNOWLEDGE REQUIREMENT UNDER § 6015(b) AND FORMER § 6013(e) INNOCENT SPOUSE RELIEF

Although § 6015(c) separation of liability relief has no statutory antecedent, there is a significant body of case law interpreting § 6015(b)’s and former § 6013(e)’s requirement that to be innocent, taxpayers must have no actual or constructive knowledge of a tax understatement. In rejecting the taxpayers’ interpretation of “item” in § 6015(c), the Cheshire and Mitchell courts relied in part on the general principle that ignorance of the law is not a defense. Courts used this same principle in interpreting the knowledge requirements of § 6015(b) and former § 6013(e).

A. Two Standards for Two Lines of Cases

The knowledge element of former § 6013(e) required spouses signing a joint return to not know, and have no reason to know, that there was a substantial understatement of tax. Courts developed two different standards for the meaning of “understatement,” drawing a line between cases involving omitted income and cases involving erroneous deductions. Courts have agreed that in omission of income cases, the claimant spouse must not know or have reason to know of the underlying transaction that produced the income (the knowledge of the transaction test). But in erroneous deduction cases, courts are divided: the Tax Court continues to apply the same knowledge of the transaction test, while some circuit courts have adopted an alternate test announced by the United States Court of Appeals for the Ninth Circuit in Price v. Commissioner. The Price standard goes beyond evaluating

---

166. See Mitchell v. Comm’r, 292 F.3d 800, 804 (D.C. Cir. 2002).
167. See Beck, supra note 1, at 352–56.
168. Mitchell, 292 F.3d at 805; Cheshire v. Comm’r, 282 F.3d 326, 336 (5th Cir. 2002).
169. See, e.g., Mitchell, 292 F.3d at 804; Cheshire, 282 F.3d at 334–35.
170. See, e.g., Price v. Comm’r, 887 F.2d 959, 964 (9th Cir. 1989).
172. See Cheshire, 282 F.3d at 333.
173. Id.
174. See id. (citing Bokum v. Comm’r, 94 T.C. 126, 151 (1990)).
175. 887 F.2d 959 (9th Cir. 1989). See Cheshire, 282 F.3d at 333.
the spouse’s knowledge of the transaction underlying a deduction, and instead considers whether a reasonably prudent taxpayer in the spouse’s position could be expected to know of the understatement.\textsuperscript{176}

Although the \textit{Price} court expanded the relief available to a spouse in erroneous deduction cases, it cautioned that “[o]f itself, ignorance of the attendant legal or tax consequences of an item which gives rise to a deficiency is no defense for one seeking to obtain innocent spouse relief.”\textsuperscript{177} Courts have consistently rejected the theory that a person’s ignorance of the tax consequences of his or her actions should be considered a factor in the application of the knowledge requirement.\textsuperscript{178} However, the other factors applied by courts vary widely.\textsuperscript{179} In determining whether the spouse has no knowledge of the understatement, courts have considered a claimant’s spouse’s level of education, involvement in the family business or finances, the culpable spouse’s deceit and ability to fool others, the culpable spouse’s indictment or conviction on a related offense, and the lavishness of the family’s expenditures compared to past spending patterns and standard of living.\textsuperscript{180}

\textbf{B. Courts Uniformly Apply the Principle That Ignorance of the Law Is Not a Defense}

While courts acknowledge that the knowledge of the transaction test conflicts with the plain meaning of the innocent spouse provision, which limits relief to spouses with no knowledge of the understatement,\textsuperscript{181} they nonetheless accept this deviation from the statute’s plain meaning because “it avoids ‘acceptance of an ignorance of the law defense.’”\textsuperscript{182} The principle that ignorance of the law is not a defense is both well-established\textsuperscript{183} and solidly rooted in common law.\textsuperscript{184}

\begin{thebibliography}{9}
\bibitem{fn176} See \textit{Price}, 887 F.2d at 963, 965.
\bibitem{fn177} See id. at 963–64.
\bibitem{fn178} See \textit{Borison}, supra note 117, at 833–34.
\bibitem{fn179} See id. at 831–34; \textit{Beck}, \textit{supra} note 1, at 351–56.
\bibitem{fn180} See \textit{Borison}, \textit{supra} note 117, at 830–34.
\bibitem{fn181} See \textit{Price}, 887 F.2d at 963 n.9; \textit{Sanders} v. United States, 509 F.2d 162, 169 n.14 (5th Cir. 1975).
\bibitem{fn182} \textit{Cheshire} v. \textit{Comm’r}, 282 F.3d 326, 333 n.16 (5th Cir. 2002) (citing \textit{Price}, 887 F.2d at 963 n.9; \textit{Sanders}, 509 F.2d at 169 n.14).
\bibitem{fn183} See \textit{Mitchell} v. \textit{Comm’r}, 292 F.3d 800, 805 (D.C. Cir. 2002).
\bibitem{fn184} See \textit{Borison}, \textit{supra} note 117, at 834.
\end{thebibliography}
Joint Tax Liability Relief

The seminal case establishing this principle in the innocent spouse context is *McCoy v. Commissioner*. In *McCoy*, the Tax Court denied innocent spouse relief to a wife who was unaware of the tax consequences of income realized by her husband when he incorporated a partnership with assumed liabilities exceeding the basis of assets transferred. The *McCoy* court rooted its decision in the legislative history of § 6013(e), which required “complete ignorance of the omission [of income]” before taxpayers could qualify for innocent spouse relief.187

Relying on the congressional mandate of complete innocence for § 6013(e) relief, other courts have followed *McCoy* and held that ignorance of the law cannot be a defense to tax violations. These courts have charged claimants with a duty to inquire into the proper tax treatment of omitted income, claimed deductions, and credits.188 The *Price* court suggested that where a spouse admitted knowing about unreported funds her husband had embezzled, but did not know that embezzled funds constituted taxable income, “she is considered as a matter of law to have reason to know of the substantial understatement and thereby is effectively precluded from establishing to the contrary.”189 Further, the courts recognized the practical problems associated with using the ignorance of the law defense in the criminal law and federal income tax law,190 such as the difficulty of refuting the defense and determining whether the claimant was at fault by not knowing the law.191

When the *Cheshire* and *Mitchell* courts rejected the item on a return interpretation of the term “item,” both courts relied in part on the principle that ignorance of the law is not a defense.192 The principle was

---

185. 57 T.C. 732 (1972).
186. See id. at 733–35 (citing 26 U.S.C. § 357(c) (1970)).
187. Id. at 734–35 (alteration in original).
188. See *Price v. Comm’r*, 887 F.2d 959, 966 (9th Cir. 1989); *accord* *Reser v. Comm’r*, 112 F.3d 1258, 1267–68 (5th Cir. 1997) (charging a spouse with a duty to inquire); *Von Kalinowski v. Comm’r*, 81 T.C.M. (CCH) 1081, 1086 (2001) (noting that “where a spouse has a duty to inquire as to the legitimacy of a deduction, failure to satisfy such duty may result in constructive knowledge of the understatement being imputed to her”).
189. *Price*, 887 F.2d at 964 (emphasis added).
190. See *Sanders v. United States*, 509 F.2d 162, 169 n.14 (5th Cir. 1975) (citing WAYNE R. LAFAVE & AUSTIN W. SCOTT, JR., HANDBOOK ON CRIMINAL LAW § 47, at 363–64 (1972)).
191. See LAFAVE & SCOTT, supra note 190, at 363–64.
previously applied as part of the innocence requirement under the innocent spouse relief provision. The Cheshire and Mitchell courts extended this principle to separation of liability cases under new § 6015(c).193

IV. THE PHRASE “ITEM GIVING RISE TO A DEFICIENCY” SHOULD BE INTERPRETED AS AN ITEM OF INCOME, DEDUCTION, OR CREDIT INCORRECTLY REPORTED ON A RETURN

Although courts have demonstrated their unwillingness to interpret the actual knowledge exception narrowly to give an electing taxpayer broader protection with separation of liability relief,194 this hesitation is unjustifiable. The structure, text, and legislative history of the actual knowledge exception indicate that an “item giving rise to a deficiency” means an item of income, deduction, or credit incorrectly reported on a return.195 Therefore, the actual knowledge exception should be limited only to those spouses who, at the time of signing a joint return, knew that the item was incorrectly reported on the return.196 Rejecting the item on a return interpretation, the Treasury’s Regulations have created three knowledge tests.197 However, these tests have no support in the text, structure, or legislative history of § 6015(c).198 Additionally, policy considerations strongly favor adopting a narrow reading of the knowledge exception, thereby permitting taxpayers to obtain separation of liability relief unless they have actual knowledge that an item was incorrectly reported on a tax return.199 Congress and the Treasury should amend § 6015(c) and the Regulations, respectively, to clearly state that the actual knowledge exception requires the Secretary to demonstrate that at the time of signing the joint return, the taxpayer knew that the item was reported incorrectly.200

194. See supra Part II.A.
195. See infra Part IV.A–B.
196. See infra Part IV.A–B.
197. See supra Part II.C.
198. See infra Part IV.C.
199. See infra Part IV.D.
200. See infra Part IV.E.
Joint Tax Liability Relief

A. Structure and Text of § 6015

When construing the actual knowledge exception to separation of liability relief under § 6015(c), the Cheshire and Mitchell courts relied in part on case law that interpreted the innocence requirement of § 6015(b) and former § 6013(e). However, the courts’ analogy is inapposite because there are several major differences between the knowledge provisions of innocent spouse relief and separation of liability relief. First, to qualify for innocent spouse relief under § 6015(b), an electing spouse has the burden of proving complete innocence—that is, the spouse must have no actual or constructive knowledge of the understatement. In contrast, the knowledge provision under separation of liability relief is not a prerequisite requirement for the taxpayer to prove, but an exception that bars the provision’s application and must be proved by the Service. This distinction, coupled with the remedial goals of § 6015, requires that courts read separation of liability relief liberally and the actual knowledge exception narrowly.

Second, the requirement of innocent spouse relief covers both a taxpayer’s actual and constructive knowledge. In contrast, the knowledge exception to separation of liability relief only bars taxpayers who have actual knowledge of items giving rise to a deficiency. To incorporate the principle that ignorance of the law is not a defense from the case law interpreting former § 6013(e) is to charge taxpayers with a duty of inquiry about the legal consequences of omitting income, claiming a deduction, or credit. However, barring taxpayers with constructive knowledge from electing separation of liability relief contradicts the text, structure, and history of § 6015(c).

201. See Mitchell v. Comm’r, 292 F.3d 800, 805 (D.C. Cir. 2002); Cheshire v. Comm’r, 282 F.3d 326, 336 (5th Cir. 2002); see also supra Part III.B.
203. See supra notes 100–02 and accompanying text.
204. See supra notes 45–48 and accompanying text.
205. See, e.g., Tcherepnin v. Knight, 389 U.S. 332, 336 (1967) (“We are guided by the familiar canon of statutory construction that remedial legislation should be construed broadly to effectuate its purposes.”); accord Peyton v. Rowe, 391 U.S. 54, 65 (1968).
206. See 26 U.S.C. § 6015(b)(1)(C); see also supra Part III.
208. See Cheshire Amici Brief, supra note 53, at 22.
209. Id.
Further, Congress specifically removed the concept of constructive knowledge from the actual knowledge exception of § 6015(c), noting that “[s]uch actual knowledge [that an item on a return is incorrect] must be established by the evidence and shall not be inferred based on indications that the electing spouse had a reason to know.” 210 The Mitchell court doubted that Congress would rely on a “subtle, ambiguous, obscure and imprecise device” like the actual knowledge exception to make ignorance of the law a defense. 211 However, Congress emphasized in both the text of the statute 212 and the accompanying legislative history 213 that only a showing of actual knowledge should bar separation of liability relief. By disallowing the Service’s use of inferred evidence 214 to prove a taxpayer’s actual knowledge, Congress eliminated the taxpayer’s duty of inquiry. 215 This restriction renders the principle that ignorance of the law is not a defense inapplicable to cases involving the separation of liability provision. 216 Thus, to give effect to this legislative intent, the principle that ignorance of the law is not a defense, developed in cases involving the innocence requirement of innocent spouse relief, should not be extended to the context of the actual knowledge exception to separation of liability relief. 217

Finally, the subject of a spouse’s knowledge is different under the two forms of relief: under innocent spouse relief, a taxpayer cannot have actual or constructive knowledge of an “understatement” of tax, 218 while under separation of liability relief, a taxpayer cannot have actual knowledge of an “item giving rise to a deficiency.” 219 In Cheshire, the Service argued that if Congress intended the knowledge exception to extend only to an “item on a return,” it could have referred instead to “an item of income that should have been reported on the return” or “actual
Joint Tax Liability Relief

knowledge of the incorrect tax reporting of an item on a return. Thus, the Service concluded that the item on a return meaning could not have been intended. Congress, however, extended the actual knowledge exception to an “item giving rise to a deficiency,” a phrase that was unknown to the pre-1998 innocent spouse relief. Congress’ choice of new language implies its rejection of the line of cases interpreting the former innocent spouse provision. If Congress intended to incorporate the judicial knowledge of the transaction standard that was developed under the knowledge of an understatement requirement of former § 6013(e) and current § 6015(b), it probably would have referred to “the actual knowledge of an understatement” in § 6015(c). But Congress did not do so, and chose instead to use new and different language in § 6015(c).

This new phrase chosen by Congress—“item giving rise to a deficiency”—is ambiguous. It can mean an income-producing transaction, as urged by the Service and adopted by the Cheshire and Mitchell courts. However, it can also mean an item on a joint return. The Code contains references to both of these potential readings. Because the meaning of the actual knowledge exception is ambiguous, the use of legislative history is appropriate to ascertain the meaning of this phrase.

220. Brief for the Appellee, Commissioner of Internal Revenue at 36, Cheshire v. Comm’r, 282 F.3d 326 (5th Cir. 2002) (No. 00-60855) [hereinafter Cheshire IRS Brief].
221. See id.
222. The term understatement under § 6015(b) and former § 6013(e) covers cases of omitted income, erroneously claimed deduction or credit. See supra text accompanying notes 27–29 and Part III.A for discussion of courts’ interpretation of this term. Thus, Congress did not have to introduce the new term “item” to cover all cases that may cause a tax deficiency; it could have done so by preserving the terminology of innocent spouse relief.
224. See Mitchell v. Comm’r, 292 F.3d 800, 806 (D.C. Cir. 2002); Cheshire, 282 F.3d at 335, 337.
225. See Cheshire, 115 T.C. at 203 (Colvin, J., dissenting).
226. For example, 26 U.S.C. § 61(a) refers to “item” as an underlying transaction by providing that gross income includes various items listed in that section. In contrast, 26 U.S.C. § 57(a) refers to various “items of tax preference” defined by reference to tax consequences. Cheshire, 115 T.C. at 205 n.1 (Colvin, J., dissenting).
227. Cheshire, 115 T.C. at 202–03 (Colvin, J., dissenting). Contra Cheshire, 282 F.3d at 336–37 (concluding that the meaning of the statute is plain and declining to give any deference to the legislative history).
228. See, e.g., Conn. Nat’l Bank v. Germain, 503 U.S. 249, 255 n.1 (1992) (Stevens, J., concurring) (noting that “[w]henever there is some uncertainty about the meaning of a statute, it is
As the Mitchell court noted, the meaning of the term “item” is defined nowhere in the Internal Revenue Code and is unremarkably general. courts should not analyze the word “item” in isolation and without reference to context.

Looking at the context and structure of § 6015, the phrase “item giving rise to a deficiency” and its variation, “item [that] gave rise to a tax benefit,” are used exclusively in the context of separation of liability relief. There are four references to this phrase in § 6015, three of which appear in the rule for how to allocate items of income, deduction, and credit between spouses, and the fourth appears in the actual knowledge exception. The text of the allocation rules implies that “an amount on a return can be allocated, i.e., split” between spouses, while “an underlying transaction or activity cannot,” which supports the item on a return interpretation. Under the rules of statutory construction, identical words or phrases used in different parts of the same act are intended to have the same meaning, especially where the word or phrase is repeatedly used therein. This principle supports interpreting the

prudent to examine its legislative history” and quoting Judge Learned Hand’s advice in Lehigh Valley Coal Co. v. Yensavage, 218 F. 547, 553 (2d Cir. 1914), that statutes “should be construed, not as theorems of Euclid, but with some imagination of the purposes which lie behind them”;

United States v. Pub. Utils. Comm’n of Cal., 345 U.S. 295, 315 (1953) (“Where the language and purpose of the questioned statute is clear, courts, of course, follow the legislative direction in interpretation. Where the words are ambiguous, the judiciary may properly use the legislative history to reach a conclusion.”); Cheshire, 115 T.C. at 204 (Colvin, J., dissenting).

229. See Mitchell, 292 F.3d at 805.
230. The word “item” is used in the Code more than two thousand times, including fifteen times in § 6015(c). See Cheshire v. Comm’r, 282 F.3d 326, 335 (5th Cir. 2002).
231. See, e.g., Owasso Indep. Sch. Dist. v. Falvo, 534 U.S. 426, 434 (2002) (stating that “[i]t is a fundamental canon of statutory construction that the words of a statute must be read in their context and with a view to their place in the overall statutory scheme”) (internal citations omitted); Deal v. United States, 508 U.S. 129, 132 (1993) (noting that a “fundamental principle of statutory construction (and, indeed, of language itself) [is] that the meaning of a word cannot be determined in isolation, but must be drawn from the context in which it is used”).
233. The three references are: (1) the subheading of the allocation rule, “Allocation of Items Giving Rise to the Deficiency,” 26 U.S.C. § 6015(d)(3); (2) the mechanics of allocation, see 26 U.S.C. § 6015(d)(3)(A); supra note 89; and (3) the tax benefit exception to the usual operation of the allocation rule, see 26 U.S.C. § 6015(d)(3)(B); supra note 92.
236. See, e.g., Cohen v. de la Cruz, 523 U.S. 213, 220 (1998) (noting the presumption that “equivalent words have equivalent meaning when repeated in the same statute”); Gustafson v. Alloyd Co., Inc., 513 U.S. 561, 570 (1995) (restating the “normal rule of statutory construction that identical words used in different parts of the same act are intended to have the same meaning”).

858
term “item” giving rise to a deficiency or tax benefit consistently throughout the separation of liability relief provision.

B. Legislative History of the Actual Knowledge Exception

The structural and textual analysis of the phrase “item giving rise to a deficiency” is supported by the history of § 6015’s adoption. The legislative history demonstrates that Congress enacted the actual knowledge exception to prevent taxpayers who knowingly sign false returns from abusing separation of liability relief. In accordance with this purpose, Congress limited the scope of the taxpayer’s knowledge to the knowledge of an error on a return.

1. Purpose of the Actual Knowledge Exception

The legislative history of separation of liability relief begins with the Senate report, in which the actual knowledge exception was listed among the three “special rules” that were adopted “to prevent the inappropriate use of the election.” The Senate Committee stated that it created these special rules to prevent taxpayers from abusing separation of liability relief by knowingly signing false returns and by transferring assets to avoid the payment of tax. Two special rules address only asset transfers in fraudulent and tax avoidance schemes. The third special rule is the actual knowledge exception. When the Senate Committee, Senator D’Amato, and the Staff of the Joint Committee on Taxation stated that taxpayers should not be allowed to abuse these rules “by knowingly signing false returns, or by transferring assets for the purpose of avoiding the payment of tax,” the reference to “knowingly signing false returns” must refer to the actual knowledge exception because it is the only special rule left that does not involve asset transfers. Thus, the legislative history shows that the intended purpose of

237. See S. REP. NO. 105-174, at 55–60 (1998); see also supra notes 52–62 and accompanying text.


239. S. REP. NO. 105-174, at 55–56, 59; see supra text accompanying note 136.


the actual knowledge exception was to prevent taxpayers who knowingly sign false returns from taking advantage of § 6015(c) relief.

Confronted with this same legislative history suggesting the item on a return meaning, the Service argued that this interpretation would render the actual knowledge exception superfluous. The Service asserted that this interpretation makes the actual knowledge exception unnecessary in light of the fraud provision which authorizes the Secretary to re-allocate items if there is fraud by one or both spouses. However, the Service’s objection is without merit for the following reasons.

The actual knowledge exception targets the taxpayer’s ability to elect § 6015(c) relief in general, unlike the fraud provision that only limits the application of allocation rules. There are five limitations on separation of liability relief. Two of them limit the allocation of items to spouses that otherwise qualify for § 6015(c) relief, including the fraud provision on which the Service relied. In contrast, the other three special rules, including the actual knowledge exception, limit the claimant’s ability to elect separation of liability relief in general. Furthermore, the fraud restriction allows the Secretary to re-allocate items in his or her discretion whereas the actual knowledge exception is a non-discretionary provision that automatically bars § 6015(c) relief for a deficiency attributable to items that a spouse knew were reported incorrectly. Because the fraud restriction provides for discretionary re-allocation of items, while the actual knowledge exception bars relief outright upon the Secretary’s proof that a claimant spouse knew about errors on a return when he or she signed it, the application of these two provisions is distinctly different, and the item on a return interpretation will not render the actual knowledge exception superfluous. Therefore, the purpose of the actual knowledge exception, derived from the provision’s legislative

243. See June 2001 IRS Hearing, supra note 155, ¶¶ 102, 107 (remarks of Judy Wall, Branch Chief, IRS Procedure and Administration); supra note 156 and accompanying text.

244. See June 2001 IRS Hearing, supra note 155, ¶¶ 102, 107 (remarks of Judy Wall, Branch Chief, IRS Procedure and Administration); supra note 156 and accompanying text; see also 26 U.S.C. § 6015(d)(3)(C); supra note 89.


246. See H.R. CONF. REP. NO. 105-599, at 253–54 (1998); S. REP. NO. 105-174, at 59; see also supra note 238 and accompanying text. The first two limitations are placed in 26 U.S.C. § 6015(d), which covers “[a]llocation of deficiency,” whereas the other three limitations are in 26 U.S.C. § 6015(c), which covers “[p]rocedures to limit liability for taxpayers no longer married or taxpayers legally separated or not living together.”

Joint Tax Liability Relief

history, can only be effectuated by giving the exception its intended item on a return meaning.

2. **Scope of the Actual Knowledge Exception**

The scope of the actual knowledge exception must be interpreted in light of the legislative intent to prevent abuse of § 6015(c) relief by spouses who knowingly sign false returns. Introducing the floor amendments to the actual knowledge exception, Senator Graham referred to its scope as actual knowledge of "the conditions within that return which led to this deficiency," the joint return’s contents, and "the circumstances in the return that led to the deficiency." The Senate report explained that the Service would be required to prove "that the electing spouse had actual knowledge that an item on a return is incorrect." The Conference Committee report, the Joint Committee General Explanation of the 1998 Tax Legislation, and another Joint Committee document prepared three years after the statutory enactment are all in accord with the Senate report.

While the Fifth Circuit in *Cheshire* rejected § 6015(c)’s legislative history as ambiguous and inconclusive, the *Cheshire* Tax Court majority noted the contradiction between its holding and the legislative history. The majority resolved the conflict by treating the legislative history as only an example. However, as Judge Colvin noted in his dissent, Congress’ statements describing the scope of the actual

---

248. See 144 CONG. REC. S4473–74 (daily ed. May 7, 1998); supra note 57.
249. 144 CONG. REC. S4474; see also supra text accompanying note 140.
250. 144 CONG. REC. S4474. Although this statement was made in conjunction with the duress portion of the actual knowledge exception amendments, the scope of the exception is the same whether it is applied to a taxpayer under duress or not.
251. Id.
252. S. REP. NO. 105-174, at 59 (1998); see also supra note 135 and accompanying text.
254. GENERAL EXPLANATION, supra note 81, at 70.
257. Cheshire v. Comm’r, 115 T.C. 183, 195 (2000) ("Arguably, this statement [that the IRS must prove that "the electing spouse had actual knowledge that an item on a return is incorrect"] conflicts with our knowledge standard for purposes of section 6015(c)(3)(C).”).
258. Id.
knowledge exception are explanations of the general rule, not examples, because Committees have a specific way of introducing examples by using the preceding phrase “for example.”

Indeed, the Conference Committee report contains an actual example in the paragraph immediately following this explanation of the statutory rule.

The allegedly inconsistent self-employment income example can be reconciled with the legislative intent Congress expressed elsewhere. First, the example’s purpose was to demonstrate how a tax deficiency is allocated between spouses when one spouse knows of a portion of the other spouse’s omitted income. The example does not address the scope of the spouse’s knowledge. The Senate report, where the example originated, placed it in a section titled “[t]ax deficiencies,” explaining how the § 6015(d) allocation rules function. During the conference, the example was moved to the “special rules” discussion immediately following the explanation of the actual knowledge provision’s scope.

These structural changes should not distort the example’s main purpose: to demonstrate the allocation of a tax deficiency if a claimant knows of some, but not all, of the other spouse’s items giving rise to a deficiency.

Second, the use of self-employment income in the example creates a presumption that if the wife knows about her husband’s self-employment income, she necessarily knows it is taxable. Therefore, she should also know that the omission of self-employment income from the return is incorrect. The Service has responded to this argument by attempting to stretch the example to cover a self-employment tax deficiency as well as an income tax deficiency. The Service’s reading of the legislative history has no textual support, given that there is no indication that Congress intended the example to illustrate anything other than the allocation rules when one spouse has actual knowledge.

259. Id. at 206 n.3 (Colvin, J., dissenting).
261. See supra notes 121–23 and accompanying text.
262. See Cheshire, 115 T.C. at 206 (Colvin, J., dissenting).
264. H.R. CONF. REP. NO. 105-599, at 253; see also supra text accompanying notes 95–101.
Joint Tax Liability Relief

about a portion of the item. 267 This dislocated example should not overcome the strong evidence that Congress intended the actual knowledge exception to cover only cases where taxpayers actually know that they have signed a return that contains an incorrect item. Thus, the scope of the actual knowledge exception should be interpreted consistently with the legislative history of § 6015(c) to mean actual knowledge that an item on a return is incorrect.

C. Treasury Regulations

The Regulations provide three distinct tests under § 6015(c) for cases that involve an omission of income, erroneous deduction, and fictitious or inflated deduction or credit. 268 However, Congress established only one standard: whether a taxpayer has “actual knowledge . . . of an[] item giving rise to a deficiency.” 269 This standard is distinct from the prerequisite of complete innocence, which taxpayers must prove to qualify for relief under § 6015(b). 270 The Treasury improperly split Congress’ single standard into three tests that roughly parallel the case law interpreting former § 6013(e). 271 This division is artificial and is not based on the structure, text or legislative history of § 6015(c). 272 When Congress intends to treat income, deduction, or credit separately, it does so. 273 The more accurate reading of the knowledge exception is that it is one standard that applies equally to all types of cases: it bars relief to a taxpayer who, at the time of signing the return, has actual knowledge that the “item” listed on a return—be it omitted income, an improperly claimed deduction or credit—gave rise to a deficiency on a return.

D. Policy Considerations

First and foremost, separation of liability relief must be interpreted liberally in light of its broad remedial purpose. 274 Courts should

268. See supra Part II.C.
270. Id. § 6015(b)(1)(C); see also supra Part III.
271. See supra Part III.
272. See supra Part IV.A–B.
274. See supra notes 45–48, 205 and accompanying text.
recognize Congress’ intent behind § 6015(c)—to provide a different, mechanical-like relief\(^{275}\) to divorced or separated taxpayers, free from the fact-dependent inquiry of the innocence and equity requirements of § 6015(b) relief.\(^{276}\) Applying case law interpreting § 6015(b) innocent spouse relief to the new separation of liability provision undermines Congress’ intent to create a new and different remedy from prior innocent spouse law.\(^{277}\)

The Service successfully argued to the Cheshire Tax Court majority that the item on a return interpretation would render the actual knowledge exception meaningless because potentially any spouse who is not a certified public accountant or tax attorney would be allowed to escape paying income tax.\(^{278}\) Yet, this argument should be rejected as having no merit because separation of liability relief was created to allow divorced or separated taxpayers to separate responsibilities for only their individual share of a tax deficiency.\(^{279}\) If one spouse obtains separation of liability relief, the Service is not left without a taxpayer from whom it can collect tax deficiencies. Instead, the Service can collect from the other spouse, whose income, deduction or credit actually produced the deficiency. The McCoy court’s fear that potentially both spouses could be found innocent, leaving the Service without a taxpayer to collect from\(^{280}\) is thus unfounded.

E. Proposed Amendment

Congress should amend § 6015(c) to clarify the meaning of the actual knowledge exception. Instead of its current ambiguous phrasing,\(^{281}\) the statute should read: “If the Secretary demonstrates that an individual making an election under this subsection had actual knowledge that an item of income, deduction or credit is incorrectly reported on a return, the election will not apply to the extent any deficiency is attributable to

---

275. See 2000 NATIONAL TAXPAYER ADVOCATE’S ANNUAL REPORT TO CONGRESS, supra note 161, at 91.
276. See Cheshire Amici Brief, supra note 53, at 14; supra text accompanying notes 36–42.
277. See Cheshire Amici Brief, supra note 53, at 18.
280. McCoy v. Comm’r, 57 T.C. 732, 735 (1972); Borison, supra note 117, at 834.
Joint Tax Liability Relief

such item. 282 Alternatively, the Treasury could amend its Regulations using similar language, thereby giving proper effect to Congress’ intent behind the actual knowledge exception.

V. CONCLUSION

The application of the innocent spouse relief provision has been heavily litigated since its inception in 1971. In 1998, Congress intended to expand the available relief to a greater number of taxpayers. However, courts have interpreted the new remedies narrowly. By expanding the scope of the actual knowledge exception, courts have barred many taxpayers from obtaining separation of liability relief. Congress should clarify its intent by amending the actual knowledge exception to give separation of liability its intended reach. Alternatively, the Treasury should amend its Regulations to bring them in line with the legislative intent, as evidenced by § 6015(c)’s history, structure, and text. The bonds of joint tax liability should not be stronger than marriage.

282 Cf. H.R. CONF. REP. NO. 105-599, at 253; S. REP. NO. 105-174, at 59 (1998) (stating that the Service is required to prove that “the electing spouse had actual knowledge that an item on a return is incorrect”).
WHEN ANIMALS INVADE AND OCCUPY: PHYSICAL TAKINGS AND THE ENDANGERED SPECIES ACT

Rebecca E. Harrison

Abstract: Government actions implementing the Endangered Species Act (ESA) on private lands have sparked extensive debate and litigation over whether such actions result in Fifth Amendment takings. To date, courts have uniformly rejected regulatory takings claims under the ESA, leading several landowners to advance a different theory—physical takings claims. Successful physical takings claims require landowners to show that government actions resulted in either per se physical takings or compensable physical invasions of their land. In two recent decisions, Boise Cascade Corp. v. United States and Seiber v. United States, courts rejected per se physical takings claims under the ESA, finding that listed species are not controlled by the government, and the presence of such species on private land does not destroy all of a landowner’s fundamental property rights. The second type of physical takings, compensable physical invasions, arises when a government action limits a landowner’s right to exclude but leaves other property rights intact. To determine if such an invasion is a taking, a court would likely employ the three-factor test the United States Supreme Court presented in Penn Central Transportation Co. v. New York City and weigh the character of the government action, the effect of the action on the landowner’s reasonable investment-backed expectations, and the economic effect of the action on the landowner. This Comment argues that government actions taken under the ESA to protect listed species on private lands do not give rise to compensable physical invasions.

Property rights organizations and private landowners have advanced two arguments for the proposition that the federal government should compensate landowners for restrictions it imposes on land pursuant to the Endangered Species Act (ESA or “the Act”). The first argument is that these restrictions result in regulatory takings under the Fifth Amendment of the United States Constitution. Regulatory takings occur when government regulations go “too far” in placing restrictions on private property. Courts have steadfastly rejected this argument.


2. U.S. CONST. amend. V (“[N]or shall private property be taken for public use, without just compensation.”).

3. See Pa. Coal Co. v. Mahon, 260 U.S. 393, 415 (1922) (stating the guiding principle behind regulatory takings as follows: “while property may be regulated to a certain extent, if regulation goes too far it will be recognized as a taking”). The United States Supreme Court has established two tests courts use to determine whether a specific regulation passes this threshold; see infra notes.
because the claims either (1) lack ripeness, 5 (2) challenge actions affecting only a portion of the property, 6 or (3) stem from a landowner’s improper investment-backed expectations for the land. 7

The second argument private landowners have advanced is that government actions under the ESA give rise to physical takings. 8 These claims avoid several of the roadblocks that prevent landowners with regulatory takings claims from receiving relief. Specifically, the claims have a lower threshold for ripeness 9 and can be brought regardless of whether the government action affects the landowner’s entire parcel. 10 Such claims have met with some success in the courts. 11

To bring a successful physical takings claim a landowner must show that a government action resulted in either a permanent physical occupation—a per se physical taking 12—or a compensable physical invasion. 13 A per se physical taking occurs when the government, or an entity authorized by the government, occupies private land 14 and destroys all of the landowner’s fundamental property rights. 15 These rights include the rights to possess the land, exclude others from possessing and using it, control its use, and dispose of it. 16 A

118–24 and accompanying text.


9. See Grant, supra note 8, at 733.

10. See id. at 734.


12. See infra Part II.

13. See infra Part II.


15. See id. at 435.

16. See id. at 435–36.
Physical Takings and the ESA

compensable physical invasion, on the other hand, stems from a
government action that causes a temporary or intermittent invasion of
private land.\textsuperscript{17} During the invasion, the landowner’s right to exclude is
limited, but his or her other fundamental property rights are left
unimpaired.\textsuperscript{18}

To date, courts have only considered whether government
implementation of the ESA amounts to a \textit{per se} physical taking.\textsuperscript{19} In two
recent decisions, the United States Court of Appeals for the Federal
Circuit and the United States Court of Federal Claims rejected such
claims.\textsuperscript{20} Both courts reasoned that government actions under the ESA
resulting in the presence of protected species on private land neither
provide for a government presence on private land nor divest a
landowner of all of his or her fundamental property rights.\textsuperscript{21} These
actions, therefore, are not \textit{per se} physical takings.

Courts have not addressed whether a government action under the
ESA may give rise to a compensable physical invasion. Given the
systematic tenacity of some landowners and property rights
organizations,\textsuperscript{22} there can be little doubt that courts soon will face such a
claim. When landowners bring this claim, it is likely that a court will
apply the United States Supreme Court’s three-factor balancing test
introduced in \textit{Penn Central Transportation Co. v. New York City}.\textsuperscript{23} The
Court has stated, in dicta, that courts should apply the \textit{Penn Central}
test to claims of compensable physical invasions.\textsuperscript{24} This test requires courts
to consider the character of the action, the effect of the action on the
landowner’s reasonable investment-backed expectations, and the
action’s economic impact on the landowner.\textsuperscript{25}

\textsuperscript{17} See id. at 435 n.12.
\textsuperscript{18} Id.
\textsuperscript{19} See, e.g., Boise Cascade Corp. v. United States, 296 F.3d 1339, 1352 (Fed. Cir. 2002), \textit{cert. denied}, 123 S. Ct. 1484 (2003); Seiber v. United States, 53 Fed. Cl. 570, 575–76 (2002); Boise
\textsuperscript{20} Boise, 296 F.3d at 1352; Seiber, 53 Fed. Cl. at 576.
\textsuperscript{21} See Boise, 296 F.3d at 1354–55; Seiber, 53 Fed. Cl. at 575–76.
\textsuperscript{22} See, e.g., Boise, 296 F.3d 1339; Tulare Lake Basin Water Storage Dist. v. United States, 49
United States, 189 F.3d 1355 (Fed. Cir. 1999); Moerman v. California, 21 Cal. Rptr. 2d 329 (Cal.
\textsuperscript{24} Loretto v. Teleprompter Manhattan CATV Corp., 458 U.S. 419, 432, 435 n.12 (1982)
(discussing \textit{Penn Central}, 438 U.S. at 124).
\textsuperscript{25} \textit{Penn Central}, 438 U.S. at 124.
This Comment argues that government actions implementing the ESA on private land do not amount to compensable physical invasions for two reasons. First, background principles of property law require private landowners to allow for government efforts to protect wildlife on their land. The enforcement of preexisting restrictions on title to property is never a taking. Second, application of the Penn Central balancing test strongly suggests that courts should reject claims arguing that ESA restrictions amount to compensable physical invasions. Legal precedent and statutory safeguards in the ESA are likely to require that affected landowners not receive compensation for these claims.

Part I of this Comment presents relevant sections of the ESA controlling its application to private land. Part II explains regulatory takings, per se physical takings, and compensable physical invasions. Part III discusses the obstacles facing ESA-based regulatory takings claims. Part IV presents per se physical takings claims under the Act, explains how these claims avoid several of the roadblocks faced by regulatory takings claims, and discusses courts’ rationale for rejecting such claims. Part V discusses compensable physical invasion takings claims and outlines the three-factor test courts may use to determine if a compensable taking is present. Finally, Part VI argues that preexisting limitations on property use, precedent, and statutory safeguards in the ESA strongly suggest that courts should reject compensable physical invasion claims under the Act.

I. THE ENDANGERED SPECIES ACT AND PRIVATE PROPERTY RIGHTS

The purpose of the ESA is to recover populations of imperiled species and conserve the ecosystems upon which such species depend. Many imperiled species occupy and depend on private land. To meet its goal, the Act provides the U.S. Fish and Wildlife Service (FWS) and the

26. See infra Part VI.B.
27. See infra Part II.
28. See infra Part VI.C.
29. See infra Part VI.C.
Physical Takings and the ESA

National Oceanic and Atmospheric Administration Fisheries Service (NOAA Fisheries) with the authority to regulate actions on private lands that may harm imperiled species or their habitat. An order to exercise this authority, the FWS and NOAA Fisheries (collectively, “the Services”) must list a species as “endangered” or “threatened.” An endangered species is “any species which is in danger of extinction throughout all or a significant portion of its range.” A threatened species is “any species which is likely to become an endangered species within the foreseeable future throughout all or a significant portion of its range.” The Act provides listed species with two protections on private land. First, it is illegal for a private landowner to harm a listed species or its habitat. Second, the federal government may not authorize, fund, or carry out actions on private land that may jeopardize the continued existence of the listed species or its habitat. To lessen the burden of these restrictions, the Act provides two means by which landowners can mitigate the effects of a listing on their land—commenting on proposed rules and applying for Incidental Take Permits (ITPs).

A. Private Land and the ESA

To meet the goals of the ESA, it is critical for the Services to provide protection for species on private land. Habitat loss is the primary cause of wildlife endangerment, and more than half of all species listed as endangered inhabit private lands. Furthermore, as of 1994, at least

32. See infra Part I.B.
34. Id. § 1532(6).
35. Id. § 1532(20).
36. See Meltz, supra note 6, at 373–76.
38. Id. § 1536(a)(1).
42. Wilcove et al., supra note 31.
eighty percent of the habitat for over half of all listed species was located on non-federal land, which includes private, tribal, and state-owned land. Private lands therefore play an important role in providing for successful species recovery efforts under the ESA.

Before the FWS or NOAA Fisheries can provide protection for imperiled species on private land, they must list the species as endangered or threatened. Section 4 of the Act requires the Services to decide whether to list a species as threatened or endangered “solely on the basis of the best scientific and commercial data available.” When a species is listed, the Services must designate its “critical habitat” when doing so is “prudent and determinable.” “Critical habitat” includes areas within the region the species occupied at the time it was listed that contain “physical or biological features (I) essential to the conservation of the species and (II) which may require special management considerations or protection.” It also encompasses certain areas outside the region the species occupied at the time it was listed that “are essential for the conservation of the species.” The establishment of critical habitat is to be based on both scientific data and potential economic impact. Critical habitat may include private as well as public land. It is intended to encompass areas “within the geographical area occupied by the species,” as well as lands “essential for the conservation of the species” that are not currently occupied.

44. Wilcove et al., supra note 31.
45. See ENDANGERED SPECIES PROGRAM, supra note 41, at 1.
46. 16 U.S.C. § 1532(6) (2000); see also supra note 34 and accompanying text.
47. 16 U.S.C. § 1532(20); see also supra note 35 and accompanying text.
49. Id. § 1533(b)(1)(A). Pursuant to § 1533(a)(1), factors the Services must consider in making listing determinations include:
   (A) the present or threatened destruction, modification, or curtailment of [the species’] habitat or range;
   (B) overutilization for commercial, recreational, scientific, or educational purposes;
   (C) disease or predation;
   (D) the inadequacy of existing regulatory mechanisms; or
   (E) other natural or manmade factors affecting its continued existence.
50. Id. § 1533(a)(3).
51. Id. § 1532(5)(A)(i).
52. Id. § 1532(5)(A)(ii).
53. Id. § 1533(b)(2).
54. See id. § 1532(5)(A).
55. Id. § 1532(5)(A)(i).
56. Id. § 1532(5)(A)(ii).
Physical Takings and the ESA

The first protection the ESA provides for a listed species is to prohibit people or agencies from effecting a “take” of the species on public and private land.57 “Take” means to “harass, harm, pursue, hunt, shoot, wound, kill, trap, capture, or collect, or to attempt to engage in any such conduct.”58 This restriction, provided for in Section 9 of the Act,59 reaches beyond injury inflicted directly on an individual animal; it also includes “significant habitat modification or degradation.”60

The ESA also provides protection for listed species and habitat on private land through the Section 761 consultation requirement.62 This provision requires that every federal agency consult with the Services to ensure actions they authorize, fund, or carry out will be unlikely to either “jeopardize the continued existence of any [listed] species”63 or “result in the destruction or adverse modification of [critical] habitat.”64 Private landowners who undertake projects that require a federal agency’s funds, authorization, or participation, and are likely to affect listed species or critical habitat, are subject to the consultation requirement.65

B. Private Landowners Can Use Two Means to Influence the ESA’s Application to Their Land

While the ESA places restrictions on the use of private land, it also provides two means by which private landowners can influence the Act’s application to their land. First, Section 4 grants the public notice of and an opportunity to comment on proposed rules to list species and designate critical habitat.66 Second, Section 1067 gives landowners the opportunity to receive ITPs.68 These permits allow landowners to take listed species or species habitat under certain conditions.69

57. Id. § 1538(a)(1)(B).
58. Id. § 1532(19).
59. Id. § 1538(a)–(g).
60. 50 C.F.R. § 17.3 (2000).
61. 16 U.S.C. § 1536(a)–(p).
62. Id. § 1536(a)(2).
63. Id.
64. Id.
65. See id. § 1536(a)(2)–(3).
66. See id. § 1533(b)(3)(B)(i).
67. See id. § 1539(a)–(j).
68. See id. § 1539(a)(1)(B).
69. See id.
Section 4 of the ESA provides landowners with an opportunity to learn about, comment on, and provide additional information regarding proposed listings and critical habitat designations.\(^70\) The Section requires the Services to notify the public of proposed rules listing species and designating critical habitat,\(^71\) and provides an opportunity for the public to comment.\(^72\) This process ensures that a landowner who may be affected by the listing of a species will know in advance that the species is proposed for listing.\(^73\) Landowners therefore have an opportunity to voice their concerns about the listing and provide additional information that may inform the Services’ final decision.\(^74\)

Once a species is listed, a landowner may avoid the Act’s prohibitions on the take of the species by applying for and receiving an ITP.\(^75\) These permits grant landowners permission to take listed species or species habitat.\(^76\) In order to receive an ITP, a landowner must submit an acceptable habitat conservation plan (HCP).\(^77\) The FWS or NOAA Fisheries must then find that the landowner’s proposed taking is incidental to, and not the purpose of, his or her proposed action.\(^78\) ITPs may be granted for a variety of activities on non-federal lands such as logging, clearing land, and development.\(^79\)

\(^70\) See id. § 1533(b)(3)(B)(ii), (b)(5)(A)–(E); see also LISTING A SPECIES, supra note 39, at 2.
\(^73\) See LISTING A SPECIES, supra note 39, at 2.
\(^74\) Id.
\(^76\) See id.
\(^77\) Id. § 1539(a)(2)(A). Pursuant to § 1539(a)(2)(A), HCPs must specify:
(i) the impact which will likely result from such taking;
(ii) what steps the applicant will take to minimize and mitigate such impacts, and the funding that will be available to implement such steps;
(iii) what alternative actions to such taking the applicant considered and the reasons why such alternatives are not being utilized; and
(iv) such other measures that the Secretary may require as being necessary or appropriate for purposes of the plan.
\(^78\) Id. § 1539(a)(1)(B).
II. REGULATORY TAKINGS, PER SE PHYSICAL TAKINGS, AND COMPENSABLE PHYSICAL INVASIONS

The Fifth Amendment to the U.S. Constitution provides that private property may not be taken for public use without just compensation. The Takings Clause was designed to prevent the government from instituting an action or regulation that forces a few individuals “to bear the public burdens which, in all fairness and justice, should be borne by the public as a whole.” A landowner affected by such an action may seek compensation through a regulatory takings claim, a physical takings claim, or both. There are two types of physical takings claims: per se physical takings and compensable physical invasions. The requirements for bringing a successful takings claim vary depending on the type of claim asserted. There is one principle, however, that applies to all claims: laws that simply repeat or enforce preexisting limitations on a landowner’s title do not cause a taking.

There are two categories of takings for which landowners may seek compensation: regulatory and physical. In general, regulatory takings arise out of government regulations that, when applied to private land, unfairly reduce or eliminate the land’s economic value or use. Physical takings, on the other hand, occur when the government, or something the government sets in motion or authorizes, physically enters and occupies private property.

The U.S. Supreme Court has recognized two types of physical takings: per se physical takings and compensable physical invasions. Per se physical takings claims stem from the permanent physical occupation of private land by either the government or an entity

80. U.S. CONST. amend. V.
83. See id. at 434–35; Robert Meltz et al., The Takings Issue: Constitutional Limits on Land-Use Control and Environmental Regulation 571 (1999).
84. See infra Parts III.B, IV.B, V.B.
86. See, e.g., Loretto, 458 U.S. at 425–40.
87. See Pa. Coal Co. v. Mahon, 260 U.S. 393, 415 (1922) (stating the guiding principle behind regulatory takings as follows: “while property may be regulated to a certain extent, if regulation goes too far it will be recognized as a taking”).
88. See Meltz et al., supra note 83, at 117.
89. See Loretto, 458 U.S. at 434–35; Meltz et al., supra note 83, at 571.
An occupation of this kind requires compensation because it destroys all of a landowner’s basic property rights. These include the rights to possess the land or exclude others from possessing and using it, control its use, and dispose of it.

Compensable physical invasions result from a temporary or intermittent invasion of private property caused by a government action. These invasions limit a landowner’s right to exclude, but the landowner retains some or all of the other basic property rights. Namely, the landowner may still possess the land, exclude others from possessing and using it, and control its use and disposal.

The U.S. Supreme Court has developed different tests to determine whether a landowner has suffered a regulatory taking, a per se physical taking, or a compensable physical invasion. In spite of these differences, there is one principle courts apply to all takings claims: laws that merely repeat or enforce preexisting limitations on a landowner’s title do not cause a taking. These limitations stem from state property and nuisance law. This rule, announced in Lucas v. South Carolina Coastal Council, recognizes that government actions do not deprive landowners of any rights simply by articulating or implementing limitations on the use of private land that applied to the land prior to private ownership. Such actions therefore do not require compensation.

---

90. See Loretto, 458 U.S. at 426, 432–33 n.9.
91. See id. at 435–36.
92. Id.
93. See id. at 435 n.12.
94. See id.
95. See id.
96. See infra Part III.B.
97. See infra Part IV.B.
98. See infra Part V.B.
100. Id. at 1029.
102. See id. at 1028–29; see also Benson, supra note 4, at 582.
103. See Benson, supra note 4, at 582.
Physical Takings and the ESA

III. REGULATORY TAKINGS CLAIMS UNDER THE ESA

Landowners affected by the ESA have argued that restrictions the Services impose on their property to protect listed species result in regulatory takings. To date, such claims have failed. There are two main reasons for this record. First, regulatory takings claims must meet an exacting ripeness standard. Claims stemming from the ESA often do not meet this standard and therefore are dismissed without adjudication on the merits. Second, these claims generally fail the two U.S. Supreme Court tests to determine whether a given regulation requires compensation.

A. Regulatory Takings Claims Under the ESA Often Fail the U.S. Supreme Court’s Exacting Ripeness Standard

Before a court determines whether a regulatory takings claim has merit, it must conclude that the claim is ripe for review. For a court to draw this conclusion, it must find that the government agency implementing the regulation has rendered a final action. One factor courts consider in determining the finality of a challenged agency action is whether the landowner has exhausted all available administrative remedies and exceptions to the regulation. Courts have established one exception to the finality requirement: the futility doctrine. Under this doctrine, a court will find a regulatory takings claim to be ripe for review if it would be futile for the landowner to pursue further agency action or relief.

105. See Benson, supra note 4, at 552–53.
106. See MELTZ ET AL., supra note 83, at 45–46.
107. See, e.g., Boise Cascade Corp. v. United States, 296 F.3d 1339, 1347 (Fed. Cir. 2002), cert. denied, 123 S. Ct. 1484 (2003) (finding regulatory takings claim was not ripe because the FWS had not denied the ITP); Seiber v. United States, 53 Fed. Cl. 570, 574–75 (2002) (concluding that the FWS’ denial of an incidental take permit, when accompanied by communications indicating a willingness to consider a modified permit application, was not a final agency action and therefore the claim was not ripe).
108. See Meltz, supra note 6, at 385–86.
109. See, e.g., Boise, 296 F.3d at 1347; Seiber, 53 Fed. Cl. at 574–75.
111. See id. at 190.
112. See MELTZ ET AL., supra note 83, at 46.
113. Id.
Regulatory takings claims under the ESA are often based on intermediate agency actions, rather than final actions, and therefore do not fulfill the ripeness requirement. In the case of a claim under the ESA, the finality requirement usually means that the landowner must have applied for and been denied an ITP. To satisfy the exhaustion of remedies factor, the landowner must then submit a revised ITP application, if doing so would not be futile, and have that application denied. Courts have repeatedly rejected regulatory takings claims under the ESA brought by landowners who have not completed this process.

B. Regulatory Takings Claims Fail the U.S. Supreme Court’s Two Tests to Determine Whether a Government Regulation Requires Compensation

If a regulatory takings claim satisfies the ripeness requirement, a court must then decide it on the merits. To determine whether a government regulation has gone “too far,” thereby requiring compensation, a court undertakes an ad hoc factual inquiry. There are two different tests a court uses to inform this inquiry. Under the first test, announced in *Lucas v. South Carolina Coastal Council*, a court determines whether the regulation has deprived the landowner’s property of its entire value or use. If it has, a *per se* regulatory taking has occurred, and the government must compensate the landowner. But, if the property retains some usefulness or value, then a court must undertake the second test and weigh three factors to determine whether the regulation requires compensation. These factors, announced in *Penn Central Transportation Co. v. New York City*, include the character of the government’s action, its effect on the landowner’s reasonable

114. See, e.g., *Boise*, 296 F.3d at 1347; *Seiber*, 53 Fed. Cl. at 575–76.
115. See *Boise*, 296 F.3d at 1349 (noting that extraordinary delay or bad faith on the part of an agency in considering an ITP application can make a claim ripe without a final agency action).
116. See id. at 1347; *Seiber*, 53 Fed. Cl. at 574–75; see also *Meltz*, supra note 6, at 386–87.
117. See, e.g., *Boise*, 296 F.3d at 1347; *Seiber*, 53 Fed. Cl. at 574–75.
119. See id. at 415–16.
120. See Benson, supra note 4, at 581–82.
122. Id.
Physical Takings and the ESA

investment-backed expectations, and its economic impact on the landowner.\(^{124}\) In its most recent decision on the merits of a regulatory takings claim under the ESA, the Court of Appeals for the Federal Circuit based its decision on only one of these factors—the landowner’s reasonable investment-backed expectations.\(^{125}\)

Landowners’ claims asserting that ESA restrictions agencies place on their property give rise to regulatory takings generally fail both tests of the regulatory takings analysis. These claims fail the first test because ESA regulations usually affect some, but not all, of the value or use of a parcel of private land.\(^{126}\) For example, if the Services designate a landowner’s entire acreage as critical habitat, he or she is not automatically prohibited from using the land to hunt and fish for animals other than the listed species.\(^{127}\) If the regulations do not eliminate the entire value or use, the landowner has not suffered a per se regulatory taking.\(^{128}\)

Regulatory takings claims under the ESA generally fail the second test of the regulatory takings analysis because they are not based on reasonable investment-backed expectations.\(^{129}\) The term “investment-backed expectations” is derived from a law review article\(^{130}\) defining it as the crystallized expectations a landowner holds for the continued use of his or her land in a certain manner.\(^{131}\) Because the U.S. Supreme Court itself has not defined the term, however, it is unclear how lower courts should apply this factor.\(^{132}\) In *Good v. United States*,\(^{133}\) the Court of Appeals for the Federal Circuit applied the factor to a regulatory takings claim under the ESA.\(^{134}\) The court determined that a landowner who, at the time of purchase, had actual and constructive knowledge of

\(^{124}\) See id. All three factors are discussed in detail in Part V.B.

\(^{125}\) See *Good v. United States*, 189 F.3d 1355, 1360–63 (Fed. Cir. 1999).

\(^{126}\) Meltz, *supra* note 6, at 385.

\(^{127}\) See *Sugameli*, *supra* note 7, at 450.


\(^{129}\) See *Sugameli*, *supra* note 7, at 447.

\(^{130}\) MELTZ ET AL., *supra* note 83, at 134.


\(^{133}\) 189 F.3d 1355 (Fed. Cir. 1999).

\(^{134}\) Id. at 1360–63.
the Act’s potential application to his land, did not have the investment-backed expectations necessary to satisfy regulatory takings claims. Thus, to satisfy the second step in the regulatory takings test, landowners must have had no knowledge at the time of purchase that the ESA may apply to their land.

IV. **PER SE PHYSICAL TAKINGS UNDER THE ESA**

Faced with courts’ consistent rejection of regulatory takings claims under the ESA, landowners have tried a different approach: asserting that government actions to protect listed species result in *per se* physical takings. *Per se* physical takings occur when the government, or an entity authorized by the government, permanently occupies private land. Claims asserting *per se* physical takings avoid the major roadblocks faced by regulatory taking claims: ripeness, the extensiveness of the injury to the landowner, and the landowner’s reasonable investment-backed expectations. In spite of these distinctions, *per se* physical takings claims have not met with much success in the courts.

In fact, courts have rejected such claims in two recent decisions: *Boise Cascade Corp. v. United States* and *Seiber v. United States*. In *Boise*, the Court of Appeals for the Federal Circuit determined that the government does not control the location of wild animals and

135. *Id.* at 1362.
136. *Id.* The court did not articulate a general principle for what constitutes “knowledge.” It did conclude, however, that the claimant’s acknowledgement of the necessity and difficulty of obtaining regulatory approval for the development, and the language in the sales contract recognizing potential problems with obtaining federal permission to develop the land amounted to adequate “knowledge” that the ESA could be applied to the land. *Id.*
139. *See* *Grant*, [*supra* note 8, at 733].
140. *Id.*
141. *See* *Sugameli*, [*supra* note 7, at 447].
144. 53 Fed. Cl. 570 (2002).
Physical Takings and the ESA

therefore did not cause the listed species to inhabit private land. This rationale is supported by the findings of two older Circuit Court of Appeals decisions: *Mountain States Legal Foundation v. Hodel* and *Christy v. Hodel*. The court in *Seiber*, on the other hand, based its conclusion largely on its finding that the government’s actions did not deprive the landowners of all of their fundamental property rights.

A. Per Se Physical Takings Claims Under the ESA Avoid the Major Roadblocks Posed by Regulatory Takings Claims

Landowners who bring per se physical takings claims for government actions implementing the ESA avoid the major roadblocks faced by those who bring regulatory takings claims. First, unlike regulatory takings claims, which are not ripe until the completion of a final agency action, per se physical takings claims are ripe as soon as the government, or an entity authorized by the government, enters or occupies private property. Proceeding with a per se physical takings claim therefore allows a landowner to avoid the often costly and laborious process of applying for an HCP and ITP prior to seeking compensation from the federal government.

Second, landowners asserting per se physical takings claims can receive compensation for any portion of property the government has occupied, even if this portion is merely a small fraction of the total property. On the other hand, a court can order compensation in a regulatory takings case only if the government action has affected the entire value or use of the property. Thus, the fact that government actions under the ESA generally affect only a portion of a parcel of private land does not present a significant barrier to successful per se physical takings claims under the Act. Finally, courts do not apply the

145. See Boise, 296 F.3d at 1354–55.
146. 799 F.2d 1423 (10th Cir. 1986).
147. 857 F.2d 1324 (9th Cir. 1988).
148. See *Seiber*, 53 Fed. Cl. at 576.
149. See supra Part III.A.
150. See McKenzie v. City of White Hall, 112 F.3d 313, 317 (8th Cir. 1997) (“[P]hysical taking is by definition a final decision.”); see also Grant, supra note 8, at 733.
151. Grant, supra note 8, at 733.
153. See supra note 121 and accompanying text.
154. See Meltz, supra note 6, at 385–86.
Penn Central three-factor balancing test to per se physical takings claims, so the landowner’s reasonable investment-backed expectations are irrelevant to the success of such claims.

B. Requirements for Per Se Physical Takings

In Loretto v. Teleprompter Manhattan CATV Corp., the U.S. Supreme Court established two requirements for a successful per se physical takings claim. First, the occupant must be either the government or an entity authorized by the government. Second, the occupation must destroy all of the landowner’s property rights. These rights include the rights to possess the land, exclude others from it, control the use of the land, and dispose of it. The occupation does not have to destroy the landowner’s property rights for his or her entire parcel of property. Instead, the government must compensate the landowner for any portion of the property it occupies, regardless of the size. If the claim fulfills both of these criteria, the landowner has suffered a per se physical taking, and the government must provide just compensation.

In Loretto, a landlord challenged a New York state law that required landlords to allow the installation of cable boxes on rental properties so tenants could access cable television. The landlord brought a class action against Teleprompter, a cable television company, claiming that the installation constituted a taking without just compensation. The Court agreed, reasoning that the government authorized the boxes’
Physical Takings and the ESA

occupation of the landlord’s property,\textsuperscript{167} and that this occupation destroyed all of the landlord’s property rights for the portion of land on which the boxes were installed.\textsuperscript{168} The landlord could not possess the space occupied by the boxes, exclude the boxes, control the use of the occupied space, or dispose of it.\textsuperscript{169} The Court reasoned that because permanent physical occupations destroy all of these rights, the government must provide compensation for such occupations, regardless of the size of the area affected\textsuperscript{170} or the public purpose served.\textsuperscript{171}

C. In Two Recent Cases, Courts Held that Government Actions Under the ESA Did Not Support Claims of Per Se Physical Takings

In two recent decisions, courts rejected claims that government actions under the ESA to protect imperiled species give rise to \textit{per se} physical takings.\textsuperscript{172} Courts in \textit{Boise Cascade Corp. v. United States} and \textit{Seiber v. United States} applied the \textit{Loretto} test and rejected claims by private landowners that government actions protecting listed species resulted in \textit{per se} physical takings.\textsuperscript{173} In \textit{Boise}, the Court of Appeals for the Federal Circuit held that the landowner had not suffered a \textit{per se} physical taking because the government does not control the location of wild animals and therefore did not cause the species to inhabit private land.\textsuperscript{174} Furthermore, in \textit{Seiber}, the U.S. Court of Federal Claims held that government actions under the ESA did not amount to a \textit{per se} physical taking because the government did not deprive the landowners of all of their property rights.\textsuperscript{175} The U.S. Court of Federal Claims hears all takings claims against the federal government for more than $10,000.\textsuperscript{176} Parties can appeal its decisions to the Court of Appeals for the Federal Circuit.\textsuperscript{177}

\begin{itemize}
  \item \textsuperscript{167} \textit{Id.} at 425–26.
  \item \textsuperscript{168} \textit{Id.} at 438.
  \item \textsuperscript{169} \textit{Id.} at 435–36, 438.
  \item \textsuperscript{170} \textit{Id.} at 438 n.16. \textit{But cf. id.} at 437 (noting that a court should consider the extent of the occupation as one factor in determining the amount of compensation due).
  \item \textsuperscript{171} \textit{Id.} at 426, 434–35.
  \item \textsuperscript{173} See \textit{Boise}, 296 F.3d at 1354–55; \textit{Seiber}, 53 Fed. Cl. at 576.
  \item \textsuperscript{174} See \textit{Boise}, 296 F.3d at 1354–55.
  \item \textsuperscript{175} See \textit{Seiber}, 53 Fed. Cl. at 576.
  \item \textsuperscript{176} Tucker Act, 28 U.S.C. § 1346 (1994).
  \item \textsuperscript{177} \textit{Id.} § 1291; see also \textsc{Meltz ET AL.}, supra note 83, at 43 ("Because few takings cases are
1. The Government Does Not Control the Location of Wild Animals That Are Protected Under Federal Statutes

In Boise, the Court of Appeals for the Federal Circuit held that the presence of wildlife protected by the ESA on private land does not constitute a government occupation of private land.\(^{178}\) This is true because the government does not control the location of wildlife protected by the ESA.\(^{179}\) The Boise case arose from a dispute between the Boise Cascade Corporation (Boise) and the FWS over the management of a 65-acre parcel of Boise’s land known as the Walker Creek Unit.\(^{180}\) This parcel hosted, and contained habitat for, the federally-listed northern spotted owl.\(^{181}\) When Boise attempted to log the parcel, the FWS conducted an inspection and determined that logging could harm owls that might otherwise use the site as nesting habitat.\(^{182}\) It then notified Boise that the corporation must secure an ITP prior to logging the parcel.\(^{183}\) Boise filed suit in federal district court seeking an injunction preventing the FWS from enforcing the ESA against it and a declaratory judgment that its logging operation would not take spotted owls.\(^{184}\) The district court dismissed Boise’s claim for lack of ripeness\(^{185}\) and instead granted the United States’ request for an injunction preventing the corporation from logging on the site without an ITP.\(^{186}\) Boise then applied for an ITP.\(^{187}\) After conducting additional surveys of the Walker Creek Unit, the FWS determined that spotted owls would not be taken by the corporation’s planned logging activities and informed

appealed from the Federal Circuit to the U.S. Supreme Court, the Federal Circuit has a leading role in articulating takings law in the many areas where the High Court has not yet spoken.”).  

178. See Boise, 296 F.3d at 1353–55.  
179. Id.  
180. Id. at 1341.  
181. Id. at 1341–42. In 1990, the FWS listed the northern spotted owl as a threatened species under the ESA, 50 C.F.R. § 17.11 (current through June 16, 2003). See also Endangered and Threatened Wildlife and Plants; Determination of Threatened Status for the Northern Spotted Owl, 55 Fed. Reg. 26,114, 26,194 (June 26, 1990).  
182. Boise, 296 F.3d at 1341–42.  
183. Id. at 1342.  
184. Id.  
185. Id. at 1342 n.3.  
186. Id. at 1342.  
187. Id.
Physical Takings and the ESA

the corporation that an ITP was no longer required.\textsuperscript{188} The district court subsequently lifted the injunction.\textsuperscript{189}

Boise returned to court, seeking compensation for the temporary taking of merchantable timber it was prevented from logging because of the district court’s injunction.\textsuperscript{190} One theory Boise presented\textsuperscript{191} was that the injunction was a \textit{per se} physical taking under \textit{Loretto}\textsuperscript{192} because the corporation was not allowed to exclude spotted owls from its property by harvesting the timber.\textsuperscript{193} The court dismissed Boise’s complaints, and Boise appealed.\textsuperscript{194}

In its decision upholding the dismissal, the Court of Appeals for the Federal Circuit relied on \textit{Loretto} and concluded that, because government actions under the ESA merely restrict what landowners can do with their land once a species has been listed, rather than actually cause the species to inhabit the land, such actions do not give rise to \textit{per se} physical takings.\textsuperscript{195} Boise had argued that \textit{Loretto} should be applied to its situation because the “occupation [of its land] by wild spotted owls [is] indistinguishable from a forced government intrusion upon its land.”\textsuperscript{196} In rejecting this argument, the court identified a critical difference between the two situations. Unlike \textit{Loretto}, in which the government regulation in question required the cable boxes to be installed on private lands,\textsuperscript{197} the government actions in \textit{Boise} were not responsible for the owls’ presence on the corporation’s land.\textsuperscript{198}

\textsuperscript{188} Id.
\textsuperscript{189} Id.
\textsuperscript{190} Id. The other claims included a categorical taking, an exaction-type taking, and a temporary regulatory taking. \textit{Id.} at 1342–43.
\textsuperscript{191} Id. The other claims included a categorical taking, an exaction-type taking, and a temporary regulatory taking. \textit{Id.} at 1342–43.
\textsuperscript{193} \textit{Boise}, 296 F.3d at 1352. Boise also argued that the requirement that it allow government agents onto its land to conduct owl surveys constituted a compensable physical taking. The court rejected this claim, finding that under \textit{Loretto}, “[t]ransient, nonexclusive entries by the [FWS] to conduct owl surveys do not permanently usurp Boise’s exclusive right to possess, use, and dispose of its property.” \textit{Id.} at 1355.
\textsuperscript{194} \textit{Id.} at 1343.
\textsuperscript{195} See \textit{id.} at 1354–55.
\textsuperscript{196} \textit{Id.} at 1354.
\textsuperscript{197} See \textit{Loretto}, 458 U.S. at 421.
\textsuperscript{198} See \textit{Boise}, 296 F.3d at 1354–55 (“The government has no control over where the spotted owls nest, and it did not force the owls to occupy Boise’s land. The government simply imposed a temporary restriction on Boise’s exploitation of certain natural resources located on its land unless Boise obtained a permit.”).
Two older cases support the Boise court’s conclusion that the government does not control wildlife protected by federal statute. First, in Mountain States Legal Foundation v. Hodel, the U.S. Court of Appeals for the Tenth Circuit considered whether damage to private land caused by wild horses and burros should be compensated as a per se physical taking.\textsuperscript{199} The federal government granted protection to the animals under the Wild Free-Roaming Horses and Burros Act,\textsuperscript{200} which prohibited landowners from removing the animals from private land.\textsuperscript{201} The landowners claimed that because the government did not respond to their requests for removal of the animals,\textsuperscript{202} the damage the animals caused through grazing and erosion should be compensated as a taking of private property.\textsuperscript{203} In rejecting this claim, the court determined that the Wild Free-Roaming Horses and Burros Act did not give the government ownership of the animals.\textsuperscript{204} The offending horses therefore were not “instrumentalities of the federal government”\textsuperscript{205} whose presence gives rise to a physical taking.\textsuperscript{206}

The U.S. Court of Appeals for the Ninth Circuit applied the Mountain States court’s reasoning to a physical takings claim arising under the ESA in Christy v. Hodel.\textsuperscript{207} The plaintiff in Christy claimed that by listing grizzly bears as threatened under the Act, the government made the bears government agents that physically took his property.\textsuperscript{208} The court rejected this claim, reasoning that the government neither owns nor controls the conduct of the species it protects.\textsuperscript{209} Thus, both the Court of Appeals for the Federal Circuit and the Ninth Circuit have based their rejection of per se physical takings claims under the ESA on the reasoning that the government does not control the location of wild animals.

\textsuperscript{199} See Mountain States Legal Found. v. Hodel, 799 F.2d 1423, 1424–28 (10th Cir. 1986).
\textsuperscript{201} See Mountain States, 799 F.2d at 1425.
\textsuperscript{202} Id.
\textsuperscript{203} Id. at 1425–26.
\textsuperscript{204} Id. at 1426–27 (“[I]t is pure fantasy to talk of ‘owning’ wild fish, birds, or animals. Neither the States nor the Federal Government . . . has title to these creatures until they are reduced to possession by skillful capture.”) (quoting Douglas v. Seacoast Products, Inc., 431 U.S. 265, 284 (1977)) (internal quotations omitted) (internal citation omitted).
\textsuperscript{205} Id. at 1428.
\textsuperscript{206} Id.
\textsuperscript{207} 857 F.2d 1324, 1334–35 (9th Cir. 1988).
\textsuperscript{208} Id. at 1334.
\textsuperscript{209} Id. at 1335.
Physical Takings and the ESA

2. The Presence of a Protected Species on Private Land Does Not Destroy All of a Landowner’s Property Rights in the Inhabited Parcel

In Seiber v. United States, the U.S. Court of Federal Claims found that the presence of the northern spotted owl on private land did not destroy all of the landowner’s property rights to the inhabited parcel. At the heart of the Seiber case were restrictions the Oregon Department of Forestry placed on the Seibers’ property to protect northern spotted owls. The department’s regulations required seventy acres of land around every nesting tree to be designated as nesting habitat and prevented logging within these areas. In 1994, the State of Oregon designated forty acres of the Seibers’ property as nesting habitat. Five years later, the Seibers sought to log the area and submitted an HCP and an ITP application. The Department of the Interior’s Regional Solicitor informed the couple that their ITP application was not adequate, offering to discuss modifications. The landowners rejected this offer, and the FWS denied the application. The Seibers then brought a takings action in the U.S. Court of Federal Claims. Under the court’s order, the FWS visited the plaintiffs’ property to evaluate the status of the habitat and determined that an ITP was no longer required for the Seibers to log the property. The plaintiffs pursued their claim, arguing that the Service’s denial of an ITP resulted in regulatory and physical takings under various theories.

One theory the Seibers asserted was that the government’s action of prohibiting logging prevented them from excluding owls from the land and therefore resulted in a per se physical taking. In rejecting this claim, the court determined that the government’s actions did not strip

211. Id. at 572–74.
212. Id. at 572.
213. Id.
214. Id.
215. Id. at 573.
216. Id.
217. Id.
218. Id.
219. Id.
220. See id. at 573–74.
221. Id. at 575–76.
the landowners of their fundamental property rights. First, the owls’ presence did not prevent the Seibers from possessing the land or excluding others from it. Second, the government’s actions did not take away the landowners’ rights to control the property and put it to other uses, even though it prevented them from logging the property. Thus, under Seiber, a landowner who retains his or her fundamental property rights following a government action under the ESA has not suffered a per se physical taking.

In sum, although per se physical takings claims avoid several of the procedural and substantive hurdles that regulatory takings claims face, courts have consistently rejected them for two reasons. First, the government does not control the location of wild animals protected by federal statutes. Second, the presence of a protected species on private land does not destroy all of a landowner’s fundamental property rights.

V. COMPENSABLE PHYSICAL INVASIONS UNDER THE ESA

The final type of takings claim is the compensable physical invasion. Compensable physical invasion claims result from temporary or intermittent invasions of private property caused by a government action. These invasions limit a landowner’s right to exclude but leave some or all of the other basic property rights intact. A landowner may suffer a compensable physical invasion and still possess the land, retain the ability to prevent others from possessing and using it, and control its use and disposal. Compensable physical invasions are physical takings. Thus, as with per se physical takings claims, compensable physical invasion claims under the ESA avoid several of the roadblocks faced by claims of regulatory takings. In determining whether a physical invasion claim will receive compensation, it is likely that a court will undertake the three-factor

222. Id. at 576.
223. Id.
224. Id.
225. See id. at 576. The court did not discuss whether the government’s actions prevented the landowners from disposing of their land because the plaintiffs had not raised the claim. Id.
227. See id.
228. See id.
229. See id.
230. See supra notes 89–95 and accompanying text.
Physical Takings and the ESA

balancing test the U.S. Supreme Court developed in *Penn Central Transportation Co. v. New York City.*231 This test requires a court to weigh the character of the government action, the effect of the action on the landowner’s reasonable investment-backed expectations, and the action’s economic impact on the landowner.232

A. Compensable Physical Invasion Claims Under the ESA Avoid the Roadblocks Faced by Regulatory Takings Claims

Compensable physical invasions are another type of physical taking.233 Consequently, landowners’ claims of compensable physical invasions have several of the same advantages over regulatory takings claims as those enjoyed by *per se* physical takings claims.234 First, compensable physical invasion claims are ripe as soon as the invasion occurs.235 Thus, a landowner may seek compensation from the federal government without having completed the process of applying for an HCP and ITP.236 Second, landowners may receive compensation for any portion of property the government has occupied in the invasion, regardless of whether this area is merely a small fraction of the total property.237

While *per se* physical takings and compensable physical invasions share these two advantages over regulatory takings, there is one advantage *per se* physical takings claims enjoy that compensable physical invasion claims do not. Namely, courts do not consider the landowners’ reasonable investment-backed expectations when evaluating *per se* takings claims.238 For compensable physical invasion claims, however, these expectations are one factor the courts consider in determining whether a physical invasion requires compensation.239


232. See *Penn Cent.,* 438 U.S. at 124.

233. See *Loretto,* 458 U.S. at 435 n.12.

234. See supra Part IV.A.

235. See, e.g., Sinaloa Lake Owners Ass’n v. City of Simi Valley, 882 F.2d 1398, 1402 (9th Cir. 1989) (stating that the “final decision requirement is inapplicable in cases of physical invasion”); see also Grant, supra note 8, at 733.

236. See Grant, supra note 8, at 733.

237. See *Loretto,* 458 U.S. at 438 n.16.

238. See supra Part IV.A.

239. See infra Part V.B.
B. Courts Consider Three Factors to Determine Whether a Physical Invasion Requires Compensation

The Loretto Court explicitly excluded physical invasions from its rule that permanent physical occupations of private land due to government actions require compensation as *per se* physical takings.240 In dicta, however, the Court suggested that lower courts use the *Penn Central* balancing test to determine whether a compensable physical invasion has occurred.241 Under this test, courts weigh three major factors—the character of the government action, the effect of the action on the landowner’s reasonable investment-backed expectations, and the economic impact of the action on the landowner.242

1. In Assessing the Character of a Government Action, Courts Consider the Degree to Which the Action Allows Physical Government Interference With Private Property

To determine whether a government action resulting in a physical invasion is compensable, courts first consider the character of the action itself.243 In *Penn Central*, the U.S. Supreme Court stated that the primary consideration courts are to make in evaluating the character of the government action is whether the action is physical in nature.244 If the action includes a physical invasion by the government, a court is more likely to find the landowner has suffered a taking.245 One example of the application of this factor is *United States v. Causby*.246 In *Causby*, the U.S. Supreme Court concluded that government flights over private property at low altitudes, which caused direct and immediate harm to the landowners, were physical invasions of private property by the government.

240. *Loretto*, 458 U.S. at 435 n.12 (“The permanence and absolute exclusivity of a physical occupation distinguish it from temporary limitations on the right to exclude. Not every physical invasion is taking . . . . [T]emporary limitations are subject to a more complex balancing process to determine whether they are a taking.”) (emphasis in original) (citations omitted).
241. See id. at 434–35.
243. Id.
244. See id. (“A ‘taking’ may more readily be found when the interference with property can be characterized as a physical invasion by government . . . than when interference arises from some public program adjusting the benefits and burdens of economic life to promote the common good.”) (citation omitted).
245. Id.
246. 328 U.S. 256 (1946); see also *Penn Central*, 438 U.S. at 124 (citing *Causby*).
Physical Takings and the ESA

government. Such actions therefore required compensation for a taking. Thus, the more the action results in the physical presence of the government on private land, the more likely a court is to find that the landowner has suffered a taking.

2. A Government Action Resulting in a Physical Invasion That Destroys a Landowner’s Reasonable Investment-Backed Expectations Requires Compensation If Those Expectations Were Based on Assurances Provided by the Government

The second factor in the *Penn Central* balancing test is the effect of the government action on the landowner’s reasonable investment-backed expectations. Although the U.S. Supreme Court has not explicitly defined the term “investment-backed expectations,” it has considered this factor in assessing a takings claim stemming from a physical invasion in *Kaiser Aetna v. United States*. Thus, in *Kaiser*, the Court set the parameters for applying a takings claim stemming from a physical invasion.

In *Kaiser*, the Court determined that landowners’ investment-backed expectations for property are reasonable when they are based on assurances provided by the government. If the government causes a physical invasion that destroys these expectations, compensation is due. The dispute in *Kaiser* arose when the Army Corps of Engineers granted a developer permission to dredge a navigable channel between the Pacific Ocean and a formerly land-locked pond on Oahu, Hawaii as part of an effort to create a marina-style community around the pond. Eleven years later, after the marina was complete and Kaiser had spent millions of dollars on the effort, the government claimed the public had a right of access to the marina. The government reasoned that, because the marina was linked to navigable waters, it was subject to a

---

248. *Id.*
250. *See supra* notes 130–32 and accompanying text.
252. *Id.* at 179.
253. *Id.* at 179–80.
254. *Id.* at 167.
255. *Id.* at 170.
navigational servitude. The United States brought an action against Kaiser to resolve whether the corporation could deny the public access to the marina in spite of Kaiser’s past improvements that turned the marina into navigable waters of the United States. The Court held that the government’s approval of the original marina project made the corporation’s expectation—that it would be able to exclude the public from the marina—reasonable. Subsequent government actions asserting the public’s right of access to the marina destroyed the corporation’s right to exclude the public from its property. This combination—creating a property-expectancy and then destroying it by requiring Kaiser to acquiesce to a physical invasion—required compensation.

3. A Government Action That Does Not Provide a Landowner With a Mechanism to Control a Physical Invasion and Mitigate the Invasion’s Economic Impact May Require Compensation

The final factor in the Penn Central balancing test is the economic impact of the government action. In PruneYard Shopping Center v. Robins, the U.S. Supreme Court determined that, when considering the economic impact of a physical invasion, courts should look to the landowner’s degree of control over the invasion. If the landowner had an available mechanism with which to control the time, place, and manner of the invasion in such a way as to reduce its economic impact, courts are unlikely to award compensation.

In PruneYard, the owners of a shopping center asserted that a California Supreme Court decision requiring the center to allow students onto their property to collect signatures effected a compensable physical invasion. After students set up a booth on the center’s property to distribute literature and collect signatures in opposition to a United

256. Id.
257. Id. at 168–69.
258. Id. at 179.
259. Id. at 179–80.
260. Id.
263. See id. at 83–84.
264. See id.
265. Id. at 82.
Physical Takings and the ESA

Nations resolution, they were ordered off the premises. The students brought an action in California state court, arguing that it was a violation of their First Amendment rights not to be able to solicit signatures on PruneYard’s property. The California Supreme Court held that the state constitution protected “speech and petitioning, reasonably exercised, in shopping centers even when the centers are privately owned.” On review, the U.S. Supreme Court held that the California Supreme Court’s requirement that the students be allowed to exercise protected free expression rights did not amount to a taking. The Court reasoned that the shopping center retained the ability, through the enactment and enforcement of regulations, to restrict the “time, place, and manner” in which the students could make public expressions within the center. This control provided PruneYard’s owners with a mechanism through which to ensure that the students’ presence would have minimal economic effects on the center. Thus, the Court concluded that a compensable physical invasion had not occurred.

VI. GOVERNMENT ACTIONS UNDER THE ESA DO NOT GIVE RISE TO COMPENSABLE PHYSICAL INVASIONS

Compensable physical invasion claims are the only remaining untested option for landowners seeking government compensation for a taking under the ESA. Such claims present several advantages over regulatory and per se physical takings claims. These advantages suggest that compensable physical invasion claims are likely to be the next type of claim advanced by landowners affected by the ESA. When faced with such a claim, courts should reject it for two reasons. First, courts should find that the protection of wildlife on private land is a background principle of property. As such, government actions implementing protections for wildlife on private land can never result in

266. Id. at 77.
267. Id.
269. See PruneYard, 447 U.S. at 83.
270. Id.
271. Id.
272. Id. at 84.
273. See infra Part VI.A.
274. See infra Part VI.A.
275. See infra Part VI.B.
takings. Second, physical invasion claims under the ESA will most likely fail each prong of the Penn Central balancing test. 277

A. Compensable Physical Invasion Claims Under the ESA Are Likely To Be the Next Type of Claim Advanced By Landowners Affected By the ESA

Courts’ rejections of regulatory and per se physical takings claims under the ESA suggest that landowners seeking compensation for takings under the Act may be left with only one option—claims of compensable physical invasions. 278 Although no court has decided such a claim, there are three reasons why landowners are likely to raise them. First, because compensable physical invasions are physical takings, 279 landowners do not have to complete HCPs and have the Services take final actions on the plans prior to bringing claims. 280 Instead, compensable physical invasion claims are ripe as soon as the physical invasion occurs. 281 Second, as with per se physical takings, landowners advancing compensable physical invasion claims may seek compensation for any portion of their property affected by the invasion. 282 Successful regulatory takings claims, on the other hand, require that the government action affect the entire piece of land owned by the claimant. 283

Finally, compensable physical invasion claims do not require as extensive an interference with private land as that required for successful per se physical takings claims. A per se physical takings claim requires a landowner to prove that a government action caused a permanent physical occupation of his or her land by a listed species. 284 A landowner may assert a compensable physical invasion claim, on the other hand, merely by showing that a government action limited his or her right to exclude the species from the land. 285 Thus, an injunction or other

276. See infra Part VI.B.
277. See infra Part VI.C.
278. See supra Part V.A.
279. See supra notes 93–95 and accompanying text.
280. See supra notes 235–36 and accompanying text.
281. See Grant, supra note 8, at 733.
283. See supra notes 126–28 and accompanying text.
284. See supra notes 158–64 and accompanying text.
Physical Takings and the ESA

government action that prevents a landowner from taking actions to exclude a listed species—such as fencing land or harvesting trees—technically may result in a physical invasion.

B. The ESA Enforces Preexisting Limitations on Landowners’ Titles, and Therefore Its Implementation Does Not Result in a Compensable Physical Invasion

In *Lucas v. South Carolina Coastal Council*, the U.S. Supreme Court held that laws that place limitations on a landowner’s title never cause a taking when the limitations are grounded in background principles of property and nuisance. The common law, which defines these background principles, provides that the state holds wildlife in trust for the benefit of the public. Commentators have asserted that limits on the use of private land that are necessary to further the government’s interest in managing wildlife therefore constitute a background principle of property. As such, these limits do not give rise to takings. Thus, federal wildlife conservation statutes such as the ESA simply recognize the government’s ability to enforce preexisting limitations on a landowner’s title to protect wildlife. Claims of compensable physical invasions under the ESA therefore should fail because the Act merely enforces a limitation on a landowner’s title that is rooted in background principles of property.

C. Claims of Compensable Physical Invasions Stemming from Government Actions Implementing the ESA Fail the Penn Central Balancing Test

If a court found that a government action under the ESA went beyond simply implementing a preexisting limitation on the landowner’s title, it would consider the merits of the takings claim. No court has yet considered a claim of a compensable physical invasion arising from a government action under the Act. When such a claim arises, however, a

286. These situations are analogous to those in *Kaiser* and *PruneYard*, in which government actions provided for the intermittent presence of boats and people, respectively, on private property. See *supra* notes 254–55, 265–66 and accompanying text.
289. See, e.g., *Sugameli, supra* note 7, at 443; *Houck, supra* note 43, at 308–21.
290. See *Sugameli, supra* note 7, at 443.
291. See *id.; Houck, supra* note 43, at 320.
court should apply the *Penn Central* three-factor balancing test. This test is appropriate because in *Loretto*, the U.S. Supreme Court stated that physical invasion takings claims are subject to a “complex balancing process.”292 The Court then referenced the *Kaiser* and *PruneYard* decisions, both of which used the *Penn Central* test, as examples of the Court’s application of this balancing process to such claims.293 A consideration of the *Penn Central* decision in light of legal precedents and the ESA’s statutory safeguards strongly suggests that courts should reject these claims.

1. **Government Actions Under the ESA Are Not of the Character Necessary to Support Claims of Compensable Physical Invasions Because They Do Not Provide for a Physical Government Presence on Private Land**

   Actions the Services take to enforce the ESA do not provide for a physical government presence on private land and therefore are not of the character necessary to support claims of compensable physical invasions. Under the U.S. Supreme Court’s standard for determining whether the character of a government action bolsters such claims, a court must consider the extent to which the action results in the government’s physical presence on private land.294 Government actions to enforce the ESA do not result in a physical government presence on private land. In *Mountain States*, *Christy*, and *Boise*, three different circuit courts determined that federally-protected species are not agents of the government.295 Thus, actions under the ESA requiring landowners to allow listed species onto private land do not result in a physical governmental presence on the land.

---

293. *Id.*
Physical Takings and the ESA

2. Notification Requirements for Listing Decisions and Section 7 Consultation Requirements Preclude Claims That Government Actions Interfere with Landowners’ Reasonable Investment-Backed Expectations

Even if a court found that the government controlled a listed species, and that the species’ inhabitance of private land comprised a government presence on the property, a claim that such a presence is compensable should still fail. Applying the other two factors in the Penn Central test should lead a court to find that such invasions do not require compensation. The second factor in this test is the landowner’s reasonable investment-backed expectations. Reasonable investment-backed expectations are crystallized expectations a landowner holds for the continued use of his or her land in a certain manner. In Kaiser Aetna v. United States, the U.S. Supreme Court established that compensation for a taking is due to landowners who base their investment-backed expectations on government assurances and then have those expectations destroyed by government-authorized physical invasions. Sections 4 and 7 of the ESA protect against such an occurrence by requiring the Services to provide landowners with notice of listing events and consult with agencies that undertake or authorize activities on private land that may affect listed species.

The notice requirement of Section 4 should ensure that, once a species is listed, the investment-backed expectations landowners hold for private property are “reasonable.” Under Section 4, landowners who may be affected by a listing receive notice that either the FWS or NOAA Fisheries has listed a species or designated critical habitat. Once notified, landowners can no longer modify the species’ habitat located on their property without an approved HCP and thus must incorporate these prohibitions into their investment-backed expectations. Landowners’ expectations that they will be able to take any actions on their land, regardless of the impact the actions have on the listed species, therefore is no longer reasonable after notice has been given. For

---

296. See Michelman, supra note 131, at 1233.
297. See Kaiser Aetna v. United States, 444 U.S. 164, 179–80 (1979); see also supra Part V.B.2.
300. Id. § 1533(b)(3)(B)(ii), (b)(5); see also LISTING A SPECIES, supra note 39, at 2.
301. See 16 U.S.C § 1538(a)(1)(B).
example, in *Boise*, the landowner knew that the northern spotted owl was listed as threatened and that habitat for the species was present on the property prior to the landowner’s initiation of efforts to log the site. A court therefore should find such a landowner’s expectation that it would be able to log freely without obtaining an ITP to be unreasonable.

The Section 7 consultation requirement presents another roadblock to landowners’ assertions that government actions, which result in physical invasions of private property, destroy their reasonable investment-backed expectations. This Section requires federal agencies to consult with the FWS or NOAA Fisheries to ensure that actions they authorize, fund, or carry out will be unlikely to jeopardize the existence of any listed species, or destroy or adversely modify critical habitat. Thus, any agency that provides a landowner with approval of a project on private land must first establish that the project will not harm a listed species or its habitat. This requirement should prevent situations such as that in *Kaiser*, in which an agency provided a landowner with assurances and then destroyed those assurances by requiring the landowner to submit to a physical invasion.

3. The Availability of ITPs Under the ESA Mitigates the Economic Impact of the Government’s Actions on Landowners

Landowners’ claims of compensable physical invasions under the ESA also fail the third prong in the *Penn Central* balancing test. In *PruneYard*, the U.S. Supreme Court established that landowners who have a mechanism to mitigate the economic effects of a government invasion by controlling its time, place, and manner should not receive compensation for a physical invasion. The ESA’s ITP provision grants landowners such a mechanism.

Section 10’s ITP provision provides private landowners affected by the Act with a mechanism through which they can mitigate the economic effects of the government’s action. This provision allows landowners

304. Id., see also supra notes 62–65 and accompanying text.
305. See supra notes 252–60 and accompanying text.
308. See id.; see also supra notes 75–79 and accompanying text.
Physical Takings and the ESA

to develop and submit an HCP for their land. If the FWS or NOAA Fisheries accepts the plan, landowners may receive an ITP to “take” individuals of a listed species or harvest segments of critical habitat.309 Under this process, landowners can determine the sections of property that are most economically valuable and seek an ITP to harvest them.310 By allowing landowners to propose areas of habitat to protect and areas to eliminate, the habitat conservation planning process also gives landowners some control over where a listed species may reside on their property. If the FWS determines that the landowner has taken adequate measures to mitigate for the destruction of habitat, and for the possible take of individual owls, it will grant an ITP and allow him or her to proceed.311 While the HCP development process can have significant costs312 receiving an ITP should provide landowners with the same ability to mitigate the economic effects of maintaining suitable habitat for listed species on their land as that enjoyed by the landowners in PruneYard.313

The Boise case illustrates the difficulty landowners would face in presenting a convincing case that the government’s actions under the ESA had an economic impact severe enough to warrant compensation as a physical invasion. Boise did not file an application for an ITP until the federal district court enjoined the corporation from logging without one.314 Thus, Boise had not used the available mechanisms to reduce the economic impact of the ESA on its land. This omission would make it difficult, if not impossible, for the corporation to claim that the government’s actions implementing the Act on its land resulted in a compensable economic impact. Had Boise applied for an ITP at the outset, it could have proposed a mitigation plan, such as undertaking limited logging in the area inhabited by the owl in return for preserving habitat elsewhere on its property.

309. See 16 U.S.C. § 1539(a); see also supra notes 75–79 and accompanying text.

310. See 16 U.S.C. § 1539(a); see also supra notes 75–79 and accompanying text.

311. See 16 U.S.C. § 1539(a); see also supra notes 75–79 and accompanying text.


VII. CONCLUSION

Given the failure of regulatory and *per se* physical takings claims in the courts, it is likely that landowners affected by the ESA will soon turn to compensable physical invasion claims. When such claims are presented, courts should reject them. Existing precedent and statutory safeguards in the ESA make it difficult, if not impossible, for private landowners to argue successfully that any of the three factors in the *Penn Central* balancing test rise to the level of requiring compensation for a physical invasion. The Act’s protection of imperiled species does not come at the expense of private landowners’ fundamental property rights. Thus, courts should reject landowners’ claims that ESA restrictions result in compensable physical invasions.
FAIR AND REASONABLE COMPENSATION MEANS JUST THAT: HOW § 253 OF THE TELECOMMUNICATIONS ACT PRESERVES LOCAL GOVERNMENT AUTHORITY OVER PUBLIC RIGHTS-OF-WAY

Jennifer Amanda Krebs, M.A.

Abstract: Section 253(c) of the 1996 Telecommunications Act expressly preserves local government authority to require fair and reasonable compensation from telecommunications providers for use of public rights-of-way. Although local government authority to require compensation for franchises is based in state law, some courts have overlooked state law when evaluating the validity of franchise fees. In addition, courts have interpreted § 253(c) narrowly, allowing local governments to recover only direct costs. This narrow interpretation of § 253(c) contradicts its text and legislative history, as well as analogous United States Supreme Court precedent. Further, this interpretation could lead to unconstitutional results, by allowing a taking of public property without just compensation and permitting Congress to commandeer local governments into implementing a federal regulatory program.

A new technology emerged and transformed peoples’ lives. The technology brought remarkable advances in communications—families separated by vast distances became connected, businesses explored new markets and served previously unreachable customers. At the same time, new concerns surfaced. Who was in control of this technology? What mechanisms were in place to ensure privacy for the information transmitted through this medium? How could the majority of people access and benefit from the technology, rather than the privileged few? Understanding the importance of this technology to the nation, the federal government took action to encourage its proliferation. Local governments, eager to benefit their citizens, struggled to retain traditional control over companies doing business within their jurisdictions and to protect public rights-of-way.

What was this new technology that opened so many doors yet created so many issues for the nation? The internet? The cell phone? No—the telegraph, in the nineteenth century. Over one hundred years later, local governments face strikingly similar issues with the proliferation of telecommunications technologies.

2. See infra Part IV.
In 1866, Congress passed the Post Roads Act (Telegraph Act) to encourage the development of new telegraph technology.\(^3\) One hundred thirty years later, Congress passed the Telecommunications Act of 1996\(^4\) (1996 Act) to encourage the development of telecommunications technology, including broadband access to the internet.\(^5\) Section 253(a) of the 1996 Act prevents state and local governments from prohibiting the entry of telecommunications providers into local markets.\(^6\) At the same time, § 253(c) expressly preserves local governments’ rights to require compensation from telecommunications companies that use public rights-of-way.\(^7\) Courts have disagreed about how much compensation local governments can require, and how to calculate this compensation.\(^8\)

This Comment argues that under § 253, courts should first consider state law to determine the local government’s authority to impose a franchise fee, and then apply a “totality of the circumstances” approach to determine whether the fee is fair and reasonable. Part I of this Comment discusses public rights-of-way and franchising under state law, and the scope of federal authority over these areas. Part II explains how § 253(c) of the 1996 Act preserves local government authority over public rights-of-way, and Part III summarizes the various ways courts have interpreted this section. Part IV describes the United States Supreme Court’s interpretation of the Telegraph Act of 1866,\(^9\) which upheld the authority of local governments to require right-of-way fees. Part V argues that the correct way for courts to apply § 253(c) is to first examine local government franchise and right-of-way authority under state law, and then consider the totality of the circumstances to determine if a franchise fee is fair and reasonable. Part V further argues that unless courts interpret § 253(c) in this way, unconstitutional results could occur, such as the taking of public property without just compensation, or the federal government commandeering local governments to implement a federal regulatory program.

---

5. See City of Auburn v. Qwest Corp., 260 F.3d 1160, 1170 (9th Cir. 2001).
8. See *infra* Part III.B.
Fair and Reasonable Compensation

I. LOCAL GOVERNMENTS CAN REQUIRE COMPANIES USING PUBLIC RIGHTS-OF-WAY TO PAY COMPENSATION AND OBTAIN FRANCHISES

State law governs how local governments own and control property. Under state law, local governments control public rights-of-way, which are segments of land designated for public use. Local governments use franchise agreements to grant utilities access to these rights-of-way, and state law constrains the terms of these agreements. State governments, and not the federal government, are sovereign over local governments in the context of public rights-of-way and franchising authority. Nevertheless, federal constitutional guarantees protect local governments from congressional overreaching.

A. State Law Controls Local Government Interests in Public Rights-of-Way

Public rights-of-way are “public corridors or strips of land,” such as streets and alleys, designated for public travel. Portions of rights-of-way are often used for utility infrastructure. This infrastructure can include wires strung on poles above and through streets and alleys, wires and cables buried below streets in conduits, and utility boxes installed on public property. Public rights-of-way are a valuable

12. See id. § 30.40.
14. See MCQUILLIN, supra note 10, § 34.10.
15. See infra Part I.A–B.
16. See infra Part I.C.
18. See id.; see also MCQUILLIN, supra note 10, § 30.03.
19. See DEVANEY, supra note 17, § 3.3.
20. See MCQUILLIN, supra note 10, § 30.10.
21. See id. § 30.06.
22. See AT&T Communications v. City of Dallas, 52 F. Supp. 2d 763, 766 (N.D. Tex. 1999), vacated as moot by 243 F.3d 928 (5th Cir. 2000); Cox Communications v. City of San Marcos, 204 F. Supp. 2d 1272, 1274–75 (S.D. Cal. 2002).
asset\textsuperscript{23}—they have value as land\textsuperscript{24} and are necessary for utilities.\textsuperscript{25} The recent increase in telecommunications technologies in the United States has depended on the availability of public rights-of-way for facilities installation.\textsuperscript{26}

State law controls local governments’ property interests in public rights-of-way.\textsuperscript{27} Although there are many models of local government authority over rights-of-way,\textsuperscript{28} they fall into three general categories. First, local governments can own rights-of-way in fee,\textsuperscript{29} with an undivided ownership interest and full title.\textsuperscript{30} Second, local governments can hold easements in rights-of-way,\textsuperscript{31} giving them a possessory interest in the land\textsuperscript{32} while allowing abutting property owners to retain title.\textsuperscript{33} Finally, local governments can hold rights-of-way in trust for the public.\textsuperscript{34} Even if the local government does not hold title to the property, it has certain rights and responsibilities, such as improving and maintaining the property.\textsuperscript{35}

B. Local Governments Grant Franc hises to Private Companies for Use of Public Rights-of-Way

A franchise is a method developed by governments to grant use of public land to private companies for purposes benefiting the general public.\textsuperscript{36} Franchises are common law creations dating back to

\begin{itemize}
\item 25. See McQuillin, supra note 10, § 30.44.
\item 27. See Mayor of Baltimore v. United States, 147 F.2d 786, 788 (4th Cir. 1945).
\item 28. See McQuillin, supra note 10, § 30.32.
\item 29. See id.
\item 31. See McQuillin, supra note 10, § 30.32.
\item 32. See Nelson et al., supra note 30, at 688.
\item 33. See McQuillin, supra note 10, § 30.32.
\item 34. See Caporal v. United States, 577 F.2d 113, 116 (10th Cir. 1978); Mayor of Baltimore v. United States, 147 F.2d 786, 788 (4th Cir. 1945); N.J. Payphone Ass’n v. Town of W. New York, 130 F. Supp. 2d 631, 638 (D.N.J. 2001), aff’d, 299 F.3d 235 (3d Cir. 2002).
\item 35. See Mayor of Baltimore, 147 F.2d at 788.
\end{itemize}
Fair and Reasonable Compensation

fourteenth-century England,37 where the sovereign granted private parties use of public property under certain conditions.38 A franchise allowed a private party to provide a service that was both a public necessity and a natural monopoly.39 In the United States, local governments use franchises to give right-of-way access to companies providing services that require intensive infrastructure, such as electrical and telephone utilities.40

Telecommunications companies often need access to public rights-of-way for their equipment,41 and must obtain local government permission before installation.42 Telecommunications infrastructure is created by stringing wire to utility poles,43 burying wire in below-ground conduit,44 or, in the case of wireless facilities, by placing utility boxes in strategic locations.45 To obtain a franchise, telecommunications companies often agree to requirements such as plan notification,46 performance bonds,47 access to books and records,48 and fees.49

Local governments traditionally charge private companies franchise fees to access public rights-of-way.50 Most cities have ordinances establishing these fees.51 The ordinance can require compensation in the form of a percentage of the company’s gross revenue within the local

40. See id. at 175–76.
41. See Qwest Corp. v. City of Santa Fe, 224 F. Supp. 2d 1305, 1308 (D.N.M. 2002).
42. See id.; Qwest Corp. v. City of Portland, 200 F. Supp. 2d 1250, 1254 (D. Or. 2002); see also MCQUILLIN, supra note 10, § 34.10.
44. See TCG N.Y. v. City of White Plains, 305 F.3d 67, 71 (2d Cir. 2002); Qwest Communications v. City of Berkeley, 202 F. Supp. 2d 1085, 1086 (N.D. Cal. 2001).
45. See sources cited supra note 22.
47. See Bellsouth Telecomms. v. Town of Palm Beach, 252 F.3d 1169, 1181 (11th Cir. 2001).
48. See id.
51. See Town of Palm Beach, 252 F.3d at 1175.
franchising area,\textsuperscript{52} permit charges,\textsuperscript{53} in-kind services for the local government,\textsuperscript{54} or fees per linear foot of right-of-way used.\textsuperscript{55} The local government may use the fees to pay for right-of-way management, such as permitting, inspection, and decreased road life,\textsuperscript{56} or for general governmental purposes.\textsuperscript{57}

State law controls franchise authority,\textsuperscript{58} including the conditions local governments can place on telecommunications providers,\textsuperscript{59} and the compensation allowed for use of the public rights-of-way.\textsuperscript{60} States can directly grant communications providers access to local public rights-of-way.\textsuperscript{61} But, a local government cannot require compensation beyond that authorized by its state.\textsuperscript{62} For example, Colorado prohibits local governments from requiring compensation from telecommunications companies beyond “costs directly incurred.”\textsuperscript{63} Oregon, however, allows local governments to charge franchise fees of up to seven percent of a telecommunications provider’s gross revenue.\textsuperscript{64} Other states, such as Alabama, do not restrict franchise fees, and local governments can require fair and reasonable compensation based on a telecommunications provider’s gross revenue.\textsuperscript{65}

\textsuperscript{52} See TCG Detroit v. City of Dearborn, 206 F.3d 618, 621 (6th Cir. 2000); City of Orangeburg, 522 S.E.2d at 805.


\textsuperscript{55} See AT&T Communications v. City of Eugene, 35 P.3d 1029, 1033 (Or. Ct. App. 2001).

\textsuperscript{56} See id. at 383.

\textsuperscript{57} See City of Orangeburg, 522 S.E.2d at 805.

\textsuperscript{58} See MCQUILLIN, supra note 10, § 34.10.

\textsuperscript{59} See City of Auburn v. Qwest Corp., 260 F.3d at 1160, 1174 (9th Cir. 2001); City of Orangeburg, 522 S.E.2d at 807.


\textsuperscript{61} See, e.g., \textit{COLO. REV. STAT.} § 38-5.5-101 (2001).

\textsuperscript{62} See City of Denver v. Qwest Corp., 18 P.3d 748, 757 (Colo. 2001); People v. Kerr, 27 N.Y. 188, 213 (1863); City of Zanesville v. Zanesville Tel. & Tel. Co., 59 N.E. 781, 785 (Ohio 1901).

\textsuperscript{63} \textit{COLO. REV. STAT.} § 38-5.5-107(1)(b).

\textsuperscript{64} \textit{OR. REV. STAT.} § 221.515 (2002).

\textsuperscript{65} \textit{ALA. CODE} § 11-50B-3 (2002); \textit{see also} \textit{CONN. GEN. STAT.} § 7-130i (2002) (granting municipalities the authority to require compensation). For current information on public right-of-way laws in all fifty states, see the National Telecommunications and Information Administration website at http://www.ntia.doc.gov/ntiahome/staterow/rowtableexcel.htm (last visited July 28, 2003).
Fair and Reasonable Compensation

C. The Federal Government Cannot Infringe Local Government Property Rights or Commandeer Local Governments into Implementing Federal Regulatory Programs

The U.S. Constitution grants limited enumerated powers to the federal government. The powers that remain are reserved to the states or the people. States, therefore, have broader general powers than the federal government. The Commerce Clause is an example of a limited grant of federal power. Under the Commerce Clause, Congress can exercise broad authority over interstate commerce and related issues. But Congress’ authority is not unlimited. For example, Congress may not commandeer state or local governments into enacting legislation on behalf of a federal regulatory program. Instead, state and local governments must have a meaningful choice regarding whether to implement a federal directive. Further, the federal government may not compel local governments to subsidize private companies, and may not require a state or local government to transfer its property to benefit a private company.

Additionally, the federal government may not take local government property without compensation. The Fifth Amendment prohibits the federal government from taking private property “for public use without just compensation.” The Takings Clause prohibits the federal government from freely taking either privately owned property or the

68. See Lopez, 514 U.S. at 552.
72. See id. at 173–74.
73. See id. at 175.
74. See id. at 175.
75. See Qwest Corp. v. City of Santa Fe, 224 F. Supp. 2d 1305, 1329 (D.N.M. 2002).
77. U.S. CONST. amend. V.
public property of state and local governments. Further, the Takings Clause protects rights-of-way and even limited property rights, such as the right to exclude others, so that the federal government may not authorize a physical occupation of the property of another without paying just compensation.

The U.S. Supreme Court has held that if a private company occupies public rights-of-way, it must pay just compensation to the owner even if the occupation is authorized by the federal government. When a private company uses public rights-of-way, it asserts control over a certain portion of the space that could be used by others. Any amount of physical presence on private or public property could constitute a taking that deserves just compensation under the Takings Clause. Wires, cables, and transmission boxes installed by telecommunications companies have a permanent physical presence. Thus, their occupation of public property without just compensation could be a taking, even if it is authorized by the federal government.

Just compensation is generally calculated to be the fair market value of the subject property under the rationale that compensation should put the owner in as good a position as before the taking. The U.S. Supreme Court has measured this amount by the owner’s loss rather than the taker’s gain, and has applied this doctrine to different forms of infringed property interests, including easements and the right to

82. See Kaiser Aetna v. United States, 444 U.S. 164, 179–80 (1979) (holding that Congress had the right to grant public access through privately-owned waters, but that just compensation must be paid to the property owners).
83. See City of St. Louis, 148 U.S. at 100–01.
84. See id. at 99.
85. See Loretto, 458 U.S. at 430.
89. See Olson v. United States, 292 U.S. 246, 255 (1934).
90. See United States v. Causby, 328 U.S. 256, 261 (1946).
91. See id.
Fair and Reasonable Compensation

exclude others. For public property, the “substitute facilities” doctrine is often used to calculate just compensation if the fair market value is not readily ascertainable. Under this doctrine, compensation is the replacement cost of similar property.

The nature of the local government’s property interest under state law influences the type of compensation it may require when the federal government takes, or authorizes the taking of, a public right-of-way. If the local government owns the property in fee, it has the same rights as a private property owner, and can seek fair market value for taken property. In contrast, if the local government has an easement, compensation is typically based on the value of the easement. Finally, if the local government holds the right-of-way in trust for the public, it can require compensation based on the public’s right to control the property.

It has been argued that local governments have a duty to require compensation from private companies because not to do so would be inadequate management of a public resource. Many state constitutions support this principle by prohibiting gifting of public resources. The federal government follows this principle as well. For example, when the federal government grants access to its own rights-of-way, it requires fair rental value for the property.

In sum, state law grants local governments property interests in rights-of-way to manage on behalf of the public. Telecommunications companies pay fees and obtain franchises to access these rights-of-

93. See California v. United States, 395 F.2d 261, 266 (9th Cir. 1968).
94. See id.
95. See California, 395 F.2d at 264.
96. See McQuillin, supra note 10, § 30.32.
98. See Mayor of Baltimore v. United States, 147 F.2d 786, 789 (4th Cir. 1945).
99. See United States v. Causby, 328 U.S. 256, 265 (1946); see also Caporal v. United States, 577 F.2d 113, 117 (10th Cir. 1978).
102. See, e.g., ALASKA CONST. art. IX, § 6; CAL. CONST. art. XVI, § 6; N.Y. CONST. art. VII, § 8; WASH. CONST. art. VIII, § 7.
way. Although the state government can preempt the authority of local governments over rights-of-way, the federal government cannot because it does not hold the same sovereign position over local governments as does the state. Further, if the federal government tried to assert this power, it could result in an unconstitutional taking of local government property, or commandeering local governments into administering federal regulatory programs.

II. THE 1996 ACT PRESERVED LOCAL GOVERNMENT AUTHORITY TO REQUIRE COMPENSATION FOR USE OF PUBLIC RIGHTS-OF-WAY

With passage of the 1996 Act, Congress clarified federal regulation of telecommunications providers while simultaneously protecting the franchise authority of local governments. The text and legislative history of § 253(c) expressly preserve the franchising authority of local governments, and allow local governments to require “fair and reasonable compensation” for use of their rights-of-way. However, because Congress left “fair and reasonable compensation” undefined, litigation has ensued over the appropriate nature and extent of this compensation.

A. The Telecommunications Act of 1996

Congress passed the 1996 Act “to promote competition and reduce regulation in order to secure lower prices and higher quality service for American telecommunications consumers and encourage the rapid deployment of new telecommunications technologies.” The 1996 Act is a comprehensive federal regulatory program that clarifies federal authority over different communications providers.

The 1996 Act provides regulatory schemes for various technologies and defines the services within its purview. “Telecommunications” is

104. See MCQUILLIN, supra note 10, § 34.37.
106. See City of Auburn v. Qwest Corp., 260 F.3d 1160, 1175 (9th Cir. 2001).
108. 47 U.S.C. § 253(c); see infra Part II.B.
110. For example, Title II addresses telecommunications technologies and sets forth common carrier requirements, see 47 U.S.C. §§ 201–276, and Title VI incorporates the previous Cable
Fair and Reasonable Compensation

“the transmission, between or among points specified by the user, of information of the user’s choosing, without change in the form or content of the information as sent and received.” This definition encompasses both telephone service and the ability to access the internet over telephone lines, but not the provision of internet service itself.

“Telecommunications service” is defined as “the offering of telecommunications for a fee directly to the public . . . regardless of the facilities used.” A company that provides telecommunications services is a “telecommunications carrier.”

Despite introducing many changes to federal telecommunications regulation, the 1996 Act preserved local franchising authority. The 1996 Act distinguishes between the regulatory authority of local governments, which it largely preempts, and the proprietary rights of local governments, which it preserves. For example, the 1996 Act preempts local government franchising authority by compelling local governments to allow companies to provide telecommunications services. At the same time, the 1996 Act recognizes the authority of local governments to require franchises and to charge franchise fees of up to five percent of the gross revenue of cable television companies. Further, the 1996 Act preserves local government authority over companies that provide video programming over telephone lines (OVS). Although OVS operators are exempt from federal franchise requirements, they remain subject to local franchise authority. Thus, the 1996 Act preserves the franchise authority of local governments by ensuring their ability to assess and require fees.

Communications Act with regard to the regulation of cable television, see id. §§ 521–573.

111. 47 U.S.C. § 153(43).
112. See California v. FCC, 39 F.3d 919, 923–25 (9th Cir. 1994).
114. Id. § 153(44).
115. See City of Dallas v. FCC, 165 F.3d 341, 347 (5th Cir. 1999).
116. See City of Auburn v. Qwest Corp., 260 F.3d 1160, 1175 (9th Cir. 2001).
118. See City of Dallas, 165 F.3d at 347.
120. See id. § 541(a)(1).
121. See id. § 542(b).
122. See id. § 573(c)(2)(B).
123. See City of Dallas, 165 F.3d at 347.
B. Section 253 of the 1996 Act Preserves Local Government Authority to Require Compensation for Use of Public Rights-of-Way

Although § 253 of the 1996 Act preserves local authority over public rights-of-way, it also limits local governments’ regulatory authority over telecommunications providers.124 In relevant part, § 253 reads:

Removal of Barriers to Entry

(a) IN GENERAL – No State or local statute or regulation, or other State or local legal requirement, may prohibit or have the effect of prohibiting the ability of any entity to provide any interstate or intrastate telecommunications service.

. . . .

(c) STATE AND LOCAL GOVERNMENT AUTHORITY – Nothing in this section affects the authority of a State or local government to manage the public rights-of-way or to require fair and reasonable compensation from telecommunications providers, on a competitively neutral and nondiscriminatory basis, for use of public rights-of-way on a nondiscriminatory basis, if the compensation required is publicly disclosed by such government.125

Thus, section 253 contains both an express recognition of local government’s authority to require compensation for use of public rights-of-way, and a limitation of local government’s authority to regulate telecommunications providers.

Subsection 253(c) expressly allows local governments to obtain “fair and reasonable compensation” from telecommunications providers for their use of public rights-of-way. This provision was included in response to concerns of House126 and Senate127 members that local governments must retain their authority over public rights-of-way. In the House, the co-sponsors of the amendment that became § 253(c)128 did not want local governments to subsidize private telecommunications companies’ use of public rights-of-way.129 Instead, they preferred that

125. Id. (emphasis added). Section 253(b) addresses state regulatory authority with regard to such issues as universal service and public safety. Id.
129. See 141 CONG. REC. H8460 (daily ed. Aug. 4, 1995). Congressman Joe Barton of Texas presented statistics indicating that local governments spend approximately $100 billion per year
Fair and Reasonable Compensation

cities set the amount of compensation without interference from the federal government.\textsuperscript{130} Congressman Barton stated that the amendment “guarantees that cities and local governments have the right . . . to set the compensation level for use of [their] right-of-way.”\textsuperscript{131} The House adopted the amendment despite concern that allowing cities to charge a percentage of gross revenue for use of their rights-of-way would stifle the competition that the 1996 Act was designed to promote.\textsuperscript{132} The Senate adopted a similar amendment to the House Act.\textsuperscript{133}

After reconciling the amendments, Congress enacted § 253 without defining “fair and reasonable compensation.”\textsuperscript{134} In the absence of a clear definition, telecommunications providers have suggested that courts interpret § 253(c) in a similar manner as the section of the 1996 Act that regulates pole attachments.\textsuperscript{135} The Pole Attachment Act includes language that requires utilities to charge other utilities “just and reasonable” rates for use of their poles.\textsuperscript{136} The Federal Communications Commission (FCC) is authorized to set these rates,\textsuperscript{137} and has limited payment to the first utility’s direct costs for providing the attachment.\textsuperscript{138} Some telecommunications providers have taken the position that courts should interpret the “fair and reasonable” language of § 253(c) similarly, and have argued by analogy that cities should recover only direct costs associated with use of public rights-of-way.\textsuperscript{139}

However, courts have not reached a consensus on the basis on which to calculate fair and reasonable compensation.\textsuperscript{140} Courts have adopted either a “cost-recovery” approach or a “totality of the circumstances” approach.\textsuperscript{141} Under the cost-recovery approach, courts have limited local

\begin{enumerate}
\item maintaining public rights-of-way but only received $3 billion from the private companies that used them. See id.
\item Congressman Barton stated that in a “free market society, the companies should have to pay a fair and reasonable rate to use public property.” Id.
\item Id.
\item See id.
\item See id. 131. See 141 CONG. REC. S8431 (daily ed. June 15, 1995).
\item See id.
\item See id.
\item See City of Dearborn, 16 F. Supp. 2d 788.
\item See infra Part III.
\item See Qwest Corp. v. City of Santa Fe, 224 F. Supp. 2d 1305, 1327 (D.N.M. 2002).
\end{enumerate}

913
governments to recovering only the direct costs caused by telecommunications providers using public rights-of-way. In contrast, under the totality of the circumstances approach, courts look at various factors to determine whether the compensation required by the local government is fair and reasonable.

In sum, the 1996 Act simultaneously affirmed federal regulatory authority over telecommunications providers while preserving the power of local governments to require compensation for public rights-of-way. The text and legislative history of § 253(c) show that Congress intended to give telecommunications companies a permissive grant to access local rights-of-way subject to local government conditions, including the power to charge fair and reasonable compensation for use of the rights-of-way. However, Congress did not define “fair and reasonable” compensation in § 253(c), leaving it up to the courts to determine the extent of local government compensation setting authority.

III. CASES INTERPRETING THE 1996 ACT HAVE REACHED DIFFERENT CONCLUSIONS REGARDING LOCAL GOVERNMENT AUTHORITY

Since passage of the 1996 Act, telecommunications providers have brought suit under the Supremacy Clause and under § 253 when cities have required them to obtain franchises or satisfy other requirements in exchange for use of public rights-of-way. Decisions made under § 253 have been very fact-specific, and courts’ analyses have differed depending on state law and the technologies at issue. However, these decisions can be divided into certain categories. One
Fair and Reasonable Compensation

category includes courts that have considered the effect of state law on compensation and those that have not. Another includes courts that have interpreted § 253(c) to establish a “cost-recovery” approach, limiting cities to recovering only direct costs,\(^{150}\) and courts that have adopted a “totality of the circumstances” approach to determine if the compensation is fair and reasonable.\(^{151}\)

A. Some Courts Have Considered State Law When Deciding Whether a Franchise Fee Is Fair and Reasonable Under § 253(c)

State law can impact the analysis of a particular telecommunications franchise agreement in two ways. First, state law determines the nature of a local government’s interest in the public rights-of-way,\(^{152}\) thereby influencing a court’s determination of whether compensation is fair and reasonable.\(^{153}\) Second, a state statute can prescribe the type of compensation that local governments can require of telecommunications providers.\(^{154}\) Although some courts have grounded their § 253(c) analysis in consideration of state law, others have not.\(^{155}\)

Courts that have considered state law as part of their § 253(c) analysis have done so in different ways. Some courts have merely acknowledged state law, without relying heavily on it in their analyses. For example, in *Bell Atlantic-Maryland v. Prince George’s County*,\(^{156}\) the court noted without further discussion that the county was a home-rule county, authorized to exercise certain powers under the state constitution.\(^{157}\) Similarly, in *City of Rome v. Verizon Communications*,\(^{158}\) the court noted the possibility that the defendant telephone company had state

---

\(^{150}\) See id. at 1175; Grant County v. Qwest Corp., 169 F. Supp. 2d 1243, 1249 (D.N.M. 2001).

\(^{151}\) See City of Dearborn, 206 F.3d at 624–25.

\(^{152}\) See supra Part I.A.


\(^{154}\) See, e.g., COLO. REV. STAT. § 38-5.5-101 (2001); see also Bellsouth Telecommuns. v. City of Coral Springs, 42 F. Supp. 2d 1304, 1309 (S.D. Fla. 1999), aff’d in part and rev’d in part sub nom. Bellsouth Telecommuns. v. Town of Palm Beach, 252 F.3d 1169 (11th Cir. 2001); City of Denver v. Qwest Corp., 18 P.3d 748, 757 (Colo. 2001); People v. Kerr, 27 N.Y. 188, 213 (1863); City of Zanesville v. Zanesville Tel. & Tel. Co., 59 N.E. 781, 785 (Ohio 1901); Bellsouth Telecommuns. v. City of Orangeburg, 522 S.E.2d 804, 808 (S.C. 1999).


\(^{157}\) See id. at 807.

authorization to occupy the public rights-of-way, but then analyzed the franchise fee under § 253 without considering the issue.

In contrast, other courts have focused heavily on state law to determine the source of local government authority to require compensation under § 253(c). For example, in New Jersey Payphone Association v. Town of West New York, the town argued that it could require a fee from pay telephone providers from using the public rights-of-way based on its ownership of the property. The court rejected this argument, finding that under state law the town held an easement in the right-of-way as a public trustee, which only granted it police powers over the property. Thus, the court found the city could not exert the same control over the streets as it could if it owned them in fee. Similarly, in Qwest Corp. v. City of Santa Fe, the court considered the state constitution and the New Mexico Telecommunications Act to identify the city’s source of franchise authority before it conducted its § 253(c) analysis. After the court determined that the city owned the public rights-of-way and had the power to enter into franchises under state law, it concluded that the particular franchise provision in question was unreasonable under the circumstances.

Other courts have analyzed state law to determine the appropriate limits of local government authority without considering the nature of the property interest held by the local government. For example, in Bellsouth Telecommunications v. City of Coral Springs, the court determined that the authority of the state Public Service Commission preempted most local authority to franchise telecommunications companies. However, because both state and federal law allowed for a collection of a reasonable fee for right-of-way use, the city could require

159. See id. at 177.
161. See id. at 638.
162. See id.
163. See id.
165. See id. at 1319–20.
166. See id. at 1322.
167. See id. at 1331.
169. See id. at 1308.
Fair and Reasonable Compensation

up to a one percent franchise fee on the company’s gross revenue.170 Similarly, in City of Denver v. Qwest Corp.,171 the Colorado State Supreme Court held that, although Denver was a home-rule municipality under the Colorado Constitution,172 its franchise ordinance was preempted in its entirety by the state telecommunications act, which allowed telecommunications providers “to occupy public rights-of-way without additional authorization” from local governments.173

B. Courts Have Adopted Different Analytic Approaches to the Meaning of Fair and Reasonable Compensation Under § 253(c)

Courts are divided over the extent of local government authority under § 253(c), with some courts adopting a very narrow “cost-recovery” approach,174 and other courts allowing broader compensation under a “totality of the circumstances” approach.175 The cost-recovery approach limits the local government’s compensation to the costs it incurred from physical impacts to public rights-of-way.176 Courts adopting this approach usually preclude compensation based on gross revenue as a method for cities to recover costs.177 In contrast, the totality of the circumstances approach allows local governments to base their compensation on factors other than actual costs, including gross revenue.178

170. See id. at 1309.
171. 18 P.3d 748 (Colo. 2001).
172. See id. at 754.
173. Id. at 752.
176. See Grant County, 169 F. Supp. 2d at 1251. This analysis appears to derive from decisions that have interpreted franchise provisions as inconsistent with federal law if they attempt to regulate telecommunications providers themselves, rather than just their use of the public rights-of-way. For example, the Ninth Circuit struck down ordinances that required telecommunications companies to submit information regarding their financial qualifications as insufficiently related to right-of-way management. See City of Auburn v. Qwest Corp., 260 F.3d 1160, 1178 (9th Cir. 2001).
177. See Grant County, 169 F. Supp. 2d at 1249; Township of Haverford, 1999 WL 1240941, at *8.
178. See Qwest Corp. v. City of Santa Fe, 224 F. Supp. 2d 1305, 1328 (D.N.M. 2002).
1. Some Courts Have Adopted a Cost-Recovery Approach to Fair and Reasonable Compensation

Although later vacated on other grounds by the Fourth Circuit, Bell Atlantic-Maryland v. Prince George's County remains a significant development in the judicial interpretation of § 253(c). Prince George’s County, Maryland required a telephone company to obtain a franchise and pay a percentage of its gross revenue when it installed lines in the county’s right-of-way. The federal district court held that the county could require compensation, but the fees had to be directly related to maintenance and improvement of the public rights-of-way, otherwise the fees would be a “barrier to entry” prohibited by § 253(a). The court held that “fair and reasonable” compensation must be based on the degree of physical use by the company, rather than on a percentage of its gross revenue. Thus, the court adopted a “cost-recovery” approach to § 253(c). The decision turned on the court’s interpretation of the phrase “use of the public rights-of-way,” which it defined as “physically impact[ing] the public rights-of-way by installing, modifying, or removing telecommunications lines and facilities.”

Although the Prince George’s County decision was later vacated, it influenced other courts to hold that “cost-recovery” was the only permissible form of compensation under § 253(c), and to reject compensation based on a percentage of gross revenue. For example, in PECO Energy Co. v. Township of Haverford, a Pennsylvania federal district court struck down an ordinance requiring a telecommunications provider to pay four different fees to string fiber-optic cable on utility poles in the town’s streets. The court held that the town had not adequately demonstrated the relationship between the fees charged and

179. See Bell Atl.-Md. v. Prince George’s County, 212 F.3d 863 (4th Cir. 2000) (vacating and remanding with an admonition for the district court to consider the state law questions at issue before deciding the federal claims).
180. 49 F. Supp. 2d 805 (D. Md. 1999), vacated on other grounds by 212 F.3d 863 (4th Cir. 2000).
181. See id. at 807–08.
182. See id. at 817.
183. See id. at 814.
184. See id. at 818.
185. See id. at 819 (emphasis added).
187. See id. at *1–2.

918
Fair and Reasonable Compensation

the use of public rights-of-way. The court held that a fee based on revenue “by definition” could not be compensation for use of the public rights-of-way. A federal district court in New Mexico reached a similar conclusion in Grant County v. Qwest Corp. and invalidated a county “user fee” on telecommunications providers in the public rights-of-way because it was revenue-based. The court held that although the county was not precluded from charging a fee, the fee must “directly relate to the County’s expenses incurred in managing the actual physical use” of the right-of-way. Without further explanation, the court stated that a revenue-based fee did not meet this criterion.

2. Other Courts Have Adopted a Totality of the Circumstances Approach to Fair and Reasonable Compensation

In contrast, other courts have upheld local governments’ rights to obtain broader compensation for telecommunications providers’ use of public rights-of-way under § 253(c), adopting a “totality of the circumstances” model. Under this approach, courts consider a number of factors when deciding whether to uphold compensation as fair and reasonable. These factors include the local government’s authority under state law, what other providers are willing to pay for similar use of the public rights-of-way, and whether the telecommunications provider previously had agreed to pay similar compensation.

A year after the Grant County decision adopting the cost-recovery approach, another New Mexico district court held otherwise and adopted the totality of the circumstances approach to determine whether a franchise fee was fair and reasonable. In Qwest Corp. v. City of Santa

---

188. See id. at *7.
189. See id. at *8.
191. See id. at 1251.
192. See id.
193. See id.
195. See City of Dearborn, 16 F. Supp. 2d at 789–90.
196. See Qwest Corp. v. City of Santa Fe, 224 F. Supp. 2d 1305, 1318 (D.N.M. 2002).
197. See id. at 1329; City of Dearborn, 16 F. Supp. 2d at 790.
198. See City of Dearborn, 206 F.3d at 625; City of Portland, 200 F. Supp. 2d at 1259.
Fe, the court recognized the city’s right to charge franchise fees based on its status as a home-rule municipality. The court noted three advantages to the totality of the circumstances approach. First, compensation can be related to actual use of public rights-of-way without being cost-based. Second, the approach allows courts to give appropriate attention to variations in state and local law. Finally, it is consistent with U.S. Supreme Court jurisprudence concerning the limits of congressional authority.

The South Carolina State Supreme Court also adopted the totality of the circumstances approach, but considered different factors in Bellsouth Telecommunications, Inc. v. City of Orangeburg. The court held that the city had the authority, under both § 253 and state law, to require a telephone company to compensate it by paying a five percent gross-revenue fee to place poles and wires in the city’s streets. The court reasoned that because the company paid the fee in exchange for using the public rights-of-way, the compensation was fair and reasonable for “the franchise’s value as a business asset to the franchisee.” The court also upheld the fee because the other telecommunications franchisee in the city paid the same percentage of its gross revenue in exchange for right-of-way access.

In TCG Detroit v. City of Dearborn, the Sixth Circuit held that a franchise fee equal to four percent of the telecommunications provider’s gross revenue was fair and reasonable compensation because it was merited by the circumstances. The court considered the amount of right-of-way used, the fact that other providers had agreed to pay similar fees, and that the provider had previously agreed to pay the fee. Other

---

200. See id. at 1321. The court thus distinguished its holding from that of the Grant County court because New Mexico counties lack express authority to require such fees. See id. at 1321–22.
201. See id. at 1327. However, the court struck down the fee at issue as unreasonable when compared to the previous fee. Id. at 1329–30.
202. See id.
203. See id. at 1328.
204. 522 S.E.2d 804 (S.C. 1999).
205. See id. at 808.
206. See id.
207. See id. at 807.
208. 206 F.3d 618 (6th Cir. 2000).
209. See id. at 625.
210. See id.
Fair and Reasonable Compensation

courts have agreed with the City of Dearborn court and considered the last factor, that the telecommunications provider had been conducting business while paying the “allegedly prohibitive” fee, to be persuasive in rejecting arguments by telecommunications providers that the fees were prohibitive.211

As these cases illustrate, courts have used various approaches to analyze franchise fees under § 253(c). Not every court has looked to state law to determine the source of local government authority, and courts have differed over the proper interpretation of the words “use” and “compensation.” Some courts have concluded that compensation is limited to the costs incurred by a city for a telecommunications provider’s physical use of the public rights-of-way, such as repair of street cuts or inspection fees.212 Other courts have interpreted “use” broadly as any presence in the right-of-way, and concluded that a city can require compensation based on a percentage of the telecommunication provider’s gross revenue.213

IV. THE U.S. SUPREME COURT HAS UPHeld LOCAL GOVERNMENT AUTHORITY TO REQUIRE COMPENSATION FOR USE OF PUBLIC RIGHTS-OF-WAY

In the nineteenth and early twentieth centuries, the U.S. Supreme Court addressed challenges to local government control of public rights-of-way in cases involving telegraph companies.214 Although the companies had received a congressional grant to occupy public rights-of-way, the Court held that this grant was subordinate to the right of


214. See generally Postal Tel.-Cable Co. v. City of Richmond, 249 U.S. 252 (1919) (holding that a city’s imposition of a telephone pole fee was not a burden on interstate commerce); W. Union Tel. Co. v. Borough of New Hope, 187 U.S. 419 (1903) (holding that a city could charge telegraph companies reasonable license and permitting fees); City of St. Louis v. W. Union Tel. Co., 148 U.S. 92 (1893) (holding that a city could charge reasonable fees for use of public rights-of-way).
local governments to condition access to and receive compensation for use of their rights-of-way.\textsuperscript{215} More recent cases have affirmed this precedent with respect to modern technologies.\textsuperscript{216}

In 1866, Congress passed the Post Roads Act (Telegraph Act), granting telegraph companies “the right to construct, maintain and operate lines of telegraph through and over any portion of the public domain . . . over and along any of the military or post roads of the United States,”\textsuperscript{217} including routes in cities and towns.\textsuperscript{218} When a later Act of Congress declared “[a]ll public roads and highways . . . to be post routes,”\textsuperscript{219} it appeared that Congress had granted telegraph companies with free access to all public rights-of-way in the nation. Yet, cities continued to require telegraph companies to obtain franchises and pay compensation for access.\textsuperscript{220} Telegraph companies claimed that these fees obstructed their congressionally granted privilege\textsuperscript{221} and impermissibly interfered with interstate commerce.\textsuperscript{222} Ultimately, however, the U.S. Supreme Court interpreted the Telegraph Act as a “permissive” grant,\textsuperscript{223} subordinate to both public and private property rights.\textsuperscript{224}

In \textit{City of St. Louis v. Western Union Telegraph Co.},\textsuperscript{225} the city sued Western Union for failing to comply with an ordinance requiring telegraph companies to pay five dollars per year for each pole in the public rights-of-way.\textsuperscript{226} As a defense, Western Union claimed that the Telegraph Act of 1866 granted it free right-of-way access.\textsuperscript{227} The Court

\textsuperscript{215} See \textit{City of Richmond}, 249 U.S. at 260–61; \textit{Borough of New Hope}, 187 U.S. at 427; \textit{City of St. Louis}, 148 U.S. at 100.


\textsuperscript{217} Act of Congress of July 24, 1866, ch. 230, 14 Stat. 221, \textit{quoted in City of St. Louis}, 148 U.S. at 100.

\textsuperscript{218} See \textit{City of St. Louis}, 148 U.S. at 100 (quoting 18 Rev. Stat. 768, § 3964).

\textsuperscript{219} Act of March 1, 1884, 23 Stat. 3, \textit{quoted in City of St. Louis}, 148 U.S. at 100.

\textsuperscript{220} See \textit{Postal Tel.-Cable Co. v. City of Richmond}, 249 U.S. 252, 256 (1919); \textit{W. Union Tel. Co. v. Borough of New Hope}, 187 U.S. 419, 419 (1903); \textit{City of St. Louis}, 148 U.S. at 95.

\textsuperscript{221} See \textit{City of St. Louis}, 148 U.S. at 100.

\textsuperscript{222} See \textit{City of Richmond}, 249 U.S. at 256; \textit{Atl. & Pac. Tel. Co. v. City of Philadelphia}, 190 U.S. 160, 162 (1903).

\textsuperscript{223} See \textit{W. Union Tel. Co. v. City of Richmond}, 224 U.S. 160, 169 (1912); \textit{City of St. Louis}, 148 U.S. at 102.

\textsuperscript{224} See \textit{City of St. Louis}, 148 U.S. at 100–01.

\textsuperscript{225} 148 U.S. 92 (1893).

\textsuperscript{226} \textit{Id.} at 95.

\textsuperscript{227} See \textit{id.} at 100.
Fair and Reasonable Compensation

rejected Western Union’s argument, characterizing Western Union’s poles and wires as a physical occupation that permanently removed a portion of land from public use. The Court held that just as an act of Congress could not authorize a company to occupy private property without payment, neither could it do so for public property. Thus, local governments, like individuals, were entitled to “just compensation” for use of their property. The Court determined that compensation for the use of public rights-of-way, such as franchise fees, represented payment “in the nature of rental” to the public for use of its property. Although the fees were charged as rent, the local government could use the fees for revenue as if they were taxes, and the Court suggested in dicta that the fees could be based on the company’s gross revenue. Thus, the Court upheld the city’s authority to charge the fees, as long as the amount was not excessive.

In other cases litigated under the Telegraph Act, the Court upheld various models of compensation. In City of St. Louis, the Court affirmed basing compensation on a flat rental fee. In Western Union Telegraph Company v. Borough of New Hope, the Court upheld a city’s “license fee” for use of its streets, even though it exceeded the city’s cost for providing the service. Similarly, in Atlantic & Pacific Telegraph Company v. City of Philadelphia, the Court upheld fees for “the enforcement of local government supervision,” noting that the city did

228. See id. at 104–05.
229. See id. at 99.
230. See id. at 101.
231. Id. Although the Court did not directly refer to the Fifth Amendment, its use of the phrase “just compensation” in exchange for a property right indicates that this holding has a constitutional basis, which was later confirmed in Loretto v. Teleprompter Manhattan CATV Corp., 458 U.S. 419, 428–30 (1982).
233. See id. at 97.
234. See id. at 98.
235. See id.
236. See id. at 99, 105.
237. 187 U.S. 419 (1903).
238. See id. at 425.
239. 190 U.S. 160 (1903). Although the Second Circuit has suggested that the interpretation of the “dormant Commerce Clause” in this case has since been repudiated by the U.S. Supreme Court, the propositions cited here are still valid. TCG New York v. City of White Plains, 305 F.3d 67, 78 (2d Cir. 2002).
240. City of Philadelphia, 190 U.S. at 164.
not have to precisely estimate the cost to provide the services, but could set the charge at a reasonable amount to cover its expenses.241

In determining local governments’ authority over telegraph companies, the Court recognized the importance of state law.242 In Postal Telegraph-Cable Company v. City of Richmond,243 the Court held that the city’s “police power”244 under Virginia law245 was sufficient authority to impose an annual charge per utility pole on a telegraph company.246 Further, in City of St. Louis, the Court stated that, “under the constitution and laws of Missouri, the city of St. Louis has the full control of its streets, and in this respect represents the public in relation thereto.”247 Thus, the Court acknowledged that cities could charge rent based on their interest in holding the land in trust for the public.248

The U.S. Supreme Court and other courts have applied the principles of City of St. Louis in cases involving more modern technologies.249 In Loretto v. Teleprompter Manhattan CATV Corp.,250 the Court examined a New York state law authorizing cable television companies to install facilities on privately owned buildings, in exchange for one dollar.251 Characterizing the presence of the cable, although minimal, as a “permanent physical occupation,” the Court held that the law effected a taking under the Fifth Amendment and that the state could not prevent the payment of just compensation.252 Applying the “physical occupation” principle it recognized a century before in City of St.

241. *Id.* at 164–65.
242. Although one commentator has suggested that the Court misunderstood the nature of the cities’ interest in the public rights-of-way in these early telegraph cases and that this undermines their validity, see Gardner F. Gillespie, *Rights-of-Way Redux: Municipal Fees on Telecommunications Companies and Cable Operators*, 107 DICK. L. REV. 209, 219 (2002), as discussed in this Part, the language of the decisions indicates otherwise.
244. *Id.* at 259.
245. *Id.* at 257.
246. *Id.* at 256.
248. *Id.* at 99–100.
251. See *id.* at 423–24.
252. See *id.* at 421.
Fair and Reasonable Compensation

the Court confirmed that the action was a taking even though the law merely authorized, rather than accomplished, the occupation. Although the cable box, like the telegraph wires at issue in *City of St. Louis*, occupied “relatively insubstantial amounts of space,” the property owner, like the city in *City of St. Louis*, could require reasonable compensation.

Lower courts have also applied the principle that a private company cannot occupy public property without compensating the local government in the telecommunications context. In *TCG Detroit v. City of Dearborn*, a Michigan district court relied on *City of St. Louis* to hold that there was “nothing inappropriate” about the city charging rent for the telecommunications provider’s facilities on city property.

Similarly, in *City of Gary v. Indiana Bell Telephone Co.*, the Indiana State Supreme Court cited *City of St. Louis* in affirming a city’s authority to require compensation for private use of public property “irrespective of the label placed on the compensation.”

In sum, the U.S. Supreme Court’s interpretation of the Telegraph Act established three important principles concerning local governments’ authority over public rights-of-way. First, local governments have the authority to require compensation from private companies using public right-of-way, even if the companies are acting pursuant to a congressional grant. Second, the compensation can take many forms, including rental and license fees, and still be reasonable. Third, in determining the appropriate amount of compensation, it is important to consider state law. These cases have had a continuing vitality. The principle that a small amount of physical intrusion is a taking that requires just compensation was affirmed in *Loretto*, and the idea that

---

253. See id. at 429.
254. See id. at 429 n.6.
255. See id. at 430.
256. See id. at 421.
258. Id. at 789.
262. Postal Tel.-Cable Co. v. City of Richmond, 249 U.S. 252, 257 (1919); *City of St. Louis*, 148 U.S. at 100.
local governments can charge rental for their property has been applied by courts in the telecommunications context.264

V. LOCAL GOVERNMENTS ARE ENTITLED TO REQUIRE FAIR AND REASONABLE COMPENSATION FOR USE OF PUBLIC RIGHTS-OF-WAY

Local governments, in accordance with state law and established federal doctrine, are entitled to require compensation from telecommunications providers for use of public rights-of-way beyond the costs incurred from the physical impact and direct regulation. When evaluating a franchise fee under § 253(c), courts should examine state law for the source of local government franchising authority, and then use a totality of the circumstances approach to determine if the compensation is fair and reasonable. The text and legislative history of § 253(c)265 and analogous U.S. Supreme Court precedent support this approach.266 Courts that narrowly interpret § 253(c) to allow only direct cost-recovery potentially permit both a taking of local government property without just compensation,267 and a commandeering of local governments to implement a federal regulatory program.268

A. Approaches to the Issue of Fair and Reasonable Compensation Can Be Generally Grouped Into Four Categories

When analyzing a particular ordinance under § 253(c), courts have placed varying levels of importance on state law.269 Courts have also adopted one of two approaches: a cost-recovery approach, which limits the local government’s reimbursement to its direct costs,270 or a totality of the circumstances approach, which considers a variety of factors and allows more expansive compensation.271 These inconsistent approaches have resulted in four analytic frameworks courts use to determine if

265. See supra Part II.B.
266. See supra Part IV.
267. See supra Part I.A.
268. See supra notes 71–74 and accompanying text.
269. See supra notes 71–74 and accompanying text.
270. See supra notes 176–93 and accompanying text.
271. See supra notes 194–211 and accompanying text.
Fair and Reasonable Compensation

compensation is fair and reasonable under § 253. The following matrix represents the possible approaches and provides an example of each:

<table>
<thead>
<tr>
<th></th>
<th>Totality of the Circumstances</th>
<th>Cost Recovery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Considers State Law</td>
<td><em>City of Santa Fe</em>²⁷²</td>
<td><em>Prince George’s County</em>²⁷³</td>
</tr>
<tr>
<td>Does Not Consider</td>
<td><em>City of Dearborn</em>²⁷⁴</td>
<td><em>Township of Haverford</em>²⁷⁵</td>
</tr>
<tr>
<td>State Law</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

These inconsistent approaches create confusion for local governments and telecommunications providers. The text of § 253(c), its legislative history, and analogous U.S. Supreme Court precedent dictate an approach that first analyzes the relevant state law and then considers the totality of circumstances surrounding the franchise agreement to determine if the compensation is fair and reasonable.

B. Courts Should Examine State Law When Interpreting “Fair and Reasonable Compensation” Under § 253(c)

To analyze “fair and reasonable compensation” under § 253(c), courts should first examine state law to determine the source of the city’s franchise authority. The text of § 253(c) states that “[n]othing in this section affects the authority . . . of local government[s].”²⁷⁶ Thus, § 253(c) expressly preserves local government authority to require compensation. Such authority invariably derives from state law, because state law governs the nature of local governments’ property interest in

²⁷² 224 F. Supp. 2d 1305 (D.N.M. 2002). The *Santa Fe* court held that the local government had authority to require franchise fees under the state constitution, *id.* at 1318, and then examined the totality of the circumstances to see if the particular fee was reasonable, *id.* at 1329.

²⁷³ 49 F. Supp. 2d 805 (D. Md. 1999), *vacated on other grounds* by 212 F.3d 863 (4th Cir. 2000). The *Prince George’s County* court recognized the county’s authority to require fees from its status as a “home-rule” county, *id.* at 807, but then adopted a cost-recovery approach to invalidate the fee under federal law, *id.* at 817–18.

²⁷⁴ 206 F.3d 618 (6th Cir. 2000). The *City of Dearborn* court only examined the totality of the circumstances surrounding the franchise fee and did not consider the city’s authority to impose such a fee. *Id.* at 624–25.

²⁷⁵ No. 99-4766, 1999 WL 1240941 (E.D. Pa. Dec. 20, 1999). The *Township of Haverford* court adopted a cost-recovery approach but did not conduct the initial step of finding the authority for the town to impose the fee. *Id.* at *6–7.

their rights-of-way and their ability to franchise. An initial inquiry into state law is necessary to provide courts with a basis for interpreting whether the compensation is fair and reasonable.

State law governs local government authority over public rights-of-way. If a local government asserts a right to receive compensation based on its interest in the rights-of-way, a court should know the nature of the interest to determine whether the compensation is appropriate. In this context, it is important whether the local government owns the right-of-way in fee, has an easement, or holds the property in trust for the public. If a court attempts to analyze an ordinance without understanding the character of the local government’s property interest under state law, it will lack the necessary foundation to evaluate the reasonableness of the compensation.

Furthermore, the state may have a telecommunications statute that limits or proscribes the type or amount of compensation that the local government can require. For example, in City of Coral Springs, the court held that under state law, only franchise fees equal to or less than one percent of the telecommunication provider’s gross revenue were valid, an amount the court held was also fair and reasonable under federal law. Likewise, in City of Denver the court held that a local franchise was completely preempted by state law, so a § 253(c) analysis was unnecessary to determine whether the compensation was fair and reasonable. If the City of Denver and City of Coral Springs courts had not evaluated the fees in the context of state law, they might have held that the fee was “fair and reasonable” even if it violated state law.

Furthermore, principles of federalism require courts to examine local authority under state law for § 253(c) analysis. Cases that interpret § 253(c) without proper analysis of the local government’s authority

277. See supra Part I.A.
278. See supra Part I.B.
279. See supra Part I.A.
280. See N.J. Payphone Ass’n v. Town of W. New York, 130 F. Supp. 2d 631, 638 (D.N.J. 2001), aff’d, 299 F.3d 235 (3d Cir. 2002); see also supra notes 95–100 and accompanying text.
281. See N.J. Payphone, 130 F. Supp. 2d at 638; see also MCQUILLIN, supra note 10, § 30.37; supra notes 95–100 and accompanying text.
282. See supra notes 58–65 and accompanying text.
284. 18 P.3d 748, 758 (Colo. 2001).
285. See Qwest Corp. v. City of Santa Fe, 224 F. Supp. 2d 1305, 1317 (D.N.M. 2002).
Fair and Reasonable Compensation

under state law 286 allow the federal government to fill what is more properly the role of the states. The federal government has limited, enumerated powers, 287 and all other powers are reserved to the states or the people. 288 Local government authority over franchises and public rights-of-way is governed by state law, 289 and the federal government can neither attribute nor ignore local powers that are removed or conferred by the states. To do so could lead to unconstitutional results, because the court would be granting power to the federal government that properly resides with the state. 290

Thus, under § 253(c), courts should first identify the source and nature of local government franchise authority, and then determine if the local government is properly exercising that authority. The source of franchise authority can be found in the property rights of the local government or through express delegation by the state. Even if this authority is granted, its exercise could still be removed by a preemptive state telecommunications statute 291 or by delegation of franchise powers to a state utility commission. 292 Only if the franchise is consistent with state law can the court consider whether the compensation is fair and reasonable under federal law.

C. Courts Should Use the Totality of the Circumstances Approach to Determine Whether the Compensation Required by the Local Government Is Fair and Reasonable

Once a court has determined that a local government’s franchise fee complies with state law, the court should consider the totality of circumstances to decide if the compensation is fair and reasonable under § 253(c). The text and history of § 253(c), 293 as well as analogous U.S. Supreme Court precedent, 294 support this approach. Failure to adopt this

288. U.S. CONST. amend. X.
289. See supra Part I.A–B.
290. See supra Part I.C.
292. See, e.g., Bellsouth Telecommns. v. Town of Palm Beach, 252 F.3d 1169, 1177 (11th Cir. 2001).
293. See supra Part II.B.
294. See supra Part IV.
approach could potentially lead to an unconstitutional taking of local
government property or commandeering of local governments.\textsuperscript{295}

1. \textbf{The Text and Legislative History of § 253(c) Support a Broad}
   Interpretation of Fair and Reasonable Compensation

Courts that have limited local government compensation to costs for
actual use of the public rights-of-way, such as repair of street cuts, have
interpreted § 253(c) too narrowly. The text of § 253(c) reads in relevant
part, “[n]othing in this section affects the authority of a State or local
government to manage the public rights-of-way or to require fair and
reasonable compensation from telecommunications providers . . . for use
of public rights-of-way . . . .”\textsuperscript{296} The wording of this section strongly
suggests an affirmation, not a limitation, of local government authority.
The section begins with a broadly worded admonition, and is followed
by a general authorization, signaling that Congress did not intend to limit
local government to a narrow cost-recovery interpretation, but instead
intended to preserve traditional local government control.

In addition, Congress’ use of the word “compensation,” rather than
“costs” is instructive. The word “compensation” indicates a more
expansive interpretation than the cost-recovery method adopted by many
courts.\textsuperscript{297} In other areas of the 1996 Act, notably the Pole Attachment
Act, Congress expressly limited fees to incurred costs.\textsuperscript{298} It did not do so
for § 253(c). This suggests Congress intentionally chose the term
“compensation” rather than “costs.”\textsuperscript{299} Furthermore, Congress
authorized the FCC to set rates for pole attachments, but did not
authorize the FCC to set rates for use of public rights-of-way, indicating
its intention to leave this authority to local governments.\textsuperscript{300} If Congress
had intended to restrict local governments to recover only physical costs,
it could have included language to that effect.

Further, the legislative history of the phrase “fair and reasonable
compensation” in § 253(c) suggests that Congress intended local
governments to set their own levels of compensation for public rights-of-

\begin{itemize}
\item \textsuperscript{295} See infra Part V.C.3.
\item \textsuperscript{296} 47 U.S.C. § 253(c) (2000) (emphasis added).
\item \textsuperscript{297} See supra Part II.B.
\item \textsuperscript{298} 47 U.S.C. § 224(d)(1).
\item \textsuperscript{299} See TCG Detroit v. City of Dearborn, 206 F.3d 618, 624–25 (6th Cir. 2000); Qwest Corp. v.
City of Santa Fe, 224 F. Supp. 2d 1305, 1327 (D.N.M. 2002).
\item \textsuperscript{300} See 47 U.S.C. §§ 224(b)(1), 253.
\end{itemize}
Fair and Reasonable Compensation

way. The House floor discussion demonstrates an unwillingness by Congress to have local governments subsidize telecommunications providers, and indicates an awareness that § 253(c) would allow local governments to set their own rates for compensation by telecommunications providers. For example, Congressman Barton specifically stated that cities should be allowed to set the compensation for use of their public rights-of-way. Taken together, it is apparent that Congress intended to allow local governments to set the amount of fair and reasonable compensation for use of their rights-of-way, without limiting recovery to actual costs.

2. Analogous U.S. Supreme Court Precedent Supports the Totality of the Circumstances Approach

In cases interpreting the Telegraph Act, the U.S. Supreme Court firmly sustained the authority of local governments to require fees from private companies using public rights-of-way. Marked similarities between the telegraph and telecommunications context make this precedent applicable to telecommunications companies. Both telegraph and telecommunications companies require access to public rights-of-way to deploy their facilities. Cities offer this access in exchange for payment. At the turn of the nineteenth century, the Court approved this arrangement in City of St. Louis and did not limit the form or method of payment required of telegraph companies as long as it was reasonable. The Court upheld a variety of fees as reasonable, including those charged for rental, licensing, and general governmental supervision. In no case did the Court require that the fees be tied to the physical or actual costs of the telegraph providers’ use of the public

301. See supra notes 126–33 and accompanying text.
303. Id.
304. Id.
305. See supra Part IV.
307. See McQuillin, supra note 10, § 34.37.
rights-of-way. This precedent was expressly followed in City of Dearborn and City of Gary, which applied the principles set forth in the early telegraph cases to the telecommunications context.

The 1996 Act, like the Telegraph Act, offers communications companies a permissive grant to access local public rights-of-way. Congress designed both Acts to promote new technologies that were important to the nation’s economic development, but in both cases the importance of local control remained paramount. In City of St. Louis, the Court judicially preserved the authority of local governments to require compensation under the Telegraph Act. In § 253(c), Congress issued a permissive grant expressly and statutorily subject to local control.

In sum, U.S. Supreme Court cases interpreting the Telegraph Act recognized that local governments have property interests in public rights-of-way and can require compensation when this property is used by private companies. The Court reaffirmed the continuing vitality of this doctrine in Loretto, and lower courts have applied these principles in the telecommunications context. Because of the similarities between the Telegraph Act and the 1996 Act, this precedent provides valuable guidance to courts interpreting the 1996 Act.

3. Unconstitutional Results Can Occur if § 253(c) Is Interpreted Too Narrowly

An interpretation of § 253(c) that limits local governments to recovering only actual costs, rather than compensation for the value of its property, could create unconstitutional results. Courts interpreting federal legislation must construe statutory language to avoid

310. City of Philadelphia, 190 U.S. at 162–63; Borough of New Hope, 187 U.S. at 426; City of St. Louis 148 U.S. at 99, 104.
311. See supra notes 257–59 and accompanying text.
313. 47 U.S.C. § 253(c).
314. See supra Part IV.
317. See Qwest Corp. v. City of Santa Fe, 224 F. Supp. 2d 1305, 1328–29 (D.N.M. 2002); see also Worstell, supra note 86, at 467–69.
Fair and Reasonable Compensation

constitutional conflicts whenever possible. Under the Fifth Amendment, the federal government cannot take local government property without providing just compensation. Further, under the Commerce Clause, Congress cannot compel local governments to subsidize private companies as part of a federal regulatory scheme. To avoid these constitutional conflicts, courts should interpret § 253(c) broadly to recognize the property rights of local governments under state law and to provide adequate compensation when telecommunications companies use public rights-of-way.

In City of St. Louis, the U.S. Supreme Court held that local governments can require just compensation even if the federal government allows private access to public rights-of-way. Federal seizure of local government property is prohibited in the same way that taking private property is barred. If a local government has a property right for which it can obtain value and the federal government prevents it from doing so, the federal government could be authorizing an unconstitutional taking. Courts assume that Congress understands the constitutional implications of its statutory language. Therefore, courts should interpret § 253(c) to avoid an unconstitutional result.

Further, paying for property use is a cost of doing business for telecommunications providers. The cables, wires, and transmission boxes necessary to provide telecommunications services have a physical presence. If these providers did not have access to public rights-of-way, they would have to seek permission from private landowners and locate equipment on private property.

A narrow cost-recovery interpretation of § 253(c) potentially commandeers local governments into implementing a federal regulatory program and enacting federal policy, in violation of the Commerce

319. See supra Part I.A.
323. See California v. United States, 395 F.2d 261, 264 (9th Cir. 1968).
324. See U.S. W. Corp. v. FCC, 182 F.3d 1224, 1231 (10th Cir. 1999).
325. See BellSouth Telecomm v. City of Orangeburg, 522 S.E.2d 804, 806 (S.C. 1999).
Clause, by forcing them to support telecommunications companies at public expense. In addition, a narrow cost-recovery approach could lead to commandeering.\(^{328}\) When cities are limited to a cost-recovery model for the use of their rights-of-way,\(^{329}\) private companies receive the benefit of property held in trust for the public without paying for its full value. In effect, the company is granted a free public resource.\(^{330}\) This results in public subsidization of private companies.\(^{331}\) Congress may not compel local governments to subsidize private companies.\(^{332}\) Such commandeering is prohibited by the U.S. Constitution.\(^{333}\) To avoid creating unconstitutional results, courts should broadly interpret § 253 to allow the public to be appropriately compensated for the use of its property.

D. Local Governments May Use a Variety of Methods to Determine What Is “Fair and Reasonable Compensation”

If the state has not set limits on compensation, local governments should be able to determine the amount of compensation for use of their rights-of-way. Compensation can be based on several factors, including the nature of the local government’s interest in the rights-of-way,\(^{334}\) and what other companies pay.\(^{335}\) Further, the franchise fees can appropriately be based on a percentage of the telecommunications provider’s gross revenue.\(^{336}\)

The nature of the local government interest in the public rights-of-way can be a factor in calculating appropriate compensation. Certain property rights have value, which can be used to establish a fair and

\(^{328}\) See Qwest Corp. v. City of Santa Fe, 224 F. Supp. 2d 1305, 1328–29 (D.N.M. 2002).

\(^{329}\) See Grant County v. Qwest Corp., 169 F. Supp. 2d 1243, 1251 (D.N.M. 2001).

\(^{330}\) See City of Santa Fe, 224 F. Supp. 2d at 1329.

\(^{331}\) See id.

\(^{332}\) See New York v. United States, 505 U.S. 144, 175 (1992). In addition, many state constitutions have provisions specifically prohibiting such a gift of public resources to private companies, so such an interpretation potentially creates unconstitutional results at the state level. See, e.g., ALASKA CONST. art. IX, § 6; CAL CONST. art. XVI, § 6; N.Y. CONST. art. VII, § 8; WASH. CONST. art. VIII, § 7.


\(^{335}\) See id.

Fair and Reasonable Compensation

reasonable amount of compensation. If the local government has complete ownership of the streets, in the same manner as a private owner, the appropriate compensation is fair rental value, the method used by the federal government to determine compensation for use of its own property. If the local government holds the right-of-way as an easement, this interest has value, which has been measured in different ways depending upon the circumstances. If the public right-of-way is held in trust for the public, the local government arguably has a duty to receive compensation for the property on behalf of the public, such as fees for each lineal foot of property used. The U.S. Supreme Court has indicated that reasonable rental may be appropriate in this case as well.

If the portion of the public rights-of-way used by the telecommunications provider could bring funds to the local government, it has value for which the local government should be compensated. If the local government gave away this valuable property for free, it would not only be neglecting the best interests of its citizens, but it would also violate many state constitutions. The appropriate amount of compensation is what the market can bear, up to state-imposed limits. The underlying purpose of local government authority over public rights-of-way is to control the streets and alleys for the benefit of the public, which includes receiving compensation when the land is used by a private company.

If local governments can receive compensation from other companies for use of public rights-of-way, they should be able to require

337. See supra notes 95–100 and accompanying text.
341. See id. at 261–62.
343. See City of St. Louis, 148 U.S. at 100.
344. See California v. United States, 395 F.2d 261, 264 (9th Cir. 1968).
346. See, e.g., ALASKA CONST. art. IX, § 6; CAL CONST. art. XVI, § 6; N.Y. CONST. art. VII, § 8; WASH. CONST. art. VIII, § 7.
347. See Nat’l League of Cities, supra note 342, at 3.
348. See MCQUILLIN, supra note 10, § 30.40.
compensation from telecommunications companies as well. The government’s loss of the space appropriated by a telecommunications provider could be measured in terms of what other providers are willing to pay.\footnote{350} Utilities other than telecommunications providers, such as cable television\footnote{351} and electrical providers,\footnote{352} also use public rights-of-way. Local governments generally receive payment from these utilities in the form of a percentage of gross revenue.\footnote{353}

A telecommunications provider’s gross revenue is an appropriate basis from which to calculate compensation, even if it is not equivalent to the local government’s costs.\footnote{354} U.S. Supreme Court decisions suggest that there is nothing inherently unreasonable about basing franchise fees on the gross revenue of a communications company within a particular jurisdiction.\footnote{355} Many lower courts have expressly allowed such a calculation.\footnote{356} Even courts that have limited fees for other reasons have not disputed that fees can be revenue based. For example, a Florida district court indicated that a one percent fee based on gross revenue was reasonable if allowed by state law.\footnote{357}

Basing fees on a percentage of gross revenue is a fair approximation of the value of a company’s use of public rights-of-way. Because it is an indication of the amount of a company’s business in a particular jurisdiction, gross revenue reflects the value of their use of the right-of-way.\footnote{358} Courts have agreed that for fees to be valid, their purpose must

\begin{footnotes}
\footnotetext{350}{See TCG Detroit v. City of Dearborn, 206 F.3d 618, 625 (6th Cir. 2000).}
\footnotetext{351}{See City of Dallas v. FCC, 118 F.3d 393, 393 (5th Cir. 1997).}
\footnotetext{352}{See Alachua County v. Florida, 737 So.2d 1065, 1066 (Fla. 1999).}
\footnotetext{353}{See City of Dallas, 118 F.3d at 393; Alachua County, 737 So.2d at 1066.}
\footnotetext{354}{See Qwest Corp. v. City of Santa Fe, 224 F. Supp. 2d 1305, 1328 (D.N.M. 2002); TCG Detroit v. City of Dearborn, 16 F. Supp. 2d 785, 789 (E.D. Mich. 1998), aff’d, 206 F.3d 618 (6th Cir. 2000).}
\footnotetext{355}{For example, City of St. Louis indicates that although Western Union was charged based on the number of its poles, other telegraph companies were apparently “taxed on their gross income for city purposes.” City of St. Louis v. W. Union Tel. Co., 148 U.S. 92, 98 (1893). The Court did not suggest that such a calculation was impermissible.}
\footnotetext{356}{See City of Dearborn, 206 F.3d at 625 (upholding a franchise fee based on 4% of the telecommunication provider’s gross revenue); accord Bellsouth Telecommns. v. City of Orangeburg, 522 S.E.2d 804, 808 (S.C. 1999) (upholding a franchise fee based on 5% of the telecommunication provider’s gross revenue).}
\footnotetext{357}{Bellsouth Telecommns. v. City of Coral Springs, 42 F. Supp. 2d 1304, 1309 (S.D. Fla. 1999), aff’d in part and rev’d in part sub nom. Bellsouth Telecommns. v. Town of Palm Beach, 252 F.3d 1169 (11th Cir. 2001).}
\footnotetext{358}{See City of Santa Fe, 224 F. Supp. 2d at 1326–27; City of Gary v. Ind. Bell Tel. Co., 732 N.E.2d 149, 156 (Ind. 2000).}
\end{footnotes}
Fair and Reasonable Compensation

be expressly stated and the fees must be related to a company’s presence
in the public rights-of-way. Fees based on a percentage of gross
revenue fulfill both these conditions. Such fees in franchises predating
the 1996 Act were presumably considered “fair and reasonable.” Nothing has changed since passage of the 1996 Act that would render
these fees unreasonable.

Furthermore, the 1996 Act expressly allows the use of a percentage of
gross revenue for payment of certain franchise fees. Cities entering into
franchise agreements with cable and OVS companies are allowed to
charge revenue-based fees: local governments may charge fees of up to
five percent of the gross revenue of cable television providers, and
OVS operators are also charged based on their gross revenues. No
technological or statutory distinction compels a different result for
telecommunications franchises.

Local governments can choose from several methods of valuation
when setting franchise fees, including the value of the property as
appraised, prevailing rental value, and what other providers are
willing to pay. The local government may provide other services to
the telecommunications provider, such as supervision or licensing, for
which it can be further compensated. The method used to determine
the amount of compensation should be left to the discretion of the local
government and the free market.


360. See, e.g., City of Santa Fe, 224 F. Supp. 2d at 1308 (dating back to 1975); City of
Orangeburg, 522 S.E.2d at 805 (dating back to 1993).


362. Id. § 573(c)(2)(B).

363. See City of Santa Fe, 224 F. Supp. 2d at 1309.

F.3d 618 (6th Cir. 2000); City of Orangeburg, 522 S.E.2d at 808.

365. See City of Santa Fe, 224 F. Supp. 2d at 1329; City of Dearborn, 16 F. Supp. 2d at 790.


367. See generally Nat’l Ass’n of Telecomm. Officers and Advisors, Local Government
Principles Relating to Rights-Of-Way Management and Compensation & Ownership of
(Aug. 20, 1998) (encouraging local governments to require fair market value for use of public
rights-of-way to fulfill their duty to the public to responsibly manage public land).
VI. CONCLUSION

Section 253(c) of the 1996 Act preserves the authority of local governments to require fair and reasonable compensation for use of public rights-of-way.368 State law is the source of both local government property rights and franchising authority.369 Therefore, when evaluating whether a local government franchise fee requires fair and reasonable compensation from a telecommunications provider, courts should first consider state law.370 Only after the court has identified the nature of the local government authority can it consider if the fee is fair and reasonable under federal law.

Some courts have interpreted § 253(c) too narrowly, limiting local governments to recover only direct costs for private companies’ use of public rights-of-way.371 Courts taking this approach potentially violate the U.S. Constitution by allowing private companies to appropriate a property interest without compensation, and by commandeering local governments into federal service.372 Other courts have examined the totality of the circumstances surrounding a franchise and have allowed local governments to require greater compensation than merely direct costs.373 The text and legislative history of § 253(c) and analogous U.S. Supreme Court precedent support this approach.374 Thus, when considering the validity of a particular franchise fee, courts should first determine the source of local government authority under state law, and then apply a totality of the circumstances approach to determine if the fee is fair and reasonable.

369. See supra Part I.A–B.
370. See Qwest Corp. v. City of Santa Fe, 224 F. Supp. 2d 1305, 1318 (D.N.M. 2000).
372. See supra Part V.C.3.
374. See supra Part V.C.1–2.
ABORIGINAL TITLE OR THE PARAMOUNTCY DOCTRINE? JOHNSON V. MCINTOSH FLOUNDS IN FEDERAL WATERS OFF ALASKA IN NATIVE VILLAGE OF EYAK V. TRAWLER DIANE MARIE, INC.

Andrew P. Richards

Abstract: In Johnson v. McIntosh and its progeny, the United States Supreme Court established the principle that aboriginal title allows Indian tribes to exclusively use and occupy their territories after they come under United States sovereignty. In Native Village of Eyak v. Trawler Diane Marie, Inc., five Alaska Native villages asserted aboriginal title to areas of the seabed and ocean off Alaska. The villages argued that federal fisheries regulations violate their aboriginal title by allowing non-Natives to fish within those areas, while excluding most of the villagers. The United States Court of Appeals for the Ninth Circuit rejected the villages’ claim, holding that the paramountcy doctrine had extinguished the villages’ aboriginal title. Under the paramountcy doctrine, the federal government must control exploitation of the seabed and ocean to fulfill its duty to defend the nation and to regulate international commerce. The Eyak court held that aboriginal title would conflict with federal supremacy over the seabed and ocean off Alaska. This Comment argues that the Court of Appeals for the Ninth Circuit en banc or the U.S. Supreme Court should hold that the paramountcy doctrine did not extinguish aboriginal title to the seabed and waters off Alaska because aboriginal title does not interfere with the federal government’s ability to protect the nation or to regulate international trade.

For seven thousand years, people from five Alaska Native villages (the villages) fished and hunted along the southern coast of what is today the State of Alaska. They continued to use their traditional areas until 1995, when the Secretary of the Department of Commerce limited fishing for halibut and sablefish off Alaska. Previously, both Alaska Natives and non-Natives pursued halibut and sablefish from Southeast Alaska to the Bering Sea. The ease of entry into these two fisheries spawned a modern-day maritime gold rush in which too many people risked too much money and life for steadily diminishing profits. Fearing that over-fishing would destroy the halibut and sablefish stocks, the Secretary curtailed fishing seasons by the late 1980s from months down
to days.\textsuperscript{4} In 1995, the Secretary responded to concerns about harvesters’ dwindling profit margins, and the inherent dangers of fisheries built on wild two- and three-day openings, by limiting the number of people allowed to participate in the halibut and sablefish fisheries to those who qualified for Individual Fishing Quota shares (IFQs).\textsuperscript{5}

IFQs enable their holder to catch a certain number of pounds of halibut and sablefish each season.\textsuperscript{6} The catch is virtually guaranteed\textsuperscript{7} and fishing is allowed over a nearly nine-month season.\textsuperscript{8} The Secretary issued IFQs to the owners or lessees of vessels used to catch halibut or sablefish between 1988 and 1990.\textsuperscript{9} Thus, the government rewarded those who invested capital in the halibut and sablefish fisheries, but not necessarily those who did the fishing. The number of IFQs awarded to any individual depended on the amount of halibut or sablefish caught by that person’s vessel during the 1980s.\textsuperscript{10} For IFQ holders, the new system is a vast improvement upon the earlier, open-access model in which harvesters were out of luck if they found no fish during the brief openers. Today, anyone who wants to benefit from the IFQ system but who did not initially qualify for IFQs—hired skippers, deckhands, and those who did not fish between 1988 and 1990—must buy the right to fish, the IFQs, from someone who already owns IFQs.\textsuperscript{11}

In 1998 and 2002, the villages claimed that the IFQ regulations violated their fishing rights based on aboriginal title to areas of the Gulf of Alaska.\textsuperscript{12} Since the United States Supreme Court’s decision in

\textsuperscript{4} Id.

\textsuperscript{5} Alliance Against IFQs v. Brown, 84 F. 3d 343, 345 (9th Cir. 1996).


\textsuperscript{7} The amount of fish caught is not absolutely guaranteed because IFQ holders must still manage to catch the fish. However, it is very likely that IFQ holders will catch their quota because they now have less competition and longer seasons.

\textsuperscript{8} Fisheries of the Exclusive Economic Zone off Alaska; Sablefish Managed Under the Individual Fishing Quota Program, 68 Fed. Reg. 7719, 7719 (Feb. 18, 2003).

\textsuperscript{9} 50 C.F.R. § 679.40(a)(2)(A)-(B).

\textsuperscript{10} An individual’s vessel must have caught halibut or sablefish between 1988 and 1990 to qualify for any IFQs. Once an individual demonstrated that his or her vessel caught halibut or sablefish during that period, the amount of halibut harvested between 1984 and 1990, and the amount of sablefish harvested between 1985 and 1990, determined the number of halibut and sablefish IFQs awarded to that individual. Pacific Halibut Fisheries, 58 Fed. Reg. 59,375, 59,386 (Nov. 9, 1993) (codified at 50 C.F.R. pt. 679).

\textsuperscript{11} Buying into the IFQ system is expensive. In June 2003, halibut IFQs cost between $3.00 and $13.00 per pound, while sablefish IFQs cost between $1.75 and $13.00 per pound. PAC. FISHING, July 2003, at 37.

\textsuperscript{12} Eyak I, 154 F. 3d 1090, 1091 (9th Cir. 1998); Native Vill. of Eyak v. Evans, No. A98-0365-CV, slip op. at 8 (D. Alaska Sept. 25, 2002) [hereinafter Eyak II].
Aboriginal Title or the Paramountcy Doctrine?

Johnson v. McIntosh\(^\text{13}\) nearly two hundred years ago, the Court has recognized that Indian tribes hold aboriginal title to their territories.\(^\text{14}\) Under aboriginal title, tribes may exclusively use and occupy their territories until Congress extinguishes their title.\(^\text{15}\) The villages argued that the IFQ regulations limited their ability to fish in their traditional areas of Cook Inlet and Prince William Sound, citing as evidence the fact that the Secretary had awarded halibut IFQs to only seventeen village members, and sablefish IFQs to only one member.\(^\text{16}\) In Native Village of Eyak v. Trawler Diane Marie, Inc. (Eyak I),\(^\text{17}\) and Native Village of Eyak v. Evans (Eyak II),\(^\text{18}\) the United States Court of Appeals for the Ninth Circuit and the United States District Court for the District of Alaska, respectively, held that the paramountcy doctrine had extinguished the villages’ rights.\(^\text{19}\) Under the paramountcy doctrine, the federal government must control the exploitation of the seabed and ocean off the coast of the United States to fulfill its duty to defend the nation and to regulate international commerce.\(^\text{20}\) The Eyak I and II courts reasoned that the villages’ claimed aboriginal rights to the seabed and offshore waters were incompatible with federal sovereignty over those areas.\(^\text{21}\)

This Comment argues that aboriginal title is compatible with federal sovereignty over the seabed and ocean off Alaska. Part I details the U.S. Supreme Court’s aboriginal title jurisprudence. Part II describes the extension of federal jurisdiction over the seabed and ocean. Part III traces the history of aboriginal title claims to the seabed and ocean off Alaska. In Part IV, this Comment argues that the Ninth Circuit en banc or the U.S. Supreme Court should apply the Court’s traditional aboriginal title analysis, and not the paramountcy doctrine, to aboriginal

---

\(^{13}\) 21 U.S. 543 (1823).


\(^{15}\) Santa Fe, 314 U.S. at 347.

\(^{16}\) STEVE J. LANGDON, RESOURCE USES BY ALASKA NATIVES AND NON-NATIVES IN THE CENTRAL GULF OF ALASKA OUTSIDE THREE MILES IN THE 20TH CENTURY 126–27 (Sept. 15, 2000) (unpublished report pertaining to Eyak II) (on file with author) (noting also that other, unidentified village members may hold IFQs).

\(^{17}\) 154 F.3d 1090 (9th Cir. 1998), cert. denied, 527 U.S. 1003 (1999).


\(^{19}\) Eyak I, 154 F.3d at 1096–97; Eyak II, No. A98-0365-CV, slip op. at 28.


\(^{21}\) Eyak I, 154 F.3d at 1096–97; Eyak II, No. A98-0365-CV, slip op. at 31.
title claims to the seabed and ocean off Alaska because those claims are consistent with federal sovereignty over offshore areas.

I. ABORIGINAL TITLE ALLOWS A TRIBE TO EXCLUSIVELY USE AND OCCUPY ITS TERRITORY

Nearly two hundred years ago, the U.S. Supreme Court incorporated the concept of aboriginal title into American law in Johnson v. McIntosh.22 The Court described the division of the New World, and explained that discovering nations respected among themselves each nation’s right to take exclusive title to any new territory that it discovered.23 This title gave the discovering sovereign the exclusive right to acquire the discovered territory from resident tribes.24 Those tribes had the right under aboriginal title to exclusively occupy their territory until the sovereign acquired it, or until the sovereign exercised its exclusive power to extinguish the tribes’ title.25 In the United States, only Congress has the power to extinguish aboriginal title.26

A. The U.S. Supreme Court Incorporated the Concept of Aboriginal Title into American Law Nearly Two Centuries Ago

In Johnson, the U.S. Supreme Court introduced into American law the doctrine of discovery, the principle that guided the European division of, and protected aboriginal title to, the New World.27 The dispute in Johnson arose from Thomas Johnson’s purchase of land from the Illinois tribe in 1775 in the area that became known as the Old Northwest.28 Johnson purchased the tribe’s land in defiance of King George III’s ban on settlements after Britain won the French and Indian War in 1763.29 As successor to Britain’s interest in the Old Northwest, the United States acquired from the Illinois the same land that they had already sold to

22. 21 U.S. 543, 574 (1823).
23. Id.
24. Id. at 573.
25. Id. at 574.
27. Johnson, 21 U.S. at 574.
Aboriginal Title or the Paramountcy Doctrine?

Johnson, and in 1818 granted that land to William McIntosh. Johnson took exception to the re-conveyance of what he thought was his land and sued McIntosh in Johnson v. McIntosh, seeking to quiet title. The Johnson case provided the Court with an opportunity to define tribes’ rights to territory under federal sovereignty. The success of Johnson’s suit depended on whether the Court would recognize the validity of the title he had purchased from the Illinois.

In Johnson, the Court explained that European nations divided the New World among themselves under the doctrine of discovery. This doctrine gave to each nation “exclusive title” to the land it discovered. Under this title, the discovering nation had the exclusive right to acquire newfound territory from its Indian occupants. Exclusive title also permitted the discovering nation to convey tribal territory at will but all conveyances came subject to the “Indian right of occupancy,” now commonly known as aboriginal title. Aboriginal title allowed a tribe to exclusively use and occupy its territory after discovery. However, that right was qualified by the restriction that the tribe could convey its territory only to the nation holding exclusive title to the tribe’s land.

The Court decided that Johnson could not enforce his title in the United States’ courts because he purchased his land in 1775 from the Illinois without the consent of Britain, at the time the sovereign reigning over the Illinois’ land.

Less than a decade after Johnson, the Court emphasized the subordination of tribal sovereignty over Indian land within the United States. In Cherokee Nation v. Georgia, the Cherokee sued the State of

30. Id. at 560.
31. Id. at 571–72.
32. Id. at 572.
33. Id. at 574.
34. Id.
35. Id. at 573.
36. Id. at 574.
38. Johnson, 21 U.S. at 574.
39. Id.
40. Id. at 604–05.
41. FELIX S. COHEN’S HANDBOOK OF FEDERAL INDIAN LAW 487 (Rennard Strickland et al. eds., 1982) [hereinafter COHEN].
42. Cherokee Nation v. Georgia, 30 U.S. 1, 17 (1831).
43. 30 U.S. 1 (1831).
Georgia to prevent it from seizing and dividing their land. The Court, however, refused to entertain the Cherokee’s suit under the Court’s original jurisdiction. The Court decided that although the Constitution gave it original jurisdiction over cases between states and foreign nations, the Cherokee were not a foreign nation but a “domestic dependent nation[].” The Court reasoned that the Cherokee were “completely under [United States] sovereignty and dominion” because the United States would consider itself invaded if any foreign nation attempted to acquire land from, or form political connections with, the Cherokee. The Court described the relationship between the Cherokee Nation and the United States as resembling “that of a ward to his guardian.” After Johnson and Cherokee Nation, tribes retained their right to exclusively use and occupy their territory, but exercised that right subject to federal sovereignty.

B. U.S. Supreme Court Decisions Following Johnson Have Refined the Concept of Aboriginal Title

U.S. Supreme Court decisions after Johnson have defined the scope of aboriginal title and the conditions under which aboriginal title can be extinguished. In those cases, the Court held that all territories acquired by the United States are subject to aboriginal title and that aboriginal title allows tribes to exploit the natural resources of their land and water territories. The Court’s precedent permits only Congress to extinguish aboriginal title, but does not require Congress to compensate tribes when it extinguishes their title because

44. Id. at 15.
45. Id. at 20.
46. Id. at 17.
47. Id. at 17–18.
48. Id. at 17.
49. COHEN, supra note 41, at 489–91. In decisions following Johnson, the Court also addressed the enforceability of aboriginal title. See, e.g., United States v. Santa Fe Pac. R.R. Co., 314 U.S. 339, 347 (1941); Cramer v. United States, 261 U.S. 219, 229 (1923). Though it is not clear from these decisions, aboriginal title appears to be enforceable against all but Congress, which has the exclusive ability to extinguish aboriginal title. See generally COHEN, supra note 41, at 488–89. In any case, the Eyak I and II courts did not question the enforceability of the villages’ title.
50. Santa Fe, 314 U.S. at 346.
51. COHEN, supra note 41, at 491 (citing Santa Fe, 314 U.S. 339; United States v. Shoshone Tribe, 304 U.S. 111, 117 (1938)).
53. Santa Fe, 314 U.S. at 347.
Aboriginal Title or the Paramountcy Doctrine?

congressionally unrecognized aboriginal title is not a “property right” protected by the Fifth Amendment. 54

1. Scope of Aboriginal Title

Federal law protects aboriginal title to all territory acquired by the United States. 55 In United States v. Santa Fe Pacific Railroad Co., 56 a railroad company claimed unencumbered title to part of the land conveyed to the United States from Mexico known as the Mexican Cession. 57 Congress granted the land to the railroad company’s predecessor in 1866. 58 In 1883, President Chester Arthur established the Walapais Indian Reservation, which surrounded some of the railroad’s land. 59 Before the Ninth Circuit, the United States argued that the railroad’s land within the Reservation came subject to the Walapais’ aboriginal title. 60 The court held otherwise, deciding that aboriginal title did not exist in the Mexican Cession. 61 On appeal, the U.S. Supreme Court disagreed, holding that all territories acquired by the United States are subject to aboriginal title. 62 The Court remanded the case to allow the Walapais the opportunity to prove that their title existed in fact. 63

Within territory subject to aboriginal title, tribes possess both the rights retained and given up by treaty. As the U.S. Supreme Court explained in United States v. Winans, 64 treaties are “not a grant of rights to the Indians, but a grant of rights from them—a reservation of those not granted.” 65 Among the rights arising from aboriginal title is a tribe’s right to exploit the resources on and beneath the surface of its territory. 66

55. Santa Fe, 314 U.S. at 346.
56. 314 U.S. 339 (1941).
58. Santa Fe, 314 U.S. at 343.
59. Id. at 357.
60. Id. at 343–44.
61. Id. at 345–46.
62. Id.
63. Id. at 360.
64. 198 U.S. 371 (1905).
65. Id. at 381.
66. COHEN, supra note 41, at 491 (citing Santa Fe, 314 U.S. 339; United States v. Shoshone Tribe, 304 U.S. 111, 117 (1938)).
In *United States v. Shoshone Tribe*, the question before the Court was whether the Shoshone tribe’s aboriginal title included the right to use the timber and minerals of its territory. The Court concluded that those resources belonged to the tribe by dint of aboriginal title established by “undisturbed possession of the soil from time immemorial.”

While many aboriginal title cases involve disputes over land, the rights conferred by aboriginal title are not restricted to *terra firma*. In *Washington v. Washington State Commercial Passenger Fishing Vessel Ass’n* (*Fishing Vessel*), the U.S. Supreme Court recognized non-exclusive Indian fishing rights in the Pacific Ocean off Washington State. The Court held that tribal parties to the Stevens Treaties reserved the right to harvest up to fifty percent of the fish passing through their fishing grounds, including those in the Pacific Ocean. Five years later, in *United States v. Washington*, the Ninth Circuit affirmed the right of the Makah tribe, a signatory to one of the Stevens Treaties, to fish up to forty miles off the coast of Washington State. Although both *Fishing Vessel* and *Washington* involved treaty rights, the Court has recognized that those rights are based upon ab original title established by prior exclusive use of the waters at issue.

The U.S. Supreme Court has never considered whether aboriginal title exists in the seabed and ocean off Alaska. However, in response to conflicts between Alaska Natives and non-Indians over control of fish trap sites, the Solicitor of the Department of the Interior issued an opinion in 1942 finding that Alaska Natives had established exclusive rights to the seabed and ocean off Alaska, based on their use of those areas. Following this opinion, the U.S. Supreme Court suggested that Congress also believed that these aboriginal rights existed. In

67. 304 U.S. 111 (1938).
68. Id. at 113.
69. Id. at 117.
70. 443 U.S. 658 (1979).
71. Id. at 670 n.15, 689.
72. Id. In 1854 and 1855, many Northwest tribes signed treaties with the United States. Id. at 661–62 n.2. These treaties are known as the Stevens Treaties because they were all negotiated by Isaac Stevens, the first Governor of the Washington Territory. Id. at 666.
73. 730 F.2d 1314 (9th Cir. 1984).
74. Id. at 1318.
77. Id. at 476–77.
Aboriginal Title or the Paramountcy Doctrine?

Organized Village of Kake v. Egan, the Court held that the Alaska Statehood Act (Statehood Act) neither extinguished nor formally recognized aboriginal rights to the seabed and waters off Alaska, but instead preserved for later resolution Alaska Native claims based on aboriginal title.

2. Extinguishment of Aboriginal Title

Only Congress has the power to extinguish aboriginal title, and when it exercises this power it must act in a “plain and unambiguous” manner. In Santa Fe, the railroad company insisted that President Arthur’s establishment of the Walapais Reservation in 1883 had extinguished the Walapais’ aboriginal title. The Court agreed that the Walapais had abandoned their claims to land outside the Reservation when they accepted the Reservation, but remanded the case for a determination of whether the Walapais had occupied any of the land inside the reservation from “time immemorial.” The Court reasoned that the Walapais would hold title to that land because Congress had not extinguished their title by treaty, warfare, purchase, or “by the exercise of complete dominion adverse to the right of occupancy.” Thus, the Court in Santa Fe recognized the exclusive power of Congress to terminate a tribe’s right to use its territory.

Even when a tribe’s aboriginal title has not been extinguished, Congress may take resources from the tribe’s territory without paying compensation if it has not formally recognized the tribe’s title. In Tee-Hit-Ton Indians v. United States, the Tee-Hit-Ton tribe of southeast

79. 369 U.S. 60 (1962).
81. Kake, 369 U.S. at 65–67 (citing the Statehood Act). In Kake, the Court also held that the State of Alaska could regulate off-reservation fishing by Alaska Natives. Id. at 76. This Comment does not address the implications of that holding because the Eyak villages claim rights to waters beyond the limits of Alaska’s jurisdiction.
83. Id. at 346.
84. Id. at 343–44.
85. Id. at 357–58.
86. Id. at 360.
87. Id. at 347.
88. Id.
89. Tee-Hit-Ton Indians v. United States, 348 U.S. 272, 279, 284–85 (1955); COHEN, supra note 41, at 491.
Alaska argued that Congress owed them the value of timber that it had sold from land the tribe claimed was subject to their aboriginal title.91 The Court assumed that the Tee-Hit-Ton held aboriginal title to their land, but concluded that their title permitted them to simply use and occupy their territory but did not give them “legal rights” to their land.92 The Court decided that the Tee-Hit-Ton had no legal rights to their land because Congress had not formally set aside land for them to use, as it had for the Shoshone and other tribes with dedicated reservations.93 Because the Tee-Hit-Ton lacked legal rights to their land, the Court held that the Fifth Amendment did not require that Congress compensate the Tee-Hit-Ton for the timber taken from their territory.94

C. U.S. Supreme Court Analysis of Aboriginal Title Claims

The following three-step approach to assessing aboriginal title claims can be distilled from the U.S. Supreme Court’s decision in Santa Fe.95 First, the Court determined whether the United States had extended its sovereignty over the land at issue because all territory acquired by the United States comes subject to aboriginal title.96 After recognizing that

---

91. Id. at 275–77.
92. Id. at 279.
93. See id. at 279, 289–90. This distinction explains why the Court ordered Congress to pay compensation to the Shoshone tribe for taking their reservation land and its resources. United States v. Shoshone Tribe, 304 U.S. 111, 117–18 (1938).
95. United States v. Santa Fe Pac. R.R. Co., 314 U.S. 339 (1941). Santa Fe represents the U.S. Supreme Court’s most complete analysis of an aboriginal title claim. The cases described in supra notes 27–75 and accompanying text involve one or two of the three steps of Santa Fe’s aboriginal title analysis, but not all three. In Johnson and Cherokee Nation, the Court focused on the first step of the analysis, the relationship between aboriginal title and United States sovereignty. Johnson v. McIntosh, 21 U.S. 543, 572–74 (1823); Cherokee Nation v. Georgia, 30 U.S. 1, 16–18 (1831). The second step of the analysis, extinguishment of aboriginal title, was also at issue in Johnson, which emphasized that the power to extinguish aboriginal title lies exclusively with the government extending its sovereignty over Indian territory, and in Tee-Hit-Ton, which explained that Congress owes tribes no compensation when it extinguishes aboriginal title. Johnson, 21 U.S. at 573; Tee-Hit-Ton, 348 U.S. at 288–90. The third step of the Santa Fe analysis, proving the existence of aboriginal title, hinges on the ability of a tribe to demonstrate that it exercised the rights stemming from aboriginal title since time immemorial. The Court addressed the scope of those rights in Winans, Shoshone Tribe, and Fishing Vessel. United States v. Winans, 198 U.S. 371, 381 (1905); United States v. Shoshone Tribe, 304 U.S. 111, 116–18 (1938); Washington v. Wash. State Commercial Passenger Fishing Vessel Ass’n, 443 U.S. 658, 666 nn.6–8, 679–81 (1979).
96. Santa Fe, 314 U.S. at 346; accord Vill. of Gambell v. Hodel, 869 F.2d 1273, 1277–78 (9th Cir. 1989) [hereinafter Gambell III]; see also Holden v. Joy, 84 U.S. 211, 244 (1872); Johnson, 21 U.S. at 591.
Aboriginal Title or the Paramountcy Doctrine?

the Mexican Cession was subject to both United States sovereignty and aboriginal title, the Court examined whether Congress had extinguished aboriginal title in that territory. After concluding that Congress had not extinguished aboriginal title, the Court remanded the case to the lower courts to decide the third question, whether the Walapais had in fact exclusively used and occupied their territory. Under Santa Fe, a court must recognize a tribe’s aboriginal title if the court finds that the United States has extended its sovereignty over tribal territory, that Congress has not extinguished aboriginal title to that territory, and that the tribe has exclusively used and occupied its territory.

When aboriginal title does exist, it allows a tribe to exercise both the rights reserved and relinquished by treaty. Among these rights is the ability to develop the natural resources above and below the surface of tribal territories, including those encompassing areas of the Pacific Ocean. Tribes may continue to exclusively use and occupy their territory until Congress exercises its power to extinguish their aboriginal title. When Congress exercises this power, it need not compensate tribes because congressionally unrecognized aboriginal title is not a property right protected by the Fifth Amendment.

II. THE PARAMOUNTCY DOCTRINE AND ACTS OF CONGRESS HAVE EXTENDED FEDERAL CONTROL OVER THE SEABED AND OCEAN

The “paramountcy doctrine” stems from two U.S. Supreme Court decisions, United States v. California and United States v. Texas. In these cases, the federal government claimed ownership of and control over the seabed off the coasts of California and Texas, respectively. The Court employed sweeping language to hold in each case that the

97. Santa Fe, 314 U.S. at 347; Johnson, 21 U.S. at 584–85.
98. Santa Fe, 314 U.S. at 359.
99. Id. at 345–47.
100. See Winans, 198 U.S. at 381.
103. Santa Fe, 314 U.S. at 346–47.
107. Texas, 339 U.S. at 709; California, 332 U.S. at 22.
federal government must have the “paramount power”\textsuperscript{108} to regulate exploitation of the seabed and ocean to fulfill its duty to defend the nation and to regulate international commerce.\textsuperscript{109} Following the Court’s recognition of this power, Congress enacted the Submerged Lands Act (SLA)\textsuperscript{110} and the Outer Continental Shelf Lands Act (OCSLA).\textsuperscript{111} These acts surrendered to the states title to the seabed within three miles of their shores\textsuperscript{112} and extended federal “jurisdiction [and] control” over the seabed beyond three miles from shore.\textsuperscript{113} Later, Congress established its exclusive regulatory authority over fisheries between three and 200 miles offshore\textsuperscript{114} through the Fishery Conservation and Management Act of 1976 (Magnuson Act).\textsuperscript{115} The SLA, OCSLA, and the Magnuson Act reflect the federal government’s paramount control over the seabed and the ocean adjacent to the United States.

A. The Paramountcy Doctrine Established Federal Control of the Seabed and Ocean

Oil was discovered off the coast of California in 1894, and by 1926 both California and Texas were executing offshore oil leases\textsuperscript{116} on the assumption that earlier U.S. Supreme Court decisions had recognized that they held title to the seabed out to three miles from shore.\textsuperscript{117} Shortly after California and Texas began leasing the seabed, Congress considered bringing the seabed within the “public domain,” and the Secretary of the Department of the Interior suggested that the federal

\textsuperscript{108} Texas, 339 U.S. at 719.
\textsuperscript{109} Id.; California, 332 U.S. at 35–36.
\textsuperscript{112} 43 U.S.C. § 1311(a)(1)–(2).
\textsuperscript{113} Id. §§ 1331(a), 1332(1).
Aboriginal Title or the Paramountcy Doctrine?

government might also lease the seabed. In May 1945, the United States took the seabed controversy to the courts by suing the Pacific Western Oil Company to enjoin it from exercising a lease granted to it by California. In September 1945, President Harry Truman declared United States jurisdiction over the natural resources of the seabed and one month later the United States dropped its suit against Pacific Western and instead sued California. In that suit, the United States challenged California’s ability to issue oil leases based on its alleged ownership of the seabed off its shores.

In United States v. California, the United States argued that its constitutional responsibility to “protect this country against dangers to the security and tranquility of its people” required it to control use of the “marginal sea and land under it.” California argued that it owned the resources of the adjacent seabed because it entered the Union on “equal footing” with the original states, which allegedly held title to submerged land off their coasts. The Court disagreed with California, finding no historical support for the idea that the original thirteen colonies owned their adjacent seabed. Instead, the Court decided that the dispute was less about title to the seabed and more about which government, state or federal, should have the power to decide whether foreign or domestic

118. Hardwicke, supra note 117, at 401.
119. FITZGERALD, supra note 116, at 29.
120. Proclamation No. 2667, 10 Fed. Reg. 12,303 (Sept. 28, 1945). President Truman’s proclamation extended U.S. jurisdiction over the seabed to the exclusion of jurisdiction claimed by foreign states. However, the proclamation was not intended to resolve the conflict between the states and the federal government over ownership of the seabed within or beyond three miles from shore. FITZGERALD, supra note 116, at 28.
121. FITZGERALD, supra note 116, at 28–29.
122. Id.
124. Id. at 29–30. Before the U.S. Supreme Court’s decisions in California and Texas, states claimed that the “equal footing” doctrine gave them title to the seabed off their coasts. Created by the Court, the equal footing doctrine holds that all states, upon admission to the Union, took title to the land beneath navigable waters within their boundaries. United States v. Texas, 339 U.S. 707, 716 (1950). The English Crown asserted sovereignty over the submerged land surrounding Britain, and, according to the coastal states, conveyed similar rights to the colonies. FITZGERALD, supra note 116, at 29–30. The coastal states theorized that they should take title to the seabed off their shores because they were on “equal footing” with the original thirteen states, which had inherited the colonies’ rights to the seabed by the 1783 Treaty of Paris, and retained those rights when they formed a Union. Id. Prior to the 1940s, courts supported state ownership of the seabed. John Hanna, The Submerged Land Cases, 3 STAN. L. REV. 193, 200–07 (1951).
125. California, 332 U.S. at 31–32.
entities may extract natural resources from the “marginal sea” bordering California.126

The Court explained that historically the federal government claimed dominion over the three-mile wide marginal sea to protect the nation’s neutrality,127 and recognized that the federal government’s control of the seabed and waters bordering the United States enabled it to regulate commerce over, and fight wars on, the ocean.128 Further, the Court concluded that the United States’ control over the “the ocean or the ocean’s bottom” was as important to federal sovereignty as ownership of land beneath inland waters was to state sovereignty.129 Accordingly, the Court held that the federal government had “full dominion over” oil and other resources of the seabed and ocean off California’s coast.130

Three years later, the Court reaffirmed this holding in United States v. Texas.131 Texas, like California, advanced an “equal footing” argument to support its claim to the land and minerals underlying the Gulf of Mexico.132 Texas asserted that it owned these resources because the Republic of Texas had previously owned them.133 Texas maintained that the Republic had owned and regulated the seabed,134 and had relinquished only its regulatory powers when it joined the Union.135 The Court read Texas’ history differently, deciding that the Republic had surrendered all of its claims to the seabed when it joined the Union on equal footing with the original states.136 Invoking the language of federal sovereignty, the Court held that “national interests and national responsibilities” dictate that all property interests within the “marginal sea” must “unite in the national sovereign.”137 Although both California and Texas appeared to involve only competing claims to the seabed, the Court accepted the opportunity provided by each case to hold that the United States’ sovereignty extends over both the seabed and the ocean.

126. Id. at 29.
127. Id. at 32–33.
128. Id. at 34–35.
129. Id. at 34–36.
130. Id. at 38–39.
132. Id. at 712; see also supra note 124 and accompanying text.
133. Texas, 339 U.S. at 711.
134. Id. at 712.
135. Id. at 713.
136. Id. at 718.
137. Id. at 719.
Aboriginal Title or the Paramountcy Doctrine?

above the seabed. Furthermore, while neither case explicitly held that the United States owns offshore natural resources, the paramountcy decisions clearly awarded the United States control over those resources.140

B. Federal Regulation of the Seabed and Ocean from 1953 to the Present

Between 1953 and 1976, Congress extended federal jurisdiction over the seabed and fisheries off the coast of the United States. Congress exercised its newly won paramountcy powers over the seabed when it passed the Submerged Lands Act and the Outer Continental Shelf Lands Act in 1953. The SLA conveyed to the coastal states title to the seabed within three miles of their shores, and OCSLA extended federal jurisdiction over the Outer Continental Shelf (OCS), which is the seabed beyond three miles from coastal states’ shorelines. Both the SLA and OCSLA preserved existing rights to the seabed under “the law in effect at the time they may have been acquired.” Two decades later, the Magnuson Act of 1976 extended Congress’ regulatory jurisdiction

139. Justice Frankfurter observed that while the Court held that California did not own the seabed, the basis for the Court’s finding that California had trespassed against the United States was the United States’ “dominion” over, and not its ownership of, the seabed. California, 332 U.S. at 43 (Frankfurter, J., dissenting).
140. In two other cases, United States v. Louisiana, 339 U.S. 699 (1950), and United States v. Maine, 420 U.S. 515 (1975), the Court reaffirmed its California and Texas decisions. In Louisiana, the Court held that Louisiana’s claims to the seabed twenty-four miles beyond the three-mile belt were contrary to the national interest. 339 U.S. at 701, 705. In Maine, the Court recognized U.S. jurisdiction to the outer edge of the continental shelf, but qualified the “constitutional premise” of its earlier decisions by noting that overruling California and Texas would disrupt years of legislation and commercial activity founded upon those decisions. 420 U.S. at 517, 524, 528.
143. 43 U.S.C. § 1311(a)(1)–(2).
144. Id. § 1333(a)(1).
145. Id. §§ 1331(a), 1332(2).
146. Id. § 1342; see also id. § 1315 (identical savings clause).
over fisheries between 3 and 200 miles offshore.\textsuperscript{148} The enormous swath of ocean covered by the Magnuson Act is now known as the United States’ “exclusive economic zone” (EEZ).\textsuperscript{149} Although Congress hoped its regulations would resuscitate depleted fish stocks, and therefore boost harvesters’ income, it passed the Magnuson Act primarily in response to foreign domination of United States fisheries.\textsuperscript{150} Together, the SLA, OCSLA, and the Magnuson Act gave the coastal states title to the seabed within three miles of their shores, extended federal control over the OCS, and regulated fisheries between 3 and 200 miles offshore.\textsuperscript{151}

Nearly twenty years after Congress passed the Magnuson Act, the Secretary of the Department of Commerce limited access to the commercial halibut and sablefish fisheries in the EEZ off Alaska by instituting the IFQ program.\textsuperscript{152} The IFQ program regulates a vast area that includes sections of the Gulf of Alaska traditionally used by the Eyak villages.\textsuperscript{153} Currently, only IFQ holders can fish within those areas.\textsuperscript{154} In 1995, the government awarded IFQs to the owners or lessees of vessels that legally caught and sold halibut or sablefish between 1988 and 1990.\textsuperscript{155} People who want to fish for halibut or sablefish today, but

\textsuperscript{148} See 16 U.S.C. § 1801(b)(1).


\textsuperscript{150} Davis, supra note 3, at 283–84.

\textsuperscript{151} The Magnuson Act did not affect rights of navigation within 200 miles of shore. See 16 U.S.C. § 1801(c)(1)-(2).

\textsuperscript{152} Alliance Against IFQs v. Brown, 84 F.3d 343, 345 (9th Cir. 1996). The Magnuson Act and the Northern Pacific Halibut Act of 1982, Pub. L. No. 97-176, 96 Stat. 78 (codified as amended at 16 U.S.C. § 773-773(k) (2000)), give the Secretary the authority to limit access to these fisheries. Alliance, 84 F.3d at 345. The fact that harvesters must first catch a fish before they can claim title to it gives them an incentive to catch as many fish as possible at one time, which is why unlimited access to fisheries can destroy stocks. Id. at 344.

\textsuperscript{153} Eyak I, 154 F.3d 1090, 1091 (9th Cir. 1998).


\textsuperscript{155} Id. § 679.40(a)(2)(A)-(B).
Aboriginal Title or the Paramountcy Doctrine?

who did not receive IFQs in 1995, must purchase IFQs from someone who currently holds them.\textsuperscript{156} IFQs entitle harvesters to a share of the annual Total Allowable Catch (TAC) for a given area.\textsuperscript{157} Each year, the International Pacific Halibut Commission sets the TAC for halibut, and the federal government sets the TAC for sablefish.\textsuperscript{158} Halibut regulations limit Alaska Native villagers to a subsistence harvest of twenty fish per person per day.\textsuperscript{159} Sablefish regulations allocate all sablefish to IFQ holders, leaving none for tribal harvest, because they assume that sablefish’s preference for deep water puts them out of the reach of all but commercial gear.\textsuperscript{160}

The IFQ program is a striking example of the extent to which the federal government has extended its control over offshore resources. The U.S. Supreme Court established the foundation of this control in its \textit{California} and \textit{Texas} paramountcy decisions. In those cases, the Court held that the federal government must control the exploitation of the seabed and ocean off the coast of the United States in order to fulfill its duty to defend the nation and to regulate international commerce.\textsuperscript{161} Congress exercised its paramount power over the seabed by enacting the SLA and OCSLA. These acts gave coastal states title to the seabed within three miles of their shores\textsuperscript{162} and extended federal jurisdiction over the seabed resources beyond three miles from shore.\textsuperscript{163} Twenty years after passage of the Magnuson Act, the federal government restricted access to the commercial halibut and sablefish fisheries in the EEZ off Alaska to people who hold IFQs.\textsuperscript{164} Today, most of the Eyak villages’ members cannot participate in the halibut and sablefish fisheries because they do not hold IFQs.\textsuperscript{165}

\textsuperscript{156} \textit{Alliance}, 84 F.3d at 345.
\textsuperscript{157} 50 C.F.R. § 679.40(b).
\textsuperscript{159} Pacific Halibut Fisheries; Subsistence Fishing, 68 Fed. Reg. 18,145, 18,159 (Apr. 15, 2003) (to be codified at 50 C.F.R. pt. 300.65(g)(2)). Previously, the villagers were limited to two fish per person per day during an eleven-month season. 2001 Pacific Halibut Fishery Regulations, § 23, 66 Fed. Reg. 15,801, 15,809 (Mar. 21, 2001).
\textsuperscript{160} \textit{Eyak II}, No. A98-0365-CV, slip op. at 13 (D. Alaska Sept. 25, 2002).
\textsuperscript{163} \textit{Id.} § 1333(a)(1).
\textsuperscript{165} \textsc{Langdon}, \textit{supra} note 16, at 126–27.
III. ALASKA NATIVE VILLAGES HAVE ATTEMPTED TO ENFORCE THEIR OFFSHORE ABORIGINAL RIGHTS FOR OVER TWO DECADES

The U.S. Supreme Court has consistently held that only Congress has the power to limit or extinguish aboriginal title. Consequently, five Alaska Native villages argued in *Eyak I* and *II* that the IFQ program may not restrict their aboriginal interests in the seabed and ocean off Alaska because Congress has not extinguished those interests. The *Eyak* decisions followed litigation beginning with *Village of Gambell v. Clark* (*Gambell I*), and involving similar claims made by different Alaska Native villages in the 1980s and 1990s. Although the *Gambell* decisions implicitly recognized limited offshore aboriginal interests, the *Eyak* courts held that the paramountcy doctrine had extinguished all aboriginal interests in the seabed and ocean off Alaska.

A. The Gambell Litigation

In *Gambell I*, the Alaska Native villages of Gambell and Stebbins sued to enjoin the Secretary of the Department of the Interior from leasing the seabed off western Alaska to several oil companies. The villages argued that offshore development would negatively affect their subsistence hunting and fishing rights, but the Ninth Circuit held that the villages had sacrificed those rights for part of the $962,500,000 and 40,000,000 acres awarded to Alaska Natives under the Alaska Native Claims Settlement Act (ANCSA). Congress passed ANCSA to

---


167. This Comment uses the term “aboriginal interests” to describe collectively the claims made in *Eyak I* and *II*. In *Eyak I*, the villages claimed the exclusive right to exploit offshore areas based on unextinguished aboriginal title. 154 F.3d 1090, 1091 (9th Cir. 1998). In *Eyak II*, the same villages asserted non-exclusive rights—the alleged remnants of their aboriginal title. No. A98-0365-CV, slip op. at 1–2 (D. Alaska Sept. 25, 2002).


169. *Vill. of Gambell v. Clark*, 746 F.2d 572 (9th Cir. 1984) [hereinafter *Gambell I*].

170. *See, e.g.*, *Gambell III*, 869 F.2d 1273, 1275 (9th Cir. 1989).

171. *See, e.g.*, id. at 1277.


173. *Gambell I*, 746 F.2d at 573.

174. *Id*.

175. *Id*. at 579.
resolve the conflict between the State of Alaska and Alaska Natives over rights to the land Alaska had selected for state ownership after its statehood in 1958. ANCSA settled the dispute by giving money and land to Alaska Natives while also extinguishing the Natives’ aboriginal title to land and waters “in Alaska.” After noting that Congress passed ANCSA to “avoid further litigation of [aboriginal] claims,” the Gambell I court interpreted the phrase “in Alaska” to mean a “geographic region” rather than the “area within the strict legal boundaries of the State of Alaska.” Therefore, the court held that ANCSA extinguished the villages’ subsistence rights in waters beyond the boundaries of the State of Alaska.

However, the Gambell I court also held that the villages could still enjoin the federal government’s leases to the oil companies under section 810 of the Alaska National Interest Lands Conservation Act (ANILCA). ANILCA protects Alaska’s rural residents’ subsistence hunting and fishing on over 100 million acres of land. Section 810 allows the Secretary of the Interior to limit subsistence uses when restrictions are necessary, but only after the Secretary attempts to avoid impacting subsistence activity in the first place. Relying on the use of the phrase “in Alaska” in its “general sense” during House debates on ANILCA and on Congress’ knowledge of the offshore hunting and fishing habits of Alaska’s coastal villagers, the Gambell I court concluded that the phrase “in Alaska” carries the same meaning in ANILCA as it does in ANCSA. Because it determined that section 810 applies to waters beyond Alaska’s boundaries, the court

177. Gambell I, 746 F.2d at 574.
178. “All aboriginal titles, if any, and claims of aboriginal title in Alaska based on use and occupancy, including submerged land underneath all water areas, both inland and offshore, and including any aboriginal hunting or fishing rights that may exist, are hereby extinguished.” 43 U.S.C. § 1603(b).
179. Gambell I, 746 F.2d at 575–76.
180. Id. at 573.
181. Id. at 582.
183. COHEN, supra note 41, at 759.
185. Gambell I, 746 F.2d at 579.
enjoined the oil leases until the district court could evaluate the Secretary’s compliance with section 810.186

Following the Gambell I decision, the Secretary concluded that oil exploration would not restrict the villages’ subsistence uses.187 The villages again sued to stop the exploration, but the District Court for the District of Alaska declined to grant an injunction because the nation’s need for new energy sources outweighed the possibility that subsistence uses would suffer.188 The villages appealed to the Ninth Circuit, which reversed the district court.189 Both the Secretary and the oil companies then appealed to the U.S. Supreme Court.190 In Amoco Production Co. v. Village of Gambell (Gambell II),191 the Court held that section 810 does not apply to the seabed and waters beyond Alaska’s boundaries.192 Deciding that the phrase “in Alaska” plainly refers to the political boundaries of the State of Alaska, the Court declined to extend ANILCA’s reach based on its legislative history.193 However, the Court revived the possibility that the Gambell villages could eventually enforce their subsistence rights when it vacated the Ninth Circuit’s judgment that ANCSA extinguished aboriginal interests in the seabed and waters beyond Alaska’s boundaries.194

On remand, the Ninth Circuit in Village of Gambell v. Hodel (Gambell III)195 reconsidered the geographic scope of ANCSA in light of Gambell II’s interpretation of the phrase “in Alaska” from ANILCA.196 The Ninth Circuit reversed its Gambell I decision and held that ANCSA does not reach beyond three miles from shore and therefore did not extinguish “aboriginal subsistence rights” that may exist in the seabed and waters beyond Alaska’s boundaries.197 Although the Gambell III court eliminated ANCSA as a defense to offshore aboriginal title claims, the court did not definitively answer the oil companies’ second defense, that aboriginal title is incompatible with the United States’

186. Id. at 582–83.
188. Id. at 540.
189. Id.
190. Id. at 534 n.1.
192. Id. at 555.
193. Id. at 552–53.
194. Id. at 555.
195. 869 F.2d 1273 (9th Cir. 1989).
196. Id. at 1275.
197. Id. at 1280.
Aboriginal Title or the Paramountcy Doctrine?

sovereignty over the seabed and waters beyond three miles from shore. 198

This defense was based on the 1982 decision *Inupiat Community of
the Arctic Slope v. United States* (*Inupiat*). 199 In *Inupiat*, the District
Court for the District of Alaska rejected the Alaska Native plaintiffs' 200
claim of “sovereign power” over the Beaufort and Chuckchi Seas as
contrary to federal paramountcy over those waters. 201 In *Gambell III*, the
Ninth Circuit distinguished *Inupiat* because the *Inupiat* plaintiffs argued
that they had never succumbed to the sovereignty of the United States, 202
while the *Gambell* villages simply claimed “rights of occupancy and use
that are subordinate to and consistent with national interests.” 203 For
that reason, the court held that the paramountcy doctrine had not
extinguished the villages’ “aboriginal rights.” 204 Ultimately, the
*Gambell III* court did not foreclose the oil companies’ paramountcy
document defense because it recognized only the possible existence of
limited subsistence rights, not aboriginal title. 205 The Ninth Circuit left it
to the district court to determine whether the villages in fact held
subsistence rights to offshore areas leased to the oil companies. 206

The *Gambell* litigation ended with *Village of Gambell v. Babbitt*
(*Gambell IV*) 207 when the Ninth Circuit affirmed the district court’s
holding that the case was moot because all the oil companies had given
up their offshore leases. 208 This closure left the villages of Gambell and
Stebbins in essentially the same legal position as when they first sued the
Secretary of the Interior. The villages’ residents could continue their
subsistence harvest free from interference by oil exploration, but without
the protection afforded by aboriginal title. The next opportunity for
Alaska Natives to assert aboriginal title to the seabed and waters beyond

---

198. Id. at 1276.
(1950); United States v. California, 332 U.S. 19 (1947)).
200. The *Inupiat* plaintiffs were amici in *Gambell III*. *Gambell III*, 869 F.2d at 1276.
202. Id. at 187.
203. *Gambell III*, 869 F.2d at 1276.
204. Id. at 1277.
205. Id. at 1280.
206. Id.
207. 999 F.2d 403 (9th Cir. 1993).
208. Id. at 407.
Alaska’s boundaries arose when the IFQ program restricted the Eyak villages’ ability to fish for halibut and sablefish.209

B. The Eyak Litigation

The IFQ regulations allow non-Alaska Native commercial harvesters to fish for halibut and sablefish within what five Alaska Native villages claim are their traditional waters in Prince William Sound, lower Cook Inlet, and the Gulf of Alaska.210 The regulations also exclude villagers without IFQs from harvesting halibut and sablefish from those areas and all other regulated waters.211 The Eyak litigation began when the villages sued to enjoin enforcement of the IFQ program by the Secretary of the Department of Commerce and also for a declaration that they hold aboriginal title to, and thus the exclusive right to exploit, their traditional use areas beyond three miles from shore.212 In granting summary judgment in favor of the Secretary, the District Court for the District of Alaska held that the paramountcy doctrine had extinguished the villages’ claimed aboriginal title.213 The court explained that the villages, like the states, could not possess property rights to the seabed and ocean because both the villages and the states depend on the United States for protection.214

The villages appealed to the Ninth Circuit, arguing that the district court erred by basing its decision on the theory that the villages’ aboriginal title claim was equivalent to the state claims in the paramountcy cases.215 The villages maintained that their claim was different because aboriginal title is not a claim of sovereignty but a right of exclusive use and occupancy, qualified by Congress’ ability to extinguish that right.216 To support their argument, the villages cited the Gambell III court’s holding that “aboriginal rights may exist

209. See infra Part III.B; see also Nome Eskimo Cmty. v. Babbitt, 67 F.3d 813, 816 (9th Cir. 1995) (dismissing as moot aboriginal title claim virtually identical to that advanced in the Gambell litigation).
210. Eyak I, 154 F.3d. 1090, 1091 (9th Cir. 1998).
211. Id.
212. Id. at 1091–92. The Alaska Native Claims Settlement Act extinguished the villages’ rights within three miles of shore. 43 U.S.C. § 1603(b) (2000).
213. Eyak I, 154 F.3d. at 1092.
214. Id. at 1094.
215. Id. at 1095.
216. Id.
Aboriginal Title or the Paramountcy Doctrine?

concurrently with a paramount federal interest, without undermining that interest. 217

In *Eyak I*, the Ninth Circuit limited its earlier decision in *Gambell III* by holding that the *Gambell III* court recognized only non-exclusive offshore subsistence rights. 218 The *Eyak I* court reasoned that reading *Gambell III* to recognize aboriginal title, which allows no third-party incursions, would be inconsistent with the *Gambell III* court’s instruction to the district court to determine whether oil exploration would substantially interfere with the *Gambell* villages’ rights. 219 The *Eyak I* court could not discern “a practical difference” between the villages’ aboriginal title claim and Texas’ failed assertion in *United States v. Texas* that it owned the resources of the seabed while the United States retained otherwise “unimpaired sovereignty over the sea.” 220 Because the Texas Court’s holding that the federal government must control all property seaward of the low-tide line did not distinguish between state property and other property, the *Eyak I* court concluded that aboriginal title to the seabed and ocean is as contrary to federal sovereignty as was Texas’ claim. 221 Therefore, the court held that the paramountcy doctrine extinguished the villages’ title to offshore areas beyond Alaska’s boundaries. 222 Like the district court, the Ninth Circuit “[l]eft for another day” the question of whether the villages still possessed non-exclusive aboriginal rights independent of aboriginal title. 223

That day dawned when the *Eyak I* villages filed another complaint against the Secretary of Commerce. 224 While the villages had argued in *Eyak I* that their aboriginal title gave them the exclusive right to exploit certain areas of the seabed and ocean off Alaska, 225 in *Eyak II* the villages claimed that the IFQ regulations interfered with their non-exclusive right to participate in the Alaskan halibut and sablefish fisheries. 226 The District Court for the District of Alaska observed that

217. Id. (citing *Gambell III*, 869 F.2d 1273, 1277 (9th Cir. 1989)).
218. Id.
219. Id.
220. Id. at 1095–96 (citing *United States v. Texas*, 339 U.S. 707, 719 (1950)).
221. Id. at 1096–97.
222. Id.
223. Id. at 1092 n.4.
225. *Eyak I*, 154 F.3d at 1091.
aboriginal rights are usually exclusive rights founded upon aboriginal title, but also noted that “[a]boriginal hunting and fishing rights can exist without aboriginal title.”

The court compared the villages’ asserted non-exclusive rights to those enjoyed by another ocean-going tribe, the Makah of Washington State. By the 1855 Treaty of Neah Bay, the Makah retained their right to fish “in common with” United States citizens at the Makah’s fishing grounds as far as forty miles from shore. Unlike the Makah, Alaska Native villages never signed treaties with the United States. However, the Eyak II court concluded that the villages theoretically could still possess non-exclusive fishing rights because ANCSA had “implicitly reserved” those rights beyond Alaska’s boundaries when it extinguished the villages’ exclusive rights within Alaska.

Ultimately, the Eyak II court decided that non-exclusive rights are also repugnant to federal sovereignty over the seabed and ocean. Because the villages’ claimed rights stemmed from their former sovereignty over their territory, the court held that their rights had been extinguished when the federal government extended its own sovereignty over the seabed and ocean. According to the Eyak II court, exclusive and non-exclusive rights equally threaten the “dominance of national sovereignty” because of the Texas Court’s holding that all offshore property rights must “coalesce and unite in the national sovereign.” The court distinguished the Makah’s non-exclusive rights in the Pacific Ocean on the grounds that the Makah reserved those rights by treaty.

Citing to Ninth Circuit precedent, the district court found that the
Aboriginal Title or the Paramountcy Doctrine?

The paramountcy doctrine essentially effected a “full title extinguishment” that only non-exclusive treaty rights survived.239 On reconsideration, the district court rejected the villages’ argument that Congress preserved their aboriginal rights by a savings clause in OCSLA.240 OCSLA protects rights to areas of the seabed beyond three miles from shore under the “law in effect at the time they may have been acquired.”241 The legislative history of this provision reveals that Congress intended to extend the doctrine of discovery, which protects aboriginal title, to the seabed:

[OCSLA] asserts Federal jurisdiction and control over the Continental Shelf areas beyond original State boundaries, thus bringing the lands and resources within such areas into the same legal status as those acquired by the United States through cession or annexation; in the alternative, such lands and resources are subject to the doctrine of discovery.242

Despite OCSLA’s apparent recognition of the fact that aboriginal title could exist in the seabed, the court held that the savings clause did not protect the villages’ aboriginal rights.243 The court reasoned that those rights had been extinguished in 1950,244 the year of the Texas decision culminating the development of the paramountcy doctrine245 and before Congress enacted OCSLA in 1953.246

Almost two centuries after Johnson, both the Eyak I and II courts applied the U.S. Supreme Court’s paramountcy doctrine to offshore

239. Id. at 29–30 (citing Confederated Tribes of Chehalis v. Washington, 96 F.3d 334, 341 (9th Cir. 1996); W. Shoshone Nat’l Council v. Molini, 951 F.2d 200, 202–03 (9th Cir. 1991); Wahkiakum Band of Chinook Indians v. Bateman, 655 F.2d 176, 180 n.12 (9th Cir. 1981)).


243. Order (Motion for Reconsideration) at 5–6, Eyak II, No. A98-0365-CV.

244. Id.


246. Although the Eyak II court cited to COHEN, supra note 41, at 442, for the proposition that aboriginal rights can exist apart from aboriginal title, Cohen does not cite to any U.S. Supreme Court decisions supporting that conclusion. Furthermore, when the villages petitioned the Court following the Ninth Circuit’s decision in Eyak I, they argued that the Gambell III court must have recognized exclusive aboriginal rights because “all aboriginal rights are, by definition, exclusive.” Petition for Writ of Certiorari at 7 n.5, Eyak I, 154 F.3d 1090 (9th Cir. 1998) (No. 98-1437). This Comment agrees with the villages’ position in Eyak I and in their subsequent petition for a writ of certiorari. Therefore, this Comment will discuss only the compatibility of aboriginal title and exclusive fishing rights with federal sovereignty over the seabed and ocean.
aboriginal title claims instead of the Court’s traditional aboriginal title analysis.\textsuperscript{247} In \textit{Eyak I}, the Ninth Circuit interpreted its \textit{Gambell III} precedent to permit only non-exclusive subsistence rights in the seabed and ocean.\textsuperscript{248} In \textit{Eyak II}, the District Court for the District of Alaska extended \textit{Eyak I} by holding that the paramountcy doctrine extinguished \textit{all} aboriginal rights to the seabed and ocean not reserved by treaty.\textsuperscript{249} After \textit{Eyak I} and \textit{II}, therefore, the villages have neither exclusive nor non-exclusive rights to offshore areas, despite the \textit{Gambell III} court’s holding that “aboriginal rights may exist concurrently with a paramount federal interest, without undermining that interest.”\textsuperscript{250}

IV. ABORIGINAL TITLE SHOULD SURVIVE THE PARAMOUNTCY DOCTRINE BECAUSE IT IS CONSISTENT WITH FEDERAL SOVEREIGNTY OVER THE SEABED AND OCEAN

The Ninth Circuit en banc\textsuperscript{251} or the U.S. Supreme Court should hold that the paramountcy doctrine did not extinguish exclusive hunting and fishing rights founded upon aboriginal title to the seabed and ocean off Alaska. Neither the military nor the economic concerns of the paramountcy cases justify extinguishing aboriginal title because the Court considered those same concerns in its aboriginal title cases and still recognized tribes’ right to exclusively use their territories.\textsuperscript{252} Furthermore, federal actions acknowledging, preserving, and enforcing offshore aboriginal interests before and after the paramountcy cases support the conclusion that judicial recognition of the villages’ title is in the nation’s interest and consistent with federal sovereignty over the seabed and ocean.\textsuperscript{253}

\textsuperscript{248}. \textit{Eyak I}, 154 F.3d at 1095.
\textsuperscript{249}. \textit{Eyak II}, No. A98-0365-CV, slip op. at 27.
\textsuperscript{250}. \textit{Gambell III}, 869 F.2d 1273, 1277 (9th Cir. 1989).
\textsuperscript{251}. The Ninth Circuit can recognize that aboriginal title is consistent with federal sovereignty over the seabed and ocean only by reversing its \textit{Eyak I} decision en banc.
\textsuperscript{253}. Congress must pay to extinguish a tribe’s aboriginal title if it has formally recognized that title. Congress is under no constitutional obligation to pay to extinguish aboriginal title recognized only by the courts. See \textit{Tee-Hit-Ton}, 348 U.S. at 288–89.
Aboriginal Title or the Paramountcy Doctrine?

A. The Eyak Courts Misapplied the U.S. Supreme Court’s Aboriginal Title Analysis

As illustrated by its decision in Gambell III, the Ninth Circuit closely followed the Santa Fe three-part analysis prior to Eyak I. The Gambell III court first recognized that the United States had extended its sovereignty over the seabed and ocean. Next, the court determined that offshore aboriginal title had not been extinguished. Lastly, the court remanded the case to the district court to complete the third step of the analysis, an inquiry into whether the Gambell villages’ rights existed in fact. The Eyak I and II courts erred by conflating the first and second steps of the aboriginal title analysis. The courts held that extending federal sovereignty over the seabed and ocean via the paramountcy doctrine amounted to an extinguishment of offshore aboriginal title. This truncated analysis is contrary to U.S. Supreme Court precedent holding that federal law protects aboriginal title, and its associated exclusive use rights, only after extension of federal sovereignty over new territory.

B. The U.S. Supreme Court’s Aboriginal Title Analysis Should Control Claims of Aboriginal Title to the Seabed and Ocean

The Ninth Circuit en banc or the U.S. Supreme Court should apply the Court’s traditional three-step aboriginal title analysis to offshore aboriginal title claims because the government’s interest in maintaining control over the territories it acquires was at the root of both the paramountcy and the aboriginal title cases. In California and Texas, the Court held that the federal government must control the ocean and seabed in order to fulfill its sovereign duties to protect the nation and to

255. Gambell III, 869 F.2d at 1278.
256. Id. at 1278–80.
257. Id. at 1280. The shortcoming of the court’s analysis is its failure to realize that the extension of federal sovereignty should protect aboriginal title and not merely subsistence rights, id. at 1278, for the reasons set forth infra Part IV. B–C.
regulate international commerce.260 The Court feared that state
ownership of offshore territory and disposal of its resources would
interfere with the nation’s ability to fight wars on, and conduct
commerce over, the ocean.261 These frontier concerns, however, are not
uniquely maritime. As Justice Reed noted in his dissent in Texas,
“[n]ational responsibility is no greater in respect to the marginal sea than
it is toward every other particle of American territory.”262 The doctrine
of discovery developed in Johnson, and applied to aboriginal title cases
ever since, accommodates the same fears stressed by the Court in the
paramountcy cases.263

National security concerns do not justify extinguishing aboriginal title
to the seabed and ocean because aboriginal title does not give to tribes
the power to admit foreign entities within their territory. In California,
the Court worried that states would allow foreign agencies to exploit
offshore resources and that this would severely undermine the federal
government’s ability to keep the peace.264 The U.S. Supreme Court’s
aboriginal title jurisprudence, on the other hand, anticipated and
accounted for the possibility that foreign interests would attempt to
infiltrate United States territory through Indian country.265 Johnson
allowed post-“discovery” Indians to convey their territory only to the
United States because conveyance to any other entity would have lead to
the outcome the doctrine of discovery seeks to avoid: “conflicting
settlements and consequent war” among nations.266 Today, the federal
government could defend the nation against foreign attempts to
politically engage the villages, or to acquire their rights to offshore
areas, because, as the Court explained in Cherokee Nation, such attempts
“would be considered by all as an invasion of [the United States’] territory and an act of hostility.”267

Economic concerns also fail to support the conclusion that the
paramountcy doctrine extinguished offshore aboriginal title. The ability

19, 35–36 (1947).
261. California, 332 U.S. at 36.
262. Texas, 339 U.S. at 723 (Reed, J., dissenting).
264. California, 332 U.S. at 29, 35.
265. Cherokee Nation v. Georgia, 30 U.S. 1, 17–18 (1831); Johnson v. McIntosh, 21 U.S. 543,
573–74 (1823).
266. Johnson, 21 U.S. at 573–74.
Aboriginal Title or the Paramountcy Doctrine?

of the United States to conduct “world commerce” over the Pacific Ocean would remain paramount even with judicially recognized aboriginal title to the seabed and ocean. Aboriginal title would not deprive the United States of any mineral or fisheries resources it desires because Congress can extinguish congressionally unrecognized aboriginal title without paying compensation to the affected tribe. Furthermore, aboriginal title would have no unwanted impact on foreign or interstate trade because tribal trade relations are the exclusive province of Congress.

While the California and Texas Courts held that the federal government must have the power to determine “in the first instance” who may exploit the seabed and ocean off the United States, the Court’s aboriginal title jurisprudence demonstrates that the “first” federal policy toward acquired lands is that they come “subject only to the Indian right of occupancy.” In Santa Fe, the Court enforced that policy by upholding aboriginal title in the face of the cross-country expansion of the railroad system. Similarly, in Shoshone Tribe the Court explained that the Shoshone were entitled to compensation for minerals taken from land set aside for them by Congress because the tribe held those resources by dint of aboriginal title established by “undisturbed possess[ion] of the soil from time immemorial.”

The Santa Fe and Shoshone Tribe decisions demonstrate that tribes retain their exclusive right to use and occupy their territory after it falls under United States sovereignty. By holding that offshore aboriginal interests would frustrate the federal government’s ability to control exploitation of seabed and ocean resources, the Eyak courts failed to appreciate U.S. Supreme Court precedent protecting aboriginal title to natural resources within acquired territories. Because the Court has found neither national security nor economic concerns reason enough to reverse its aboriginal title precedent, the Ninth Circuit en banc or the

---

268. California, 332 U.S. at 35.
270. Oneida II, 470 U.S. 226, 234 (1985); Johnson, 21 U.S. at 574; see also U.S. CONST. art. I, § 8, cl. 3 (Indian commerce clause).
271. California, 332 U.S. at 29.
273. Johnson, 21 U.S. at 574.
Washington Law Review

Vol. 78:939, 2003

U.S. Supreme Court should apply the Court’s traditional aboriginal title analysis to aboriginal title claims to the seabed and ocean.

C. There Is a Strong Federal Interest in Recognizing Aboriginal Title to the Seabed and Ocean

In Texas, the Court held that control of the seabed and ocean involves “national interests and national responsibilities.” 276 The Eyak I and II courts decided that this language, which swept aside state claims to the seabed in the paramountcy cases, also extinguished the villages’ offshore aboriginal interests. 277 However, a series of federal actions before and after the paramountcy cases reveal that aboriginal title to the seabed and ocean is in the nation’s interest. As demonstrated by a 1942 opinion by the Solicitor of the Department of the Interior, 278 the Alaska Statehood Act, 279 OCSLA, 280 and federal efforts to enforce the offshore fishing rights of the Makah of Washington State, 281 the federal government has acknowledged, preserved, and fought for aboriginal interests in the seabed and ocean. This dedication demonstrates that those interests are national interests that survived the paramountcy doctrine.

1. At the Time of the Paramountcy Cases, the Federal Government Acted to Protect Aboriginal Title to Offshore Areas

Before and after the paramountcy decisions, the federal government acknowledged the existence of aboriginal interests in the seabed and ocean off Alaska and acted to preserve them. For example, in 1942 the Solicitor of the Interior observed that Alaska Natives recognized, between themselves, exclusive rights to exploit ocean areas and the seabed off Alaska. 282 The Solicitor concluded that those rights survived the extension of Russian, and later American, sovereignty over Alaska Native territory. 283 Sixteen years after the Solicitor’s opinion, Congress preserved those rights when it admitted the State of Alaska into the

283. *Id.* at 464, 476.
Aboriginal Title or the Paramountcy Doctrine?

Union.284 As the U.S. Supreme Court noted in *Kake*, the Alaska Statehood Act neither extinguished nor recognized as compensable Alaska Natives’ rights, but left them “unimpaired.”285

Similarly, through OCSLA Congress protected aboriginal rights to the seabed beyond three miles from shore under the “law in effect at the time they may have been acquired.”286 The legislative history of this provision reveals that Congress intended to protect aboriginal title by subjecting the seabed to the “doctrine of discovery.”287 This is consistent with the Court’s holding in *Santa Fe* that all territory acquired by the United States comes subject to aboriginal title.288

The *Eyak II* court held that the Solicitor’s opinion was irrelevant because it was written before the Court developed the paramountcy doctrine.289 For purposes of interpreting the paramountcy doctrine’s effect, it does not matter that the Solicitor’s opinion pre-dated the paramountcy doctrine. The Solicitor’s opinion indicates that before the Court developed the paramountcy doctrine it was in the nation’s interest to acknowledge aboriginal rights to the seabed. The paramountcy doctrine invalidated only those claims that were contrary to the nation’s interest.290 The Solicitor’s opinion simply shows that aboriginal title was not a claim contrary to the nation’s interest.

The *Eyak II* court also held that the Statehood Act and OCSLA do not support the villages’ claims because they were enacted after the paramountcy doctrine, and thus could not preserve extinguished rights.291 This holding, however, renders the acts’ savings clauses inconsequential, an effect contrary to that required by the traditional canon of statutory construction “that a statute should be interpreted so as not to render one part inoperative.”292 In sum, the Solicitor’s opinion, the Statehood Act, and OCSLA support the conclusion that extension of federal sovereignty over the seabed and ocean did not extinguish the

---

285. Id.
291. Order (Motion for Reconsideration) at 5–6, *Eyak II*, No. A98-0365-CV.
villages’ offshore aboriginal title because that title is a “national interest and national responsibilit[y].”

2. Thirty Years After the Paramountcy Decisions, the Federal Government Intervened to Protect the Makah’s Fishing Rights in the Pacific Ocean

The Makah’s fishing rights provide another example of federal protection of offshore aboriginal interests. Rather than resisting the Makah’s post-paramountcy fishing, the United States successfully sued on behalf of the Makah to enforce their right to fish out to forty miles from shore. According to the \textit{Eyak II} court, federal paramountcy accommodated the Makah’s fishing rights, but not the villages’ rights, because the Makah, unlike the villages, reserved their rights by treaty. Yet the decisions upholding the Makah’s rights do not discuss the paramountcy doctrine or why treaty rights survived it. The primary distinction between treaty and non-treaty rights is that the government must compensate tribes when it abrogates the former but not the latter. The fact that the federal government would go to court to enforce aboriginal rights that it must pay to abrogate is evidence that aboriginal title, which the government need not pay to extinguish, is also compatible with federal sovereignty over the seabed and ocean.

V. CONCLUSION

The \textit{Eyak I} and \textit{II} courts invoked the paramountcy doctrine to explain why five Alaska Native villages no longer have aboriginal interests in the seabed and ocean off Alaska. In \textit{United States v. California} and \textit{United States v. Texas}, the U.S. Supreme Court held that the federal government must have exclusive control over offshore areas to fulfill its sovereign duties to protect the nation and to regulate international commerce. Neither national security nor economic concerns justify extinguishing aboriginal title to the seabed and ocean because aboriginal

\footnotesize{293. \textit{Texas}, 339 U.S. at 719.  
296. \textit{Washington}, 730 F.2d at 1318 (discussing Makah’s historical fishing in waters out to forty miles offshore); see also Midwater Trawlers Co-operative v. Evans, 282 F.3d 710, 718 (9th Cir. 2002) (finding nothing in Treaty of Neah Bay that explicitly limited geographic extent of Makah’s fishing rights).  
297. \textit{Tee-Hit-Ton Indians v. United States}, 348 U.S. 272, 279, 284–85 (1955); \textit{COHEN, supra} note 41, at 491.}
Aboriginal Title or the Paramountcy Doctrine?

title accommodates these concerns. Tribal enforcement of aboriginal title would not threaten the nation because tribes may not convey their territory to, or form political connections with, foreign entities. Furthermore, offshore aboriginal title would not complicate congressional regulation of international oceanic trade because the Fifth Amendment does not require that Congress pay to extinguish congressionally unrecognized aboriginal title. For these reasons, aboriginal title is fully compatible with federal sovereignty over the seabed and ocean. Therefore, the Ninth Circuit en banc or the U.S. Supreme Court should apply the Court’s traditional aboriginal title analysis, rather than the paramountcy doctrine, when analyzing offshore aboriginal title claims.
THIS ISSUE OF THE WASHINGTON LAW REVIEW
IS DEDICATED TO THE MEMORY OF
JOAN FITZPATRICK
JOAN FITZPATRICK: IN MEMORIAM*

Roland L. Hjorth†

Joan Fitzpatrick graduated from Harvard Law School in 1975. Women were then beginning to enter the legal profession in increasing numbers, but role models were still important in encouraging women to become equal partners in our profession. Joan was an especially effective role model for our students. I think she realized that. It was one of the things that drove her to excel in everything she did. Joan told me—more than once in fact—that she earned every penny she made. It was a point of pride to her. She was a hard worker whose work yielded very important results. And she was the kind of teacher who would make students think: If Professor Fitzpatrick can do that, then I can also do great and important things. Joan joined our faculty in 1984. In her eighteen years with us, she became an internationally known and respected authority on human rights. She was a primary author or editor of six books, the author or co-author of fourteen book chapters, and the author or co-author of about forty scholarly articles. Joan spoke on issues of international human rights throughout North America and Europe. In the words of one of her admirers, she was “brilliant, eloquent, and internationally renowned.”

Joan had a great impact on our students. They recognized her commitment to scholarship and to the furtherance of fundamental human rights. They also knew Joan as one of the truly outstanding teachers in the Law School. Some have told me that Joan was the most significant teacher they had in their entire life. Joan received the Law School’s Professor of the Year Award in 1986 and again in 1988. But as I have talked to former students in the past two weeks I have come to realize that Joan was more than a popular teacher. One of her students told me, “It is not so much that Professor Fitzpatrick was a popular teacher as she was an honored and revered teacher.” Many referred to her fondly—and I suspect behind her back—as “Saint Joan.”

* Remarks made at a memorial service for Joan Fitzpatrick.
† Dean Emeritus and Garvey, Schubert & Barer Professor of Law, University of Washington School of Law.
Joan was active in service to this Law School, this University, and to the world outside this community. She served as associate dean of the Law School and as acting vice provost of the University. She was active in the national and international organizations that were important to her work. For many years, Joan served on the Executive Committee of Amnesty International and on the Executive Council of the American Society of International Law. She was Rapporteur for the Enforcement of Human Rights Committee of the International Law Association. Her accomplishments in scholarship, teaching, and service caused many of us to raise our own aspirations for ourselves.

I have tried, as I know many have tried, to make some sense out of the tragedy of Joan’s death. Her lifetime of professional achievement was a source of good in the world. She cared deeply and passionately about human rights, and did all she could to champion the cause of oppressed peoples. And her achievements were recognized. The Law School Executive Council, with my enthusiastic concurrence, recommended that Joan should be the inaugural holder of the Jeff and Susan Brotman Professorship, and I know that she prized that Professorship as one of the great achievements of her life.

Joan’s achievements should have been a source of great satisfaction to her. Many of us on the faculty thought of Joan as somewhat of a loner, but as I observed her publications during her tenure here, I was always impressed by the number of people with whom Joan collaborated in her scholarly efforts. At least seventeen other people appear as co-authors with Joan on scholarly publications. These co-authors ranged from professors to practicing lawyers, to students. As I have talked with students I have become increasingly impressed by how Joan regarded scholarly activity as something that should involve all of her students. It may be somewhat of an exaggeration, but it was often said that Joan supervised as many student writing projects as the rest of the faculty combined. In her work with students and colleagues in other schools, Joan was a co-worker—not a loner.

And yet, all of Joan’s achievements and work with other people did not lift the veil of loneliness and depression that fell over her. We who knew her saw the pain. We also saw sparks of anger that were often symptoms of that pain. In later years we learned of the disease that was tormenting her. We were continually amazed at Joan’s stellar accomplishments despite the pain and the disease. But speaking for myself, I am not sure I was sufficiently aware of the extent of Joan’s pain, or that I took enough effort to concern myself with doing
something about that pain. I have been acquainted with Joan for eighteen years. But I did not really know her until a little more than two years ago. Joan’s disease and the pain accompanying that disease required me to talk with Joan about the situation. We had a long talk that was initially difficult for me and undoubtedly more difficult for Joan. But as our talk went on, the conversation became easier for both of us. At the end of that talk, I turned to leave. And as I was headed for the door, I hear Joan’s voice say, “Ron.” I turned around. Joan stood up from her desk and gave me a heartfelt hug. I will remember that hug for the rest of my life.

I believe Joan’s death should remind all of us in the law school community that we are a community. At our best, we should be a community that genuinely rejoices in the accomplishments of our members. We should be a community that recognizes pain in our fellow members in the same way we recognize pain in members of our own families.

Joan’s death saddens us deeply. Our most sincere sympathy goes out to Joan’s son Devin and to other members of her family. Joan was a member of our community for eighteen years. Our students were much richer because of her presence here. The cause of human rights throughout the world was advanced by Joan’s work. We are proud of her accomplishments, and we are deeply grateful for her presence in our community for these many years. But our pride and gratitude will always be touched by a tinge of sorrow.

For an unaccountable reason, Joan’s death reminds me of lines from a song about Vincent Van Gogh:

Now I think I know what you tried to say to me,
How you suffered for your sanity,
How you tried to set them free.
They would not listen, they did not know how.
Perhaps they’ll listen now.1

1 DON MCLEAN, Vincent (Starry, Starry Night), on DON MCLEAN: STARRY, STARRY NIGHT (Don McLean’s Starry Night Music Inc. 1999).
IN MEMORY OF JOAN FITZPATRICK: A FORMER STUDENT'S PERSPECTIVE

Elizabeth J. Kane

When I first met Joan Fitzpatrick, it was immediately apparent that she possessed a seemingly limitless understanding of law, politics, and history. She combined this knowledge with a passion for teaching and promoting global human rights. As a scholar, Joan was brilliant; as a person, she was compassionate, dignified, and honest. Throughout the five years that I was lucky enough to have known her, her dedication to students, education, and above all action, was in my eyes heroic.

In 1998, as an undergraduate research assistant at the University of Washington School of Law, Joan allowed me to take her graduate level seminar in international human rights law. Although I was initially reticent to participate in an advanced course full of law students taught by a world-renowned expert, I was reassured by Joan’s practice of treating everyone with a high degree of respect and consideration. In class, Joan applied her comprehensive knowledge of the field with an informed and precise analysis. Among the students, she was able to spread a feeling of inspired confidence that made even the most dismal human rights crisis seem surmountable. Like other students who met Joan, I left her class feeling informed, and motivated to make a change in the world.

Human rights law attracts many students and scholars who aspire to positively impact the world, however pursuing this field can be disillusioning on many levels. For practitioners, it can be an arduous struggle with extremely high stakes and an emotional drain. For students, it can be a seemingly unattainable career goal that promises mainly, perhaps only, altruistic rewards. Despite these negative aspects, Joan encouraged many students to persevere in the field. One way that she did so was to tell students that when she graduated law school in 1975, the field of human rights law scarcely existed—Joan was left to her own initiative and determination to help legitimate human rights law as a legal specialization.

* University of Washington School of Law Juris Doctorate graduate 2003 with honors; associate at Wilson Sonsini Goodrich & Rosati.
Joan’s personal success story and her willingness to help moved many students to dedicate themselves to human rights work. For example, at the University of Washington, many students have been able to find jobs with immigration and human rights organizations, including the Office of the United Nations High Commissioner for Refugees, and participate in projects such as brief writing, research, symposia, and human rights fact-finding delegations. Most, if not all, of these projects would not have been possible if not for the support of Joan. She knew or had contacts with every significant organization promoting human rights; and equally, everyone knew Joan. Although always busy, she consistently and selflessly made time to get students involved in her work. Joan was an excellent resource for students.

Joan’s devotion to individual students was matched by her dedication to educating the public about human rights issues. Immediately after September 11, 2001, I worked with Joan to organize a panel to discuss the impact of terrorism on law and human rights.1 When the day of the event arrived, the room overflowed with concerned people, even on a sunny Friday afternoon the first week of classes. At the same time that she was preparing for the panel discussion, Joan swiftly digested hundreds of pages of legislative materials concerning the Patriot Act and other legal documents being generated. It was always important to her to use her exceptional skills and information to educate the community, and this calling must have been as important as ever in the period of drastic change and international conflict following the attacks on the World Trade Center. Through speaking and educating, Joan instilled in others a sense of duty to get involved and to think critically about the impact of world events on human lives and freedom.

Joan was not a simple academic locked in an ivory tower of unrealistic dreams. She strived to make a real world impact on human rights every day. For example, when her recent book, Human Rights Protection for Refugees, Asylum-Seekers, and Internally Displaced Persons,2 was published, she donated copies to nonprofit agencies that could use it in their everyday work. She once told me that she did not want her work to sit on a shelf somewhere gathering dust; she wanted to

---

get her books directly into the hands of human rights workers who could benefit from them. To me, this was perhaps the most moving aspect of her character: Joan was not content to study merely a situation. She had the talent and courage necessary to tackle the problem, and thereby change the world.

Joan’s combination of scholarly integrity and drive to make an impact on a complex and evolving body of international human rights law made her an outstanding asset to the legal and academic community. We simply will never be able to understand how her loss has changed the realm of law and academics. In this world, where there are so very few true heroes of human rights, Joan will be sorely missed—she was the real thing. However, it is unquestionable that her legacy and knowledge will survive in the minds and hearts of the many students, scholars, and professionals who she taught and inspired.
OBITUARY: JOAN FITZPATRICK: HUMAN RIGHTS LAWYER AND ACTIVIST, SHE FOUGHT FOR REFUGEE RIGHTS∗

Irene Khan

The brilliant, eloquent, and internationally renowned American human rights advocate Professor Joan Fitzpatrick, who has died suddenly aged 52, was a practicing lawyer, academic, and campaigner. She will be remembered for her razor-sharp analysis, clear vision, and deeply principled approach to every issue she took on.

And those issues were many: they included protection of refugees, promotion of women’s rights and respect for human rights during states of emergency. She also worked passionately against torture and the death penalty.

I first came to know Joan through her work on refugee law. She had the rare and remarkable ability to combine academic analysis with a deep compassion for those forced to flee their homes and a practical understanding of the problems facing refugee advocates.

Joan’s commitment to Amnesty International (AI) started in the late 1980s. She played a critical role at the international council meeting in Yokohama, Japan, in 1991, where the organization sought to adjust its mandate to the new global realities in the aftermath of the cold war. Throughout the week she worked tirelessly to build consensus so that AI could start addressing new forms of government repression, and begin work to protect the human rights of gay men and lesbians.

Always known for her thoughtful, meticulous, and extremely thorough way of working, Joan helped lay the foundations for future developments. Through her leadership of AI’s main policy committee, she encouraged further advances on cutting-edge issues in international human rights.

Born in El Paso, Texas, she was the daughter of a U.S. Army officer father. She gained a history degree from Rice University, Texas (1972), a law degree from Harvard Law School (1975), and a law diploma from

Oxford University (1980). From 1975 to 1977 she was a trial attorney in the Bureau of Competition at the U.S. Federal Trade Commission, and from 1977 to 1979 an attorney in the Civil Rights Division of the U.S. Department of Justice.

She then became a professor at the University of Arkansas School of Law, from 1980 to 1984, before moving to the faculty of the University of Washington School of Law, in Seattle, where she taught for the rest of her career. She defied establishment academia by declaring herself a feminist; many students referred to her as “Saint Joan.”

One of the world’s foremost experts on legal protections during states of emergency, Joan was appalled by the impact on international human rights of President George Bush’s “war on terrorism.” She dedicated her recent efforts to articulating these concerns, her most recent piece being Speaking Law to Power: The War on Terrorism and Human Rights, soon to be published in the European Journal of International Law.

Never one to limit herself to academic debate, Joan also engaged on these issues through the judicial system. In a recent lawsuit filed by eleven Kuwaitis being detained in Guantanamo Bay by the U.S. government (the Al Odah case), she submitted an expert statement on behalf of the plaintiffs, briefing the U.S. Court of Appeals for the District of Columbia Circuit on the relevant international legal standards on incommunicado detention and prolonged detention without charge or trial. When the court ruled that it could not protect the detainees, she maintained that the opinion was utterly unsupported by law, and announced her intention to write an article on the jurisdiction with respect to territory of human rights treaties.

The deaths of two captives at an American interrogation facility at Bagram airbase in Afghanistan in December 2002 prompted her to address the use of torture by US forces in Torture As an Impeachable Offense, which she wrote on behalf of the San Francisco-based Center for Justice and Accountability. As a member of the legal advisory council of the center, she helped victims of torture pursue legal remedies against the perpetrators.

Her legal work included filing briefs as an amicus curiae, an adviser to the court, in cases before the U.S. Supreme Court addressing the imposition of the death penalty on children, the turning away of Haitian asylum seekers on the high seas by U.S. agents, and refugee protection for women fleeing violence by private individuals.

In addition to numerous articles, she wrote five books on subjects ranging from states of emergency to refugees, and a widely used
Tributes to Professor Fitzpatrick

textbook on international human rights law which she coauthored with Professor David Weissbrodt.
She is survived by her son Devin, two brothers and a sister.
UNPACKING NEW POLICING: CONFESSIONS OF A FORMER NEIGHBORHOOD DISTRICT ATTORNEY

Alafair S. Burke*

Introduction......................................................................................986
I. The New Policing Models and New Discretion Scholarship...988
   A. The New Policing.................................................................992
   B. New Policing and Police Discretion .................................996
   C. The New Discretion Scholarship .......................................999
II. Critiquing the New Discretion Scholarship ...............................1003
   A. Idealizing Notions of “Community” ..................................1004
   B. Determining Constitutional Parameters by Community    
      Sentiment ...........................................................................1009
      1. Assessing Community Support for the New Policing ....1009
      2. Permitting Communities To Waive Constitutional        
         Protections ......................................................................1013
   C. Political Process: An Inadequate Safeguard ......................1015
   D. The Virtues of Vagueness Review.....................................1017
   E. Elasticity of the New Discretion Alternatives....................1023
III. The Programmatic Purpose Approach: Distinguishing          
     Between Punitive and Non-Punitive Policing.......................1028
     A. Abandoning Rhetoric and Developing Meaningful        
        Distinctions ......................................................................1030
     B. Drug Free Zones: An Example ..........................................1032
        1. No Affirmative Restraint ............................................1036
        2. Not the Traditional Aims of Punishment ....................1039
        3. Connection To Civil Purposes ......................................1042
        4. Not Historically Regarded as Criminal Punishment ......1048
        5. Not Excessive in Relation to Civil Purposes ...............1052

* Associate Professor, Hofstra Law School. B.A., Reed College; J.D., Stanford Law School. I would like to thank Robin Charlow, Nora Demleitner, Peter Spiro, and Robert Weisberg for their helpful comments on previous drafts. I am grateful to Cynthia Leigh, Reference Librarian of the Hofstra Law Library, and Daniel Venditti for their dedicated assistance.
C. Recharacterizing Morales ..................................................1052
D. Applying the Programmatic Purpose Approach ............1059
IV. Conclusion ..............................................................................1065

Abstract: This Article attempts to reframe a burgeoning scholarly debate about the appropriateness of neighborhood self-governance as both a means to local crime control and a normatively worthy end in itself. On one side of the existing debate stands an emerging and influential group of “new discretion” scholars, who defend the delegation of discretion to police officers attempting to enforce social norms that are often ambiguous. These scholars argue that the support and involvement of so-called “communities” in such law enforcement efforts can be an adequate substitute for traditional judicial scrutiny of police discretion, particularly the prohibition against vague criminal laws. On the other side of the debate are traditional civil libertarians who view norm-based policing and the theories of self-governance underlying it as thinly disguised forms of majoritarianism.

This Article has two primary goals. One goal is to use the author’s experience as a community-based prosecutor to critique the new discretion scholars’ reliance upon malleable notions of community to determine the legality of police programs. The second goal is to develop a more meaningful distinction among new policing efforts. Specifically, this Article advocates a distinction between civil and criminal initiatives. This approach would retain the existing prohibition against vague criminal laws. However, it would permit cities to implement strategies requiring police discretion, as long as those strategies avoid traditional criminal investigation, prosecution, and punishment. Such an approach would force cities either to adopt nontraditional responses to public safety problems or to be scrutinized under the traditional rules governing criminal law and procedure.

INTRODUCTION

Community policing is the bandwagon of modern law enforcement1 and has quickly led to a burgeoning debate regarding the appropriateness of neighborhood self-governance and the enforcement of social norms as local crime control methods. An emerging and influential group of scholars defends the delegation of discretion to police officers attempting to enforce often ambiguous community norms.2 These scholars argue that at least some forms of community support and involvement in police programs serve as adequate substitutes for traditional judicial scrutiny of police discretion, particularly the

1. See WESLEY G. SKOGAN & SUSAN M. HARTNETT, COMMUNITY POLICING, CHICAGO STYLE at vi (1997) (“The concept [of community policing] is so popular with the public and city councils that scarcely a chief wants his department to be known for failing to climb on this bandwagon.”); Tracey L. Meares, A Colloquium on Community Policing: Praying for Community Policing, 90 CAL. L. REV. 1593 (2002) (noting that the term community policing “has become ubiquitous among law-enforcement practitioners and scholars” and collecting sources that use the term).
2. See infra Part I.B–C for a discussion of the role of police discretion in the new policing models and a summary of the new discretion scholarship defending that discretion.

986
Unpacking New Policing

prohibition against vague criminal laws. On the other side of the debate are traditionalists who view the new policing and the theories of self-governance underlying it as threats to traditional civil libertarian values protected by well-established constitutional law.3

This Article takes an initial step at reframing the debate by attempting to reconcile at least some new policing forms with existing constitutional jurisprudence governing the criminal justice system. The Article has two primary goals. First, the Article critiques the new policing scholarship. Drawing on my own experiences as a “Neighborhood District Attorney” in Portland, Oregon, I argue that the rhetoric of community can be used to disguise similarities between traditional law enforcement models and at least some new policing initiatives. By using the malleable concept of “community” to distinguish the traditional from the new, some new policing advocates would permit cities to avoid the constitutional rules that usually govern law enforcement by dressing up traditional initiatives as “community-oriented.”

A second goal is to develop a more meaningful distinction among new policing programs. Where the current advocates of the new policing attempt to draw lines between programs based on the role of neighborhood governance, this Article advocates a distinction between civil and criminal initiatives. Whether one’s notion of criminal punishment is retributive or utilitarian, traditional criminal law operates by punishing individual offenders for past acts. Much of the new policing, in contrast, involves the implementation of long-term programs that seek to prevent violations of the criminal law through means other than criminal punishment. Distinguishing between programs based upon their “programmatic purpose”4 would result in increased judicial tolerance for some forms of police discretion, but would do so without eradicating the jurisprudence that appropriately limits discretion in the criminal context.

This Article critiques the new discretion scholarship and develops the alternative “programmatic purpose” model in three sections. Part I

---

3. See infra notes 193–200 and accompanying text for an overview of the scholarship criticizing the new policing scholarship.

4. I use the term “programmatic purpose” as it is used by the U.S. Supreme Court in its recent “special needs” jurisprudence. See Ferguson v. City of Charleston, 532 U.S. 67, 82 (2001) (holding that a public hospital’s drug testing program for pregnant women did not advance a special need because its programmatic purpose was to generate evidence to be used in a criminal prosecution); City of Indianapolis v. Edmond, 531 U.S. 32, 45–48 (2000) (holding that the special needs doctrine requires judicial examination of programmatic purpose and striking down a drug interdiction checkpoint because its programmatic purpose was general crime control).
provides an overview of the new policing models and the emerging scholarship defending police discretion. Part II draws upon my own experiences in a community prosecution unit to argue that the new policing models do not warrant an overhaul in the constitutional rules that govern the criminal justice system. While retention of the current regime would limit police discretion in enforcing laws with a punitive purpose, criminal justice jurisprudence would not hinder new policing efforts reflecting a non-punitive programmatic purpose.

Part III fleshes out the programmatic purpose model in more detail. It does so first through example. Portland’s Drug Free Zone ordinance enables police to control public spaces in high-crime neighborhoods by excluding targeted individuals from discrete geographic areas. Applying the U.S. Supreme Court’s developing jurisprudence on criminal punishment, I argue that this form of neighborhood exclusion is not a criminal sanction because individuals are excluded only temporarily and from a relatively small portion of the city, and are not arrested unless they violate the exclusion order. After concluding that neighborhood exclusion does not necessarily constitute criminal punishment, I turn to an analysis of the U.S. Supreme Court’s decision in City of Chicago v. Morales and argue that the Court should have determined as a threshold matter whether Chicago’s anti-gang loitering ordinance imposed a criminal or civil restraint before holding that it was intolerably vague. Finally, Part III discusses the advantages of the programmatic purpose model over the current new discretion scholarship.

I. THE NEW POLICING MODELS AND NEW DISCRETION SCHOLARSHIP

To understand what is “new” about the new policing, it is helpful to contrast it with its immediate predecessor. The prevailing model of policing over the past forty years has been dominated by an emphasis on rapid responses to 911 calls and subsequent criminal case creation. Anyone who has seen a few episodes of the television series “Law & Order” is familiar with the traditional role of police in what this Article calls the rapid-response model of policing.

7. The long-running television series “Law & Order” commences with the familiar
Unpacking New Policing

In rapid-response policing, a crime’s occurrence is what triggers the involvement of law enforcement. Once the crime is reported, the criminal justice system seeks to identify, arrest, prosecute, and punish the crime’s perpetrator. The rules of criminal procedure, investigatory techniques, and criminal intelligence are means to that end. In pursuit of that common end, each actor within the criminal justice system maintains her own, distinct role. Police get involved after the crime occurs, prosecutors join in only after an arrest is made, and corrections officers play a role only post-conviction.

American cities have increasingly departed from rapid-response policing over the past decade by turning to forms of policing that emphasize an ongoing, proactive role in maintaining order, rather than simply responding to crimes as they occur. Although policy makers commonly describe the new policing trend as “community policing,” one purpose of this Article is to question the meaningfulness of community participation in these emerging forms of policing. Accordingly, this Article uses the alternative term of “new policing models.”

Looking beyond the rhetoric of “community” is especially important in light of that term’s steadily increasing influence on the criminal justice system. The concept of community policing blossomed during the 1990s as the centerpiece of President Clinton’s anti-crime efforts.


9. To some degree, this trend is “new” only when viewed in a relatively narrow historical context. Prior to the emergence of rapid-response policing, the United States experienced an earlier era of law enforcement when police practiced an ongoing order-maintenance function. For a history of modern policing environments, see Livingston, supra note 6, at 565–73 (documenting the “reform era” at the turn of the twentieth century, which brought increased professionalism and autonomy to law enforcement) and George L. Kelling & Mark H. Moore, From Political To Reform To Community: The Evolving Strategy of Police, in COMMUNITY POLICING: RHETORIC OR REALITY 3–11 (Jack R. Greene & Stephen D. Mastrofski eds., 1988) (noting law enforcement’s twentieth century shift away from social welfare work toward a narrow focus on crime control).

10. In his 1994 State of the Union Address, President Clinton vowed to place 100,000
By 1999, a majority of police departments boasted community policing programs, and a fifth of all local police officers in the nation were designated as community policing officers.11

The scholarship examining the new policing models assumes that the models are defined by partnerships between law enforcement and community.12 However, were community policing simply a term to describe increased cooperation between police and residents, there might be little to debate among criminal and constitutional law scholars. At the heart of the new community-oriented crime control efforts lies more than just an increased emphasis on community involvement. Much of the new policing is premised on the claim that neighborhoods should be permitted to establish governing norms of behavior, enforced by the police.13 In an attempt to enforce social norms in public spaces, cities have enacted or increased the enforcement of ordinances establishing curfews and prohibiting panhandling, public intoxication, public

11. This change was a quick one. In 1997, only 4% of local police officers were designated community policing officers, and approximately one third of local police departments had full-time officers engaged in community policing activities. In 1999, 21% of all police officers were designated community policing officers, and 64% of local police departments—representing 86% of the nation’s population served by local police—had full-time officers working on community policing projects. MATTHEW J. HICKMAN & BRIAN A. REAVES, BUREAU OF JUSTICE STATISTICS SPECIAL REPORT: COMMUNITY POLICING IN LOCAL POLICE DEPARTMENTS, 1997 & 1999 2 (Feb. 2001, rev. Mar. 2003).

12. See Heymann, supra note 8, at 420 (noting that new policing models use the help of those in a neighborhood); Livingston, supra note 6, at 575 (noting community policing’s emphasis on community involvement); Meares, supra note 1, at 1600 (noting that community involvement is one of the “key forces” shaping community policing).

13. See Dan M. Kahan, Social Influence, Social Meaning, and Deterrence, 83 VA. L. REV. 349, 367–77 (1997) (discussing the enforcement of social norms in new policing efforts); Livingston, supra note 6, at 578–84 (same); Schragger, supra note 8, at 382 (noting that the new policing is based on “a claim that communities— . . . at the level of the neighborhood—not only can but should effect basic changes in the fundamental rules that govern relations between the state and the individual, as well as among individuals”); Sarah E. Waldeck, Cops, Community Policing, and the Social Norms Approach to Crime Control: Should One Make Us More Comfortable with the Others?, 34 GA. L. REV. 1253, 1256–58 (2000) (discussing the enforcement of social norms in new policing efforts).
Unpacking New Policing

camping, graffiti, unlicensed street vending, and loitering, often in the name of the “community.”¹⁴

The fascination with the enforcement of community norms is not limited to police departments. Prosecutors’ offices around the country have developed so-called “community prosecution” units, moving prosecutors out of courthouses and into neighborhoods to address public safety concerns directly.¹⁵ Court systems have created specialized community courts to dispose of the many low-level offenses resulting from the increasing criminalization of disorder.¹⁶

From the new policing has emerged a growing body of legal scholarship questioning whether current constitutional jurisprudence is appropriate for evaluating the constitutionality of the new policing efforts. This Part provides an overview of the new policing approaches and the responding scholarship.


A. The New Policing

Although distinctions can—indeed, should—be drawn among various new policing models, a handful of characteristics separate these models from rapid-response policing. First, new policing models tend to address the needs of specific neighborhoods, rather than apply the same general policing approach throughout the department’s jurisdiction. For example, police might depart from the traditional rapid-response model by stepping up enforcement against a specific type of crime, such as prostitution, that plagues only one district in the city. A nearby district, however, might suffer more than its fair share of car thefts, requiring a different strategy a few miles away within the same jurisdiction.

Second, as part of this focus on the problems of specific neighborhoods, the new policing models tend to incorporate increased involvement of citizens and enhanced citizen-police partnerships. The rapid-response model looks to citizens as crime reporters and witnesses. While traditional policing methods value citizen assistance in identifying, locating, and prosecuting offenders, they do not look to citizens to shape the priorities of the police department. New policing models, in contrast, tend to look to community “stakeholders,” like neighborhood residents and business owners, to identify the public

17. See Archon Fung, Beyond and Below the New Urbanism: Citizen Participation and Responsive Spatial Reconstruction, 28 B.C. ENVTL. AFF. L. REV. 615, 629 (2001) (discussing variations among community groups in selecting priorities and responsive approaches); Heymann, supra note 8, at 421, 423–24 (discussing the concentration of police resources by neighborhoods and “the fact that police are accountable to neighborhoods as well as to cities”).


19. Of the terms that emerge repeatedly from the jargon that surrounds community justice programs, the term “stakeholder” appears to be a favorite, referring not just to residents but to “practically everyone” who might care about what happens to a neighborhood. See Jeffrey Fagan & Garth Davies, Street Stops and Broken Windows: Terry, Race, and Disorder in New York City, 28 FORDHAM URB. L.J. 457, 502 (2000) (discussing the role of “stakeholders” in shaping norms under new policing approaches); Joan W. Howarth, Toward the Restorative Constitution: A Restorative Justice Critique of Anti-Gang Public Nuisance Injunctions, 27 HASTINGS CONST. L.Q. 717, 720 (2000) (noting the importance of “stakeholder” agreement in restorative justice programs); Tracey L. Meares, Norms, Legitimacy and Law Enforcement, 79 OR. L. REV. 391, 410 (2000) (noting that participation of community “stakeholders” legitimizes government action); Thacher, supra note 8, at 765 (noting that community policing attempts to incorporate “practically everyone”).
Unpacking New Policing

safety problems specific to that community and to develop and implement strategies to solve those problems.

A third trend among the new policing models, emphasis on low-level criminal offenses, may follow from the second. Few citizens are troubled on a day-to-day basis by the murder, rape, and major assault cases that police, prosecutors, and courts traditionally prioritize. What bothers them are the street-level drug dealers, prostitutes, drunks, noise makers, and vandals who remind them on a daily basis that their neighborhood is not what they wish it to be. The new policing models tend to address these so-called “quality of life” offenses.

The new policing’s emphasis on low-level offenses stems in part from James Wilson and George Kelling’s influential 1982 essay, “Broken Windows.” The broken windows theory maintains that low-level disorder—such as loitering, public intoxication, and littering—contributes to more serious crime if left uncorrected. Wilson and Kelling suggested that disorder can contribute to overall crime in two separate ways. First, the appearance of disorder can create fear among community residents, who will begin to stay indoors, withdraw from their community, and eventually stop trying to assert control over the neighborhood. Second, low-level crime and disorder may lead directly to an increase in crime by breaking down community standards and signaling to lawbreakers that “no one cares” enough about the community to enforce the law. Under the direct component of the theory, then, disorder is essentially contagious, and one broken window,

20. See Heymann, supra note 8, at 444 (noting that traditional policing prioritized offenses by their severity); Kelling & Moore, supra note 9, at 60 (maintaining that the historical prioritization of serious offenses stems from those “who think the enduring social interest in non-intrusive and fair policing can best be served by focusing attention on a few serious and visible crimes”); Livingston, supra note 6, at 578 (noting that citizens’ highest law enforcement priorities are low-level disorders such as abandoned buildings, vandalism, gangs, loitering juveniles, and unsafe parks); Thacher, supra note 8, at 776 (noting disjoint between traditional police prioritization of serious crimes and community groups’ concern about “soft crime” such as rowdy youths, barking dogs, and physical decay); George L. Kelling, Policing and Communities: The Quiet Revolution, PERSP. ON POLICING (Nat’l Inst. of Justice, U.S. Dep’t of Justice, Washington, D.C.), June 1988, at 2 (also noting low-level nature of community concerns).

21. See Livingston, supra note 6, at 558–59 (noting the new police focus on “promoting the ‘quality of life’ in public spaces”).


23. Id.

24. Id. at 32–33.

25. Id. at 31.
if left unrepaired, can lead to the rest of the windows being broken. To
heighten law-abiding residents’ confidence in their neighborhoods and to
prevent the progression of disorder, Wilson and Kelling advocated a
police function in reducing minor but visible social disorder.26

Wilson and Kelling’s article has influenced not just criminal justice
theory, but also police strategies in the field.27 For example, based on the
broken windows theory, former New York City mayor Rudolph Giuliani
employed an aggressive, “zero-tolerance” policing approach, directing
officers to increase arrests for such low-level crimes as subway turnstile
jumping, public intoxication and urination, jaywalking, unlicensed street
vending, and window-squeegeeing.28 Although critics question the
efficacy of the zero-tolerance approach,29 the notable decrease in New
York City’s crime rate during the 1990s has been widely attributed to the
Giuliani strategy.30

Finally, a fourth defining characteristic of the new policing models is
the desire to develop programmatic responses to specific problems in
order to prevent future occurrences. The rapid-response model’s
immediate objective is to arrest and punish an individual offender.
Through specific and general deterrence, the punishment of that
individual offender and others like him might lead to an overall decrease
in crime, but the immediate objective is case-creation. In contrast, the

26. Id. at 38.
27. See Livingston, supra note 6, at 583–85 (discussing the influence of the broken windows
theory on contemporary policing).
City, 45 CRIME & DELINQUENCY 171, 172 (1999); Thompson, supra note 16, at 84; Waldeck, supra
note 13, at 1275–78.
29. See BERNARD E. HARCOURT, ILLUSION OF ORDER: THE FALSE PROMISE OF BROKEN
WINDOWS POLICING 90–104 (2001) (questioning the efficacy of order-maintenance policing);
Greene, supra note 28, at 177–81 (suggesting alternative explanations for New York City’s crime
rate decrease during the 1990s).
30. See, e.g., WILLIAM BRATTON WITH PETER KNOBLER, TURNAROUND: HOW AMERICA’S TOP
COP REVERSED THE CRIME EPIDEMIC 152–56 (1998) (discussing New York City’s implementation
of the broken windows theory); Bernard E. Harcourt, Reflecting on the Subject: A Critique of the
Social Influence Conception of Deterrence, The Broken Windows Theory, and Order-Maintenance
Policing New York Style, 97 MICH. L. REV. 291, 293 (1998) (noting that “it is today practically
impossible to find a single scholarly article that takes issue with the quality-of-life initiative”);
Kahan, supra note 13, at 368–69 (noting the decline in New York City’s crime rates and stating that
“city officials and at least some criminologists credit the larger reduction in crime rates to [the]
emphasis on ‘order maintenance’”). But see Jeffrey Fagan et al., Declining Homicide in New York
City: A Tale of Two Trends, 88 J. CRIM. L. & CRIMINOLOGY 1277, 1322 (1998) (maintaining that
“the pattern in New York City is much more consistent with gun-oriented policing than with
indiscriminate quality-of-life interventions as a cause of decline”).
new policing models incorporate the influential work of Herman Goldstein by treating the prevention of future occurrences as the primary objective. Goldstein called into question the prevailing rapid-response model of policing, where police resources were squandered on chasing down individual suspects after crimes had already occurred. He argued that police would be more effective if they spent more time seeking to prevent or reduce long-term, recurring problems instead of responding to individual crime occurrences. In Goldstein’s view, the response should be tailored to the particular problem and should look beyond formal law enforcement methods. Adopting this concept of “problem-oriented” policing, the new policing models attempt to identify the most effective response to a community problem, which may not require increased arrests and prosecutions.

The new policing’s emphasis on problem-solving instead of case-creation may very well be inherent in its prioritization of low-level criminal offenses. Rapid-response policing is particularly ill-suited to low-level crimes. Unless a low-level offender is caught red-handed, police are unlikely to identify and locate the offender without expending resources that would outweigh that single offense’s importance. Even if a low-level offender is identified and arrested, the criminal justice system is ill-suited to remedy a neighborhood’s ongoing problems. Prostitutes, vandals, and the publicly intoxicated serve short sentences or


32. See Heymann, supra note 8, at 423 (noting Goldstein’s role in encouraging police to treat crime prevention as a primary goal, focusing more on general problems than on individual incidents); Livingston, supra note 6, at 573–75 (discussing Goldstein’s influence on modern policing environments).

33. Goldstein, Improving Policing, supra note 31, at 245.

34. See generally GOLDSTEIN, PROBLEM-ORIENTED POLICING, supra note 31, at 32–49, 102–03.

35. See id. at 103–04.

36. See Gest, supra note 10, at 762 (listing problem-oriented policing as a move away from the rapid-response model); Lawrence Rosenthal, Policing and Equal Protection, 21 YALE L. & POL’Y REV. 53, 53 (2003) (discussing problem-oriented policing). Of course, new policing methods may result in increased arrests and prosecutions if these are determined to be the most effective response to a community problem. What distinguishes problem-oriented policing from the rapid-response model is its willingness to prevent future crimes without generating additional cases for the criminal justice system. See Livingston, supra note 6, at 573–74 (noting that under problem-oriented policing, traditional law enforcement may be “only one method among many” in a response to public safety concerns).
receive probationary sentences, then return to the same corners to provoke the same community complaints.

B. New Policing and Police Discretion

The new policing models inevitably require police to exercise discretion. The police may turn to the community to identify law enforcement priorities in community policing, but the community may identify more concerns than the police are able to tackle effectively, requiring police to choose among them. Problem-solving policing\(^{37}\) calls for discretion in choosing among non-traditional alternatives to reducing systemic problems. Even so-called zero-tolerance policing\(^{38}\) calls for police discretion; it is inconceivable in a world of limited resources that police fully enforce the penal code in every instance.\(^{39}\)

The most controversial need for discretion, however, arises from a legislative inability to define with precision the “disorder” that police should seek to eradicate. Although police can reduce disorder indirectly by enforcing non-vague prohibitions against prostitution, littering, graffiti, and vandalism, any broad attempt to criminalize deviations from community norms is bound to be vague. The new policing’s inevitable reliance upon police discretion raises potential conflicts with existing judicial limitations upon such discretion, including the void for vagueness doctrine.

Particularly during the Warren Court years, the U.S. Supreme Court made limiting the exercise of police discretion a primary purpose of constitutional jurisprudence in the criminal arena.\(^{40}\) As part of this effort, the Court has held that vague criminal laws offend the Fourteenth Amendment’s guarantee of due process. For example, in *Papachristou v. City of Jacksonville*,\(^{41}\) the Court unanimously held void for vagueness a

\(^{37}\) See *supra* notes 31–36 and accompanying text for discussion of problem-solving policing.

\(^{38}\) See *supra* notes 28–30 and accompanying text for discussion of the zero-tolerance policing model.


\(^{41}\) 405 U.S. 156 (1972).
Unpacking New Policing

city ordinance that punished those “strolling around from place to place without any lawful purpose,” not to mention jugglers and men who lived off the earnings of their wives.42

Three values have been offered to justify the Court’s constitutional prohibition against vague criminal laws.43 First, a vague criminal law fails to give notice to the citizenry of the boundary between legal and criminal activity.44 That this failure of notice is seen as fundamentally unfair suggests, at least implicitly, a value placed on the citizenry’s ability to walk with confidence all the way to the line of criminal conduct without fear of crossing it.45 Second, vague criminal laws are criticized for permitting police arbitrariness.46 By allowing police officers to interpret the scope of its prohibition, a vague criminal law entrusts police officers with discretion to decide whom to arrest for a criminal offense. Finally, courts have articulated concerns that vague criminal laws could chill protected activities, such as speech or association.47 In this respect, the void for vagueness doctrine provides an

42. Id. at 170–71. The ordinance authorized police to arrest:

Rogues and vagabonds, or dissolute persons who go about begging, common gamblers, persons who use juggling or unlawful games or plays, common drunkards, common night walkers, thieves, pilferers or pickpockets, traders in stolen property, lewd, wanton and lascivious persons, keepers of gambling places, common railers and brawlers, persons wandering or strolling around from place to place without any lawful purpose or object, habitual loafers, disorderly persons, persons neglecting all lawful business and habitually spending their time by frequenting houses of ill fame, gaming houses, or places where alcoholic beverages are sold or served, persons able to work but habitually living upon the earnings of their wives or minor children.[

Id. at 156–57 n.1. See also Kolender v. Lawson, 461 U.S. 352, 357–58 (1983) (striking down as unconstitutionally vague a California statute requiring those found “loiter[ing] or wander[ing] upon the streets or from place to place without apparent reason or business” to provide a “credible and reliable” identification and reason for their presence when requested to do so by a police officer).

43. For a general discussion of the void for vagueness doctrine in criminal cases, see 1 WAYNE R. LAFAVE & AUSTIN W. SCOTT, JR., SUBSTANTIVE CRIMINAL LAW § 2.3 (1986), and Anthony G. Amsterdam, Note, The Void-for-Vagueness Doctrine in the Supreme Court, 109 U. PA. L. REV. 67 (1960).

44. Kolender, 461 U.S. at 357; Papachristou, 405 U.S. at 162; Connally v. Gen. Constr. Co., 269 U.S. 385, 391 (1926) (providing that criminal statutes must define the offense with sufficient clarity for an ordinary person to understand what conduct is prohibited).

45. The importance of fair notice underlies other criminal law rules as well. For example, consider the values reflected in the rule of lenity, providing that ambiguities in criminal statutes should be resolved by construing the statute narrowly. The Court has stated that the rule of lenity reflects in part the principle that “a fair warning should be given to the world . . . of what the law intends to do if a certain line is passed” and that “the line should be clear.” United States v. Bass, 404 U.S. 336, 347–48 (1971) (internal citations omitted).

46. See Kolender, 461 U.S. at 357 (holding that criminal prohibitions must be defined “in a manner that does not encourage arbitrary and discriminatory enforcement”).

47. In Papachristou, for example, the Court likened the perceived virtues of “strolling” and
avenue for the judiciary’s substantive review of legislative decisionmaking. 48

The Court’s decision in City of Chicago v. Morales 49 to strike down Chicago’s anti-gang loitering ordinance has raised the question of whether the void for vagueness doctrine creates an unnecessary barrier to new policing methods. Chicago enacted the ordinance in response to pervasive gang activity in its inner-city neighborhoods. 50 In contrast to rapid-response policing, the ordinance attempted to reduce gang control over neighborhoods proactively through problem-solving policing. 51 Rather than direct police officers to arrest offenders for the serious crimes that often result from gang activity, the ordinance authorized police officers to issue verbal orders to disperse if a group of two or more people were loitering in a public place with no apparent purpose, and if the officer reasonably believed that at least one person within the group was a member of a “criminal street gang.” 52 Violation of the order to disperse was a misdemeanor offense. 53 A six-member majority of the Court struck down Chicago’s anti-gang loitering ordinance as unconstitutionally vague, holding that the Chicago City Council failed to “establish minimal guidelines to govern law enforcement” when it instructed officers to issue dispersal orders to groups loitering with “no apparent purpose.” 54 The void for vagueness doctrine prohibited the Chicago City Council from protecting neighborhoods from pervasive gang intimidation by deferring to the “‘moment-to-moment judgment of the policeman on his beat.’” 55

“loafing” to the values reflected in the writings of Walt Whitman and Henry D. Thoreau, freedoms that have “encouraged lives of high spirits rather than hushed, suffocating silence.” 405 U.S. at 163–64.

48. See Ellickson, supra note 14, at 1210–11 (noting that the Papachristou decision treated the city’s attempt to control the use of public spaces as the “suppression of high spirits in favor of drab conformity”); Livingston, supra note 6, at 621–27 (discussing the void for vagueness doctrine as substantive, not procedural, review).


50. Id. at 46–47.


52. See Morales, 527 U.S. at 47 (summarizing CHI., ILL., MUNICIPAL CODE § 8-4-015 (2002)).

53. Id.

54. Id. at 60–64 (holding that anti-gang loitering ordinance entrusted too much discretion to police officers to determine which types of loitering “purposes” were lawful) (quoting Kolender v. Lawson, 461 U.S. 352, 358 (1983)).

55. Id. at 60 (quoting Kolender, 461 U.S. at 360).
Unpacking New Policing

C. The New Discretion Scholarship

Because the Court relied on the void for vagueness doctrine to strike down one of the first new policing initiatives to reach the Court, the doctrine has become the target of new policing advocates who maintain that police discretion is essential to the enforcement of social norms. Relying heavily on the new policing models, an influential group of scholars has argued that a changed political landscape mitigates the need for judicially-imposed limitations on police discretion. They argue that, where vagrancy laws were once used to marginalize racial minorities, today’s police exercise discretion to the benefit of and often at the behest of inner-city communities. Police discretion, the argument further goes, is not only inevitable, but in fact helpful if exercised to enforce social norms in a way that aids normatively-worthy communities. These “new discretion” scholars advocate a two-tiered criminal justice jurisprudence in which the level of judicial scrutiny of police discretion turns on whether the challenged program constitutes new policing.56

Debra Livingston, for example, asserts that “quality-of-life” offenses with low-level sanctions should not be reviewed under the void for vagueness doctrine.57 She argues that “rule-like,” specific laws are inevitably imperfect in capturing the disorderly conduct that a community seeks to prohibit and, therefore, police will inevitably exercise discretion in deciding against whom to enforce such ordinances.58 She argues:

When public order laws do not appear aimed at the exclusion of groups or individuals from participation in a community’s public life, but rather at the articulation of reasonable behavioral standards in the interest of preserving public spaces, communities should not be prevented from replacing the Papachristou regime with a new legal regime in which more contextualized, conduct-based prohibitions authorize police to perform order maintenance tasks.59

When laws are vague, she argues, police discretion can be managed through other means, such as political accountability, community

56. Cole, supra note 39, at 1062 (coining the term “new discretion” scholarship).
57. Livingston, supra note 6, at 562; see also Ellickson, supra note 14, at 1243–46 (advocating informal zoning of public spaces with respect to disorder and advocating discretionary enforcement in the most orderly zones of vagrancy and disorderly conduct laws).
58. Livingston, supra note 6, at 618, 647–50.
59. Id. at 647.
monitoring, internal enforcement guidelines, and sunset provisions on ordinances that delegate police authority.60

Similarly, Dan Kahan and Tracey Meares have argued that “exacting judicial scrutiny of routine policing functions” and judicial “hostility toward discretion”61 are no longer warranted in contemporary urban crime control.62 They assert that these components of modern criminal procedure jurisprudence were necessary during the Warren Court years when majority groups used selective enforcement of vague criminal statutes as a critical component of the institutionalized racism that subordinated minority groups.63 In contrast to its less benevolent predecessor, discretion invoked in the new policing is to the benefit of—and often at the behest of—minority groups in urban communities that are exercising increasing political power.64 Kahan and Meares argue that the new policing enhances liberty in inner-city neighborhoods by altering destructive and constraining norm perceptions, such as the expectation that a young man should carry a gun or join a gang,65 and by providing a less draconian law enforcement alternative to long prison sentences.66 They also argue that members of inner-city communities are “practically and morally” in a better position than civil libertarians and judges to strike the proper balance between liberty and order.67

Kahan and Meares argue that the modern regime of criminal procedure should be abandoned in favor of a model that considers the emergence of African-American political strength.68 Specifically, they

60. Id. at 650–70.

61. Kahan & Meares, supra note 39, at 1158–59 (noting that two central features of the modern criminal procedure regime are “its authorization of exacting judicial scrutiny of routine policing functions” and “its hostility toward discretion”).

62. Id. at 1166–71.

63. Id.

64. Id. at 1163–64 (noting the support for community policing among African-Americans and maintaining that the “new community policing is an outgrowth” of demand within African-American communities for increased protections of law enforcement).

65. Id. at 1168–69, 1181–82; see also Lawrence Rosenthal, Gang Loitering and Race, 91 J. CRIM. L. & CRIMINOLOGY 99, 129 (2000) (arguing that gang influence on “community mores” undermines the inner cities’ chances for revitalization). For further discussion of the reinforcement of social norms to stabilize neighborhoods, see Kahan, supra note 13, at 367–77, Livingston, supra note 6, at 578–84, and Waldeck, supra note 13, at 1299–308.

66. Kahan & Meares, supra note 39, at 1169; see also Rosenthal, supra note 36, at 58; Schragger, supra note 8, at 441; William J. Stuntz, Race, Class, and Drugs, 98 COLUM. L. REV. 1795, 1836–38 (1998).


68. Id. at 1171.
Unpacking New Policing

advocate a political process theory of criminal procedure jurisprudence, in which courts would defer to the political process and uphold the delegation of discretion to police as long as the burden of the police procedures falls on the average member of the community or those with a “linked fate” to the governing majority. Applying this political process theory, Kahan and Meares would permit some community-oriented programs to avoid not only the prohibition against vague laws, but also other rules limiting police discretion, such as the Fourth Amendment’s general requirement that searches be conducted pursuant to a warrant based on probable cause.

Randall Kennedy shares Kahan and Meares’s belief that traditional liberal concern about the effects of law enforcement on African-American defendants is misplaced in light of the disproportionate effect of crime on African-American victims. Because the African-American community suffers the most from the commission of criminal offenses, he argues, the community’s problem with the criminal law is not its over-enforcement, but its under-enforcement.

Robert Ellickson shares many of the same concerns as Livingston, Kahan, and Meares, but approaches urban quality-of-life problems from a zoning perspective. He argues that communities could be zoned pursuant to their tolerance for disorder. In the most orderly zones, prohibitions against panhandling and loitering would be enforced, thereby relegating such activity to the less orderly zones. Like Livingston, Kahan, and Meares, Ellickson criticizes judicial

69. Id. at 1171–75.
70. Id.
72. Id.; see also Rosenthal, supra note 65, at 130 (emphasizing the need for enhanced policing as part of inner-city revitalization). Like Kahan and Meares, Kennedy believes that victimization by crime is now a larger burden on the everyday lives of African-Americans than the mistreatment of suspects and criminal defendants. KENNEDY, supra note 71, at 19. Although Kennedy shares Kahan and Meares’s belief that African-Americans benefit from enforcement of criminal law, he does not share their view on one of the first new policing efforts to be reviewed by the U.S. Supreme Court. While Kahan and Meares worked to support Chicago’s gang loitering ordinance, invalidated as vague in Morales, Kennedy co-authored an amicus brief opposing the ordinance. See Brief of Amici Curiae Chi. Alliance for Neighborhood Safety et al. in Support of Respondents, City of Chicago v. Morales, 527 U.S. 41 (1999) (No. 97-1121); see also Cole, supra note 39, at 1069 & n.59 (observing that while Kennedy has not directly assailed the vagueness doctrine, Kahan and Meares have relied critically on his claim that black communities need more, not less, law enforcement).
73. Ellickson, supra note 14, at 1219–46.
74. Id. at 1219–26.
75. Id. at 1221–22.
decisionmaking during the late 1960s and early 1970s, including the void for vagueness doctrine, to the extent that the Court “seemed blind to the fact that their constitutional rulings might adversely affect the quality of urban life and the viability of city centers.”\textsuperscript{76} He maintains that federal courts should refrain from restricting the authority of state and local legislative bodies to craft laws responding to local conditions.\textsuperscript{77}

Mark Rosen emphasizes geographic boundaries from a different perspective, arguing for increased judicial tolerance of self-governance through the creation of geographic nonuniformity of constitutional requirements.\textsuperscript{78} In Rosen’s view, courts should apply the void for vagueness doctrine contextually for at least some communities. For example, he maintains that the Morales Court, instead of pondering the possibility that the anti-gang loitering ordinance would be applied to innocent people waiting for a taxi or resting from a jog, could have inquired whether the ordinance “was sufficiently definite to persons living in a city thick with street gangs.”\textsuperscript{79}

Underlying the new discretion scholars’ willingness to relax constitutional doctrines that they perceive to be barriers to new policing is their shared belief that community groups support and play an integral part in the new policing models. Kahan and Meares, for example, emphasize the popularity of the new policing efforts within inner-city communities and argue that the support of racial minorities demonstrates an even-handed application of the police initiatives.\textsuperscript{80} Livingston similarly relies on community support when she defends many of the public order laws at the heart of the new policing models by observing that they do not aim to exclude outsiders from “participation in a community’s public life.”\textsuperscript{81}

\textsuperscript{76} Id. at 1213.
\textsuperscript{77} Id. at 1213–14.
\textsuperscript{79} Id. at 1174.
\textsuperscript{80} Kahan & Meares, supra note 39, at 1160–66; see also Dan M. Kahan, Privatizing Criminal Law: Strategies for Private Norm Enforcement in the Inner City, 46 UCLA L. REV. 1859, 1863–65 (1999) (noting that government can gain legitimacy by working in partnership with private organizations respected within the community). Although Kennedy has not challenged the vagueness doctrine, he appears to share some of Kahan and Meares’s views about the desire of African-Americans for increased police protection. See Kennedy, supra note 71, at 19 (arguing that African-Americans desire more, not less, enforcement of criminal laws); see also Rosenthal, supra note 36, at 56 (arguing that traditional law enforcement leaves the poor and vulnerable subject to a higher risk of crime than the wealthy and powerful).
\textsuperscript{81} Livingston, supra note 6, at 647.
Unpacking New Policing

New discretion scholars assume not only community support, but also community involvement in the new policing. 82 Livingston, for example, argues that internal police regulations and community monitoring of the police department can be more effective checks on police discretion than the void for vagueness doctrine. 83 Similarly, Sarah Waldeck emphasizes the importance of “partnerships” between police and community as an effective method of ensuring that police do not hide attempts to increase felony arrests beneath the cover of enforcing community norms. 84

II. CRITIQUING THE NEW DISCRETION SCHOLARSHIP

New discretion scholars have brought a much-needed shift to criminal law scholarship. Whereas criminal law scholars have traditionally criticized the impact of judicial decisionmaking on offenders, the new discretion scholars are concerned also about the consequences to victimized inner-city residents. 85 They recognize that the traditional hostility of inner-city residents toward police arises not just from police aggression toward their community, but also from the manifestation among law enforcement of low expectations about the quality of life in poor neighborhoods and increased tolerance for illegal activity against minority residents. 86 The new discretion scholars rightly enlarge the

82. See generally Randolph M. Grinc, “Angels in Marble”: Problems in Stimulating Community Involvement in Community Policing, 40 CRIME & DELINQUENCY 437, 440–42 (1994) (discussing why community participation is thought to be important in the new policing models). Grinc notes that law enforcement seeks the participation of the community to enhance the community’s perception that police are responding to its concerns, to increase the community’s sense of safety, to improve community-police relations, and to decrease crime. Id. at 440–41. Grinc also observes that most theorists stress the importance of community involvement even though they disagree about why it is important. Id. at 440.

83. Livingston, supra note 6, at 659–67 (discussing internal regulation and civilian monitoring of police discretion); see also Heymann, supra note 8, at 454–55 (maintaining that officer discretion that is unlikely to be monitored judicially should be governed by departmental regulations).

84. Waldeck, supra note 13, at 1301. For example, Waldeck praises the Day Labor Project in Glendale, California, where police work with activists, businesses, and social service providers to create a central site for day laborers as an alternative to quality-of-life policing in a neighborhood affected by the gathering of day laborers. Id. at 1302–04. Similarly, she discusses partnerships between police and schools to combat truancy and juvenile crime. Id. at 1304–06.

85. As William Stuntz has recently noted, “everyone does and should care about both” police overreaching and victimization by private parties, because “the state of nature—a world free of all risks of police coercion—is an unhappy place in many of the same ways that a police state is an unhappy place.” William J. Stuntz, Local Policing After the Terror, 111 YALE L.J. 2137, 2146 (2002) (arguing that the scope of constitutional rights does and should change in response to broad crime trends).

86. This view dates back to one of the earliest attempts to understand the relationship between
concept of liberty beyond traditional civil libertarian notions to include a basic sense of safety that privileged Americans take for granted. In doing so, they recognize that a model government does more than simply refrain from interference with the citizenry; it plays a role in guaranteeing its safety.

Unfortunately, the new discretion scholarship reflects its concern for real-world quality of life by concluding that community involvement in law enforcement obviates the need for traditional constitutional limitations upon police. That conclusion is flawed in several respects. First, it gives constitutional significance to notions of community that are often idealized and easily manipulated. Second, it assumes that a community’s support for and participation in a challenged police initiative can be quantified accurately and will not fluctuate significantly over time. A third flaw is the new discretion scholarship’s emphasis upon majority rule, which would deprive political minorities of substantive judicial review of rights determinations resulting from the political process. A fourth flaw is the failure to respect the values served by traditional vagueness review. Finally, the reliance upon community support as a substitute for the void for vagueness doctrine would create a potentially limitless tolerance for police discretion. The remainder of this Part discusses each of those flaws in turn.

A. Idealizing Notions of “Community”

New discretion scholars’ reliance on malleable notions of community support and involvement to justify radical jurisprudential paradigm shifts is troubling for several reasons. As a threshold matter, it is not obvious that residents of a geographically-defined region comprise a “community” in anything other than the most superficial sense. Granted, in inner-city neighborhoods, one’s address shares a high correlation with one’s income and skin color, and at least some of the police and minority communities. See, e.g., REPORT OF THE NATIONAL ADVISORY COMMISSION ON CIVIL DISORDERS 161–62 (1968). The commission, popularly known as the Kerner Commission, was established after the 1967 summer riots to study factors that contributed to the riots and to suggest means of preventing reoccurrences. Id. at 1. The commission concluded that the riots commenced in part because of antagonism between police and inner-city neighborhoods, attributable not only to aggressive police tactics within those neighborhoods, but also to police apathy about crimes against those neighborhoods. Id. at 157–62; see also Livingston, supra note 6, at 571 (discussing Kerner Commission).

87. See infra Part II.E for further discussion regarding the difficulty of defining communities.
88. Richard Ford, for example, has explained how the continuing correlation between race and income contributes to the maintenance of “racially identified spaces.” See Richard Thompson Ford,
new discretion scholarship relies upon the notion of a meaningful community among African-Americans. 89 Nevertheless, to assume the existence of an identifiable community view within inner-city neighborhoods is to ignore both the presence of other racial groups and recent immigrants in those neighborhoods and the tensions that can exist between them and African-Americans. 90 It also ignores the reality that there is diversity of opinion among African-Americans regarding crime, police, and competing policy priorities.91

Some new discretion scholars would argue that, even in the absence of a community consensus, it is inner-city residents’ improved opportunity to participate fairly in the political process that is critical in the new policing efforts.92 Although representative forms of government undoubtedly play a role in determining the “will” of the community, a political process account of constitutional rules ignores the reality that every community, however defined, has its outsiders “whose complaints are least likely to be heard by the rest of the community.”93

Moreover, even if there were something resembling an empirically identifiable consensus within a meaningfully defined community, it is
less than clear that new policing efforts actually identify that consensus. Community policing efforts, for example, frequently look to neighborhood organizations for support and partnerships. However, the limited empirical research about community organizations indicates that these groups reflect only a small proportion of residents and tend to be dominated by homeowners and white residents in racially-mixed neighborhoods. 94

Indeed, one of the few empirical studies of community participation in community policing projects found nearly universal difficulties stimulating and continuing citizen participation in community policing projects, even though the projects were markedly different in their approaches. 95 Studying eight early pilot community policing projects, Randolph Grinc reported that “ordinary” residents within the areas affected by the pilot programs usually had either no knowledge at all of the programs or a vague idea that there was a new government program operating in their neighborhoods. 96 Many tended to see the programs as opportunities for neighborhood picnics and other social gatherings. 97 Even among leaders of neighborhood groups, who demonstrated the highest level of knowledge about the programs, few appeared to understand fully the role of the community in the programs. 98 Importantly, Grinc reported that only a “small core group of residents” was involved in the programs. 99

When only a small percentage of residents participate in “partnerships” with police, there is no guarantee that these squeaky

94. See, e.g., Michael E. Buerger, A Tale of Two Targets: Limitations of Community Anticrime Actions, in COMMUNITY JUSTICE: AN EMERGING FIELD, supra note 15, at 137–38 (noting that membership in community crime-prevention groups is often low and tends to be dominated by homeowners and by white residents in racially diverse neighborhoods); Skogan, supra note 18, at 68 (concluding that community organization is more likely to occur in “homogeneous, better-off areas of cities”).

95. See Grinc, supra note 82, at 442–64 (surveying data from eight pilot Innovative Neighborhood-Oriented Policing programs funded by the Bureau of Justice Assistance).

96. Id. at 442–44.

97. Id. at 444.

98. Id. at 456.

99. Id. at 445. Grinc noted shared hurdles to police partnerships with community across the programs. In addition to the community’s lack of understanding about the programs, Grinc noted problems presented by preexisting hostilities between police and the members of poor and minority communities now being asked to participate, residents’ fear of retaliation by neighborhood gangs and drug-dealers, community distrust created by previous, short-lived, failed experiments in their neighborhoods, police frustration with what they perceive as apathy within the community, and diverse interests and infighting within the community. Id. at 446–60.
Unpacking New Policing

wheels represent the will of the broader community. They may very well be a vocal minority that happens to share law enforcement’s priorities. The following anecdote demonstrates the risks of placing too much emphasis on the perception of community sentiment. From 1997 to 1999, I served as a member of the Multnomah County District Attorney’s Neighborhood District Attorney program, a nationally-recognized program in Portland, Oregon, at the forefront of “community prosecution.” Although listening to community groups was a major component of our agenda, we were not above massaging public perception of the community. For example, “Richard” was an African-American retiree who could be counted on to vocalize a predictably pro-police stance at critical city council hearings and other decisionmaking sessions. Preparing for such meetings always involved a phone call to Richard to ensure his participation. We respected Richard for his views and his activism, but we appreciated the political reality that his voice would go further with a city council struggling to gauge the desires of inner-city residents than the voices of white liberals opposing police measures.

It is not surprising, therefore, that the term “community” has been called “imprecise” and potentially “idealized.” The political popularity of community-oriented programs already provides a convenient rhetoric that could potentially mask aggressive police initiatives that run counter to the concept at the heart of true community policing—improved relations between the police and the rest of the community. A suburban police chief once confided in me that he was willing to go along with some forms of community policing in light of its popularity and the availability of grant funding for community-

100. See Thompson, supra note 15, at 365 n.171 (touting Portland’s Neighborhood District Attorney program as one of three “excellent community prosecution efforts that have benefited from self-reflection, and from evaluation and adjustment”); see also Boland (Portland), supra note 15, at 253–77 (discussing Portland’s Neighborhood District Attorney program); Boland (NIJ), supra note 15, at 35–40 (same).


102. Livingston, supra note 6, at 577.

103. Academics have warned “that a bewildering and sometimes inappropriate variety of police initiatives could well be implemented in community policing’s name.” Livingston, supra note 6, at 578 (collecting cites); see also Law and Disorder: Is Effective Law Enforcement Inconsistent with Good Police-Community Relations?, 28 FORDHAM URB. L.J. 363, 366 (2000) (comments of Paul Chevigny) (“So-called community policing that does not mean participation by the people isn’t really community policing.”); Robert Weisberg, Foreword: A New Agenda for Criminal Procedure, 2 BUFF. CRIM. L. REV. 367, 370 (1999) (noting that a “somewhat sentimental notion of ‘community’ norms masks a dangerously majoritarian anti-Constitutionalism”).

1007
oriented programs, but that his version of community policing simply meant that “we police the community.”

The problem opposing an initiative labeled as community-oriented, of course, is, as Professor Schragger has stated, “No one can be against community.” George W. Bush, for example, has come under fire for proposing drastic cuts to the budget of the office of Community Oriented Policing Services (COPS), the federal program developed under President Clinton to promote community-oriented policing. Rather than terminate the program entirely, Bush’s 2004 request retains the program in name only with a proposal to reduce spending from $1.4 billion in 2003 to $164 million in 2004. Even before the budget-cut proposal, the administration had shifted the focus of the program toward the provision of technological assistance to local police departments and placing officers in public schools, nevertheless retaining the community-oriented label. In light of the current potential for the community policing trend to amount at least in some programs to politically popular rhetoric, scholars should be reluctant to permit constitutional principles to turn on the supposed involvement and support of the “community.”

104. The availability of federal grant money for projects considered to be community-based provides an incentive for cities to engage in creative labeling of programs. Heymann notes that Chicago and New York City adopted markedly different approaches, but both call them “community policing,” a label that Heymann notes is often required in order to obtain federal grant money that is ultimately used to increase the number of officers within the department. Heymann, supra note 8, at 422.

105. Schragger, supra note 8, at 403.


108. See Gest, supra note 10, at 762 (discussing the direction of COPS under the Bush administration).

109. Professors Alschuler and Schulhofer have warned about the need “to be on guard against the appealing but highly manipulable rhetoric of ‘community,’ a rhetoric that is increasingly prevalent in contemporary discourse.” Albert W. Alschuler & Stephen J. Schulhofer, Antiquated Procedures or Bedrock Rights?: A Response to Professors Meares and Kahan, 1998 U. CHI. LEGAL F. 215, 216 (1998); see also supra notes 100–08 and accompanying text (discussing potential of community rhetoric to mask more traditional, and even aggressive, forms of policing).
Unpacking New Policing

B. Determining Constitutional Parameters by Community Sentiment

Even if there were general consensuses within communities about policing efforts, and even if those consensuses were accurately measured, a second problem with the new discretion scholarship is its assumption that communities should be permitted to determine their own minimum protections from government intrusion. There are two problems with this assumption. First, it can be difficult to determine whether the articulated consensus of the community truly reflects the desires of the residents. Second, even if a community’s true desire is to submit to a particular police practice as an alternative to systemic neighborhood crime problems, the Constitution should provide minimum protections that cannot be waived by the political process.

1. Assessing Community Support for the New Policing

As an initial matter, claims that inner-city communities favor the new policing models should be viewed cautiously, especially in light of the limited alternatives for which those communities may feasibly opt. Programs seemingly supported by an inner-city community may very well be viewed by the community as the lesser of policy evils in a world where politically feasible alternatives are limited by the broader majority. Kahan and Meares, for example, observe that African-Americans favor order-maintenance policing because it is less harmful than other forms of law enforcement, such as long prison sentences for drug offenders. Even if their claim is empirically correct, inner-city communities do not—even with the increased political power that Kahan and Meares attribute to them—have the ability to shape the governance of their communities as they might truly see fit.

Rather, their options—at least those that are realistically obtainable—are limited by the broader majority whose values shape the quality of public schools, banks’ lending policies, the availability of job training, and so forth.

110. Some have questioned the support for new policing models in inner-city communities, even as a choice among limited and unfavorable options. See Alschuler & Schulhofer, supra note 109, at 217–20.

111. Of course, even if the new discretion scholars are correct in their assertion that African-Americans are better represented in the political process than they were a few decades ago, they still remain a minority in the vast majority of political districts.

112. Kahan & Meares, supra note 39, at 1165. Cole has criticized Kahan and Meares for failing to provide any empirical support for their assertion that African-Americans favor order-maintenance policing. See Cole, supra note 39, at 1085 & n.142.
and the general economic composition of inner-city neighborhoods. The new policing models might give the community increased participation within the narrow sphere of law enforcement, where discretionary responses to low-level disorder might be preferred over draconian responses to more serious offenses. However, the community’s true desire might be the less realistic alternative of diverting money from law enforcement in order to improve educational and job opportunities.113

Moreover, by the time a public safety concept emerges as a concrete policy proposal, it can be difficult to determine whether individual community groups played a critical role in the development of the proposal. Consider, for example, Chicago’s anti-gang loitering ordinance. Although inner-city residents may have been responsible for identifying the mitigation of gang activity as a city priority, it is at best unclear whether the community actually gave birth to the anti-gang loitering ordinance as a response to their concerns. Alschuler and Schulhofer maintain that the ordinance was drafted by a white alderman with cooperation from a predominantly white neighborhood association and the city’s attorneys.114 Kahan and Meares, on the other hand, maintain that minority residents not only provided the impetus for the ordinance, but also played a critical role in drafting it.115 In short, it is unclear whether residents viewed the anti-gang loitering ordinance as an ideal solution or whether they supported it only after it was drafted as an alternative to crime in their neighborhoods.116

As a community prosecutor, I played my role several times in the familiar process of proposing a new initiative. Generally, residents would identify their problems, then community police and prosecutors would devise a potential solution, using the community’s complaints about the status quo as political support for the proposal. Invariably, residents would prefer something else, usually expensive, infeasible solutions like deterrence through round-the-clock police presence. Community support for the only proposed alternative to a crime-ridden

113. See Cole, supra note 39, at 1088 (observing that inner-city residents might prefer expensive alternatives that the larger community is unwilling to pay for); Erik G. Luna, The Models of Criminal Procedure, 2 BUFF. CRIM. L. REV. 389, 453 (1999) (maintaining that “[i]inner-city minorities have opted for discretionary policing techniques not on the merits but because society at large refuses to provide adequate resources to safeguard urban communities”).


116. Id.
Unpacking New Policing

neighborhood does not necessarily amount either to full community support or to full community participation in the decisionmaking process.

Furthermore, it is less than obvious that residents fully understand what they are getting when they sign on to support police programs. For example, African-American communities might support curfews and anti-gang loitering ordinances, believing that these laws provide a method to intervene in their youths’ criminality before it becomes more serious.117 However, because few lay people are well-versed in the often counter-intuitive parameters of criminal procedure, residents may not understand that enforcement of low-level criminal prohibitions against quality of life offenses can lead to the very type of traditional law enforcement that they disfavor. Increased police-citizen interactions on the street lead to more frisks,118 which may reveal drugs or weapons that trigger lengthy prison sentences. Similarly, a custodial arrest, even for the most trivial offense, triggers the broad power to search incident to arrest.119 Finally, even if a low-level misdemeanor arrest does not lead to prosecution for a more serious offense, rampant arrests and convictions within a community are stigmatizing120 and can undermine the community’s long-term relationship with police.121

117. Kahan & Meares, supra note 39, at 1169.
118. See Terry v. Ohio, 392 U.S. 1, 21–27 (1968) (authorizing police to stop individuals based on reasonable suspicion that criminal activity has occurred or is afoot and to frisk stopped individuals for weapons based on reasonable suspicion that the person is armed and presently dangerous).
119. The lawful custodial arrest of an individual authorizes police to search the person and any area within her immediate control. See Chimel v. California, 395 U.S. 752, 768 (1969) (delineating the spatial scope of the search incident to arrest exception to the probable cause and warrant requirements of the Fourth Amendment). When an individual is arrested from a vehicle, the vehicle’s passenger compartment is considered within the arrested individual’s immediate control and therefore can be searched pursuant to the search incident to arrest power. See New York v. Belton, 453 U.S. 454, 460 (1981) (providing a bright-line rule that the passenger compartment of a vehicle falls within the scope of a search incident to arrest of one of the vehicle’s passengers). The search incident to arrest authority applies even if the police cannot articulate probable cause, or even reasonable suspicion, to believe that the arrested person is carrying a weapon, contraband, or items of evidentiary value. See United States v. Robinson, 414 U.S. 218, 235 (1973) (holding that the search incident to arrest power applies automatically upon arrest). Moreover, police can initiate custodial arrests for the most minor criminal offenses, even those that do not trigger a potential sentence of imprisonment upon conviction. See Atwater v. City of Lago Vista, 532 U.S. 318, 354 (2001) (holding that custodial arrest for failure to use a seatbelt did not violate the Fourth Amendment).
120. Critics of order-maintenance policing have argued that widespread enforcement of low-level prohibitions stigmatizes those targeted by the crackdown on disorder, who are frequently minorities. Accordingly, the policing efforts may contribute to stereotypes of criminality within the communities affected by order-maintenance policing and may hinder attempts to establish non-
Moreover, community sentiment can be molded. As a Neighborhood District Attorney in Portland, I frequently “sold” proposed programs to both the communities who would be subject to them and the police officers who would implement them. It was not unusual to use different talking points for the two audiences. For example, selling the enforcement of a local curfew ordinance to community groups would generally entail an emphasis on protecting children. It was better for a minor to be removed from the streets by police, we said, than by a dangerous predator. We also emphasized the strong likelihood that police would contact the juvenile’s parents to retrieve their child from the police department before the child would be transported to the juvenile detention center. Seeking the community’s support in advance was critical, in our view, not so much to shape the end-product substantively, but for strategic purposes. Community cooperation legitimized the ultimate proposal and could prevent a political confrontation down the road with interested parties who felt excluded from the decisionmaking process.

Selling police officers on the concept of curfew enforcement entailed a strikingly different rhetoric. Rather than emphasizing the benefits that inured to children, we stressed the ability of police to use their discretion to take custody of a child at night based merely upon the child’s presence in a public place. To police, curfew enforcement was an...
Unpacking New Policing

opportunity to search and seize local youths whom police continually suspected of residential burglaries and drug crimes, even if the police lacked probable cause to do so. Because the message behind the new policing programs can be spun as necessary to garner support, communities may support new policing programs without fully understanding every aspect of them, including their larger implications under current Fourth Amendment jurisprudence.

2. Permitting Communities To Waive Constitutional Protections

Perhaps more importantly, even if the majority of citizens truly desire to subject themselves to intrusive government programs to protect their security, the federal Constitution should provide at least some limitation on their ability to bind the entire citizenry through the political process. The aftermath of September 11 demonstrates the difficulties of preserving the value of abstract constitutional liberties in the face of immediate fears about individual and collective security. In the first months following the attacks on the World Trade Center and Pentagon, public sentiment was high that individual rights should defer to the government’s war on terrorism; however, as time passes, the public’s concern about civil liberties has increased, even as the country remains concerned about “homeland security.” Although judicial scrutiny of is used to “win over an entire organization to a concern for disorder”).

125. The role that pretextual stops and arrests play in order-maintenance policing cannot be overlooked. The U.S. Supreme Court has made clear that a seizure or search is lawful as long as there is an objective basis to support it, even though the individual police officer initiating the seizure or search used that objective basis as a pretext to mask a subjective purpose for which he lacked lawful justification. Whren v. United States, 517 U.S. 806, 812–19 (1996). As lawmakers authorize police to arrest for offenses further attenuated from immediate harms, there is an increased likelihood that police will invoke this authority to justify a search of the person on the chance that they will discover evidence of a more serious offense.

126. For a more thorough discussion of the “value conflicts” between police and the groups and individuals with whom they “partner” in community policing enterprises, see Thacher, supra note 8, at 766–71. I raise that issue here for the more limited purpose of establishing that the community and police may support a particular program for different reasons, and that the program may not always serve the community’s values to the extent that the community is led to believe.

127. Luna, supra note 113, at 452 (criticizing Kahan and Meares’s political process model for turning individual rights over to the political majority).

128. For example, in January 2002, three months after the attacks in Manhattan and at the Pentagon, 47% of Americans said that the government should take steps to prevent terrorism, even if those steps required violations of basic civil liberties. Joan Biskupic, Attention Turns Back to Liberty, USA TODAY, Nov. 1, 2002, at 17A (summarizing data from USA Today/CNN/Gallup polling conducted in January and September of 2002). By September 2002, only one-third of Americans expressed that view. Id.
government action can take into account the extent of the need for the action, the governing constitutional first principles should not depend on the public’s current perception of the need to place security over traditional concepts of individual liberty.  

In that context, consider Kahan and Meares’s argument that inner-city residents’ support for the new policing measures “reflects their judgment that in today’s political and social context, the continued victimization of minorities at hands of criminals poses a much more significant threat to the well-being of minorities than does the risk of arbitrary mistreatment at the hands of the police.” To advocate that the judiciary defer to that judgment is no different than arguing, post-September 11, for judicial deference to the broader majority’s desire to place homeland security over civil liberties.

Just as many Americans have been willing to exchange civil liberties for a chance at increased security, at least some inner-city communities have been long willing to make that same trade at the neighborhood level. As Maya Angelou has said about the day that is thought to have changed the way Americans think about their security, “Living in a

129. Current constitutional jurisprudence makes room for the judiciary to evaluate the need for the challenged government conduct. For example, the warrant requirement of the Fourth Amendment does not apply if the police act under exigent circumstances presented by the risk of destruction of evidence, the escape of a suspect, or danger to the public. See Warden, Md. Penitentiary v. Hayden, 387 U.S. 294, 298–99 (1967) (holding that police were entitled to enter the suspect’s home without a warrant because they entered in hot pursuit of the fleeing suspect). Moreover, several rule-based exceptions to the probable cause and warrant requirements are grounded in the legitimate needs of law enforcement. See United States v. Robinson, 414 U.S. 218, 234–35 (1973) (grounding the search incident to arrest power upon the exigencies that apply in an arrest situation); Terry v. Ohio, 392 U.S. 1, 21–27 (1968) (approving minimally intrusive stops and frisks based on mere reasonable suspicion in light of the needs of patrol officers on the street); Carroll v. United States, 267 U.S. 132, 154–56 (1925) (permitting searches of automobiles based on probable cause but without a warrant in light of the need to prevent the automobile from leaving the scene before the search is conducted). The Court’s “special needs” doctrine also leaves room to consider the need for warrantless searches. See infra note 354 and accompanying text for an overview of the special needs doctrine.

130. William Stuntz, however, has argued that the scope of constitutional rights should be—and is—responsive to changes in the public’s prioritization of crime concerns, and he has suggested ways in which the events of September 11 may affect criminal procedural rules. See Stuntz, supra note 85, at 2144–90. However, Stuntz appears to refer to the law reflecting broad social changes, not the particular desires of an individual neighborhood or even short-lived changes in the priorities of the collective citizenry. See, e.g., id. at 2155–57 (discussing the tendency of courts to be influenced by large social trends over time and suggesting that September 11, even though only a one-day salient event, may influence judicial decisionmaking because of its indication of an ongoing threat).


132. September 11, 2001, is widely seen as the first day since Pearl Harbor that Americans felt
Unpacking New Policing

state of terror was new to many white people in America, but black people have been living in a state of terror in this country for more than 400 years.”

In light of the widespread desire after September 11 to set aside civil liberties so that we can all “feel safe again,” it should come as no surprise that African-American communities in inner-city neighborhoods—“terrorized” regularly by random street robberies, drive-by shootings, and drug- and gang-related violence—were willing to make that compromise long ago. If the new discretion scholarship entrusts African-American communities to strike that compromise through local political processes, it presumably would permit the broader majority to set aside traditional liberties in favor of homeland security.

C. Political Process: An Inadequate Safeguard

Another flaw in the new discretion scholarship is its reliance on the political process to ensure constitutional rights. Although African-Americans have more political power and representation than they did during the Warren Court years, they still do not have the same representation as white voters. Indeed, the disproportionate representation of African-Americans at every phase of the criminal justice process indicates that the increased political power of African-Americans has not been sufficient to eradicate the adverse impact of law enforcement on African-American communities.

vulnerable. See, e.g., Kenneth Chang & Andrew Pollack, On Many Fronts, Experts Plan for the Unthinkable: Biowarfare, N.Y. TIMES, Oct. 23, 2001, at F1 (stating that it was not until September 11 that “it collectively occurred to Americans how vulnerable they were”); Julia Keller, The Doctor—As in Kissinger—Is Still In, CHI. TRIB., Oct. 19, 2001, at C1 (quoting Henry Kissinger as stating that September 11 brought “a change in the consciousness of people” because “the sense of vulnerability is now a fact” instead of “something you never used to think about”).


134. Biskupic, supra note 128, at 17A (“Rattled by the Sept. 11 terrorist attacks, Americans rallied around the flag and accepted the idea that the government needed to take extraordinary measures to make people feel safe again.”); David Winston, It’s a Show of Strength, Not a Political Ploy, THE WASH. POST, Sept. 22, 2002, at B1 (noting that, in light of September 11, Americans support military action against Iraq because “[t]hey want to feel safe again”).

135. Cole, supra note 39, at 1080–81. Moreover, because of Kahan and Meares’s emphasis on the black-white dichotomy, it is unclear how their model would apply to neighborhoods with significant percentages of recent immigrants, who do not have the kind of political power that Kahan and Meares ascribe to African-Americans.

136. For a powerful critique of the new discretion scholars for failing to consider the adverse
Moreover, even if one were confident that African-Americans had adequate access to and opportunities within the political process, the new discretion scholarship incorrectly assumes that the only role of constitutional law in the criminal justice system is to ensure that police practices apply across the governed equally. Some constitutional guarantees protect the individual only comparatively, but others protect citizens absolutely. This contrast is most familiar in the context of the Fourteenth Amendment, where the equal protection clause is concerned only with disparities in a state’s treatment of similarly situated individuals, whereas the due process clause provides certain absolute protections to the individual from the state.  

A similar distinction can be drawn between constitutional rights afforded to criminal suspects. Some rights of criminal procedure are only comparative. For example, a state need not provide any appeal at all to criminal defendants; if, however, a state does create appellate procedures, it cannot arbitrarily deny those procedures to indigents while affording them to the affluent.138 “The question is not one of absolutes, but one of degrees.”139

However, most rights of criminal procedure are absolute, not merely comparative. States could not, for example, evenhandedly prohibit all suspects from having counsel during custodial interrogation, because the Fifth Amendment’s right to counsel is absolute, not comparative.140 Similarly, in drafting the Fourth Amendment, the Framers did not create “a novel ‘evenhandedness’ requirement;”141 rather, they generally limited the government’s ability to engage in searches and seizures.

impact of law enforcement on African-Americans, see Cole, supra note 39, at 1074–79.

137. See, e.g., Ross v. Moffitt, 417 U.S. 600, 609 (1974) (“‘Due process’ emphasizes fairness between the State and the individual dealing with the State, regardless of how other individuals in the same situation may be treated. ‘Equal protection,’ on the other hand, emphasizes disparity in treatment by a State between classes of individuals whose situations are arguably indistinguishable.”).

138. Id. at 606–07. Applying that rule, the Court has held that an indigent defendant has a right to appointed counsel for any first appeal of right created by the state, but does not have such a right for discretionary appeals. Id. at 616–18 (refusing to extend right to counsel on first appeals of right to discretionary appeals); Douglas v. California, 372 U.S. 353, 355 (1963) (holding that states must provide counsel to indigents for first appeals of right).

139. Ross, 417 U.S. at 612.

140. Miranda v. Arizona, 384 U.S. 436, 471–73 (1966) (holding that states cannot use statements obtained during custodial interrogation unless defendant is first advised of and waives the right to remain silent and to have counsel present).

Unpacking New Policing

absent probable cause and a warrant.\textsuperscript{142} If the only purpose of the Fourth Amendment were to invalidate discriminatory searches and seizures, then communities could vote to allow suspicionless home searches to find evidence of criminal activity, as long as the police searched all homes and not just some.\textsuperscript{143} Although a political process approach is useful in applying comparative rights,\textsuperscript{144} the separate rules of criminal procedure do more than protect against discrimination. In short, if the political process were sufficient to constrain governmental conduct, courts would not regularly interpret the federal Constitution to limit governmental conduct.

D. The Virtues of Vagueness Review

Another flaw in the new discretion scholarship is its seeming indifference to the virtue of notice protected by the void for vagueness

\textsuperscript{142} For example, one of the few areas of Fourth Amendment jurisprudence where the evenhandedness of a search can substitute for the usual requirement of probable cause is the “special needs” doctrine, where a search advances some governmental concern other than ordinary law enforcement. See infra note 354 and accompanying text for an overview of the special needs doctrine. Even in that limited context, the notion that government intrusions are acceptable as long as they are not discriminatory has faced resistance. See 

\textit{Vernonia}, 515 U.S. at 668–76 (O’Connor, J., dissenting) (criticizing majority’s reliance upon evenhandedness in upholding random drug-testing of school athletes, where the Court found a “special need” justifying a departure from the usual probable cause and warrant requirements); 

\textit{Delaware v. Prouse}, 440 U.S. 648, 664 (1979) (Rehnquist, J., dissenting) (dissenting from majority’s holding that police officer’s discretionary stop of driver was unlawful and suggestion of nondiscretionary roadblocks as a solution, and noting that the majority had “elevate[d] the adage ‘miserable loves company’ to a novel role in Fourth Amendment jurisprudence”).

\textsuperscript{143} See 

\textit{City of Indianapolis v. Edmond}, 531 U.S. 32, 45–47 (2000) (striking down drug interdiction checkpoint, even though it was administered evenhandedly, because it did not fall within the “special needs” exception to usual Fourth Amendment requirements); 

\textit{Vernonia}, 515 U.S. at 673 (O’Connor, J., dissenting) (noting that, despite the majority’s emphasis on evenhandedness in the special needs context for suspicionless searches, “it remains the law that the police cannot, say, subject to drug testing every person entering or leaving a certain drug-ridden neighborhood in order to find evidence of crime”).

\textsuperscript{144} See 

\textit{United States v. Carolene Prods. Co.}, 304 U.S. 144, 152–53 n.4 (1938); 


\textit{Caroline Products}, calls for the protection of procedural rather than substantive rights in order to avoid the judiciary’s imposition of its own value judgments on society. See 

\textit{Luna}, supra note 113, at 443 (discussing Ely’s theory); see also 

\textit{Erik G. Luna, Sovereignty and Suspicion}, 48 DUKE L.J. 787, 812 (1999) (describing Kahan and Meares’s theory as “neo-political process theory of criminal procedure”). According to Professor Ely, the Court should remain uninvolved, in Fourteenth Amendment terms, unless and until (1) the powerful impede access to the “channels of political change” or (2) laws are enacted that systematically disadvantage a discrete and insular minority. 

\textit{Ely, supra}, at 103; 

\textit{Luna, supra} note 113, at 443–44.
doctrine. Vague laws do not provide notice of what should be a clear boundary between criminal and non-criminal activity. By requiring substantive criminal law to be specific, the void for vagueness doctrine attempts to ensure that citizens do not offend criminal prohibitions without at least the ability to know in advance the scope of the criminal law.\textsuperscript{145} Requiring criminal laws to be specific also limits law enforcement’s ability to interpret vague laws in a manner to justify the arrest of anyone police might arbitrarily identify.\textsuperscript{146}

The new discretion scholars dispute the virtues of the void for vagueness doctrine by emphasizing the discretion that police retain even when they enforce specific laws. For example, Livingston notes that legislators forced to articulate specific standards tend to articulate rule-based prohibitions rather than standard-based norms.\textsuperscript{147} Rules attempt to articulate prohibited conduct with particularity, while standards articulate the policies and goals underlying the law.\textsuperscript{148} Livingston notes that because rule-based prohibitions are imperfect reflections of the underlying standard, they can be both overbroad and underbroad.\textsuperscript{149} As an example, Livingston cites laws that attempt to curtail youths from automobile “cruising” by prohibiting driving past a traffic control point

\textsuperscript{145} See \textit{supra} Part I.B for a discussion of the void for vagueness doctrine. Of course, criminal law does not require actual knowledge by an individual defendant that his conduct violated the law; it requires only that criminal laws provide fair notice to individuals about their scope. \textit{Compare} Papachristou v. City of Jacksonville, 405 U.S. 156, 162 (1972) (providing that criminal laws must require notice), \textit{with} Lambert v. California, 355 U.S. 225, 228 (1957) (stating the general rule that “ignorance of the law will not excuse”) (internal quotations and citation omitted). Rather than contradict one another, these two well-accepted criminal law principles work hand in hand. Because criminal laws are required to be “definite and knowable,” one can argue that there is no true reasonable ignorance of the criminal law. See \textsc{Joshua Dressler, Understanding Criminal Law} 165–66 (3d ed. 2001) (discussing mistake of law claims); \textit{see also} United States v. Baker, 807 F.2d 427, 430 (5th Cir. 1986) (stating that “a defendant could be convicted without any knowledge whatsoever of the law making his conduct criminal”).

\textsuperscript{146} See \textit{supra} Part I.B (discussing anti-vagueness doctrine’s role in limiting arbitrary police discretion).

\textsuperscript{147} Livingston, \textit{supra} note 6, at 610–18.

\textsuperscript{148} See generally Louis Kaplow, \textit{Rules Versus Standards: An Economic Analysis}, 42 DUKE L.J. 557 (1992) (offering an economic analysis to the question of whether laws should be promulgated as rules or as standards); Kathleen M. Sullivan, \textit{Foreword: The Justices of Rules and Standards}, 106 HARV. L. REV. 22, 57 (1992) (describing “the rules and standards debate in a nutshell”). Sullivan summarizes the debate succinctly. “A [law] is standard-like when it tends to collapse decisionmaking back into the direct application of the background principle or policy to a fact situation.” Sullivan, \textit{supra}, at 58. “A [law] is rule-like when it binds a decisionmaker to respond in a determinate way to the presence of delimited triggering facts.” \textit{Id.} For example, a posted speed limit is a rule, while a prohibition against careless driving is a standard.

\textsuperscript{149} Livingston, \textit{supra} note 6, at 614.
in a designated area during specified night hours or more than a specified number of times within a fixed period. While such laws effectively limit police discretion by specifically defining the prohibited conduct, the statutes’ reach is not limited to those against whom enforcement of the legislation was intended. When such laws are overbroad, Livingston argues, police will use their discretion to decide not to enforce it. Similarly, Kahan and Meares maintain that judicial invalidation of vague ordinances actually forces communities to tolerate police strategies that involve even more discretion. They point out, for example, that Chicago could substitute its gang-loitering efforts with New York City’s zero-tolerance approach, where public order laws are specifically defined, but where police officers retain wide discretion about where, when, and whether to enforce them. The void for vagueness doctrine places no restraint, they note, on this type of discretion.

In other words, the new discretion scholars equate officers’ discretion to choose whom to arrest among those who are arrestable under a specific law with the discretion to decide who is arrestable under a vague law. Such a comparison overlooks a critical distinction between the discretion to be lenient and the discretion to sanction. When criminal laws are specific, police undoubtedly retain discretion to reduce criminal liability by opting not to enforce the law where it applies. This form of police discretion is similar to the discretion that other participants in the criminal justice system exercise. For example, prosecutors have discretion not to file the most serious criminal charges that might apply to a given defendant and to pursue a less serious charge instead or not to file any charges at all, but they do not have the discretion to file a charge for which there is no probable cause. Similarly, upon conviction,

150. Id. at 616.
151. Id. at 616–18.
152. Id.
154. Id.; see also Heymann, supra note 8, at 442–43 (noting that the denial of discretion is “a fiction” because police exercise discretion in determining whether to enforce criminal laws); Livingston, supra note 6, at 615 (noting that police officers can enforce specifically defined prohibitions in discriminatory ways).
judges exercise discretion to sentence the defendant within the applicable sentencing range, but do not have the discretion to go beyond the maximum sentence. To permit vague criminal laws, in contrast, would permit not just police but also prosecutors and judges to select whom to punish because a non-specific law could be interpreted at any stage in the criminal justice system to encompass the entire population.

The distinction between discretion to sanction and discretion to show leniency is a meaningful one, despite its formality. Whereas the former creates the potential to sweep unwitting actors within the scope of a criminal law defined vaguely, the latter involves discretion only over those who step first across the criminal law’s clearly defined borders. The importance of this difference can be seen in the U.S. Supreme Court’s recent sentencing jurisprudence. The Court has held that while a trial court judge can determine sentencing factors by a preponderance of the evidence, the elements of the substantive criminal offense must be pled in the indictment and proven to the jury beyond a reasonable doubt.

In Apprendi v. New Jersey, the Court held that any fact, other than perhaps recidivism, that increases what would otherwise be the applicable maximum sentence must be treated as an element of the offense and be proven to the jury beyond a reasonable doubt. In

157. See id. at 474–97. Apprendi and its progeny are based on the longstanding rule that due process requires the prosecution to prove each element of a criminal offense beyond a reasonable doubt. See In re Winship, 397 U.S. 358, 361–64 (1970). Prior to Apprendi, the Court had suggested that state legislatures could avoid Winship by labeling relevant factors as sentencing factors rather than as elements of the offense. See Almendarez-Torres v. United States, 523 U.S. 224, 228–47 (1998) (upholding federal unlawful reentry statute, under which the maximum sentence increased if the judge found that the defendant’s prior deportation was for an aggravated felony, because this recidivism finding was a sentencing factor and not an element of the offense); McMillan v. Pennsylvania, 477 U.S. 79, 84–93 (1986) (upholding statute specifying a mandatory minimum sentence if the trial judge found by a preponderance of the evidence that the defendant “visibly possessed a firearm” during the offense, because the factor was not an element of the offense under Winship). After an initial hint that Winship was not so easily avoided, see Jones v. United States, 526 U.S. 227, 243 (1999), the Court in Apprendi held that the government must plead and prove to the jury beyond a reasonable doubt any factor that increases what would otherwise be the applicable maximum sentence. Apprendi, 530 U.S. at 490.
158. 530 U.S. 466 (2000).
159. The Court left open the possibility that the legislature could assign factual findings regarding recidivism to be determined by the trial court by a preponderance of the evidence, rather than by the jury by proof beyond a reasonable doubt. Id. at 496. This qualification permitted the Court to reconcile its holding with Almendarez-Torres. Id.; see Almendarez-Torres, 523 U.S. at 228–47 (holding that recidivism was a sentencing factor, not an element of the offense).
160. Apprendi, 530 U.S. at 490.
Unpacking New Policing

contrast, as long as the fact does not result in a sentence beyond the maximum penalty provided for the offense of conviction, the legislature can specify the fact as a sentencing factor to be determined by the trial court by a mere preponderance of the evidence. The Apprendi regime has been criticized for creating “a meaningless and formalistic difference.” For example, while Apprendi prohibits using a so-called “sentencing factor” to increase the defendant’s sentence beyond the otherwise applicable statutory maximum, it permits the legislature to establish a very high statutory maximum—triggered by the jury’s finding of guilt on the substantive offense—and then to establish “sentencing factors” that govern the judge’s determination of the actual sentence beneath the statutory maximum. This criticism echoes the new discretion scholars’ sentiment about the distinction between vague and specific laws: that legislatures can avoid the void for vagueness doctrine and yet continue to vest discretion in police by establishing clear, broadly defined prohibitions, and then permitting the police to exercise discretion in determining when, where, and against whom to enforce the law.

Although the distinction between authority to be lenient and authority to sanction is a formalistic one, recognizing the distinction preserves the important virtue of notice to the citizenry about potential criminal penalties. As Justice Scalia has written in the sentencing context, it is “not unfair to tell a prospective felon that if he commits his

161. Harris v. United States, 536 U.S. 545, 556–69 (2002) (upholding against an Apprendi challenge a statute specifying a mandatory minimum sentence that was shorter than the maximum allowable sentence and triggered by a trial court’s finding by a preponderance of the evidence that the defendant brandished a firearm); McMillan, 477 U.S. at 89–91 (upholding statute specifying a mandatory minimum sentence if the trial judge found by a preponderance of the evidence that the defendant “visibly possessed a firearm” during the offense).

162. Apprendi, 530 U.S. at 541 (O’Connor, J., dissenting) (stating that the Constitution should not “require a state legislature to follow such a meaningless and formalistic difference in drafting its criminal statutes”); see Harris, 536 U.S. at 569 (Breyer, J., concurring) (concurring in Court’s holding that Apprendi does not apply to mandatory minimums, but stating that he “cannot easily distinguish” the statute struck down in Apprendi from a mandatory minimum sentencing statute and “cannot agree with the plurality’s opinion insofar as it finds such a distinction”).

163. Justice O’Connor has demonstrated the ease with which legislatures can avoid Apprendi problems. For example, in Apprendi, the Court held that New Jersey could not extend the maximum sentence of a felony from ten years to twenty based on a judicial finding that the defendant was motivated to intimidate the victim based on race. 530 U.S. at 494–97. However, as Justice O’Connor points out, New Jersey could cure its sentencing scheme by establishing twenty years as the maximum sentence for the underlying felony and then using the defendant’s motivation as a sentencing factor to establish the actual sentence imposed within the theoretical sentencing range. Id. at 540 (O’Connor, J., dissenting).
contemplated crime he is exposing himself to a jail sentence of 30 years—and that if, upon conviction, he gets anything less than that he may thank the mercy of a tenderhearted judge.”\textsuperscript{164} Although such a regime might lead to disparities, “the criminal will never get more punishment than he bargained for when he did the crime.”\textsuperscript{165}

Notice of substantive criminal prohibitions is even more important than notice of the applicable potential penalties. Just as the \textit{Apprendi} doctrine requires legislatures to be clear when defining factors that affect the extent of liability, the void for vagueness doctrine requires legislatures to be clear when defining the underlying substantive criminal prohibitions. And just as \textit{Apprendi} can be avoided by inflating maximum sentences, anti-vagueness prohibitions can be avoided by defining substantive prohibitions broadly but clearly.\textsuperscript{166} However, a legislature that avoids either \textit{Apprendi} or the void for vagueness doctrine by manipulating the doctrines’ formalities is more accountable politically than an unrestricted legislature. For example, if legislators avoid \textit{Apprendi} by inflating all statutory maximum sentences beyond reason, it is more likely that they will be held accountable politically than if inflated sentences are imposed only at the courtroom level.\textsuperscript{167} Similarly, if the legislature avoids the void for vagueness doctrine by clearly but overbroadly defining the scope of substantive criminal prohibitions, their decision to do so at least has the potential to trigger political debate.\textsuperscript{168} In contrast, permitting vague criminal laws invites little political oversight because members of the citizenry will tend to

\begin{itemize}
\item \textsuperscript{164} Id. at 498 (Scalia, J., concurring) (noting that defendant similarly “may thank the mercy of a tenderhearted parole commission if he is let out inordinately early, or the mercy of a tenderhearted governor if his sentence is commuted”).
\item \textsuperscript{165} Id.
\item \textsuperscript{166} See supra notes 147–54 and accompanying text for a discussion of the new discretion scholars’ perspective that police continue to exercise discretion even when legislatures avoid the anti-vagueness doctrine.
\item \textsuperscript{167} Justice Stevens, in defending the majority’s opinion in \textit{Apprendi}, noted that “structural democratic constraints exist to discourage legislatures” from exposing all defendants convicted of any given offense “to a maximum sentence exceeding that which is, in the legislature’s judgment, generally proportional to the crime.” \textit{Apprendi}, 530 U.S. at 490 n.16. From this perspective, the \textit{Apprendi} doctrine enhances transparency by requiring politically accountable legislatures to expose their sentencing priorities through statutory maximum sentences. “So exposed, ‘[t]he political check on potentially harsh legislative action is then more likely to operate.’” Id. at 491 n.16 (quoting Patterson v. New York, 432 U.S. 197, 228–29 n.13 (1977) (Powell, J., dissenting)).
\item \textsuperscript{168} Professor Sunstein has made the point before that invalidation of vague criminal laws can enhance democracy by forcing legislatures to be clear. Cass R. Sunstein, \textit{Foreword: Leaving Things Undecided}, 110 HARV. L. REV. 4, 25 (1996).
\end{itemize}
read vague criminal laws as applying only to what they think of as criminal behavior, and therefore, as not applying to them or to people like them.\footnote{Like the void for vagueness doctrine, the well-established rule of lenity reflects the importance of legislative accountability. By mandating strict construction of ambiguous criminal statutes, the rule of lenity attempts to ensure that legislatures, not the judiciary, define criminal prohibitions. See United States v. Bass, 404 U.S. 336, 348 (1971) (explaining important functions served by the rule of lenity). Kahan has observed that the rule in practice departs from its theory. He argues that courts have been “sporadic and unpredictable” in the application of lenity principles out of recognition that legislatures operate more effectively if permitted to delegate lawmaking at the interstices to the judiciary. See Dan M. Kahan, \textit{Lenity and Federal Common Law Crimes}, 1994 \textit{Sup. Ct. Rev.} 345, 346–54.} Consider, for example, the considerable political debate that surrounds proposed laws to prohibit smoking in public places. Those laws are controversial because they clearly criminalize conduct that many citizens consider to be innocent. If, in contrast, the proposals were worded as prohibitions against “offensive” conduct, smokers might be unlikely to read the proposed ban as applying to them, at least until the statute was actually enforced in a way that was considered objectionable. By equating the discretion not to enforce specific laws with the discretion to choose whom to punish under a vague law, the new discretion scholars undervalue the role that political oversight can play as a check on legislative decisionmaking, even as they tout the power of the political process.

\subsection*{E. Elasticity of the New Discretion Alternatives}

Finally, in addition to the theoretical problems posed by the new discretion scholarship, the alternative decisionmaking models suggested by the new discretion scholars provide little predictive value in application. Although some new discretion scholars concede that application of their proposed alternatives “could result in legitimate differences of opinion at the margins,”\footnote{See Livingston, \textit{supra} note 6, at 647 (acknowledging that it may be difficult to determine whether a law is targeted at “outsiders,” but stating that other constitutional standards require similar judgments).} the reality is that the models can be manipulated to justify upholding just about any police conduct.

One problem in applying the new discretion scholarship is in defining the relevant community, the starting point to determining whether the “community” has burdened itself evenhandedly. Richard Schragger has discussed the tendency of the new policing models to use the term “community” in a way that masks contradictory notions of what
community means. He summarizes three accounts of community: contractarian, deep, and dualist. Under a “contractarian” account, a community is defined by its members’ agreement to join the community. Chess clubs and housing associations are communities from this perspective because members can withdraw from the group if they disagree. A “deep” account of community suggests that community identity is imposed, not chosen. On this account, community inures from a sense of connectivity and reciprocity inherent in human social relationships. Meares, for example, discusses a sense of “linked fate” shared among all humanity, but particularly by African-Americans, whose life circumstances have been shaped historically by race. Finally, Schragger develops a “dualist” model of community that emphasizes both the intentionality and connectivity of group identity. This account of community emphasizes not who is involved in a decision, but the quality of the decisionmaking process. Under the dualist model, community involves grassroots participation by individuals in small, local settings, where decisions are the results not of a fair majority vote, but of “true conversations” where each “stakeholder” is heard.

The new policing models, in contrast, tend to emphasize geographic boundaries when defining community, an identity that may not correlate with any meaningful sense of “community.” Because of exit

171. Schragger, supra note 8, at 387–403.
172. Id.
173. Id. at 387–93.
174. Id. at 393–97.
175. Id.
176. As Meares explains the concept of “linked fate,” it has two aspects: one that exists generally among all humanity, and one that describes African-Americans specifically. All people have a bond with family and friends and tend to consider the effect of government policies upon them in formulating their own policy positions. African-Americans, however, share this empathy even with African-Americans who are strangers, because their life circumstances have been shaped historically by race. Tracey L. Meares, Social Organization and Drug Law Enforcement, 35 AM. CRIM. L. REV. 191, 215–17 (1998) (discussing the concept of a “linked fate” among African-Americans); Meares, supra note 91, at 682–83 (same).
177. Schragger, supra note 8, at 398–403.
179. See Schragger, supra note 8, at 403–16 (warning that the rhetoric of community can be used to reinforce geographic and spatial boundaries); see also McELROY, supra note 101, at 3–4
Unpacking New Policing

costs, residence within an inner-city neighborhood does not indicate a social contract. Spatially-defined neighborhoods also do not seem to meet a deep account of community. Although residence might be a proxy for race and socioeconomic status, a deep account of community suggests that *some* neighborhood residents would not share a sufficient connection to the “community” to speak on its behalf. From this perspective, spatial attempts to define community fail to explain which “communities” trigger social attachments worthy of moral deference, and which do not. Similarly, geographic boundaries are inconsistent with a dualist notion of community because there is no guarantee that an initiative favored by a spatially-defined neighborhood is the result of a process in which neighbors spoke to each other locally and openly.

Moreover, even if one accepts the notion of defining the relevant “community” by residents’ addresses, the new discretion scholars have not articulated a standard to determine the size of the relevant community. By manipulating the geographic boundaries of the relevant “community,” one can determine whether police action is seen as burdening the community as a whole or only targeted outsiders. Consider, for example, Kahan and Meares’s political process theory. Kahan and Meares maintain that a Chicago Housing Authority policy authorizing suspicionless searches of units in a public housing project would pass muster under their political process theory because representatives of the housing project tenants supported the policy and

("Virtually all commentators agree that the concept of ‘community’ as used in the rhetoric of community policing is imprecise . . . and largely uninformed by a century of sociological usage and study."); Carl B. Klockars, *The Rhetoric of Community Policing*, in *COMMUNITY POLICING: RHETORIC OR REALITY*, supra note 9, at 239, 247–50 (noting that genuine communities are rare).

180. Schragger, for example, notes that a white inner-city shopowner losing business because of minority youths loitering outside would lack a normative claim for redress. See Schragger, supra note 8, at 433.

181. Rosen, for example, suggests that new policing could be accommodated by creating geographic nonuniformities in constitutional law, but recognizes that minimal constitutional safeguards would be imperiled if the "wrong communities" were evaluated with constitutional nonuniformities. See Rosen, supra note 78, at 1193. However, Rosen does not attempt to articulate which communities are the “right” ones. He suggests that nonuniformity should be permitted only for "those communities whose existence or creation is not otherwise incompatible with a well-ordered liberal society,” but ultimately concludes that the question of which communities warrant nonuniformity is an “important and relevant” one that “must await another day.” Id. at 1193–94.

182. See Kahan & Meares, supra note 39, at 1171–75 (maintaining that courts should defer to the political process as long as the burden of challenged police procedures falls upon average members of the community or those who share a “linked fate”); Livingston, supra note 6, at 647 (arguing that vague public order laws should be tolerated if they "do not appear aimed at the exclusion of groups or individuals from participation in a community’s public life") (emphasis added).
because the burden of the searches fell “on everyone who lived in the projects.”

Although they concede that some project residents undoubtedly disapproved of the search policy, they argue that courts should not second-guess the policy because any dissenting tenants had the opportunity to participate in the political process, the majority had the dissenters’ interests in mind, and the majority agreed to subject themselves to the searches as well.

However, the nature of the “representation” of the public housing authority tenants in Chicago was unusual and uniquely local in that tenants were permitted to select representatives from their housing development to act on a Local Advisory Council. Typically, however, representation takes a less specific form, and housing project tenants would be seen as “represented” by the person elected from the district containing the affected housing project. When one defines the community more broadly as the entire neighborhood, residents of the housing projects are among the least powerful in a broader community that contains businesses, homeowners, and private landlords and tenants. Seen from this perspective, a public housing search policy does not burden average members of the community at all; rather, support for the policy by the “community’s” elected officials could be seen as the majority singling out a powerless minority for intrusive searches for law enforcement purposes.

A second problem in applying Kahan and Meares’s political process theory is in determining whether the community is burdening itself evenhandedly. Kahan and Meares would uphold not only those police practices that apply to all residents of a community equally, but also those that apply to only a subsegment of the population, as long as the targeted minority is not “despised.”

This particular aspect of their theory appears especially difficult in application. For example, Kahan and Meares maintain that juvenile curfews and gang-loitering provisions would pass their political process theory because inner-city teenagers and gang members “are linked to the majority by strong social and familial ties.” In Kahan and Meares’s view, inner-city communities

183. Kahan & Meares, supra note 39, at 1175.

184. Id.

185. See id. (arguing that curfews and gang loitering laws would pass the political process test, “albeit in a less straightforward fashion” than laws that burden the entire community); see also Livingston, supra note 6, at 647 (arguing that courts should tolerate vagueness in public order laws if they do not exclude outsiders “from participation in a community’s public life”).

186. Kahan & Meares, supra note 39, at 1175.
support public order laws “precisely because they care so deeply about the welfare” of at-risk youths and share a “linked fate” with them.\footnote{(187)}

However, this argument presumes not only that a majority of inner-city residents support such laws, but also that the majority shares a specific subjective motivation for that support. As an initial matter, Kahan and Meares cite no clear empirical support for their claim.\footnote{(188)}

Moreover, it is unclear how courts can determine whether participants in the political process are subjectively motivated by feelings of a “linked fate” or good old-fashioned animus. Even if African-Americans feel a sense of empathy toward other African-Americans generally,\footnote{(189)} they may nevertheless disfavor a minority within that larger group.\footnote{(190)} Kahan and Meares’s argument suggests that whenever African-Americans support police conduct, courts should defer to the political process as long as the people affected by the conduct are African-American. As Professor Cole has noted, this approach should presumably allow communities to single out short or overweight people for loitering prohibitions (or, as Kahan and Meares concede, gang members or youths more generally), as long as the broader community shared a “linked fate” at a more abstract level.\footnote{(191)}

Finally, even if courts could determine as an evidentiary matter the subjective motivations underlying public support for police conduct, Kahan and Meares’s emphasis on those motivations is inconsistent with the general notion that the permissibility of police conduct generally does not turn on subjective motivations. For example, an arrest for a low-level offense based on probable cause is lawful, even if the arresting officer had a subjective motive to use the arrest as a pretext to

\footnote{187. \textit{Id.} at 1175–76; see also Meares, supra note 176, at 215–17 (discussing the concept of a “linked fate” among African-Americans); Meares, \textit{supra} note 91, at 682–83 (same).}

\footnote{188. See Kahan \& Meares, \textit{supra} note 39, at 1165 n.78. To support their notion of a “linked fate” among African-Americans, Kahan and Meares cite to Meares’s prior scholarship and to articles discussing the interplay between gang members and drug dealers and law-abiding residents in inner-cities. \textit{Id.} To say that gang members and drug dealers have law-abiding family members, or that there are occasions for mutual reliance between law-abiding elements of the community and street gangs, is not inconsistent with the notion that the majority of the community may single out youths in an attempt to control crime in their communities.}

\footnote{189. Meares, \textit{supra} note 176, at 215–17 (discussing the concept of a “linked fate” among African-Americans); Meares, \textit{supra} note 91, at 682–83 (same).}

\footnote{190. For example, Suzanne Meiners has explored how community policing programs frequently target youth. Suzanne Meiners, \textit{A Tale of Political Alienation of Our Youth: An Examination of the Potential Threats on Democracy Posed by Incomplete “Community Policing” Programs}, 7 U.C. DAVIS J. JUV. L. \& POL’Y 161 (2003).}

\footnote{191. Cole, \textit{supra} note 39, at 1084.}
investigate a different offense, because there is an objective basis for the arrest that renders the arrest reasonable. If suspicious motivations do not undermine police conduct with an objectively reasonable basis, then compassionate motivations should not salvage police conduct lacking an objectively reasonable basis.

III. THE PROGRAMMATIC PURPOSE APPROACH: DISTINGUISHING BETWEEN PUNITIVE AND NON-PUNITIVE POLICING

As set forth in the previous Part, new discretion scholars advocate the new policing by encouraging a retreat from established jurisprudence. This Part challenges the new discretionists’ assumption that current doctrinal rules necessarily conflict with the new policing models. In doing so, this Part explores a distinction virtually ignored by current scholarship between punitive and non-punitive programs within the new policing models.

A handful of scholars have responded to the new discretion scholarship by defending current limitations on police discretion. However, they have done so without entertaining a threshold question of whether all new policing models necessarily fall within the applicable scope of the rules that they defend. For example, responding primarily to Kahan and Meares, Professors Cole, Alschuler, and Schulhofer have defended the void for vagueness doctrine. They maintain that Kahan and Meares overstate both the political power and the support for order-maintenance policing among African-Americans. They also criticize the malleability of the Kahan and Meares thesis, questioning whether it has any limits at all. Cole further attacks the Kahan and Meares position that freedom from crime is a co-equal component of liberty with freedom from government intrusion. Cole argues that official misconduct is a qualitatively-different, more destructive harm—

192. See Whren v. United States, 517 U.S. 806, 808–19 (1996) (vehicle stop based on police officers’ observation of several traffic offenses was lawful, even if the officers initiated the stop to facilitate investigation of drug offenses for which they lacked reasonable suspicion or probable cause).
195. Alschuler & Schulhofer, supra note 109, at 242–43 (wondering if the “next step might be to say if you hate the sin but love the sinner, you can share her burden”) (internal footnote omitted); Cole, supra note 39, at 1084 (criticizing concept of “linked fate” as a limit upon Kahan and Meares’s political process theory).
Unpacking New Policing

particularly to African-Americans—than harm from private, unofficial criminal offenders.\(^{196}\) By arguing to retain vagueness review of criminal laws without addressing which categories of new policing initiatives should be subject to such review, Cole, Alschuler, and Schulhofer appear to assume—at least implicitly—that all new policing initiatives are punitive.

Other critics of the new policing appear to share this assumption without squarely entering the debate over the void for vagueness doctrine. For example, Bernard Harcourt has been critical of the new policing models, particularly the broken windows theory underlying them.\(^{197}\) He argues that the theory lacks empirical support and, in any event, emphasizes only order-maintenance policing’s benefits, without taking into account the resulting harms.\(^{198}\) To Harcourt, an accurate cost-benefit calculation must also consider the hardship of misdemeanor arrests on a largely harmless population of arrestees, the increased risks of police misconduct that accompany increased arrests, the disparate impact that order-maintenance policing has on racial minorities, and the delegation of authority to police about whom to arrest.\(^{199}\) Similarly, Fagan and Davies have compiled statistical evidence indicating that zero-tolerance policing, at least in New York City, adversely affects racial minorities.\(^{200}\) Both Harcourt and Fagan and Davies critique order-maintenance policing for cracking down on disorder, but neither distinguishes between punitive and non-punitive responses to disorder.

Viewed as non-criminal programs, much of the new policing might conform with current constitutional jurisprudence. This Part pursues the distinction between punitive and non-punitive policing models in four subparts. Subpart A develops a distinction at a programmatic level between policies that pursue criminal law’s traditional objectives of retribution and deterrence through punishment, and those that pursue non-punitive objectives. Although existing criminal procedural rules should not apply to many of the policies implemented under the new

---

\(^{196}\) Cole, supra note 39, at 1089–90.

\(^{197}\) HARcourt, supra note 29, at 90–104.

\(^{198}\) Id. at 213.

\(^{199}\) See id. at 57–89, 212–14; Bernard E. Harcourt, The Collapse of the Harm Principle, 90 J. CRIM. L. & CRIMINOLOGY 109, 154–61 (discussing order-maintenance policing); see also Thacher, supra note 8, at 770 (agreeing with Harcourt that the potential “side effects” of new policing models, such as distortion of ideals about appropriate law enforcement, must be taken into account when determining their efficacy).

\(^{200}\) Fagan & Davies, supra note 19, at 475–96.
policing models, the level of scrutiny should be determined by examining the programmatic purpose of the policy, not by manipulating the rhetoric of community.

To develop further the distinction between punitive and non-punitive programs, Subpart B examines neighborhood exclusion ordinances as an example, incorporating existing jurisprudence regarding police conduct that advances a “special need” and the developing jurisprudence defining the heart of criminal punishment. Subpart C then reexamines Chicago’s anti-gang loitering ordinance as a non-punitive response to gang activity. Finally, Subpart D discusses the advantages of the programmatic purpose model over the current new discretion scholarship.

A. Abandoning Rhetoric and Developing Meaningful Distinctions

At least implicit in the new discretion scholarship is the assumption that existing constitutional jurisprudence governing criminal law and procedure must be changed before cities can take full advantage of the promise of the new policing models. Examination of the programs approved by the new discretion scholars, however, reveals that at least some of these programs fall well outside of any traditional notion of the criminal justice system and should therefore be free from scrutiny under the jurisprudence that the new discretion scholars seek to overhaul. Responding effectively to the low-level, frequent occurrences that trouble neighborhood residents often involves creating long-term, preventative solutions rather than responding to individual crimes after the fact through arrest and prosecution. Other new policing initiatives, however, clearly implicate the traditional constitutional concerns raised by citizen-police encounters, despite the “community policing” labels. The current academic literature regarding the new policing models fails to explore the fuzzy middle between these two extremes, focusing entirely on the rhetoric of “community” policing and ignoring an important distinction between punitive and non-punitive initiatives in the new policing models.

Consider, for example, police collaboration with community groups to pick up garbage, paint over gang graffiti, or plant flowers.201 These

201. See, e.g., Waldeck, supra note 13, at 1271–77 (discussing New York City Transit Authority’s Clean Car Program, involving immediate removal of graffiti from subway trains, and contrasting it with the New York Police Department’s zero-tolerance approach). Neal Katyal has called attention to the potential of architecture as instrumental in crime control, an approach that
Unpacking New Policing

efforts are made under the new policing approach, are favored by the new discretion scholars, and clearly do not raise constitutional questions under existing jurisprudence. In contrast, some new policing techniques simply resort to the traditional law enforcement model, with a shifted emphasis on low-level, quality of life offenses. For example, New York City’s zero-tolerance campaign devoted street-level policing efforts to the enforcement of low-level criminal prohibitions such as turnstile jumping, aggressive panhandling, and squeegeeing windows. Other than the emphasis on low-level offenses, zero-tolerance policing mirrors rapid-response policing: officers become aware of an offense, arrest an individual offender, and refer the case for prosecution on criminal charges.

New policing efforts can fall somewhere between the extremes of flower planting and zero tolerance. Consider, for example, a hypothetical noise ordinance that, instead of criminalizing unreasonable noise, authorizes police to seize any devices used to make unreasonable noise pending an administrative hearing to determine whether the devices should be returned. Unlike community clean-up and flower planting, this hypothetical ordinance responds to an individual problem directly. The ordinance would give police leverage to negotiate compliance with norms at the street level; indeed, even a resistant noise-maker who is unwilling to negotiate could ultimately be silenced through seizure. However, unlike traditional criminal law, the ordinance does not seek to deter future noise violations by convicting and punishing the offender.


203. See, e.g., Fagan & Davies, supra note 19, at 471 (noting that the focus of New York City’s policing approach was controlling disorder, and “the tactic to achieve it was arrest, the most traditional of law enforcement tools”); Greene, supra note 28, at 175 (noting that New York City’s zero-tolerance approach is “grounded in traditional law enforcement methods and in relentless crackdown campaigns to arrest and jail low-level drug offenders and other petty perpetrators”).

204. Ordinances resembling this hypothetical are not uncommon. In Chicago, police are authorized to impound any vehicle used to violate the city’s sound device restrictions. If a hearings officer determines that there is probable cause that the vehicle was used to violate the restrictions, the vehicle is not returned until the owner pays a cash bond of $500 plus towing fees. CHI., ILL., MUNICIPAL CODE § 11-4-1115 (2002). It is increasingly common for cities to authorize police to impound devices used to make excessive noise, primarily vehicles containing booming stereos, as evidence of an unreasonable noise violation. See, e.g., SAN DIEGO, CAL., MUNICIPAL CODE § 59.5.0502(b)(3) (2000) (authorizing those who enforce the prohibition against unreasonable noise to seize as evidence any component transmitting or amplifying the noise and authorizing police to impound vehicles containing such a component if the component cannot readily be removed).
Accordingly, constitutional rules that govern criminal law and procedure are arguably inapplicable. Implicit in the new discretion scholarship is the assumption that any law giving authority to police should be analyzed as a criminal statute. For example, Livingston recognizes that the community clean-up aspect of the new policing approaches does not implicate any constitutional concerns. However, she appears to assume that if police go beyond these types of clean-up efforts and attempt to enforce norms of orderliness, their conduct necessarily presents legal problems under current jurisprudence.

Rather than focus on the unique programmatic purposes often (but not always) pursued by the new policing approaches, the new discretion scholars have attempted to mark these approaches as unique by emphasizing the role of the community. To some extent, it is not surprising that the early scholarship in this area has fallen into this trap, as the police approaches themselves are cloaked in community rhetoric and do not expressly highlight their non-traditional operational mechanisms. A more meaningful basis upon which to distinguish much (but not all) of the new policing from traditional criminal law is its underlying programmatic purpose of solving public safety problems other than through the traditional mechanism of deterring and punishing through arrests, convictions, and criminal sentences.

B. Drug Free Zones: An Example

If the hypothetical noise ordinance above demonstrates the possibility of using non-punitve ordinances to solve criminal law problems, actual ordinances across the country show just how far this approach can be taken. This subpart explores neighborhood exclusion under “drug free zone” ordinances as an example of how cities can gain control over the use of public spaces without resorting to criminal punishment.

The neighborhood exclusion concept originated in Portland, Oregon, with the enactment of a city ordinance designating discrete geographical areas as “drug free zones” based on their high incident rates of drug

---

205. One of Livingston’s examples of the new policing involved police officers working with other municipal agencies to improve the physical deterioration of housing projects by removing trash and abandoned cars from the neighborhood, an effort that was followed by a notable decrease in residential burglaries. Livingston, supra note 6, at 575. This clearly does not raise any criminal procedural questions, just as new policing efforts avoid any legal questions when they eradicate graffiti by painting over it instead of attempting to catch suspects with paint cans in hand. Id. at 584.

206. Id. at 584–85.
If a person is arrested for committing a drug crime in a public area of a drug free zone, police exclude the individual from the zone for ninety days. Entry into a drug free zone in violation of an exclusion order constitutes criminal trespass, a misdemeanor defined by the state penal code. The exclusion orders are effective automatically unless the excluded individual invokes one of two important procedural protections. First, the individual has the right to file an administrative appeal to challenge the basis for the exclusion. Once the appeal is filed, the exclusion is stayed while the appeal is pending.

Second, even if the excluded individual does not challenge the basis for the exclusion itself, he may seek a variance from the police to enter the zone for a broad variety of purposes. Variances can be granted if the individual lives, works, goes to school, or needs access to social services or “essential needs” within the zone. Additionally, the ordinance authorizes “general” variances that may be issued for “any reason” to “an excluded person who presents a plausible need to engage in any non-criminal activity.” The ability to obtain a variance is so broad, in fact, that the ordinance arguably acts more like a targeted loitering ordinance than an exclusion ordinance. In other words, the practical effect of a neighborhood exclusion order is not an absolute prohibition against the individual entering the zone, but a restriction on

---

207. PORTLAND, OR., CITY CODE §§ 14B.20.010–.070 (2002). The City Council determines the drug free zone designations based on the number of drug arrests made in an area in the previous twelve months, revisiting the zone designations every few years. Id. §§ 14B.20.010–.020; see State v. James, 978 P.2d 415 (Or. Ct. App. 1999). Portland also has a prostitution free zone ordinance that enables police to exclude individuals from high-vice neighborhoods. PORTLAND, OR., CITY CODE §§ 14B.30.010–.070.

208. PORTLAND, OR., CITY CODE § 14B.20.030(A). The ordinance does not apply if the drug offense is committed within a private residence. Id. If the actor is eventually convicted of the crime for which he is arrested, he is issued a separate exclusion of one-year duration. Id. § 14B.20.030(B).

209. See, e.g., OR. REV. STAT. § 164.245 (2001) (defining criminal trespass in the second degree as entering or remaining unlawfully in or upon premises); see also MODEL PENAL CODE § 221.2 (1985) (prohibiting criminal trespass).

210. PORTLAND, OR., CITY CODE § 14B.20.060(A)(6) (providing that if excluded individual does not file a notice of appeal, the exclusion takes effect on the eighth day following the initial notice of exclusion).

211. Id.

212. Id.

213. Id. § 14B.20.060(B).

214. Id.

215. Id.
his ability to move within the zone without an articulated purpose approved in advance by police.\(^{216}\)

Unlike Chicago’s anti-gang loitering ordinance, drug free zone ordinances have not been the subject of considerable academic commentary.\(^{217}\) In lower courts, opponents of neighborhood exclusion ordinances have had limited success arguing that exclusion orders constitute punishment.\(^{218}\) However, the argument that neighborhood exclusion constitutes punishment appears to assume that \textit{any} impairment of an individual’s status quo constitutes punishment, an assumption clearly at odds with current law.

After a brief flirtation with aggressive judicial scrutiny of supposedly civil legislation,\(^{219}\) the U.S. Supreme Court made clear in \textit{Hudson v. United States}\(^{220}\) that whether a penalty constitutes criminal punishment

\(^{216}\) Andrew Leipold has suggested that one cure for loitering ordinances is to target only those who have already broken the law. Andrew D. Leipold, \textit{Targeted Loitering Laws}, 3 U. PA. J. CONST. L. 474, 483 (2001). In some respects, Portland’s drug free zone ordinance conforms to Leipold’s model by excluding those who have been arrested for a drug-related offense, but then permitting excluded individuals to obtain variances permitting them to move within the zone, but only for specified purposes.

\(^{217}\) \textit{But see} Schragger, \textit{supra} note 8, at 408 n.131 (citing Cincinnati’s drug exclusion zone ordinance as an example of a law that essentially zones a targeted geographic area as drug free); Kim Strosnider, \textit{Anti-Gang Ordinances After City of Chicago v. Morales: The Intersection of Race, Vagueness Doctrine, and Equal Protection in the Criminal Law}, 39 AM. CRIM. L. REV. 101, 129 (2002) (primarily discussing anti-gang loitering ordinances, but citing Cincinnati’s drug free zone ordinance as demonstrating “the lengths to which cities will go in order to carve out zones where different—and more restrictive—laws apply to combat perceived crime problems”).

\(^{218}\) The issue of whether neighborhood exclusion constitutes punishment has been litigated in the double jeopardy context, with excluded individuals arguing that they may either be excluded from a neighborhood or prosecuted criminally for the underlying conduct, but not both. One federal district court agreed with opponents of neighborhood exclusion ordinances that neighborhood exclusion constitutes criminal punishment, but the Court of Appeals affirmed the district court’s decision on separate grounds. See \textit{Johnson v. City of Cincinnati}, 119 F. Supp. 2d 735, 747–49 (S.D. Ohio 2000) (holding that neighborhood exclusion ordinance violated defendant’s right to be free from double jeopardy by allowing him to be convicted of the underlying drug offense and to be excluded from neighborhood), \textit{aff’d on other grounds}, 310 F.3d 484 (6th Cir. 2002), \textit{cert. denied}, 123 S. Ct. 2276 (2003). In contrast, Oregon’s courts have held that neighborhood exclusion does not constitute criminal punishment. \textit{State v. Lhasawa}, 55 P.3d 477, 561 (Or. 2002) (concluding that neighborhood exclusion is a civil sanction that does not implicate double jeopardy concerns); \textit{State v. James}, 978 P.2d 415, 417–21 (Or. Ct. App. 1999) (same).

\(^{219}\) \textit{See United States v. Halper}, 490 U.S. 435, 446–52 (1989) (holding that supposedly civil penalties for false Medicare reimbursement claims constituted punishment because they were not related to the actual damages from the fraud); \textit{United States v. Ursey}, 518 U.S. 267, 283–92 (1996) (refusing to apply \textit{Halper} analysis to \textit{in rem} civil forfeitures). The Court has expressly abandoned \textit{Halper} as being too quick to label a penalty as criminal punishment. \textit{Hudson v. United States}, 522 U.S. 93, 100–03 (1997).

Unpacking New Policing

is a matter of statutory construction requiring a two-part inquiry. The first question is whether the legislature enacting the penalty expressly labeled it as civil or criminal. Second, even if the penalty is designated as civil, the court must ask whether the penalty is so punitive in either its purpose or effect as to transform what was purportedly intended as a civil penalty into a criminal one. Seven factors, originally set forth by the Court in *Kennedy v. Mendoza-Martinez*, serve as guideposts for answering the second question. These factors ask whether the scheme is in response to criminal conduct, requires scienter, imposes an affirmative restraint, promotes the traditional aims of criminal punishment, is rationally connected to a non-punitive purpose, has historically been regarded as punishment, or is excessive in relation to its supposed civil purpose.

Municipalities enacting drug free zone ordinances have designated neighborhood exclusion as a civil penalty. Accordingly, the relevant inquiry under *Hudson* is whether, in light of the seven *Mendoza-Martinez* factors, exclusion is so harsh in purpose or effect as to reveal a contrary nature. Two of the seven factors do suggest punishment. Exclusion orders are issued in response to conduct constituting a crime, such as drug possession. These crimes, in turn, require proof of at least some mens rea.

---

221. Id. at 99.
222. Id. ("A court must first ask whether the legislature, ‘in establishing the penalizing mechanism, indicated either expressly or impliedly a preference for one label or the other.’") (quoting United States v. Ward, 448 U.S. 242, 248 (1980)).
223. Id.
225. Id. at 168–69.
226. Smith v. Doe, 538 U.S. 84, 123 S. Ct. 1140, 1149 (noting the relevance of the seven *Mendoza-Martinez* factors), reh’g denied, 123 S. Ct. 1925 (2003); *Hudson*, 522 U.S. at 99 (same); *Ward*, 448 U.S. at 249 (same).
227. PORTLAND, OR., CITY CODE § 14B.20.030 (2002) (entitled “Civil Exclusion”); State v. James, 978 P.2d 415, 419 (Or. Ct. App. 1999) (“It is undisputed that [the exclusion provisions] were expressly designated as civil.”).
228. See *Hudson*, 522 U.S. at 99.
229. *James*, 978 P.2d at 420 (concluding that the *Hudson* factors of scienter and whether conduct constitutes a crime supported the defendant’s argument that exclusion constituted criminal punishment).
230. See PORTLAND, OR., CITY CODE § 14B.20.030(A) (setting forth the drug crimes that trigger neighborhood exclusion under Portland’s drug free zone ordinance). Under the prostitution-free zone ordinance, neighborhood exclusions are issued based on an arrest for prostitution-related offenses. See id. §§ 14B.30.010–.070.
However, such suggestion of punishment is not necessarily dispositive in the case of neighborhood exclusion orders. As an initial matter, future ordinances employing the neighborhood exclusion concept could be drafted to trigger police authority in response to conduct that does not constitute a crime and does not require proof of a culpable mental state. More importantly, under *Hudson*, no one factor is determinative—all must be given consideration.\(^{231}\) As demonstrated below, the five remaining *Mendoza-Martinez* factors all suggest that neighborhood exclusion does not constitute punishment.

1. **No Affirmative Restraint**

First, neighborhood exclusion is not an “affirmative restraint” under the Court’s apparent conception of that term.\(^{232}\) Unfortunately, the Court has given little guidance about what constitutes an “affirmative restraint,” suggesting obliquely that the term as “normally understood” provides sufficient clarity.\(^{233}\) The Court has indicated that the definition of affirmative restraint might be extremely narrow, including only those sanctions that approach the “‘infamous punishment’ of imprisonment.”\(^{234}\) The Court has held, for example, that the civil commitment of sexual offenders constitutes an affirmative restraint,\(^{235}\)

---

231. Under the U.S. Supreme Court’s current jurisprudence, the seven *Mendoza-Martinez* factors are guideposts; no single factor is determinative. See *Hudson*, 522 U.S. at 101 (rejecting *Halper* analysis in part because it rendered one of several relevant factors dispositive).

232. But see *James*, 978 P.2d at 419 (concluding with little analysis that a drug free zone exclusion was an affirmative restraint because it limited an individual’s freedom to enter a specific area).

233. *Hudson*, 522 U.S. at 104 (noting that money penalties and occupational disbarment did “not involve an ‘affirmative disability or restraint,’ as that term is normally understood”).

234. *Id.* (quoting Flemming v. Nestor, 363 U.S. 603, 617 (1960)); see also State v. Lhasawa, 55 P.3d 477, 486 (Or. 2002) (noting that recent U.S. Supreme Court decisions “appear to suggest that nothing short of imprisonment would qualify as an affirmative disability or restraint”). In holding that a prohibition against participation in the banking industry did not constitute an affirmative restraint, the Court in *Hudson* simply noted that the petitioners’ sanction was “certainly nothing approaching the infamous punishment of imprisonment” and did not constitute an affirmative restraint “as that term is normally understood.” *Hudson*, 522 U.S. at 104 (internal citations and quotations omitted). The Court has not made clear whether “affirmative restraints” encompass only those things that approach imprisonment, or whether the concept of an affirmative restraint as “normally understood” is broader. At least one circuit court has adopted the former, narrow notion of affirmative restraint, asking only whether the sanction approaches imprisonment. See Cutshall v. Sundquist, 193 F.3d 466, 474 (6th Cir. 1999) (noting that the question of whether something is an affirmative restraint or disability depends on whether the sanction approaches imprisonment).

235. *Kansas v. Hendricks*, 521 U.S. 346, 360–69 (1997) (holding that, although civil commitment of sexual offenders who were rendered incurably dangerous by mental disorder constituted an
but that mandatory registration as a sexual offender does not.\textsuperscript{236} Exclusion from a discrete geographic area within a city does not involve imprisonment.\textsuperscript{237}

Some of the Court’s decisions, however, suggest that a sanction short of literal confinement can constitute an affirmative restraint. Most recently, the Court explained that the appropriate inquiry asks “how the effects of the . . . [scheme] are felt by those subject to it. If the disability or restraint is minor and indirect, its effects are unlikely to be punitive.”\textsuperscript{238} Applying that test, the Court concluded that the mail-in registration required by Alaska’s sex offender registration and notification scheme did not impose an affirmative restraint, but appeared to entertain the notion that an in-person registration requirement might be considered an affirmative restraint.\textsuperscript{239}

Even if there is room within the affirmative restraint concept for sanctions other than confinement, it clearly does not encompass every sanction that restricts the activities of an individual.\textsuperscript{240} For example, in

affirmative restraint, it did not constitute criminal punishment that implicated the double jeopardy clause).

\textsuperscript{236} Smith v. Doe, 538 U.S. 84, 123 S. Ct. 1140, 1151–54 (upholding Megan’s Law against an ex post facto challenge and holding that registration and notification scheme did not impose an affirmative restraint), \textit{reh’g denied}, 123 S. Ct. 1925 (2003).

\textsuperscript{237} \textit{Cf. Hendricks}, 521 U.S. at 363 (holding that civil commitment, although not necessarily punishment, is an affirmative restraint); United States v. Salerno, 481 U.S. 739, 749 (1987) (discussing pretrial incarceration).

\textsuperscript{238} \textit{Smith}, 123 S. Ct. at 1151.

\textsuperscript{239} \textit{Id.}; see also Kennedy v. Mendoza-Martinez, 372 U.S. 144, 168–69 (1963) (concluding that deportation of draft evaders amounted to punishment, without determining whether it constituted an affirmative restraint).

\textsuperscript{240} Even courts using a notion of affirmative restraint that goes beyond literal confinement have conceded that sanctions of limited scope do not constitute affirmative restraints under current punishment jurisprudence. For example, in an opinion ultimately reversed by the U.S. Supreme Court, the Ninth Circuit created a circuit court split by using a broad notion of affirmative restraint (and also of punishment) to strike down as double punishment Alaska’s program for registering convicted sex offenders and notifying communities regarding their presence. Doe I v. Otte, 259 F.3d 979, 983–95 (9th Cir. 2001), \textit{rev’d sub nom. Smith v. Doe, 538 U.S. 84, 123 S. Ct. 1140, \textit{reh’g denied}, 123 S. Ct. 1925 (2003)}. Despite the broad standards for punishment applied in \textit{Doe I}, even the \textit{Doe I} court conceded that a more limited sanction would not constitute punishment. \textit{Id.} at 988–89. In striking down the Alaska program, the Court distinguished the program from a less onerous program previously upheld as neither an affirmative restraint nor punishment. \textit{Id.} at 987. While the prior program required only a one-time registration, the Alaska program required multiple registrations per year for the offender’s lifetime as well as community notification. \textit{Id.} These more onerous requirements went beyond the “limited” scope of the previously upheld program and rendered the program, in the Ninth Circuit’s view, an affirmative restraint. \textit{Id. But see Burr v. Snider, 234 F.3d 1052, 1054 (8th Cir. 2000)} (upholding as reasonable state court’s decision that sex offender registration and notification requirements were not affirmative restraints or punishment);
Hudson, the Court held that lifetime disbarment from the banking industry was neither an affirmative restraint nor criminal punishment.241 Like disbarment from a profession, a neighborhood exclusion requires the affected individual to refrain from specified conduct, but does not require the individual to engage in any affirmative conduct. The impact of exclusion might be harsh in some cases. However, the impact is no less onerous than barring an individual from participating in what had been his lifetime career, a prohibition that was upheld in Hudson as neither an affirmative restraint nor criminal punishment.242

In fact, neighborhood exclusion orders are arguably a narrower limitation on an individual’s status quo than a lifetime prohibition against working in one’s trained profession. Current neighborhood exclusion ordinances authorize exclusion from only discrete geographic areas within the city, comprising only a small percentage of their respective metropolitan areas.243 The exclusions are also for a limited duration.244 Moreover, the ordinances authorize the issuance of variances.245

Femedeer v. Haun, 227 F.3d 1244, 1250 (10th Cir. 2000) (holding that sex offender registration and community notification program was neither an affirmative restraint nor punishment); Cutshall v. Sundquist, 193 F.3d 466, 474–77 (6th Cir. 1999) (same).

242. Id. at 103–05; see also Brewer v. Kimel, 256 F.3d 222, 228 (4th Cir. 2001) (noting harsh effect of prohibiting individual from lifetime career in Hudson, and holding by comparison that revocation of driver’s license was not an affirmative restraint or punishment). But see Johnson v. City of Cincinnati, 119 F. Supp. 2d 735, 748 (S.D. Ohio 2000), aff’d on other grounds, 310 F.3d 484 (6th Cir. 2002), cert. denied, 123 S. Ct. 2276 (2003). In Johnson, a federal district court held that neighborhood exclusion amounted to an affirmative restraint simply because “[i]t is a restraint against the liberty of those arrested or convicted for the drug-abuse crimes.” Id. This reasoning is wholly inconsistent with Hudson’s express recognition that a restriction might be burdensome and yet nevertheless constitute neither an affirmative restraint nor criminal punishment. See Hudson, 522 U.S. at 99 (noting that the Double Jeopardy Clause does not prohibit multiple civil punishments).
244. The ordinances usually authorize exclusion for ninety days, and none of the ordinances authorizes exclusion for more than a year. See id. § 14B.20.030 (defining duration of Portland’s drug free zone exclusion orders); Johnson v. City of Cincinnati, 310 F.3d 484, 487–88 (6th Cir. 2002) (describing duration of exclusion under Cincinnati’s drug free zone ordinance), cert. denied, 123 S. Ct. 2276 (2003).
245. PORTLAND, OR., CITY CODE § 14B.20.060(1B) (describing procedures to obtain variances under Portland’s drug free zone ordinance); Johnson, 310 F.3d at 488 (describing procedures to obtain variances under Cincinnati’s drug free zone ordinance).
Unpacking New Policing

2. Not the Traditional Aims of Punishment

Secondly, under *Mendoza-Martinez*, neighborhood exclusion zones do not promote the “traditional aims of punishment.” When properly understood, neighborhood exclusion ordinances do not promote retribution and deterrence, at least not in the manner in which criminal punishment traditionally does. Rather, as explained further in this Subpart, neighborhood exclusion seeks to protect a targeted geographic area from future prohibited conduct by civilly restraining and redistributing past offenders.

As an initial matter, it is worth noting that the role of police officers in issuing exclusion orders and variances does not itself render the orders criminal punishment.\(^{246}\) Moreover, although police officers issue the exclusion orders and variances under drug free zone ordinances, all appeals are handled as administrative hearings outside of the police department and criminal courts.\(^{247}\)

The scope of neighborhood exclusion orders suggests that they are not intended as retribution. Under any retributive notion of punishment, the duration or geographic scope of the exclusion would be linked to the severity of the wrongfulness of the excluded actor’s underlying conduct.\(^{248}\) One would also expect exclusion orders designed to extract retribution to prohibit the wrongful actor from entering his own neighborhood, so that the punishment in each case would deprive the offender of his own community. Moreover, a retributive exclusion presumably would not permit variances for individuals who lived within the community. In contrast, civil neighborhood exclusion orders are identical, regardless of the severity of the underlying criminal offense;

\(^{246}\) Cf. *New York v. Burger*, 482 U.S. 691, 717–18 (1987) (holding that inspection by police officers of automobile junkyard premises should be evaluated by standards applicable to administrative searches, and not as a traditional search for criminal evidence, even though the administrative process was implemented by law enforcement officers).

\(^{247}\) See *Hudson*, 522 U.S. at 103 (holding that Congress’s decision to vest disbarment authority in administrative agencies suggested that Congress intended disbarment as a civil sanction, not a criminal one).

apply only to pre-designated high-crime zones, regardless of the offender’s residence; and permit variances to accommodate offenders who live within those zones.

Although neighborhood exclusion orders are not intended as retribution, they certainly are intended to reduce criminal activity within the zones, a purpose that falls within any broad conception of deterrence.249 However, the U.S. Supreme Court has recognized that both civil and criminal penalties can have deterrence goals.250 For example, sexual offender registration and notification requirements are undoubtedly intended to deter future crimes; nevertheless, the Court has held that such schemes do not impose criminal punishment.251

In State v. James,252 the Oregon Court of Appeals held that neighborhood exclusions did not promote the aims of traditional criminal punishment. Although the James court reached the right result, it did so with flawed reasoning. The court reasoned that the deterrent goals of neighborhood exclusions were civil rather than criminal, because the city’s ultimate legislative concerns were not about drug activity per se, but rather about the collateral consequences of that activity on neighborhood property values, business activity, and general quality of life.253

By looking at the legislative concerns motivating the imposed penalties, rather than the nature of the penalties themselves, the court made it too easy for lawmakers to evade criminal procedural rules by doctoring their legislative history. If what matters are the goals of legislators (regardless of the penalty they design to reach those goals), any penalty will fall on the civil side of this Hudson factor as long as lawmakers avoid tough-on-crime rhetoric in their legislative history and

---

249. See generally DRESSLER, supra note 145, at 15 (discussing deterrence).

250. Smith v. Doe, 538 U.S. 84, 123 S. Ct. 1140, 1152 (“Any number of governmental programs might deter crime without imposing punishment.”), reh’g denied, 123 S. Ct. 1925 (2003); Hudson, 522 U.S. at 105 (noting that a civil penalty’s deterrent goals do not render the sanction criminal, “as deterrence ‘may serve civil as well as criminal goals’”) (quoting United States v. Ursery, 518 U.S. 267, 292 (1996)).

251. Smith, 123 S. Ct. at 1152 (upholding retroactive enforcement of Megan’s Law against an ex post facto challenge and noting that “[t]o hold that the mere presence of a deterrent purpose renders such sanctions criminal . . . would severely undermine the Government’s ability to engage in effective regulation”) (internal citations and quotations omitted).


253. Id. at 420.
Unpacking New Policing

speak instead of broader concerns that are traditionally the aims of civil law.254

The U.S. Supreme Court has already made clear in a separate context that the rules of criminal procedure are not so easily avoided. In Ferguson v. Charleston,255 the Court addressed the constitutionality of testing pregnant women for drugs at public hospitals. The ultimate purpose behind the drug testing program was to reduce the number of drug-influenced babies by using the threat of criminal prosecution to persuade drug-addicted pregnant women to obtain substance abuse treatment.256 The government argued that this ultimate goal was a special need beyond traditional law enforcement, and, therefore, pregnant women could be tested without a warrant or probable cause.257 Nevertheless, the Court held that the drug testing program violated the Fourth Amendment.258 In doing so, the Court rejected the government’s special needs analysis by scrutinizing the program’s actual content, rather than its ultimate goal. When viewed for its “programmatic purpose” rather than its ultimate objectives, the drug testing program operated by generating evidence to be used in a criminal prosecution. That the ultimate objective revealed a kinder, gentler form of law enforcement was insufficient to demonstrate a special need for Fourth

254. Existing neighborhood exclusion ordinances contain such language within the introductory subsections, typically with a statement from the city council that the higher incidences of criminal activity within the designated zones are deteriorating the quality of life in those neighborhoods. See, e.g., Cincinnati, Ohio, Ordinance 229-1996 (Sept. 6, 1996) (city council finding that the city “has a substantial and compelling interest in restoring the quality of life and protecting the health, safety, and welfare of citizens using the public ways in [the zones]”); Portland, Or., Ordinance 170913 (Feb. 12, 1997) (city council finding that drug activity within zones “contribute[d] to the degradation of these areas and adversely affect[ed] the overall quality of life for the areas’ residents, businesses and visitors”).


256. Id. at 82–83 (noting that “the ultimate goal of the program may well have been to get the women in question into substance abuse treatment and off of drugs”).


258. Ferguson, 532 U.S. at 81–84.
Amendment purposes, “[b]ecause law enforcement involvement always serves some broader social purpose or objective.”

3. Connection To Civil Purposes

Although the James court applied a test that makes avoidance of criminal procedural requirements too easy, it reached, I believe, the correct result. Stated in Mendoza-Martinez terms, neighborhood exclusions are civil in nature not because they are motivated by a civil purpose, but because of the way in which they connect to that civil purpose. In other words, what matters most is not that neighborhood exclusion ordinances seek to improve neighborhood quality of life, but that they do so through a remedial penalty rather than through traditional criminal punishment.

As an initial matter, the nature of the restraint is similar to a civil restraining order. For example, assault victims routinely obtain restraining orders preventing their assailants from coming within a certain distance of them, and this does not constitute criminal punishment. These orders can restrain respondents from contacting not just the immediate victims of assaults, but related individuals, and they often require respondents to move from their homes. Nevertheless, they are seen as civil restraints. Drug free zone exclusions are an extension of the restraining order concept, treating the neighborhood as the victim of the individual’s drug activity.

A second way of understanding neighborhood exclusion as a civil remedy, despite its deterrent purpose, requires an understanding of the

259. Id. at 84 (emphasis added).

260. Kennedy v. Mendoza-Martinez, 372 U.S. 144, 168–69 (1963) (holding that whether a sanction has an assignable non-criminal purpose is one of seven factors in determining whether sanction is punitive in nature).

261. See Peter Finn, Statutory Authority in the Use and Enforcement of Civil Protection Orders Against Domestic Abuse, 23 FAM. L.Q. 43, 44 (1989) (noting that double jeopardy does not prohibit a subsequent prosecution for the same act for which a family abuse civil protection order is issued); Catherine F. Klein & Leslye E. Orloff, Providing Legal Protection for Battered Women: An Analysis of State Statutes and Case Law, 21 HOFSTRA L. REV. 801, 1122 (1993) (same).

262. See Mary M. Cheh, Constitutional Limits on Using Civil Remedies To Achieve Criminal Law Objectives: Understanding and Transcending the Criminal-Civil Law Distinction, 42 HASTINGS L.J. 1325, 1406 & n.431 (1991) (noting that almost all domestic abuse civil protection order statutes permit courts to evict the respondent from his own residence); Klein & Orloff, supra note 261, at 921–22 (noting that civil restraining orders have required respondents to stay away from the petitioner’s residence even when he, rather than the petitioner, was the lawful owner).

263. Finn, supra note 261, at 44; Klein & Orloff, supra note 261, at 1122.
Unpacking New Policing

theory on which the program is premised. Traditional criminal punishment (e.g., imprisonment) is thought to deter in two familiar ways.\footnote{264} The goal of specific deterrence is to deter an individual actor’s future misconduct both by incapacitating him and by teaching him that he will suffer further punishment if he reoffends upon release.\footnote{265} General deterrence aims to dissuade the general population from engaging in criminal conduct by notifying the population of the consequences of such conduct through the punishment of individual actors.\footnote{266}

These traditional means of deterrence play at most a minor role in the goals of neighborhood exclusion orders. Exclusion from individual neighborhoods does not prevent the excluded actor from engaging in criminal conduct outside of the zone. The most that can be said is that the order aims to prevent the individual from committing offenses within the zones. Moreover, it is difficult to imagine that the prospect of a short-term order of exclusion significantly tips a would-be offender’s cost-benefit calculation, either at a specific or general level, when the relevant conduct is already subject to criminal penalties. Finally, even if neighborhood exclusion orders have some incidental effects that resemble traditional deterrence, the same could be said of civil fines, and yet those are not necessarily considered criminal penalties.\footnote{267}

Although the goal of neighborhood exclusion orders, like traditional punishment, is to reduce the number of drug offenses (at least within the designated zones), they do so in an operational manner that is entirely different from traditional deterrence functions. When distinguishing between traditional law enforcement and non-criminal enforcement in the Fourth Amendment context, the U.S. Supreme Court has recognized that civil and criminal law regimes might share the same ultimate

\footnote{264. A utilitarian theory of punishment assumes that actors commit crimes when it is in their interest to do so. Accordingly, the optimal punishment is one that is just sufficient to make the perceived costs of the undesired act outweigh the anticipated advantages of the act, considering for example the likelihood and likely severity of the punishment. See Steven Klepper & Daniel Nagin, The Deterrent Effect of Perceived Certainty and Severity of Punishment Revisited, 27 CRIMINOLOGY 721 (1989). See generally JEREMY BENTHAM, AN INTRODUCTION TO THE PRINCIPLES OF MORALS AND LEGISLATION (J.H. Burns & H.L.A. Hart eds., 1982) (1789); JOHN STUART MILL, UTILITARIANISM (Longmans, Green, & Co. 1907) (1863); FRANKLIN E. ZIMRING & GORDON J. HAWKINS, DETERRENCE: THE LEGAL THREAT IN CRIME CONTROL (1973).

265. See DRESSLER, supra note 145, at 15 (discussing specific deterrence).

266. See id. (discussing general deterrence).

267. For example, one claimed benefit of our compensatory tort law system is to alter the cost-benefit balance applicable to negligent conduct. When the potential costs of a tort law suit are taken into account, repairing broken sidewalks and designing goods safely are not just kindhearted acts; they are rational. Yet, such damages are considered civil, not criminal, in nature.
purposes, but can nevertheless be distinguished on the basis of their methodologies. In *New York v. Burger*,\(^\text{268}\) for example, the Court upheld a New York City administrative program that authorized police to conduct warrantless searches of businesses involved in the dismantling of automobiles.\(^\text{269}\) Although traditional law enforcement searches are presumed to be unreasonable and in violation of the Fourth Amendment in the absence of probable cause and a warrant,\(^\text{270}\) the Court held that searches pursuant to New York’s administrative scheme were reasonable because they were justified by a “special need” that was separate from the ordinary needs of traditional law enforcement.\(^\text{271}\)

In doing so, the Court expressly rejected the notion that a government program’s purpose was the sole determinant of whether civil or criminal procedural rules governed its constitutionality. New York’s highest court had struck down the statute authorizing the searches in part because the statute targeted criminal activity.\(^\text{272}\) The Court reversed, noting that a state can choose to address major social concerns through both administrative and penal regimes.\(^\text{273}\) “Administrative statutes and penal laws may have the same ultimate purpose of remedying the social problem, but they have different subsidiary purposes and prescribe different methods of addressing the problem.”\(^\text{274}\) New York’s administrative scheme sought to prohibit the sale of stolen vehicle parts by setting forth how businesses should be operated, while the penal law emphasized the punishment of individuals for specific criminal acts.

As demonstrated by the contrasting results of *Burger* and *Ferguson*, there is a critical distinction between the government’s use of traditional


\(^{269}\) Id. at 702. The statute required businesses involved in dismantling automobiles to maintain records of all transactions and to permit police to inspect those records and any automobile parts in the business’s possession. Id. at 694 n.1. The failure to maintain the required records or to cooperate with a requested inspection was punishable as a misdemeanor. Id.

\(^{270}\) See, e.g., Almeida-Sanchez v. United States, 413 U.S. 266, 269–70 (1973) (reiterating previous holdings that probable cause is the Fourth Amendment’s minimum requirement); Katz v. United States, 389 U.S. 347, 357 (1967) (equating the lack of a warrant with unreasonableness); Johnson v. United States, 333 U.S. 10, 13–14 (1948) (discussing the importance of the warrant requirement).

\(^{271}\) *Burger*, 482 U.S. at 701.

\(^{272}\) See id. at 712 (noting that Court of Appeals struck down the New York statute as violative of the Fourth Amendment because it “had no truly administrative purpose but was ‘designed simply to give the police an expedient means of enforcing penal sanctions for possession of stolen property’”) (quoting New York v. Burger, 493 N.E.2d 926, 929 (N.Y. 1986)).

\(^{273}\) Id.

\(^{274}\) Id.
Unpacking New Policing

criminal law methods to accomplish a non-criminal ultimate objective (as in *Ferguson*), and the use of methods outside of the criminal law to accomplish objectives that are shared by the criminal law (as in *Burger*). Just as the administrative regime upheld in *Burger* shared the same ultimate objective as criminal sanctions for possession of stolen property, neighborhood exclusion ordinances share the same ultimate objective as criminal drug laws.\(^{275}\) However, neighborhood exclusion seeks to prevent drug activity through a method that differs qualitatively from arrest and prosecution.

Neighborhood exclusion orders are thought to work not because offenders fear being subject to them, but because they operate to disrupt a retail market. In order for drug transactions to occur, buyers and sellers must locate each other. Once a given area is known for drug activity, individual sellers and buyers know to congregate there to offer and accept the product. Displacing market participants from known retail locations is intended to prevent new transactions not by incapacitating individual participants from attempting to enter into transactions (traditional deterrence), but by making it difficult for the participants to locate each other to transact.

Some might argue that neighborhood exclusion orders will simply move the drug market into surrounding neighborhoods, not disrupt it.\(^{276}\) However, it seems probable that neighborhood exclusion orders both disrupt and displace the targeted activities.\(^{277}\) The drug trade (like gang and vice activity) depends on group behavior. As one criminologist has suggested, drug dealing is most successful when transacted within a well-known marketplace associated with reliable participants.\(^{278}\) With only word of mouth to rely on, participants in the drug trade cannot

\(^{275}\) With respect to prostitution free zones, the ordinances creating the zones share the same ultimate objective as criminal laws against prostitution.

\(^{276}\) Schragger, for example, argues that the power to zone empowers one neighborhood to displace unwanted people and activities to surrounding neighborhoods, “a probably limited and short-lived success.” Schragger, supra note 8, at 429.

\(^{277}\) In Portland, for example, drug-related arrests within a defined zone decreased by an average of thirty-five percent after implementation of the exclusion program. However, on average, drug arrests in areas surrounding the targeted zone increased twelve percent. Multnomah County Neighborhood DA Unit, at http://www2.co.multnomah.or.us/da/NDAP/index.cfm?fuseaction=strategies&menu=19&title=Drug %20Free%20Zones%20%28DFZ%29 (last visited Sept. 30, 2003) (on file with author).

collectively and immediately open shop in any single new location. In the process of relocating, at least some transactions must be lost. 279

Moreover, even if displacement were the only effect of neighborhood exclusion ordinances, displacing crime out of high-crime neighborhoods may be both an effective and normatively appropriate crime-control strategy. If one believes that low-level crimes reach a “tipping point” at which they lead to more crime, such offenses may be less detrimental to a community if maintained beneath a threshold at which they begin to deteriorate community confidence and overall quality of life. 280

Furthermore, from a normative perspective, there is utility in sparing a neighborhood from bearing a disproportionately heavy burden of either a jurisdiction’s crimes 281 or its rigorous police activities. By disbursing a city’s crimes more equitably among neighborhoods, displacement can prevent relatively sheltered residential pockets from writing off street-level crimes as a problem only for the “inner-city.” And if policing efforts must spread into other neighborhoods to follow the crime, that move could repeal what Randall Kennedy has called the disproportionate “tax” that racial minorities currently pay for our various wars against criminality. 282

Note that both the “neighborhood victimization” and “market disruption” arguments, which distinguish neighborhood exclusion orders from traditional punishment, apply only when the exclusions are limited to neighborhoods with significant, visible activity. To argue that exclusion orders are intended to disrupt a market, the city must show the existence of such a market within the zone. Similarly, it is difficult to argue that an entire neighborhood is victimized by a single, isolated drug

279. See Leipold, supra note 216, at 497 (arguing that anti-gang loitering laws do more than move gang activity, at least in the short term, because groups need time to reform and relocate).

280. See supra notes 22–26 and accompanying text for a discussion of the broken windows theory, which posits that visible disorder can lead to further neighborhood deterioration.

281. There is a stronger normative argument for displacing crime out of high-crime neighborhoods than out of relatively trouble-free areas. Cf. Robert W. Helsley & William C. Strange, Gated Communities and the Economic Geography of Crime, 46 J. URB. ECON. 80 (1999) (using an economic model to show that gated communities divert crime to other neighborhoods).

282. See Randall Kennedy, Suspect Policy, THE NEW REPUBLIC, Sept. 13–20, 1999, at 30, 34 (arguing that the cessation of racial profiling would repeal the “racial character” of the current “tax” for the wars against illegal immigration and drugs); see also DAVID COLE, NO EQUAL JUSTICE: RACE AND CLASS IN THE AMERICAN CRIMINAL JUSTICE SYSTEM 54 (1999) (arguing that “well-to-do white people” would be more likely to pressure police to regulate themselves if they were routinely subject to police-citizen encounters); Leipold, supra note 216, at 497 (arguing that moving gang activity from one neighborhood into another “should help spread the cost of law enforcement more equitably”).
Unpacking New Policing

transaction within an otherwise drug-free neighborhood, and it is therefore difficult to liken members of that neighborhood to the traditional categories of victims protected by civil restraining orders. 283

The broken windows theory posits that low-level disorder, if left uncorrected, leads to serious crime. 284 Whether one believes the link between disorder and crime is direct or indirect, the disorder must be visible under the broken windows theory in order to affect the conduct of either law-abiding members of the community or would-be troublemakers. Unlike a broken window, isolated drug transactions can go unnoticed. In order for drug activity to destabilize the community through either the direct or indirect mechanisms proffered by Wilson and Kelling, the activity would need to reach a critical mass of visibility. 287 When that point has been reached, neighborhood exclusion for offenders can be justified as a remedy to the neighborhood. Immediate and temporary exclusion of former drug market participants prevents potential future actors from perceiving an ongoing market and believing that no one in the community cares if further criminal activity occurs there. Moreover, with drug offenders excluded from the neighborhood, residents will feel safer using their neighborhood in lawful ways. 288

283. See supra notes 261–63 and accompanying text (comparing neighborhood exclusion orders to abuse prevention restraining orders).

284. See supra notes 22–26 and accompanying text (summarizing the broken windows theory).

285. The direct mechanism of Wilson and Kelling’s theory posits that a broken window, if left unrepaired, signals to lawbreakers that “no one cares” and leads to the rest of the windows being broken. See supra note 25.

286. As an indirect cause of serious crime, low-level disorder can create fear among community residents, who will begin to stay indoors, withdraw from their community, and eventually stop trying to assert control over the neighborhood. See supra note 24.

287. In this respect, crime can follow epidemiological patterns observed for public health risks. In his influential article, The Tipping Point, Malcolm Gladwell theorized that crime, like a low-level flu outbreak, can develop non-linearly. Applied to crime, the tipping point theory suggests that crime might develop linearly until a small incremental increase brings it to a critical point leading to a severe escalation in crime rates. Malcolm Gladwell, The Tipping Point: How Little Things Can Make a Big Difference (2000); Malcolm Gladwell, The Tipping Point, The New Yorker, June 3, 1996, at 32; see also Greene, supra note 28, at 181 (discussing the tipping point phenomenon); George L. Kelling, Why Did People Stop Committing Crimes? An Essay About Criminology and Ideology, 28 Fordham Urb. L.J. 567, 578 (2000) (maintaining that crime decreased in New York City because it declined to a tipping point at which crime plummeted).

contrast, it would be difficult to justify neighborhood exclusion as a remedy to an entire community in a low crime area where occasional drug activity occurs below the public’s threshold of perception.289

4. Not Historically Regarded as Criminal Punishment

For many of the same reasons that neighborhood exclusion serves a civil purpose, neighborhood exclusion does not resemble traditional criminal punishment. Opponents of neighborhood exclusion have argued, and at least one court has agreed, that neighborhood exclusion amounts to banishment, a traditional criminal punishment.290 Traditional banishment was intended to incapacitate the banished individual and protect citizens within the banishing jurisdiction.291 Traditional banishment is undoubtedly a criminal punishment, one that generally is considered unlawful, even as a criminal sentence.292 Banishment has been criticized for its failure to rehabilitate the offender,293 for permitting

289. Debra Livingston has also observed that isolated occurrences of disorder are not themselves destructive of a community. For example, she notes that a single person loitering on a corner does not threaten the neighborhood’s quality of life, but the loitering of many on a single street is likely to trigger decline. See Livingston, supra note 6, at 559; see also Wilson & Kelling, supra note 22, at 35 ("Arresting a single drunk or a single vagrant who has harmed no identifiable person seems unjust, and in a sense it is. But failing to do anything about a score of drunks or a hundred vagrants may destroy an entire community.").

290. See Johnson v. City of Cincinnati, 119 F. Supp. 2d 735, 748–49 (S.D. Ohio 2002) (finding a double jeopardy violation where individual was both convicted of a criminal offense and excluded from a drug free zone for the same underlying conduct), aff’d on other grounds, 310 F.3d 484 (6th Cir. 2002), cert. denied, 123 S. Ct. 2276 (2003).

291. See Snider, supra note 288, at 481 (noting that banishment was most likely intended to further the goals of incapacitation and community protection).

292. See generally Leipold, supra note 216, at 490–91 (noting that many states have abolished the practice of banishment); Snider, supra note 288, at 484–99 (discussing bases for disallowing banishment); Jason S. Alloy, Note, “158-County Banishment” in Georgia: Constitutional Implications Under the State Constitution and the Federal Right To Travel, 36 GA. L. REV. 1083, 1087 (2002) (noting that courts have held banishment to be illegal because it does not rehabilitate the banished, because it permits the banishing jurisdiction to unload its undesirables upon other jurisdictions, and because it has been held to violate various state constitutions). At least fifteen states have provisions in their constitutions that expressly prohibit the banishment of individuals from the state. Snider, supra note 288, at 465 & n.70 (citing state constitutional provisions prohibiting banishment). Even in the absence of an express prohibition against banishment from a state, other state courts have held that sentences of banishment are unlawful on public policy grounds. Id. at 465–66 & n.78 (citing cases finding banishment conditions unlawful and opining that the absence of express constitutional prohibitions against banishment in the applicable state constitutions may explain the courts’ use of public policy grounds to strike down the banishment conditions).

293. Toni M. Massaro, Shame, Culture, and American Criminal Law, 89 Mich. L. Rev. 1880, 1895 n.75 (1991) (noting that permanent exile appears reserved for those beyond rehabilitation and
jurisdictions to dump their criminals on neighboring jurisdictions, for depriving the banished individual of any community, and as a punishment unauthorized by the legislature.

Any resemblance between contemporary neighborhood exclusion and traditional banishment is superficial at best. This point is best made by comparing both sanctions to prison, the quintessential criminal that the only possible rehabilitative justification for such banishment would be if the offender were so disappointed in his banishment from his original community that he would behave in the future so that he would not also lose the community to which he was moved; Snider, supra note 288, at 478 ("It is highly unlikely that banishment has either as its intended effect, or actually effectuates, any sort of rehabilitation."); see also State v. Franklin, 604 N.W.2d 79, 83 (Minn. 2000) (excluding defendant from entire city was not reasonably related to goal of rehabilitating him); McCreary v. State, 582 So. 2d 425, 428 (Miss. 1991) (observing that banishment from a large geographical area does not serve a rehabilitative function); Johnson v. State, 672 S.W.2d 621, 623 (Tex. Ct. App. 1984) (noting that banishing defendant from county would not assist in rehabilitating him); Ray v. McCoy, 321 S.E.2d 90, 93 (W. Va. 1984) (noting that banished persons are not likely to be rehabilitated because they lack supervision while banished).


295. The U.S. Supreme Court has held that deprivation of citizenship constitutes cruel and unusual punishment prohibited by the Eighth Amendment, in part because it “strips the citizen of his status in the national and international political community.” Trop v. Dulles, 356 U.S. 86, 101 (1958). It is perhaps due to similar concerns about the loss of community that courts have generally struck down any sentence that banishes an offender from an entire or most of a state, while upholding sentences of “banishment” that apply to only a small geographic area. See Snider, supra note 288, at 473 (noting that appellate courts have upheld probation conditions that restrict the probationer’s ability to enter a geographic area “if the banishment is confined to a small geographic area”); see also United States v. Cothran, 855 F.2d 749, 752 (11th Cir. 1988) (upholding probation condition prohibiting defendant from entering the county without the permission of his probation officer); Dukes v. State, 423 So. 2d 329, 331 (Ala. Crim. App. 1982) (upholding probation condition prohibiting probationer from entering the store from which she stole); State v. Morgan, 389 So. 2d 364, 366 (La. 1980) (upholding probation condition that required defendant, convicted of attempted prostitution, to stay out of New Orleans’s notorious French Quarter); State v. Harrington, 336 S.E.2d 852, 857 (N.C. Ct. App. 1985) (upholding probation condition restricting drunken driver’s ability to enter establishments selling alcoholic beverages).

296. Some courts have refused to impose banishment because the power to impose banishment belongs to the legislature and is not inherent in the judiciary. See, e.g., Cooper v. Telfair, 4 U.S. (4 Dall.) 14, 19 (1800) (holding that “the power of confiscation and banishment does not belong to the judicial authority . . . and yet, it is a power, that grows out of the very nature of the social compact, which must reside somewhere, and which is so inherent in the legislature”); Rutherford, 468 F. Supp. at 1360 (stating that “the power to banish, if it exists at all, is a power vested in the Legislature”); Weigand v. Commonwealth, 397 S.W.2d 780, 781 (Ky. Ct. App. 1965) (holding that it is not within court’s authority to impose banishment in lieu of prison sentence); People v. Wallace, 124 N.Y.S.2d 201, 204 (N.Y. Co. Ct. 1953) (holding that banishment cannot be imposed by judiciary). Neighborhood exclusions, on the other hand, result from legislative action.
punishment. Even traditional banishment is distinguishable from prison in that it does not literally incapacitate; it simply prohibits the individual from offending within the banishing jurisdiction. However, traditional banishment does serve the same function as prison if viewed solely from the self-interested perspective of the banishing jurisdiction. As long as the banished individual offends outside the geographic boundaries of the banishing jurisdiction, the citizens within the banishing jurisdiction are as protected from him as if the individual were in prison.

In contrast, neighborhood exclusion does not attempt to segregate the offender from the entire constituency of the banishing political entity. Rather, it simply seeks to move the offender within the jurisdiction from one geographic subsection into others. Although the offender may be incapacitated with respect to a discrete geographic neighborhood, the offender still remains a part of the citizenry of the larger governmental entity that issues the neighborhood exclusion. Moreover, if the offender does reoffend during the exclusion term, it is the jurisdiction that issued the exclusion order that bears the burden of the offense, not some other jurisdiction whose citizens were not represented in the decision to order the exclusion.

Neighborhood exclusion is also distinguishable from traditional banishment because it does not deprive the excluded individual of her community. As one commentator wrote recently about traditional banishment,

In its most general and benign form, banishment is the punishment of one who has incurred the displeasure of a group to which one had previously enjoyed full membership status. As a means of expressing displeasure with the conduct of the banished, the community takes the ultimate step and declares that the banished individual is no longer a part of the community. . . . [T]he banished individual presumably retains no rights of membership in the community from which he/she was banished.

297. For the same reason that neighborhood exclusion orders do not further incapacitation, a traditional goal of criminal punishment, the orders do not deprive offenders of the political process. Although excluded from a geographic subsection of the city, excluded individuals retain their membership within recognized political entities. But see Snider, supra note 288, at 495 (arguing that traditional banishment deprives individuals of their ability to participate in the political process).

298. See supra notes 276–82 and accompanying text for the potential of neighborhood exclusion ordinances to displace crime from a high-crime neighborhood into surrounding areas.

299. Snider, supra note 288, at 476.
Unpacking New Policing

When a neighborhood exclusion order is issued, the geographic area from which the affected individual is excluded is not necessarily that individual’s own neighborhood. If the intention of the ordinance were to punish the excluded individual by depriving him of his community, then the scope of the exclusion would be determined by the individual’s place of residence. In contrast, neighborhood exclusion orders attempt to displace crime from high-crime areas primarily by closing the neighborhood off to community outsiders who enter the area and commit crimes there. For example, a wealthy suburbanite who leaves her own community to purchase drugs in a high-crime area cannot claim that an exclusion from the high-crime area deprives her of her community. When the affected individual does happen to reside within the zone, she can apply for a variance permitting her to move between essential locations within the zone during the period of the exclusion.300

Finally, neighborhood exclusion orders are far less burdensome than traditional banishment. Traditional banishment was permanent and required the individual to leave the country,301 an entire state, or a large part of a state.302 Neighborhood exclusion orders, on the other hand, are of a limited duration303 and apply to only discrete neighborhoods within a city.304

In Johnson v. City of Cincinnati,305 the federal district court held that neighborhood exclusion was criminal punishment, likening it to banishment.306 However, the court’s analysis was inconsistent with Hudson. For example, the court reasoned that neighborhood exclusion was just as punitive as banishment because it prevented excluded individuals from seeking employment within designated zones during the time of exclusion.307 However, in Hudson, the Court held that an

300. See State v. James, 978 P.2d 415, 419 (Or. Ct. App. 1999); see also Leipold, supra note 216, at 491 (arguing that targeted loitering laws do not equate to banishment).
301. See James, 978 P.2d at 419.
302. See Snider, supra note 288, at 466.
303. See James, 978 P.2d at 419.
304. See id.
306. Id. at 748. The Sixth Circuit affirmed the district court, but on the grounds that the ordinance violated the plaintiffs’ rights to intrastate travel and to association. See Johnson v. City of Cincinnati, 310 F.3d 484, 506 n.10 (6th Cir. 2002) (noting that the court need not decide whether neighborhood exclusion constitutes criminal punishment in light of the case’s disposition of plaintiffs’ other issues), cert. denied, 123 S. Ct. 2276 (2003).
individual’s lifetime disbarment from his chosen profession did not amount to criminal punishment. Even courts adopting a seemingly expansive concept of criminal punishment have distinguished between impairment of employment—which is not criminal punishment under *Hudson*—and restrictions that render an individual entirely unemployable.\(^{308}\) A temporary exclusion from small geographic areas of a city does not significantly impair employment activities, let alone render the individual unemployable.

5. **Not Excessive in Relation to Civil Purposes**

Finally, the exclusion orders are not excessive in light of the market-disrupting purpose of neighborhood exclusion ordinances.\(^{309}\) The nature of the exclusions demonstrates that they are intended not to punish the excluded individuals by depriving them of access to the zone, but simply to disrupt their ability to participate in the retail market within the zone. The exclusions are short-term, and even during the period of exclusion, individuals may receive variances to enter the zones for legitimate purposes.

In sum, a short-term neighborhood exclusion order does not appear to constitute criminal punishment under current jurisprudence, at least not when applied in high-crime neighborhoods where exclusion serves non-punitive purposes.\(^{310}\) Accordingly, neighborhood exclusion ordinances should not be subject to the constitutional jurisprudence governing criminal law and procedure.

C. **Recharacterizing Morales**

Although the ordinances are concerned with qualitatively different harms, neighborhood exclusion ordinances and the Chicago anti-gang

\(^{308}\) See Doe I v. Otte, 259 F.3d 979, 988 (9th Cir. 2001), rev’d sub nom. Smith v. Doe, 538 U.S. 84, 123 S. Ct. 1140, rehe’d denied, 123 S. Ct. 1925 (2003). In Doe I, the Ninth Circuit struck down Alaska’s sex offender registration program, holding that it constituted criminal punishment, in part because the registrants’ offender status was easily discoverable by the public and therefore likely to prevent registrants from obtaining employment of any kind. In doing so, the court distinguished this aspect of the program at issue in Doe I from the disbarment from a single profession found to be civil in *Hudson*. Id. The U.S. Supreme Court subsequently reversed the Ninth Circuit’s decision, in part because sex offenders remained free to change employment. *Smith*, 123 S. Ct. at 1151.

\(^{309}\) See *James*, 978 P.2d at 420–21.

\(^{310}\) See *supra* text accompanying notes 283–89 for a discussion of the reasons why neighborhood exclusions should be limited to high-crime areas in order to be considered non-punitive.
Unpacking New Policing

loitering ordinance at issue in Morales share some fundamental similarities. Both ordinances seek to protect neighborhoods by authorizing police to constrain the movement of individuals seen as disruptive to those neighborhoods. 311 Chicago’s ordinance authorized the movement of suspected gang members and those loitering with them, while Portland’s ordinance authorizes the exclusion of suspected drug offenders. 312 Moreover, under both ordinances, police orders did not result in arrest and prosecution unless the targeted individual disobeyed the orders by failing to move along or by re-entering a drug free zone, respectively. 313

Part III.B notes the strong argument that neighborhood exclusion should be treated as a civil restraint. Additionally, at least some of the few lower courts that have addressed that issue have held that neighborhood exclusion does not constitute criminal punishment. 314 From that perspective, the exclusion order is a civil restraint that should be evaluated as such, and the rules governing criminal laws do not come into play until the excluded individual is arrested for criminal trespass for violating the terms of the exclusion order. Applying that model to Chicago’s anti-gang loitering ordinance, a police order to disperse from the site of gang-loitering should be evaluated as a civil restraint, while an arrest or prosecution for failing to obey that order should be analyzed under the rules governing the criminal justice system.

Nevertheless, the U.S. Supreme Court was quick to analyze the entirety of Chicago’s anti-gang loitering ordinance as a criminal statute. 315 For example, in concluding that it was “clear” that the plaintiffs could bring a facial challenge to the ordinance, the Court stated simply that the ordinance was “a criminal law.” 316 The Court provided little analysis to support this conclusion. The city argued that only the “arrest” component of the statute should be scrutinized for vagueness,

311. See text accompanying supra notes 50–53 and 207–16 for a summary of Chicago’s anti-gang loitering ordinance and Portland’s drug free zone ordinance, respectively.
313. See id. (summarizing CHI., ILL., MUNICIPAL CODE § 8-4-015 (2002) and PORTLAND, OR., CITY CODE § 14B.20.030(A)).
314. State v. Lhasawa, 55 P.3d 477, 488 (Or. 2002); James, 978 P.2d at 417–21.
315. See Debra Livingston, Gang Loitering, the Court, and Some Realism About Police Patrol, 1999 SUP. CT. REV. 141, 184–85 (criticizing the Court’s failure in Morales to credit Chicago’s ordinance for requiring a “move along” order prior to arrest).
316. Morales, 527 U.S. at 55 (stating that the ordinance was “a criminal law that contains no mens rea requirement,” rendering a facial challenge permissible).
and the arrest portion of the ordinance clearly set forth that it was a crime to disobey a dispersal order issued under the ordinance. The majority rejected the city’s argument because the ordinance gave insufficient guidance to police officers deciding whether to issue the dispersal orders in the first place. However, the Court failed to explain why it was demanding more particularity to govern the issuance of dispersal orders than would be required generally in any other non-criminal context.

A plurality of Justices at least attempted to articulate a governing principle for determining whether an ordinance should be scrutinized for vagueness, stating that it was the loitering component of the statute that should be scrutinized for vagueness, not just the refusal to obey a dispersal order, because “the loitering is the conduct that the ordinance is designed to prohibit.” But, as the dissenting Justices correctly noted, what should matter for purposes of vagueness review “is not what the ordinance is ‘designed to prohibit,’ but what it actually subjects to criminal penalty.” The ultimate objective of a statute cannot be what determines whether it is treated as a criminal law; otherwise, in the special needs context, administrative search provisions aimed at chop shops would be treated as criminal, and drug-testing programs designed to encourage substance abuse treatment would be treated as civil.

317. Id. at 58.
318. See id. at 62 (noting that the ordinance’s authorization for police offers to arrest only after a dispersal order has been disobeyed “does not provide any guidance to the officer deciding whether such an order should issue”).
319. The U.S. Supreme Court is more tolerant of imprecision in civil ordinances than those that impose criminal penalties “because the consequences of imprecision are qualitatively less severe.” Vill. of Hoffman Estates v. Flipside, Hoffman Estates, Inc., 455 U.S. 489, 498–99 (1982). Similarly, because of the qualitative differences between criminal and civil punishment, the rule of lenity applies only to criminal statutes. See United States v. Bass, 404 U.S. 336, 348 (1971) (noting that ambiguous criminal statutes should be construed narrowly “because of the seriousness of criminal penalties, and because criminal punishment usually represents the moral condemnation of the community, legislatures and not courts should define criminal activity”) (citations omitted).
320. See Morales, 527 U.S. at 58 (Stevens, J., joined by Souter and Ginsburg, J.J.).
321. Id. at 90 (Scalia, J., dissenting).
Unpacking New Policing

One could argue that the Court treated Chicago’s dispersal orders as criminal in nature because Chicago placed both the authority to issue the orders and the prohibition for violating the orders in the same criminal ordinance. The question of whether a law is criminal turns on legislative intent, and the Chicago City Council included dispersal orders in the same ordinance as a criminal penalty. However, it does not appear that the Court envisioned the ordinance’s constitutional flaw as a mere labeling problem. As an initial matter, because the Court wholly omitted any analysis of whether the ordinance was civil or criminal, it certainly did not state that it was relying on the inclusion of the authority for the dispersal orders in the same law as a criminal penalty. Moreover, the Court suggested ways in which Chicago might cure the ordinance’s flaws, but never suggested it would be so simple as enacting two separate ordinances: a civil provision authorizing police to order loiterers to disperse and a criminal ordinance prohibiting the failure to heed a dispersal order.

The problem in Morales, then, appears to have been the lawfulness of the underlying dispersal orders. Although the Court relied on the void for vagueness doctrine to find the dispersal orders unlawful, it did so without articulating a convincing basis for treating dispersal orders as criminal sanctions. Alternatively, the Court could have treated the dispersal orders in Morales as civil sanctions, unlike the criminal law that prohibited disobeying the dispersal orders. Had the Court done so, the constitutionality of the dispersal orders could still have been questioned, but the void for vagueness doctrine would not be seen as a

323. See supra notes 221–26 for a summary of the two-prong test of statutory construction that governs whether a law imposes criminal punishment. The first step in the analysis requires the court to ask whether the legislature indicated a preference for either the civil or criminal label in establishing the penalty. Hudson v. United States, 522 U.S. 93, 99 (1997).

324. See Morales, 527 U.S. at 62 (noting that the ordinance would be lawful if it “only applied to loitering that had an apparently harmful purpose or effect, or possibly if it only applied to loitering by persons reasonably believed to be criminal gang members”).

325. It is unclear whether the new discretion scholars view dispersal orders as civil or criminal in nature. For example, Debra Livingston argues that some effective public safety efforts could be made by using civil sanctions and injunctions, but then appears to assume that a law requiring police to ask citizens to cease their behavior before issuing a citation is criminal in nature. Livingston, supra note 6, at 637 (noting that discretion to enforce criminal prohibitions could be limited by “requiring that people be requested to cease specified behavior before citation or arrest is authorized”). On the other hand, Livingston maintains that cities could treat loitering as “a civil infraction for which arrest and prosecution are permissible only when a loiterer refuses to comply with a police request to move along or fails to respond to a citation for civil infraction.” Id. at 639 (noting the potential to treat school loitering and loitering with intent to engage in prostitution or in drug distribution in this manner).
threat to new policing approaches, at least those that depart from
criminal punishment. In other words, legislatures would be permitted to
entrust police with discretion as long as that discretion did not include
decisions about whom to arrest on criminal charges.

By relying improperly on the void for vagueness doctrine, the Court
permitted itself to avoid deciding the difficult issues that would have
been presented if the Court had evaluated Chicago’s anti-gang loitering
ordinance as a civil sanction. For example, unlike the neighborhood
exclusion orders issued under drug free zone ordinances, the orders to
disperse authorized by Chicago’s anti-gang loitering ordinance were
delivered orally and informally and could not be challenged. Under a
due process inquiry, the Court would have had to balance the nature of
the private interest involved, the risk of erroneous deprivation of such an
interest given the nature of the procedure used, and the nature of the
governmental interest involved.

Moreover, the Court would have had to determine whether laws
affecting an individual’s freedom of movement in public places
implicate a protected constitutional right to travel or, more apt in
Morales, a right not to travel (i.e., to loiter). Laws that implicate
fundamental rights are subject to strict judicial scrutiny. Moreover,

appealing neighborhood exclusion order under Portland’s drug free zone ordinance); see also State
procedures did not deprive the defendant of due process, even though the exclusion notice did not
explain that an appeal of the order would stay the order, because the notice explained the right to
appeal and process for appealing); State v. James, 978 P.2d 415, 421 (Or. Ct. App. 1999) (upholding
drug free zone ordinance against due process challenge because “[i]t is difficult to imagine a greater
procedural protection than a predeprivation hearing”).

327. Mathews v. Eldridge, 424 U.S. 319, 335 (1976). Laurence Tribe has also noted the difficulty
of identifying the right protected by the Court in Morales. To Tribe, the Court’s ruling could not
have been a matter of procedural due process, because procedural concerns would have been
implicated only if the individual was denied an adequate hearing on the factual issue of whether he
had disobeyed a dispersal order. Laurence H. Tribe, Comment, Saenz Sans Prophecy: Does the
Privileges or Immunities Revival Portend the Future—Or Reveal the Structure of the Present?, 113
Harv. L. Rev. 110, 192 (1999). His analysis gives short shrift to the argument that the problem was
the absence of any adequate procedural protections concerning the factual issue of whether the
individual should have been the subject of a dispersal order for loitering.

328. Laws implicating a fundamental liberty are subject to strict scrutiny. See, e.g., Planned
Parenthood v. Casey, 505 U.S. 833, 848 (1992) (finding unconstitutional a state statute that required
women to notify their husbands before having an abortion, except in limited circumstances); Moore
v. City of E. Cleveland, 431 U.S. 494, 500 (1977) (invalidating zoning ordinance limiting
occupancy to members of a single “family,” defined narrowly); Griswold v. Connecticut, 381 U.S.
479, 485–86 (1965) (striking down a state statute that made it illegal to use or counsel others to use
contraceptives); Skinner v. Oklahoma, 316 U.S. 535, 541 (1942) (declaring state statute authorizing
courts are less tolerant of vagueness in laws—civil or criminal—that might restrain constitutionally protected activities. Accordingly, if the Court had found a fundamental right to loiter in *Morales*, it could have struck down the law without suggesting that a dispersal order constitutes a criminal sanction.

Lower courts are divided over the issue of whether the constitutional right to travel protects a freedom of movement intrastate. To date, the U.S. Supreme Court’s jurisprudence regarding the right to travel has been limited to cases involving interstate travel. The Court has never recognized a right to travel within a state, and the nature of the interstate right to travel suggests that the right might not be implicated by restrictions on intrastate travel.

Although the Court bypassed the opportunity to make a clear pronouncement in *Morales* regarding the existence of a fundamental right *not* to travel, the myriad of opinions in that case suggests that a majority of the Court would permit cities to restrict individuals’ freedom to travel.
of intrastate movement. Although six Justices voted in *Morales* to strike down Chicago’s anti-gang loitering ordinance on vagueness grounds, the defendants in *Morales* failed to gain the support of a majority of the Court for their argument that the Chicago ordinance violated a constitutional right to loiter. Only three members of the Court were prepared to recognize a fundamental “freedom to loiter for innocent purposes.” Three dissenting Justices made clear their rejection of such a right.

While the three concurring Justices did not expressly address the issue of a right to loiter, Justice O’Connor’s concurring opinion, in which Justice Breyer joined, suggests at least two additional votes against a fundamental right to loiter. While O’Connor and Breyer purported to “express no opinion about” issues other than the ordinance’s vagueness, they made a point of noting that “there remain open to Chicago reasonable alternatives to combat the very real threat posed by gang intimidation and violence.” Specifically, the Justices indicated that an anti-loitering ordinance would pass muster if it targeted only gang members or applied only in specific geographic areas. If such ordinances would be lawful, then either there is no fundamental right to loiter, or this supposed right to “loiter for innocent purposes” is an entirely different kind of fundamental right than those that trigger the strictest judicial scrutiny under substantive due process or equal protection jurisprudence. Indeed, even the Justices who supposedly

---

334. City of Chicago v. Morales, 527 U.S. 41, 60–64 (1999) (six-member opinion of the Court, holding that anti-gang loitering ordinance entrusted too much discretion to police officers to determine which types of loitering “purposes” were lawful).
335. *Id.* at 53 (Stevens, J., with Souter and Ginsburg, J.J., joining).
336. *See id.* at 102–06 (Thomas, J., dissenting, with Rehnquist and Scalia, J.J., joining).
337. *See id.* at 64–69 (O’Connor, J., concurring in part, with Breyer, J., joining).
338. *Id.* at 67.
339. *Id.*
recognized a constitutional right to loiter did not appear to give this right the same weight as those previously recognized as fundamental, suggesting that the right could be limited by narrower—but nevertheless quite restrictive—ordinances. 342

The goal of this Article is not to answer the ultimate question of whether Chicago’s anti-gang loitering ordinance would pass constitutional muster if analyzed as a civil statute. Rather, the Article recasts the dispersal orders authorized by the ordinance as civil restraints for the purpose of demonstrating that the Court could have preserved a judicial role in scrutinizing the dispersal orders without insisting inappropriately on a level of legislative clarity typically reserved for criminal statutes and laws that implicate fundamental constitutional rights.

D. Applying the Programmatic Purpose Approach

The political process theory offered by Kahan and Meares and my recommended programmatic purpose approach sometimes will lead to the same result, albeit for different reasons. For example, Kahan and Meares attribute the government’s ability to erect sobriety checkpoints343 and to search individuals at airports344 and in government buildings345 to the fact that these searches burden average members of the community.346 However, the evenhanded application of the searches is defendants access to state courts).

342. See Morales, 527 U.S. at 62 (noting that the ordinance would be lawful if it applied only to loitering with an apparently harmful purpose or effect, “or possibly if it only applied to loitering by persons reasonably believed to be criminal gang members”).

343. Mich. Dep’t of State Police v. Sitz, 496 U.S. 444, 455 (1990) (upholding brief, suspicionless seizures of motorists at sobriety checkpoints aimed at removing drunk drivers from the road); see also Delaware v. Prouse, 440 U.S. 648, 657 (1979) (suggesting that roadblocks could be used to verify drivers’ licenses and vehicle registrations).

344. United States v. Edwards, 498 F.2d 496, 498 (2d Cir. 1974) (upholding airport security checkpoints as reasonable by balancing individual liberty interests against government’s interest in preventing hijacking); United States v. Davis, 482 F.2d 893, 908–12 (9th Cir. 1973) (analyzing search at airport screening checkpoint as an administrative search); United States v. Moreno, 475 F.2d 44, 48–49 (5th Cir. 1973) (noting that most airport searches are justified by the exigent circumstances presented by the need for on-flight safety).

345. See United States v. Bulacan, 156 F.3d 963, 967–73 (9th Cir. 1998) (recognizing that searches of people entering government buildings can be justified if limited in scope to fulfill the government’s legitimate need to discover weapons and explosives); Downing v. Kunzig, 454 F.2d 1230, 1232–33 (6th Cir. 1972) (upholding search at government building as reasonable in light of government’s interest in protecting federal employees and property).

346. See Kahan & Meares, supra note 39, at 1172–73; see also Meares & Kahan, supra note 115,
just one factor in determining its programmatic purpose. Another critical factor in all of these examples is the close connection between the government action and an interest other than traditional criminal investigation. For example, searches occur at the entrances of government buildings and airport terminals to ensure that no one enters with weapons or explosives, not for the primary purpose of discovering evidence to be used in a criminal prosecution.347 Similarly, stopping motorists on roadways is a way to ensure that drivers are sober and licensed.348 In contrast, a roadblock erected simply to determine whether any cars contain evidence of a crime does not serve any government objective other than traditional criminal investigation.349 Therefore, under the programmatic purpose theory, the roadblock would be subject to normal Fourth Amendment requirements.350 Under Kahan and Meares’s political process approach, such a roadblock would presumably be permissible as long as it burdens average members of the community.

Similarly, consider Kahan and Meares’s approval of the Chicago Housing Authority’s (CHA) building search policy.351 Under the policy, CHA police were permitted to authorize warrantless sweeps of the homes of residents of Housing Authority developments under certain circumstances, including to locate weapons after shooting incidents.352

at 255 (explaining that roadblocks and searches at airports and government buildings are best explained under their political process theory because “insofar as these policies do burden average members of the community, there is much less reason for courts to doubt the determination of politically accountable officials that these policies strike a fair balance between liberty and order”).

347. See Bulacan, 156 F.3d at 967–73 (holding search for drugs at the entrance of a federal building to be outside the lawful scope of an administrative search justified by the government’s interest in locating weapons and explosives and not by a general interest in finding evidence of criminal activity).

348. See supra note 343.

349. City of Indianapolis v. Edmond, 531 U.S. 32, 44 (2000) (holding that a roadblock erected to facilitate drug interdiction did not advance a special need and was “primarily for the ordinary enterprise of investigating crimes”).

350. This is the current approach under the U.S. Supreme Court’s jurisprudence. Id.


352. Under the policy, Chicago Housing Authority police were permitted to conduct warrantless searches of the homes of residents of Housing Authority developments under certain circumstances. Id. at 178. The requisite preconditions included “random gunfire from building to building and/or intimidation at gunpoint or by shooting if weapons were taken into buildings.” Id. If the police could not determine into which apartment the weapons were taken, the policy authorized extensive “searches” of all residential apartment units in the building, which included looking in closets and drawers and searching residents’ personal effects. Id.
Unpacking New Policing

Applying their political process approach, Kahan and Meares argue that the CHA building search policy should be upheld because the burden of unannounced searches fell on everyone in the projects, not just on persons suspected of wrongdoing, and because their representatives had approved the policy. 353 Under the programmatic purpose approach, in contrast, the central inquiry is whether the policy advances a purpose other than ferreting out evidence of criminal activity to be used in criminal proceedings. For example, if the government used evidence discovered during the searches against tenants in criminal prosecutions, but did nothing with respect to their tenancy, the searches would appear to promote traditional criminal investigation and prosecution. If, on the other hand, evidence from the searches was used to support eviction from public housing, the searches would appear designed to ensure that housing units were given to tenants who were not engaged in criminal activity on the property. 354 The fact that the CHA searches were triggered by individual occurrences of crime suggests a primary motive to gather evidence to support an arrest and prosecution, not a “special need” to maintain the safety of the developments on an ongoing basis.

Unlike the political process theory, the programmatic purpose approach would permit cities to implement strategies requiring police discretion, as long as those strategies avoided traditional criminal investigation, prosecution, and punishment. The discussion of Portland’s

---

353. See Kahan & Meares, supra note 39, at 1175. See supra notes 183–84 and accompanying text for further discussion of Kahan and Meares’s analysis of the CHA building search policy.

354. The heart of the special needs exception to the Fourth Amendment’s warrant and probable cause requirements is the government’s need to conduct searches for purposes other than simply discovering evidence of a crime. See, e.g., Bd. of Educ. v. Earls, 536 U.S. 822, 838 (2002) (upholding as reasonable drug testing of students to ensure that drug-affected students were not involved in extracurricular activities); Vernonia Sch. Dist. v. Acton, 515 U.S. 646, 658 (1995) (upholding suspicionless drug tests of student athletes partly because the results of the tests were not turned over to law enforcement authorities); Nat’l Treasury Employees Union v. Von Raab, 489 U.S. 656, 666 (1989) (upholding suspicionless drug testing of certain sensitive Customs Agents partly because it was “clear that the . . . program [was] not designed to serve the ordinary needs of law enforcement”); Skinner v. Ry. Labor Executives’ Ass’n, 489 U.S. 602, 621 n.5 (1989) (upholding a suspicionless drug testing scheme aimed at railroad employees because it was not designed as a pretext to gather evidence for law enforcement purposes); O’Connor v. Ortega, 480 U.S. 709, 721–21 (1987) (suggesting that Fourth Amendment’s warrant requirement should not apply when the government searches its employees’ workspace for evidence of work-related misconduct); New Jersey v. T.L.O., 469 U.S. 325, 341 n.7 (1985) (upholding a warrantless search on school grounds when justified by “special needs” and not conducted for the advancement of law enforcement). The strongest argument to support searches in public housing projects is that the searches are necessary to ensure that the government allocates scarce subsidized housing to those who do not use it for criminal purposes. Cf. Dep’t of Hous. & Urban Dev. v. Rucker, 535 U.S. 125, 135 (2002) (upholding policy permitting evictions from public housing for criminal activity).
drug free zones in Part III.B demonstrates one way a civil regime can enable government to exercise control over public spaces. Current restrictions on access to public housing properties present similar civil consequences. While the programmatic purpose approach would permit many such civil sanctions, it would not permit vague criminal laws. This approach has several advantages. First, it preserves the important value of notice to the citizenry regarding the scope of criminal prohibitions. It recognizes that, even for low-level criminal offenses, a criminal sanction is qualitatively different from civil penalties because it reflects society’s moral condemnation. Accordingly, criminal punishment should not be imposed unless the defendant had at least a fair opportunity to know that his conduct was criminal.

Second, the programmatic purpose approach will give municipalities an incentive to avoid using criminal laws to regulate disorder. If forced to resort to non-criminal laws to permit discretionary responses to disorder, cities might develop more effective, flexible approaches rather than outright prohibitions. Moreover, civil responses to disorder permit cities to address the concerns of communities without imposing

355. This approach is not unique to American cities. For example, in the United Kingdom, the Football Spectators Act authorizes “banning orders” that prohibit individuals who engage in banned activities at football matches from attending future games. See Football (Disorder) Act, 2000, c. 25 (Eng.).


357. See supra Part II.D for a discussion of the importance of forcing legislatures to define clearly the scope of criminal laws.

358. See Vill. of Hoffman Estates v. Flipside, Hoffman Estates, Inc., 455 U.S. 489, 498–99 (1982) (noting that vagueness is more permissible in civil ordinances than in criminal laws “because the consequences of imprecision are qualitatively less severe”); United States v. Bass, 404 U.S. 336, 348 (1971) (discussing the rule of lenity and noting that “because of the seriousness of criminal penalties, and because criminal punishment usually represents the moral condemnation of the community, legislatures and not courts should define criminal activity”) (citations omitted).

359. See Livingston, supra note 6, at 638 (“Civil sanctions, moreover, are generally preferable to penal sanctions when the community wishes to reduce the level of an activity—to regulate it—but not prohibit the activity entirely.”); see also John C. Coffee, Jr., Paradigms Lost: The Blurring of the Criminal and Civil Law Models—And What Can Be Done About It, 101 YALE L.J. 1875, 1887–92 (1992) (encouraging use of civil sanctions to respond to conduct that is not egregious). For example, under Ellickson’s zoning model, disfavored activities are not criminalized, but are treated as “nonconforming uses” if committed within orderly zones of the city. See Ellickson, supra note 14, at 1232–39.
Unpacking New Policing

the often ignored costs of enlarging the scope of criminal liability. After all, the purported goal of new policing efforts is not to penalize violators of norms, but to “negotiate compliance on the streets.”

Civil responses to disorder, for example, avoid stigmatizing the disorderly with arrest and convictions for offenses that do not meet the traditional requirements of criminal law. Bernard Harcourt has set forth the hidden costs of order-maintenance policing, including the reinforcement of stereotypes about black criminality and increased tensions between at least some African-American residents and the police due to the proliferation of arrests for low-level offenses. The programmatic purpose approach would also help ensure that new policing efforts were not used as a pretext to engage in old-fashioned searches for evidence.

The programmatic purpose approach also avoids the pitfalls of giving malleable notions of community support and involvement constitutional significance. Under the programmatic purpose approach, political majorities (or handfuls of citizens purporting to represent majorities) are not permitted to waive the constitutional rights of the entire community. Accordingly, there is no risk that rights will be lost because the relevant community is defined improperly, because the community’s level of support was assessed inaccurately, or because the community supported a police program only because it had no other realistic alternatives.

Most importantly, the programmatic purpose approach would require courts to scrutinize new policing efforts on their individual merits. Applying the void for vagueness doctrine to any police action that might ultimately form the basis of a criminal arrest and prosecution, as the Court appeared to do in *Morales*, casts too wide a net over prohibited delegations of discretion. On the other hand, the new discretion scholars are too quick to conclude that courts should permit police discretion, even with respect to traditional criminal law efforts. The programmatic purpose approach, in contrast, would require courts to engage in the fact-

---

360. Livingston, *supra* note 6, at 638.
361. *Id.* (noting that civil sanctions can be sufficient to handle community problems without bearing “the hallmarks of blameworthiness associated with criminal law”).
363. See *supra* notes 118–26 and accompanying text for a discussion of how police could use new policing models as a pretext to search for evidence of more serious crimes.
364. See *supra* Part II.A–B, E for discussions criticizing the new discretion scholars’ reliance on community sentiment.
specific, “contextual” inquiry sought by new discretion scholars, without giving up the void for vagueness doctrine entirely.

A fact-specific inquiry would be required at two levels. First, courts would need to decide whether a challenged police program served the traditional functions of criminal law or whether it was qualitatively different. For example, Part III.B of this Article argues that neighborhood exclusion orders issued pursuant to drug free zone ordinances do not constitute criminal punishment. However, the classification of neighborhood exclusion as different from traditional criminal procedure does not turn solely on majority support for the exclusions, as it would under the political process model. Indeed, classification of neighborhood exclusion as a civil restraint does not even turn solely on the nature of the imposed burden. Rather, that classification depends on the limited scope and duration of the restraint, the availability of variances, and, importantly, the extent of the existing problem within the zones.

If, after scrutinizing the parameters of a specific program and the needs of the jurisdiction where it applies, a court determines that the program amounts to traditional criminal law, then the court should evaluate it as such by applying the rules that govern traditional criminal law programs. For example, searches conducted as part of a traditional criminal investigation should generally be conducted pursuant to a warrant based on probable cause. Similarly, substantive criminal prohibitions should be subject to the void for vagueness doctrine.

If, however, a court determines that the challenged program does not promote the traditional aims of criminal law, the court should evaluate the program as a civil restraint. In the civil context, the warrant and probable cause requirements may not apply to searches, and the void for vagueness doctrine does not apply as rigidly to substantive provisions. On the other hand, weighing the constitutionality of a non-criminal restraint requires the court to determine whether the civil program implicated a fundamental right. The court must then weigh

365. See Livingston, supra note 6, at 635.
366. See supra text accompanying notes 284–89 for a discussion of drug free zone ordinances and the importance that they apply only to neighborhoods plagued by more than their proportional share of criminal activity.
367. See supra note 270.
368. See supra notes 255–59, 268–74 and accompanying text.
369. See supra note 319.
370. See supra notes 326–29 and accompanying text for a discussion of how improper judicial
Unpacking New Policing

the government interest served by the restraint and the degree to which the restraint serves that interest against the individual liberty interests at stake.\(^{371}\) In this respect, judicial evaluation of the program as a non-criminal restraint provides a second opportunity to engage in a contextualized review of the program. For example, in determining that a Cincinnati neighborhood exclusion ordinance violated due process, the Sixth Circuit refused to consider evidence demonstrating the ordinance’s effectiveness in Portland, suggesting that judicial inquiry should focus on whether a particular program sufficiently advances the government’s interest in a particular location.\(^{372}\)

IV. CONCLUSION

Early scholarship focusing on the new policing broke ground by recognizing the importance of protecting not just the individual liberties of offenders brought into the criminal justice system, but also the interests of communities struggling to gain minimum controls over their neighborhoods—controls that more privileged communities take for granted. The new discretion scholars have argued that new policing approaches should be subject to less stringent judicial scrutiny than typical law enforcement efforts. To distinguish the new from the traditional, they have relied on claimed inner-city support for new policing approaches and on a political process theory in which inner-city communities should be permitted to exchange traditional liberty for safety, as long as they do so evenhandedly.\(^{373}\)

However, in their attempt to empower inner-city neighborhoods to establish necessary controls, the new discretion scholars have been too quick to discard constitutional rules that preserve and respect critical distinctions between criminal and non-criminal programs. This Article reliance on the void for vagueness doctrine permits courts to avoid determining whether a law implicates a fundamental right.


\(^{372}\) Johnson v. City of Cincinnati, 310 F.3d 484, 504 (6th Cir. 2002) (suggesting that it may not be appropriate to extrapolate “that what worked in Portland would likely work in [the targeted Cincinnati zone]”), cert. denied, 123 S. Ct. 2276 (2003). The Johnson court refused to consider empirical evidence relating to the Portland ordinance also because the court subjected the ordinance to strict scrutiny. Because the court recognized a fundamental right to intrastate movement, it was concerned not only with the degree to which the ordinance advanced the government’s interest, but also whether the government could advance its purposes with less restrictive means. Id. As set forth in supra notes 330–42 and accompanying text, there are reasons to believe that the U.S. Supreme Court would not recognize a constitutional right to loiter.

\(^{373}\) See supra Part I.C for a discussion of the new discretion scholarship.
has argued that the balance between traditional civil liberties and the need for urban crime control would be better struck if cities were forced to choose either nontraditional responses to public safety problems or to be scrutinized under the traditional rules governing criminal law and procedure. To separate “new” public safety responses from the traditional, this Article encourages scrutiny of the challenged response’s programmatic purpose to determine whether it serves the traditional purposes of criminal law.
WHO CAN DEFEND A FEDERAL REGULATION? THE NINTH CIRCUIT MISAPPLIED RULE 24 BY DENYING INTERVENTION OF RIGHT IN KOOTENAI TRIBE OF IDAHO V. VENEMAN

Stephanie D. Matheny

Abstract: In Kootenai Tribe of Idaho v. Veneman, the United States Court of Appeals for the Ninth Circuit misapplied Rule 24 of the Federal Rules of Civil Procedure by denying intervention of right to organizations that had protectable interests in the adoption and implementation of the Roadless Rule. The court based its decision to deny intervention of right on its federal defendant rule, which bars intervention of right by parties other than the federal government to defend a challenge brought under the National Environmental Policy Act (NEPA). The Kootenai decision extended the reach of the federal defendant rule to include environmental organizations that had actively participated in the challenged NEPA administrative rulemaking process. This extension contradicts Rule 24’s focus on the practical effects of litigation and the Ninth Circuit’s precedent of liberally granting intervention in public law cases. This Note argues that the Ninth Circuit should abandon the federal defendant rule and instead apply Rule 24 by individually evaluating whether absentees have a protectable interest within NEPA’s zone of concern for the environment or have actively participated in the process of adopting the challenged regulation.

In January 2001, the United States Forest Service issued its final Roadless Rule, a landmark conservation regulation that would prohibit nearly all logging and roadbuilding in 58.5 million acres of National Forest lands. The Roadless Rule was the culmination of a lengthy administrative rulemaking process undertaken pursuant to the National Environmental Policy Act of 1969 (NEPA). The Forest Service received over one million public comments on the Roadless Rule, ninety-six percent of which called for strong conservation measures. The Clinton administration adopted the Roadless Rule just two weeks before the inauguration of President George W. Bush. When opponents filed a NEPA challenge in the District Court of Idaho seeking to

2. Kootenai Tribe of Idaho v. Veneman, 313 F.3d 1094, 1105–06 (9th Cir. 2002) [hereinafter Kootenai II].  
3. Id. at 1104–05.  
5. Kootenai II, 313 F.3d at 1116 n.19, 1119.  
6. Id. at 1105–06.
overturn the Roadless Rule, the new administration announced that it had its own concerns about the Roadless Rule and declined to defend it in court.\textsuperscript{7} In response, several environmental organizations sought to intervene under Rule 24 of the Federal Rules of Civil Procedure to defend the NEPA challenge.\textsuperscript{8} The District Court of Idaho alternatively granted both intervention of right and permissive intervention,\textsuperscript{9} but later rejected the arguments of the environmental intervenors by concluding that the Forest Service had likely violated NEPA\textsuperscript{10} and temporarily enjoined implementation of the Roadless Rule.\textsuperscript{11}

In \textit{Kootenai Tribe of Idaho v. Veneman},\textsuperscript{12} the U.S. Court of Appeals for the Ninth Circuit concluded that the environmental organizations could not intervene of right under Rule 24(a)(2) to defend a federal regulation even though they had helped to promulgate it.\textsuperscript{13} The court applied its federal defendant rule, reasoning that because the federal government is the only party capable of violating NEPA, no other entities have the right to intervene to defend a NEPA challenge.\textsuperscript{14} Instead, the court held that district courts have discretion to grant defendant-side intervention in NEPA cases under Rule 24(b)(2)’s standards for permissive intervention.\textsuperscript{15} The Ninth Circuit then affirmed the district court’s grant of permissive intervention and reached the merits of the intervenors’ defense of the Roadless Rule.\textsuperscript{16} Ultimately, the Ninth Circuit agreed with the arguments presented by the environmental intervenors and lifted the injunction that had blocked implementation of the Roadless Rule.\textsuperscript{17} Although the \textit{Kootenai} case was at least a

\textsuperscript{7} \textit{Id.} at 1106, 1111.
\textsuperscript{8} \textit{Id.} at 1106.
\textsuperscript{11} \textit{Kootenai II}, 313 F.3d at 1106–07.
\textsuperscript{12} \textit{313 F.3d 1094 (9th Cir. 2002).}
\textsuperscript{13} \textit{Id.} at 1104.
\textsuperscript{14} \textit{See id.} at 1108. This Note will use the term “federal defendant rule” to denote the Ninth Circuit’s rule barring defendant-side intervention in NEPA cases by parties other than the federal government. See \textit{Portland Audubon Society v. Hodel}, 866 F.2d 302, 308–09 (9th Cir. 1989), for a detailed discussion of the federal defendant rule.
\textsuperscript{15} \textit{Kootenai II}, 313 F.3d at 1111.
\textsuperscript{16} \textit{See id.} at 1104.
\textsuperscript{17} \textit{Id.}
Rule 24 Intervention in NEPA Litigation

temporary victory for the Roadless Rule, this win might have come at a high cost: leaving defendant-side intervention in NEPA cases to the discretion of district courts may limit the ability of absentees to protect their interests in future government decisions.

The Kootenai case is a prominent example of a recent litigation trend in which parties opposed to conservation-oriented plans and endangered species protections have used NEPA and other federal environmental statutes to challenge these programs in court. This reverse form of conservation litigation, which seeks to achieve primarily economic goals using environmental statutes to provide the cause of action, can be a particularly effective method to overturn regulations that were adopted by a previous administration but are at odds with the policies of a new administration. The new administration might decline to defend a regulation promulgated by its predecessor, and may find it politically expedient to accede to an “adverse” court judgment or settle a case rather than independently undertake a public rulemaking process to change the regulation.

Intervention under Rule 24 provides an opportunity for public interest organizations to protect regulations that the federal government adopts.

---


19. The Ninth Circuit has already applied the Kootenai rule to deny intervention of right to conservation groups seeking to defend a NEPA challenge to forest rules in Alaska. Alaska Forest Ass’n v. United States Dep’t of Agric., No. 01-35549, 2003 U.S. App. LEXIS 10880, at *3 (9th Cir. May 29, 2003).

20. See, e.g., id. at *2–3 (applying the Kootenai rule to deny intervention of right to conservation groups seeking to appeal a district court decision finding that the government had violated NEPA when the government elected not to appeal); Wyoming v. United States Dep’t of Agric., 201 F. Supp. 2d 1151, 1153 (D. Wyo. 2002) (challenging the Roadless Rule for NEPA and other procedural violations).


22. See, e.g., id.


24. See id.
but later chooses not to defend in court. Intervenors become parties to the litigation. They can submit briefs, examine witnesses at trial, and appeal adverse decisions. Absentees may intervene of right under Rule 24(a)(2) when they have a protectable interest in the subject matter of the suit that may be adversely affected by the outcome of the case and that is not adequately represented by the existing parties. Alternatively, courts may grant permissive intervention under Rule 24(b)(2) if absentees assert an independent basis for federal court jurisdiction and raise claims that share common questions of law or fact with issues in the litigation. Rule 24 is a flexible and practical rule intended to protect absentees whose interests might be affected by the outcome of pending litigation.

This Note argues that the Ninth Circuit misapplied Rule 24 in Kootenai by denying intervention of right to groups with protectable interests in defending the Roadless Rule. The Ninth Circuit’s federal defendant rule contravenes Rule 24(a)(2)’s requirement that courts assess the practical effects of litigation on the protectable interests of absentees, including the stare decisis effect of overturning a federal regulation. This Note proposes that absentees satisfy the protectable interest requirement of Rule 24(a)(2) by asserting injuries within NEPA’s zone of interest for the environment and by actively participating in the rulemaking process. If they satisfy the other requirements of the rule, they have a right to intervene. Part I briefly reviews the function and environmental purpose of NEPA and introduces the zone of interest test under the Administrative Procedure Act.
Rule 24 Intervention in NEPA Litigation

Act (APA). Part II describes intervention under Rule 24, focusing on the protectable interest requirement for intervention of right as interpreted by the U.S. Supreme Court, its application to public law cases, and its relationship to the requirements for Article III standing. Part III describes the federal defendant rule for NEPA cases and reviews Ninth Circuit public law intervention cases in non-NEPA contexts. Part IV reviews the Kootenai decision. Finally, Part V argues that the Ninth Circuit should have allowed the environmental groups to intervene of right under Rule 24(a)(2), and that the court should abandon the federal defendant rule in NEPA challenges in favor of a practical and flexible inquiry into the asserted interests of the proposed intervenors as mandated by Rule 24.

I. NEPA: A PROCEDURAL STATUTE ENACTED TO PROTECT THE ENVIRONMENT

Congress enacted NEPA with the goal of protecting the environment by requiring federal agencies to evaluate significant environmental impacts prior to undertaking major actions. NEPA established a national environmental policy directing the federal government to use “all practicable means” to “fulfill the responsibilities of each generation as trustee of the environment for succeeding generations” and to “attain the widest range of beneficial uses of the environment without degradation.” NEPA has profoundly influenced decisionmaking by federal agencies and has made administrators more accessible and accountable to the public. The centerpiece of NEPA is its requirement that federal agencies prepare an environmental impact statement (EIS) prior to undertaking “major Federal actions significantly affecting the quality of the human environment.” A federal agency must follow NEPA’s procedural requirements when it proposes legislation or

36. See U.S. CONST. art. III, § 2 (requiring a case or controversy for federal court jurisdiction).
37. See 42 U.S.C. § 4321 (2000) (declaring the goal of promoting efforts that “will prevent or eliminate damage to the environment and biosphere and stimulate the health and welfare of man”).
38. Id. § 4332(2).
39. Id. § 4331.
42. 42 U.S.C. § 4332(2)(c).
undertakes an administrative rule making process to adopt a new
regulation, such as a forest management plan or a pollution control rule,
that might have a significant effect on the environment.\textsuperscript{43} Although
NEPA establishes procedural rather than substantive requirements,\textsuperscript{44}
these procedural requirements can lead federal agencies to abandon
environmentally damaging proposals\textsuperscript{45} by requiring administrators to
consider the environmental effects of their actions and engage in a
public process that exposes these effects.

Because NEPA does not create a private cause of action, parties
seeking to challenge a NEPA decision in court must allege violations of
the Administrative Procedure Act (APA).\textsuperscript{46} Under the APA, citizens may
file suit alleging that a final agency decision was arbitrary and capricious
or otherwise not in accordance with law.\textsuperscript{47} A NEPA challenge brought
pursuant to the APA may include charges that the federal government
should have prepared a full EIS, or did prepare an EIS but failed to
consider an adequate range of alternatives.\textsuperscript{48} A typical result of a
successful NEPA challenge is an injunction against the federal action
pending NEPA compliance.\textsuperscript{49} On remand, the agency may either modify
its proposed action or correct analytical deficiencies.\textsuperscript{50}

The APA requires plaintiffs to assert injuries “arguably within the
zone of interests” of the underlying statute to establish prudential
standing.\textsuperscript{51} In applying the APA’s zone of interest test, federal courts
interpret NEPA as protecting environmental, but not economic,
interests.\textsuperscript{52} Thus, only parties who assert an injury within NEPA’s

\begin{itemize}
  \item 43. See id.
  \item 44. See Tillamook County v. United States Army Corps of Eng’rs, 288 F.3d 1140, 1143 (9th Cir.
             2002).
  \item 45. Blumm, supra note 41, at 452.
             Mgmt., 150 F.3d 1132, 1135 (9th Cir. 1998).
  \item 47. See 5 U.S.C. § 706.
  \item 48. See 42 U.S.C. § 4332(2)(c).
  \item 49. See MANDELKER, supra note 40, § 4:54.
  \item 50. See id. § 4:61.
             (granting a right of review to persons suffering legal wrong as a result of agency action “within
             the meaning of the underlying statute”).
  \item 52. See Taubman Realty Group v. Mineta, 320 F.3d 475, 481 (4th Cir. 2003) (noting that NEPA
does not require preparation of an environmental impact statement for economic or social effects
alone); Rosebud Sioux Tribe v. Mcdiivitt, 286 F.3d 1031, 1037–38 (8th Cir 2002); ANR Pipeline
United States Forest Serv., 8 F.3d 713, 715–17 (9th Cir. 1993); MANDELKER, supra note 40, § 4:22.
\end{itemize}
Rule 24 Intervention in NEPA Litigation

environmental zone of interest have standing to bring a NEPA challenge.\(^\text{53}\)

II. RULE 24 ALLOWS ABSENTEES TO INTERVENE TO DEFEND THEIR INTERESTS THAT MAY BE AFFECTED BY PENDING LITIGATION

Intervention under Rule 24 allows an absentee with an interest in a pending lawsuit to join the litigation as a party on its own motion.\(^\text{54}\) The rule distinguishes between intervention of right under Rule 24(a)(2)\(^\text{55}\) and permissive intervention under Rule 24(b)(2).\(^\text{56}\) Although both forms of intervention grant party status, there are significant practical differences between the two in criteria, degree of participation, and scope of review on appeal. The U.S. Supreme Court has established a flexible and practical approach to evaluating intervention requests.\(^\text{57}\) As a result, many federal courts have become “more receptive” to granting intervention in the context of public law litigation.\(^\text{58}\) In spite of this acceptance, Article III standing requirements\(^\text{59}\) may limit the ability of absentees to intervene in some situations.

A. Rule 24 Allows Absentees To Intervene To Protect Their Interests in Existing Litigation

Intervention allows absentees with an interest in pending litigation to enter the lawsuit on their own motion.\(^\text{60}\) A primary purpose of intervention is to protect the rights of people who may potentially be affected by litigation to which they are not parties.\(^\text{61}\) This policy is consistent with the nature of the adversary process, which requires the

\(^{53}\) See Nevada Land Action, 8 F.3d at 715–17.
\(^{54}\) 7C CHARLES A. WRIGHT ET AL., FEDERAL PRACTICE & PROCEDURE § 1901 (2d ed. 1986).
\(^{56}\) Id. 24(b)(2).
\(^{57}\) See infra Part II.B.
\(^{59}\) See U.S. CONST. art. III, § 2.
\(^{60}\) WRIGHT ET AL., supra note 54, § 1901.
\(^{61}\) Kleissler v. United States Forest Serv., 157 F.3d 964, 971 (3d Cir. 1998).
court to “hear from both sides before the interests of one side are impaired by a judgment.”

Federal Rule of Civil Procedure 24 governs intervention in federal district courts. Rule 24(a)(2) requires a federal district court to allow an absentee to intervene in an action when (1) the absentee makes a timely motion to intervene, (2) the absentee has a significantly protectable interest relating to the property or transaction that is the subject of the action, (3) the applicant is so situated that the disposition of the litigation may as a practical matter impair its ability to protect this interest, and (4) none of the current parties adequately represents the absentee’s interest. The protectable interest requirement is often the critical factor upon which courts decide whether or not an absentee can intervene of right.

Alternatively, a court may grant permissive intervention under Rule 24(b)(2). This Rule provides that a court “may” permit absentees to intervene when their “claim or defense and the main action have a question of law or fact in common.” The district court has discretion to grant or deny intervention when an applicant meets these requirements and establishes independent grounds on which the court can assert jurisdiction.

There are several significant practical differences between intervention of right and permissive intervention. Most importantly, intervention of right is not subject to district court discretion: if absentees meet the requirements for intervention of right, the district court must allow them to join the litigation. In addition, although both permissive intervenors and intervenors of right become parties to a

64. Wetlands Action Network v. United States Army Corps of Eng’rs, 222 F.3d 1105, 1113 (9th Cir. 2000). In relevant part, Rule 24(a) provides: “Upon timely application anyone shall be permitted to intervene in an action . . . when the applicant claims an interest relating to the property or transaction which is the subject matter of the action and the applicant is so situated that the disposition of the action may as a practical matter impair or impede the applicant’s ability to protect that interest, unless the applicant’s interest is adequately represented by existing parties.” F ED. R. CIV. P. 24(a)(2).
65. See, e.g., Northwest Forest Res. Council v. Glickman, 82 F.3d 825, 837 (9th Cir. 1996) (holding that an absentee was not entitled to intervene of right because it lacked a significantly protectable interest).
68. See F ED. R. CIV. P. 24(a)(2) (“anyone shall” be granted intervention).
Rule 24 Intervention in NEPA Litigation

litigation, the district court has broad discretion to impose even “highly restrictive” conditions on the involvement of permissive intervenors. For example, the court can preclude a permissive intervenor from raising new claims for relief, prevent the intervenor from participating in some of the existing claims, and limit the intervenor’s participation in discovery. By contrast, the district court has less discretion to limit the scope of participation under Rule 24(a)(2) because an intervenor of right has an interest in the litigation that cannot be adequately protected without joining the litigation. Finally, the scope of review on appeal depends on whether intervention is of right or permissive. An appellate court reviews a decision to grant or deny intervention of right de novo but reviews a decision of permissive intervention only for abuse of discretion.

B. The U.S. Supreme Court Has Adopted a Flexible, Practical Approach To the Protectable Interest Requirement for Intervention of Right Under Rule 24(a)(2) in Public Law Litigation

The U.S. Supreme Court has established a pragmatic approach to granting intervention of right under Rule 24(a)(2). This approach requires courts to focus on the practical effects of the litigation when determining whether an absentee’s protectable interest will be affected by the ongoing litigation. The liberalized interest requirement has led to increased flexibility for intervention in public law litigation by directing courts to consider the practical effect of the outcome of litigation. Based on this pragmatic approach, courts in public law cases often consider the stare decisis effect of a judgment, which, while not

70. Id. at 382 n.1 (Brennan, J., concurring).
71. See id. at 373.
72. Id. at 381–82.
73. Wetlands Action Network v. United States Army Corps of Eng’rs, 222 F.3d 1105, 1113 (9th Cir. 2000).
74. Greene v. United States, 996 F.2d 973, 978 (9th Cir. 1993).
75. See Appel, supra note 58, at 215–16.
76. See Fed. R. Civ. P. 24 advisory committee’s notes to 1966 Amendments (noting that the prior rule 24 was unduly restrictive and that this amendment reflects a more practical focus of liberalized joiner requirements).
77. See Appel, supra note 58, at 215–16.
78. See Kleissler v. United States Forest Serv., 157 F.3d 964, 971 (3d Cir. 1998).
directly binding non-parties, constitutes a significant practical impairment in later litigation of the same issue.  

For example, in *Cascade Natural Gas Corp. v. El Paso Natural Gas Co.*, the U.S. Supreme Court established that some public concerns constitute protectable interests sufficient to intervene of right under Rule 24(a)(2), particularly when these concerns are protected by the statutes under consideration in the case. The Court allowed the State of California to intervene in an antitrust suit involving the regional natural gas market, holding that the state’s public interest in a competitive natural gas market was the same interest that the antitrust laws at issue in the case sought to protect. In reaching this conclusion, the Court reasoned that Rule 24(a)(2) focuses on the practical effects of litigation, and stated that if “an absentee would be substantially affected in a practical sense by the determination made in an action, he should, as a general rule, be entitled to intervene.”

Thus, the Court implied that the purpose of the underlying statute is relevant to determining whether absentees have a protectable interest sufficient to intervene of right. Several circuits have applied a comparable analysis to hold that the citizen suit provisions of the Endangered Species Act (ESA) establish a legally protected interest sufficient for a private citizen to intervene of right to defend a federal agency’s decision to list a species under the act. This inquiry is also

79. See infra notes 91–95 and accompanying text.
80. 386 U.S. 129 (1967).
81. See *Diamond v. Charles*, 476 U.S. 54, 68 (1986). The Court has provided little guidance on what public concerns are sufficient for intervention of right. See Appel, *supra* note 58, at 256, 266.
82. See *Diamond*, 476 U.S. at 66 n.17.
83. *Cascade Natural Gas Corp.*, 386 U.S. at 132.
84. *Id.* at 135–36.
85. *Id.* at 134 n.3.
86. *Id.* (emphasis omitted).
87. See *id.* at 135–36.
89. See, e.g., Coalition of Ariz./N.M. Counties for Stable Econ. Growth v. United States Dep’t of the Interior, 100 F.3d 837, 841–42 (10th Cir. 1996) (granting intervention of right to a private party who had successfully petitioned and litigated for the listing of the Mexican Spotted Owl); Idaho Farm Bureau Fed’n v. Babbitt, 58 F.3d 1392, 1395–98 (9th Cir. 1995) (granting intervention to organizations that had successfully sued to secure listing of the Bruneau Hot Springs Snail); see also Carl Tobias, *Rethinking Intervention in Environmental Litigation*, 78 WASH. U. L.Q. 313, 316 (2000) (noting that the “touchstone of analysis will be the purpose of the environmental statute which underlies the litigation, and how this purpose relates to the interest requirement for intervention”).
Rule 24 Intervention in NEPA Litigation

closely analogous to the APA’s zone of interest test for prudential standing, which considers whether the underlying statute arguably protects the plaintiff’s interest.90

Courts often base the grant of intervention to absentees in public law cases on the potential stare decisis effect of a court’s judgment determining the legality of an agency decision.91 Such a determination can create a significant practical barrier to future litigation of the issue.92 The U.S. Supreme Court has noted that although persons who were not parties to a particular adjudication are not bound by strict rules of res judicata, “subsequent litigation serves little practical value to the potential intervenor” because later courts would likely invoke the doctrine of stare decisis and defer to the original decision.93 For example, the precedential effect of a decision that a particular forest management statute bars even-age logging could impair the ability of absentees to protect their interests in future lawsuits challenging the application of that statute to other timber sales.94 Similarly, the stare decisis effect of a judgment regarding disputed aboriginal land rights constitutes a sufficient practical impairment to justify intervention of right by absentee tribes claiming rights to the same land.95 Granting intervention in such situations avoids the problem of a court effectively binding parties without first giving them an opportunity to be heard.96

As a result of Rule 24’s liberalized interest requirements and the effect of stare decisis, most federal courts have been generally receptive to intervenors in public law cases.97 Granting intervention in these cases ensures that all sides are represented in litigation that implicates issues of broad social concern.98 It can also reduce future litigation of the same issues while broadening access to the courts.99

90. See Ass’n of Data Processing Serv. Orgs. v. Camp, 397 U.S. 150, 153 (1970); see also supra notes 52–53 and accompanying text.

91. See, e.g., Sierra Club v. Espy, 18 F.3d 1202, 1207 (5th Cir. 1994) (concluding that the potential stare decisis effect of the judgment in the case would impair the interests of the absentees).

92. See id.


94. Espy, 18 F.3d at 1207.

95. Oneida Indian Nation of Wis. v. New York, 732 F.2d 261, 265 (2d Cir. 1984).

96. Appel, supra note 58, at 299.

97. See id. at 282.


99. Forest Conservation Council v. United States Forest Serv., 66 F.3d 1489, 1496 n.8 (9th Cir.
C. Article III Standing May Limit the Ability of Absentees To Intervene

The federal circuits disagree about whether a party seeking to intervene of right at the district court level must satisfy Article III standing requirements. Although the concepts of intervention of right and standing have different origins and purposes, they have similar requirements. An applicant for intervention of right must demonstrate a protectable interest that may be affected by the outcome of the lawsuit. To establish standing, a plaintiff must allege an “invasion of a legally protected interest” that might be redressed by the litigation. The similarity between these two standards has led some federal courts to conflate the tests. For example, the Tenth Circuit has concluded that meeting the requirements of standing is one way to demonstrate that an absentee has a protectable interest sufficient to intervene of right. In some circuits, however, requirements for standing pose a significant barrier to intervention.

Although the federal circuits differ on whether standing is required to intervene at the district court level, the U.S. Supreme Court has established that an intervenor must demonstrate independent Article III standing in order to appeal a judgment without the party on whose side

100. See U.S. CONST. art. III, § 2; Diamond v. Charles, 476 U.S. 54, 68–69 (1986); Tyler R. Stradling & Doyle S. Byers, Comment, Intervening in the Case (or Controversy): Article III Standing, Rule 24 Intervention, and the Conflict in the Federal Courts, 2003 BYU L. REV. 419, 424–35 (noting that of the eight circuits to consider the issue, three require standing and five do not). Compare Portland Audubon Soc’y v. Hodel, 866 F.2d 302, 308 n.1 (9th Cir. 1989) (noting that the Ninth Circuit resolves intervention questions without reference to standing doctrine), with United States v. 36.96 Acres of Land, 754 F.2d 855, 859–60 (7th Cir. 1985) (holding that an organization’s aesthetic and environmental interests were not sufficient under Rule 24(a)(2), which requires a “direct and substantial” legally-protectable interest).

101. Carl Tobias, Standing To Intervene, 1991 WIS. L. REV. 415, 416–17 (noting that standing entails whether a plaintiff can initiate a lawsuit in order to meet Article III’s case or controversy requirement, while intervention entails whether an absentee may participate in ongoing litigation in which the plaintiff already has standing). This Note will not address standing in intervention cases. For more information about the application of standing requirements to intervention analysis, see generally id.

102. See Portland Audubon Soc’y, 866 F.2d at 308 n.1.

103. See id. at 308.


105. See Coalition of Ariz./N.M. Counties for Stable Econ. Growth v. United States Dep’t of the Interior, 100 F.3d 837, 842 (10th Cir. 1996).

106. See, e.g., United States v. 36.96 Acres of Land, 754 F.2d 855, 859–60 (7th Cir. 1985).
Rule 24 Intervention in NEPA Litigation

they intervened.\textsuperscript{107} In \textit{Diamond v. Charles},\textsuperscript{108} the Court held that a private individual whose conduct is not directly implicated by a state abortion law could not appeal a decision finding the law unconstitutional because the state did not appeal.\textsuperscript{109} The intervenor lacked a judicially cognizable interest in the state’s prosecution of another person under this criminal statute.\textsuperscript{110} The Court reasoned that because only the state could adopt a legal code, only the state has standing to defend the code against a constitutional attack.\textsuperscript{111} However, the Court left open the possibility that a private party could establish standing to defend the constitutionality of a statute that creates a protectable private interest.\textsuperscript{112}

In sum, the U.S. Supreme Court has established a pragmatic test for granting intervention under Rule 24. Courts must grant intervention of right when absentees meet the requirements of Rule 24(a)(2), guaranteeing the opportunity to defend interests at stake in pending litigation. By contrast, courts have broad discretion to grant, deny, or limit the scope of intervention under Rule 24(b)(2). Federal courts are generally receptive to intervenors in public law cases, particularly when the absentees seek to vindicate interests consistent with the purpose of the underlying statute. However, Article III standing requirements may limit the ability of public interest organizations to intervene in some circumstances.

III. THE NINTH CIRCUIT LIBERALLY GRANTS INTERVENTION OF RIGHT IN PUBLIC LAW CASES EXCEPT TO THOSE DEFENDING NEPA CHALLENGES

The Ninth Circuit liberally approves intervention requests in most public law cases.\textsuperscript{113} The court broadly grants intervention of right to private groups to defend procedural or constitutional challenges to laws or regulations other than NEPA,\textsuperscript{114} but it applies the federal defendant rule to deny intervention of right by parties other than the federal

\begin{flushright}
109. \textit{Id.} at 56.
110. \textit{Id.} at 64–65, 71.
111. \textit{Id.} at 65.
113. See Vreeland, supra note 98, at 289.
114. See infra Part III.A.
\end{flushright}
government to defend NEPA challenges.\textsuperscript{115} In reaching these contradictory results, the court has not explained how NEPA differs from other procedural or constitutional rules that a private party cannot violate.\textsuperscript{116}

\textbf{A. The Ninth Circuit Interprets Rule \textit{24(a)(2)} in Favor of Granting Intervention, and Generally Allows Public Interest Groups To Defend Government Decisions}

The Ninth Circuit generally interprets Rule 24(a)(2)’s protectable interest requirement broadly.\textsuperscript{117} The court views Rule 24(a)(2)’s protectable interest test as “primarily a practical guide to disposing of lawsuits by involving as many apparently concerned persons as is compatible with efficiency and due process.”\textsuperscript{118} Applying these principles to non-NEPA cases, the Ninth Circuit implicitly rejects the federal defendant rule by granting intervention of right to groups “directly involved in the enactment of the law or in the administrative proceedings out of which the litigation arose” to defend even procedural and constitutional challenges to government actions.\textsuperscript{119}

When private groups have sought to intervene to defend non-NEPA challenges to laws for which they had advocated, the Ninth Circuit has demonstrated broad flexibility in granting intervention of right.\textsuperscript{120} The

\begin{itemize}
  \item[\textsuperscript{115}] See infra Part III.B.
  \item[\textsuperscript{116}] Compare Sagebrush Rebellion, Inc. v. Watt, 713 F.2d 525, 526–27 (9th Cir. 1983) (granting intervention of right to organizations that had participated in the administration designation of a conservation area), Wash. State Bldg. & Constr. Trades Council v. Spellman, 684 F.2d 627, 630 (9th Cir. 1982) (granting intervention of right to an initiative sponsor to defend a constitutional challenge), and Idaho v. Freeman, 625 F.2d 886, 887 (9th Cir. 1980) (granting intervention of right to the National Organization for Women to defend a constitutional challenge), with Kootenai II, 313 F.3d 1094, 1108 (9th Cir. 2002) (denying intervention of right to organizations that had promoted adoption of the Roadless Rule).
  \item[\textsuperscript{117}] See Spellman, 684 F.2d at 630.
  \item[\textsuperscript{118}] Forest Conservation Council v. United States Forest Serv., 66 F.3d 1489, 1496 (9th Cir. 1995) (quoting Nuesse v. Camp, 385 F.2d 694, 700 (D.C. Cir. 1967)).
  \item[\textsuperscript{119}] Northwest Forest Res. Council v. Glickman, 82 F.3d 825, 837 (9th Cir. 1996).
  \item[\textsuperscript{120}] See id. Several other federal courts have followed the Ninth Circuit’s lead in granting intervention of right to private groups to defend the constitutionality of laws for which they had advocated. See Utah Ass’n of Counties v. Clinton, 255 F.3d 1246, 1256 (10th Cir. 2001) (granting intervention of right to environmental groups to defend a monument designation); Mich. State AFL-CIO v. Miller, 103 F.3d 1240, 1245–47 (6th Cir. 1997) (citing Ninth Circuit cases in granting intervention to the Chamber of Commerce to defend a constitutional challenge to a campaign finance reform law that the Chamber had previously advocated for at the state legislature); Mausolf v. Babbitt, 85 F.3d 1295, 1296, 1303–04 (8th Cir. 1996) (granting intervention of right to conservation groups that had been involved in prior litigation and administrative proceedings to...
Rule 24 Intervention in NEPA Litigation

court has granted intervention in non-NEPA cases even when the private
groups could not themselves have enacted, enforced, or violated the law
at issue in the case.\textsuperscript{121} For example, the Ninth Circuit granted
intervention of right to the National Organization for Women to defend a
constitutional challenge to the ratification procedures of the proposed
Equal Rights Amendment because it had a continuing interest in the
amendment that would be impaired by an adverse decision.\textsuperscript{122} The court
has also allowed the private sponsor of a ballot initiative to intervene of
right to defend a Commerce Clause\textsuperscript{123} challenge because of the
sponsor’s direct involvement in the adoption of the law.\textsuperscript{124} Similarly, the
court recently allowed the private sponsors of an anti-trapping initiative
to intervene to defend a constitutional challenge involving claims of
federal preemption under the Supremacy Clause\textsuperscript{125} as well as due
process and Commerce Clause claims.\textsuperscript{126}

The Ninth Circuit has also granted intervention of right to defend
procedural challenges to administrative decisions in environmental
litigation involving statutes other than NEPA. In \textit{Sagebrush Rebellion, Inc. v. Watt},\textsuperscript{127}
the court granted intervention of right to environmental
group has a significantly protectable
interest in the conservation of birds and their habitats that would, as a
practical matter, be impaired by an adverse judgment in the case.\textsuperscript{130}

\begin{itemize}
\item[121.] See, e.g., \textit{Freeman}, 625 F.2d at 887 (granting intervention of right to a private organization
to defend a challenge to procedures for adopting a constitutional amendment).
\item[122.] \textit{Id}.
\item[123.] Wash. State Bldg. & Constr. Trades Council v. Spellman, 684 F.2d 627, 629–30 (9th Cir.
1982).
\item[124.] \textit{Id}.
\item[125.] See \textit{Nat’l Audubon Soc’y, Inc. v. Davis}, 307 F.3d 835, 841–42 (9th Cir.), \textit{as amended on denial of reh’g en banc}, 312 F.3d 416 (9th Cir. 2002).
\item[126.] See \textit{id}.
\item[127.] 713 F.2d 525 (9th Cir. 1983).
\item[128.] \textit{Id} at 526, 529.
\item[129.] \textit{Id} at 526.
\item[130.] \textit{Id} at 526–28.
\end{itemize}
Similarly, absentees who petition to list a species under the ESA can intervene of right to defend litigation seeking to eliminate or reduce protections for that species because of the government’s alleged procedural violations. For example, in *Idaho Farm Bureau Federation v. Babbitt*, plaintiffs challenged the procedures that led to the listing of the Bruneau Hot Springs Snail under the ESA, including the alleged failure by the federal government to comply with time limits and notice requirements. The intervening environmental groups had actively participated in the administrative process leading to the decision, including filing the original petition for listing and suing the agency to force it to act. In determining that the conservation groups had a protectable interest sufficient to intervene of right under Rule 24(a)(2), the Ninth Circuit stated that “a public interest group is entitled as a matter of right to intervene in an action challenging the legality of a measure it has supported.”

In *Didrickson v. United States Department of the Interior*, the Ninth Circuit allowed permissive intervenors to appeal a district court decision that overturned a wildlife regulation when the original federal government defendant declined to appeal. Although this case involved permissive intervention, the court’s reasoning is relevant to determining whether intervenors have a right to appeal a district court decision invalidating a government action. The *Didrickson* court considered the practical effects of denying the request, noting that by acquiescing to the district court’s judgment overturning the regulation, the government had effectively promulgated a new regulation adverse to the intervenors’ position. Moreover, the intervenors participated in the rulemaking proceedings leading to the regulation, and could have challenged the government’s new position under the APA had this been the agency’s original decision. Thus, the lower court judgment created a sufficient

---

131. See *Idaho Farm Bureau Fed’n v. Babbitt*, 58 F.3d 1392, 1397 (9th Cir. 1995).
132. Id. at 1392 (9th Cir. 1995).
133. Id. at 1395.
134. Id. at 1398.
135. Id. at 1397 (emphasis added).
136. 982 F.2d 1332 (9th Cir. 1992).
137. Id. at 1337–39.
138. See id.
139. Id. at 1339.
140. Id. at 1338–39.
Rule 24 Intervention in NEPA Litigation

interest to justify permissive intervention. In addition, the court expressly rejected the government’s argument that, based on the Supreme Court’s decision in *Diamond*, it was the only party that could have standing to defend its own regulation in court.

In granting intervention to private groups to defend these diverse government actions, the Ninth Circuit has repeatedly noted the importance of the intervenors’ participation in the process of adopting the challenged law. By contrast, the Ninth Circuit has held that a group’s generalized environmental interests were not sufficient to intervene of right in defense of legislation that the group had not promoted. This participation interest continues to be a basis for granting intervention of right under Rule 24(a)(2) to non-governmental organizations to defend non-NEPA challenges.

B. The Ninth Circuit Has Severely Restricted Intervention of Right by Groups Other Than the Federal Government To Defend NEPA Challenges

In contrast to its decisions in non-NEPA cases, the Ninth Circuit has strictly limited the ability of private groups to intervene to defend NEPA challenges. In a series of NEPA cases beginning with *Portland Audubon Society v. Hodel*, the Ninth Circuit has stated two discrete rationales for denying defendant-side intervention. First, the court has adopted a blanket federal defendant rule, holding that because only the federal government can violate NEPA, it is the only proper defendant in a NEPA challenge. Second, in many of the same cases, the court also denied intervention because the applicants asserted purely economic

143. Id. at 1339–40.
144. See, e.g., id. at 1339 (noting that the proposed intervenors had participated in the relevant rulemaking process as a basis for granting permissive intervention); Sagebrush Rebellion, Inc. v. Watt, 713 F.2d 525, 526–27 (9th Cir. 1983) (citing the fact that “[h]oth groups participated actively in the administrative process . . . to establish the Birds of Prey Conservation Area”).
146. See, e.g., Nat’l Audubon Soc’y v. Davis, 307 F.3d 835, 842 (9th Cir. 2002) (allowing the sponsor of an anti-trapping initiative to intervene to defend a constitutional challenge).
147. 866 F.2d 302 (9th Cir. 1989).
148. Wetlands Action Network v. United States Army Corps of Eng’rs, 222 F.3d 1105, 1114 (9th Cir. 2000); Churchill County v. Babbitt, 150 F.3d 1072, 1082 (9th Cir. 1998); Sierra Club v. Envtl. Prot. Agency, 995 F.2d 1478, 1485 (9th Cir. 1993); see *Portland Audubon Soc’y v. Hodel*, 866 F.2d 302, 309 (9th Cir. 1989) (citing Wade v. Goldschmidt, 673 F.2d 182, 185 (7th Cir. 1982)).

1083
injuries that fall outside of NEPA’s zone of interest for the environment.149 These two competing rationales have led to ambiguity in the Ninth Circuit’s application of Rule 24,150 particularly when absentees assert environmental injuries. Several other federal courts have rejected the Ninth Circuit’s federal defendant rule as being inconsistent with the purposes of Rule 24 and its emphasis on practical considerations.151

The Ninth Circuit bases its federal defendant rule on the rationale that because NEPA establishes procedural requirements for the federal government, only the federal government can be liable for failing to comply with NEPA.152 For example, the court applied the federal defendant rule in *Wetlands Action Network v. United States Army Corps of Engineers*153 to deny intervention of right to a developer who sought to defend a NEPA challenge to a federal permit that allowed it to fill wetlands on its property.154 Because a private party cannot violate NEPA, the court ruled that the developer could not defend the NEPA challenge.155 Similarly, the court in *Sierra Club v. Environmental Protection Agency*156 stated that interests that might be significantly protectable in other circumstances may not suffice for intervention of right to defend a NEPA challenge.157

The Ninth Circuit has also applied the APA’s zone of interest test to determine whether prospective intervenors satisfy the protectable interest requirement for intervention of right in a NEPA case.158 In applying this test to determine whether a private party has standing to

---

149. *Wetlands*, 222 F.3d at 1109; *Sierra Club*, 995 F.2d at 1485; *Portland Audubon Soc’y*, 866 F.2d at 309.

150. *Kootenai II*, 313 F.3d 1094, 1108 (9th Cir. 2002).

151. Kleissler v. United States Forest Serv., 157 F.3d 964, 971 (3d Cir. 1998) (noting that the Ninth Circuit’s rigid rule in NEPA cases contravenes the major purpose of intervention, which is to protect third parties affected by pending litigation); *Sierra Club v. Espy*, 18 F.3d 1202, 1207 (5th Cir. 1994) (concluding that the stare decisis effect of a decision in a NEPA challenge could impair the intervenors’ property interests); *Wilderness Soc’y v. Babbitt*, 104 F. Supp. 2d 10, 18 (D.D.C. 2000) (concluding that the purposes of Rule 24 are best served by allowing interested parties to participate in all aspects of the litigation).

152. *See, e.g., Wetlands*, 222 F.3d at 1114.

153. *Id.* at 1105.

154. *Id.* at 1109.

155. *Id.* at 1114.

156. 995 F.2d 1478 (9th Cir. 1993).

157. *Id.* at 1485.

158. *See Forest Conservation Council v. United States Forest Serv.*, 66 F.3d 1489, 1497 (9th Cir. 1995) (noting that the intervenors asserted non-economic interests “within NEPA’s zone of concern for the environment”) (quoting Douglas County v. Babbitt, 48 F.3d 1495, 1497 (9th Cir. 1995)).
Rule 24 Intervention in NEPA Litigation

bring a NEPA challenge, the court has concluded that “the purpose of NEPA is to protect the environment, not the economic interests of those adversely affected by agency decisions.”159 Applying this reasoning to questions of intervention of right, the court has denied intervention to absentees who assert purely economic injuries because they do not have a protectable interest under NEPA.160 For example, in *Portland Audubon Society*, the court denied intervention to logging companies seeking to defend a NEPA challenge that would enjoin a group of timber sales.161 The court distinguished *Sagebrush Rebellion*, noting that the logging companies seeking to intervene in *Portland Audubon Society* asserted purely economic claims that were not protected by NEPA whereas the *Sagebrush Rebellion* intervenors asserted interests that were protected by the federal law at issue in the case.162

However, other statutes can create an interest sufficient to intervene of right to defend a NEPA challenge even when the absentees do not assert environmental interests. For example, in *Sierra Club v. Environmental Protection Agency*, the court concluded that the holder of a Clean Water Act discharge permit had a protectable interest sufficient to intervene of right to defend a NEPA challenge to that permit.163 Although NEPA did not protect the intervenor’s economic interest, the Clean Water Act did.164

The Ninth Circuit has granted a limited form of intervention to absentees asserting purely economic interests.165 The court denied intervention of right to defend the government’s compliance with NEPA in the liability phase of the trial.166 However, it granted permissive intervention under Rule 24(b)(2) to allow intervenors to participate in the remedial phase of the trial when the court decides whether to impose an injunction.167

159. Nev. Land Action Ass’n v. United States Forest Serv., 8 F.3d 713, 716 (9th Cir. 1993) (noting that a plaintiff who asserts purely economic injuries falls outside NEPA’s zone of interest).
161. *Id.*
162. *Id.*
164. *Id.*
165. *Wetlands Action Network v. United States Army Corps of Eng’rs*, 222 F.3d 1105, 1109, 1114 (9th Cir. 2000); Churchill County v. Babbitt, 150 F.3d 1072, 1083 (9th Cir. 1998).
166. *Wetlands*, 222 F.3d at 1109, 1114; *Churchill County*, 150 F.3d at 1083.
167. *See Wetlands*, 222 F.3d at 1109, 1114; *Churchill County*, 150 F.3d at 1083.
The Ninth Circuit combined its use of the federal defendant rule and NEPA’s zone of interest test in *Forest Conservation Council v. United States Forest Service*. The court held that entities asserting environmental injuries have a protectable interest sufficient to intervene of right in a NEPA challenge, but could only participate in the remedial phase of the trial. The court concluded that the State of Arizona’s interests in preventing forest fires and pest infestations on state land adjacent to the National Forests affected by the litigation constituted a concrete, plausible conservation concern within NEPA’s zone of concern for the environment. However, under the federal defendant rule, the state lacked a protectable interest in the government’s compliance with NEPA, and thus could participate only in the remedial phase of the trial.

Several other federal courts have considered and rejected the Ninth Circuit’s federal defendant rule for NEPA cases as unduly formalistic given the flexible and practical approach of Rule 24. In *Kleissler v. United States Forest Service*, a NEPA challenge involving timber sales that would provide revenue to local communities, the Third Circuit expressly rejected the “narrow approach” of the Ninth Circuit’s federal defendant rule. The court stated that the Ninth Circuit’s approach “minimizes the flexibility and spirit of Rule 24 as interpreted in *Cascade Natural Gas*. The *Kleissler* court granted intervention of right to local school districts and municipalities to participate in both the liability and the remedial phases of the trial, noting the practical effects of a potential injunction on the intervenors’ interests. Similarly, the Fifth Circuit granted intervention of right to defend both the liability and remedial phases of a NEPA challenge in *Sierra Club v. Espy*, because

168. 66 F.3d 1489 (9th Cir. 1995).
169. *Id.* at 1496, 1499.
170. *Id.* at 1497.
171. *Id.* at 1499 & n.11.
172. *Kleissler* v. United States Forest Serv., 157 F.3d 964, 971 (3d Cir. 1998); *Sierra Club v. Espy*, 18 F.3d 1202, 1207–08 (5th Cir. 1994); *Wilderness Soc’y v. Babbitt*, 104 F. Supp. 2d 10, 18 (D.D.C. 2000). However, at least one other federal court follows the federal defendant rule, applying reasoning very similar to the Ninth Circuit’s. See *Wade v. Goldschmidt*, 673 F.2d 182, 185 (7th Cir. 1982).
173. 157 F.3d 964 (3d Cir. 1998).
174. *Id.* at 971.
175. *Id.*
176. *Id.* at 972–73.
177. 18 F.3d 1202 (5th Cir. 1994).
Rule 24 Intervention in NEPA Litigation

the stare decisis effect of the decision would impair the intervenor’s protectable interests.\textsuperscript{178} Finally, the District Court for the District of Columbia noted that its circuit had not adopted the federal defendant rule and concluded that the purposes of Rule 24 would best be served by full participation by affected parties in a NEPA challenge.\textsuperscript{179}

In sum, the Ninth Circuit’s federal defendant rule in NEPA cases contrasts with its liberal grant of intervention when other statutes or regulations are at issue. Outside of the NEPA context, the Ninth Circuit has granted intervention to defend even constitutional and procedural violations that only the government could commit. The Ninth Circuit’s more restrictive rule for intervention in NEPA cases is also at odds with several other circuits’ NEPA cases.

IV. IN KOOTENAI TRIBE OF IDAHO V. VENEMAN, THE NINTH CIRCUIT APPLIED THE FEDERAL DEFENDANT RULE TO DENY INTERVENTION OF RIGHT

In Kootenai Tribe of Idaho v. Veneman, the Ninth Circuit substantially limited the ability of environmental groups to intervene to defend NEPA challenges. Although the Kootenai court upheld a far-reaching conservation law that would protect nearly sixty million acres of National Forests from logging and roadbuilding, it extended the federal defendant rule beyond its prior application by denying intervention of right to organizations that both asserted environmental injuries\textsuperscript{180} and participated actively in the rulemaking process.\textsuperscript{181} Instead, the court granted only permissive intervention.\textsuperscript{182} By so holding, the court left the environmental groups’ ability to intervene to defend future NEPA challenges to the discretion of the district court under Rule 24(b)(2).\textsuperscript{183}

The Forest Service adopted the Roadless Rule after numerous environmental groups worked for years to secure strong protections for

\textsuperscript{178}. Id. at 1207.
\textsuperscript{180}. Kootenai II, 313 F.3d 1094, 1104, 1109–10 (9th Cir. 2002).
\textsuperscript{181}. See Wyoming v. United States Dep’t of Agric., 201 F. Supp. 2d 1151, 1153–54 (D. Wyo. 2002) (hearing a parallel challenge to the Roadless Rule in which plaintiffs alleged that the federal government had violated the Federal Advisory Committee Act by working too closely with environmental groups to formulate the Roadless Rule).
\textsuperscript{182}. Kootenai II, 313 F.3d at 1110.
\textsuperscript{183}. See Northwest Forest Res. Council v. Glickman, 82 F.3d 825, 839 (9th Cir. 1996).
these remote wildlands in the National Forests. The Forest Service began to study and inventory roadless areas larger than five thousand acres in the 1970s. In response to active persuasion by environmental groups during the Clinton administration, the Forest Service initiated a NEPA rulemaking process in October 1999 to develop a nationwide plan to protect these roadless areas. The Forest Service issued a comprehensive EIS in May 2000 examining the potential impacts of implementing the Rule. Just days before leaving office, the Clinton administration issued a final rule that would ban most new roadbuilding and logging in 58.5 million acres of inventoried roadless areas in National Forests throughout the United States.

Environmental organizations participated extensively throughout the NEPA rulemaking process, actively seeking adoption of the Roadless Rule. As a result of a broad campaign by these and other organizations, the Forest Service received more than 1.15 million comments on the Roadless Rule EIS, ninety-six percent of which favored strong protections. In response to this public support, the Forest Service issued its final rule in January 2001, prohibiting most roadbuilding and increasing the area to be "committed to pristine wilderness" by seven million acres.

Three days later, the Kootenai Tribe of Idaho, the State of Idaho, and snowmobile, timber, and livestock groups filed suit in the District Court of Idaho to challenge the Roadless Rule. They alleged procedural violations of NEPA, including failure to consider an adequate range of alternatives. The newly-inaugurated President George W. Bush temporarily suspended implementation of the Roadless Rule. The Forest Service then announced that it would initiate a new rulemaking

184. See Wyoming, 201 F. Supp. 2d at 1153.
185. Kootenai II, 313 F.3d at 1104.
186. See Wyoming, 201 F. Supp. 2d at 1153.
187. Kootenai II, 313 F.3d at 1105.
188. Id.
189. Id.
190. Wyoming, 201 F. Supp. 2d at 1153.
191. See Kootenai II, 313 F.3d at 1116 n.19, 1119.
192. Id. at 1105–06.
193. Id. at 1104.
194. Id. at 1120.
195. Id. at 1106.
Rule 24 Intervention in NEPA Litigation

process because of the administration’s concerns over how the rule had been promulgated.196

Anticipating that the Bush administration might not vigorously defend the Roadless Rule, several environmental groups sought to intervene to defend the NEPA challenge.197 The district court granted intervention of right under Rule 24(a)(2) and concluded that the groups satisfied the requirements of having a protectable interest that could be impaired by an adverse judgment and that was not adequately represented by the existing parties.198 The court reasoned that the environmental groups asserted conservation interests protected by NEPA and fell within the Sagebrush Rebellion rule allowing groups that actively supported a federal regulation during the administrative rulemaking process to intervene of right to defend the regulation.199 The court further reasoned that the federal defendants may not adequately represent these conservation interests because the federal defendants must represent the broad public interest rather than the specific environmental concerns of the intervenors.200 In addition, the court alternatively granted permissive intervention under Rule 24(b)(2).201 However, after allowing the groups to intervene, the court rejected their NEPA arguments and held that the Forest Service had considered an inadequate range of alternatives and provided an inadequate comment period.202 The court later issued a temporary injunction blocking implementation of the Roadless Rule.203 The Forest Service declined to appeal this decision, and the environmental intervenors appealed without the original defendant.204

On appeal, the Ninth Circuit invoked the federal defendant rule to reject the district court’s grant of intervention of right under Rule 24(a)(2).205 The court first considered whether intervenors asserting environmental interests could defend the federal government’s alleged violations of NEPA’s procedural requirements without the government’s

196. Id.
197. Id.
199. Id.
200. Id.
201. Id.
203. Kootenai II, 313 F.3d 1094, 1106–07 (9th Cir. 2002).
204. Id. at 1107.
205. Id. at 1108.
participation. While admitting that its precedent on this issue was perhaps not “crystal clear,” the court determined that under the federal defendant rule, these groups lacked the significantly protectable interest in the rulemaking process required to intervene of right. In so holding, the court did not distinguish either its application of NEPA’s zone of interest test to grant intervention of right to groups asserting environmental injuries in *Forest Conservation Council* or the *Sagebrush Rebellion* rule that groups can intervene of right to defend a federal regulation they supported.

Although the Ninth Circuit denied intervention of right, it held that the district court had not abused its discretion by granting permissive intervention under Rule 24(b)(2). The court reasoned that the environmental groups raised common issues of law or fact with the primary claims in the case by asserting defenses directly responsive to the claim for an injunction and, while lacking a direct interest in the rulemaking process, had interests in the use and enjoyment of roadless areas that would be protected by the Roadless Rule. The Ninth Circuit noted that the federal government had declined from the beginning to defend the Roadless Rule and cited with approval the district court’s reasoning that participation by the intervenors would “contribute to the equitable resolution of this case.”

The Ninth Circuit also held that the intervenors had standing to appeal the decision without the federal government based on their environmental interests in the implementation of the rule. Under the U.S. Supreme Court’s holding in *Diamond*, the intervenors were required to establish independent Article III standing because the government did not appeal. In holding that the intervenors had standing, the Ninth Circuit reasoned that they had suffered an imminent

206. *Id.* at 1107.
207. *Id.* at 1108.
208. *Id.*
209. See *id.* at 1107–08.
210. *Forest Conservation Council v. United States Forest Serv.*, 66 F.3d 1489, 1497 (9th Cir. 1995).
212. *Kootenai II*, 313 F.3d at 1111.
213. *Id.* at 1110–11.
214. *Id.* at 1111.
215. *Id.* at 1109–10.
216. *Id.* at 1109; see supra Part II.C.
Rule 24 Intervention in NEPA Litigation

invasion of a “legally-protected interest” in their use of the public lands that would receive less protection if the Roadless Rule were not implemented.217 This injury was caused by the injunction and could be redressed by a decision to reinstate the Roadless Rule.218

After deciding these procedural issues, the Ninth Circuit held that the government had complied with NEPA and lifted the injunction.219 The court agreed with the intervenors’ arguments that the government had followed the correct procedures and considered an adequate range of alternatives.220 Noting that the agency was “now governed by a new presidential administration which is perhaps less sympathetic to the Roadless Rule,” the court expressly rejected the Forest Service’s argument that the court should leave the injunction in place because implementing the rule would cause irreparable harm.221 The court’s decision to reinstate the Roadless Rule thus depended solely on the participation of the intervenors, who presented the only defense to the NEPA challenge while the government argued against implementing its own regulation.222

In summary, the Kootenai court extended the federal defendant rule to disallow intervention of right by environmental groups that had actively participated in the NEPA process. These groups are limited to permissive intervention under Rule 24(b)(2). Thus, the Kootenai court limited future defendant-side intervention by environmental groups in NEPA cases to the discretion of district courts.

V. THE NINTH CIRCUIT ERRED IN DENYING INTERVENTION OF RIGHT TO ENVIRONMENTAL GROUPS SEEKING TO DEFEND THE ROADLESS RULE

The Kootenai court misapplied Rule 24 and contradicted the environmental purpose of NEPA by denying intervention of right to conservation groups that had actively participated in the adoption of the Roadless Rule. This case demonstrates that the Ninth Circuit’s federal

217. Kootenai II, 313 F.3d at 1109.
218. Id. at 1110. This standing analysis differs from the more typical case where the plaintiff must demonstrate Article III standing. Here, because the intervenors are defendants, their injury cannot be caused directly by the defendants as is typically required.
219. Id. at 1126.
220. Id. at 1121–22.
221. Id. at 1124.
222. See id. at 1111.
defendant rule contravenes Rule 24(a)(2) by failing to account for the practical impairment of the protectable interests of absentees. The Ninth Circuit should abandon its blanket rule denying defendant-side intervention of right in NEPA cases.\textsuperscript{223} Instead, the court should apply Rule 24(a)(2) by focusing on whether the interests asserted by absentees are protected by NEPA,\textsuperscript{224} whether the absentees established a protectable interest by actively participating in the administrative rulemaking process,\textsuperscript{225} and whether the outcome of litigation might impair these interests.\textsuperscript{226} Under this approach, the \textit{Kootenai} intervenors satisfied the requirements for intervention of right because they asserted protectable conservation interests under NEPA\textsuperscript{227} and had actively supported the Roadless Rule throughout the NEPA administrative rulemaking process.\textsuperscript{228}

\textbf{A. The Ninth Circuit’s Federal Defendant Rule Misapplies Rule 24}

The Ninth Circuit’s federal defendant rule misapplies Rule 24. The primary purpose of Rule 24 is to allow absentees to protect their interests by permitting them to participate in existing litigation that might impair these interests.\textsuperscript{229} By applying the federal defendant rule to deny intervention of right to environmental organizations, the Ninth Circuit has prevented these groups from joining litigation that might have a profound practical effect on their protectable interests unless the district court exercises its discretion to grant permissive intervention. This application of the federal defendant rule is inconsistent with the environmental purpose of NEPA because it renders participation by organizations asserting conservation interests less likely. In addition, the

\textsuperscript{223} The Ninth Circuit does not currently apply the federal defendant rule to statutes other than NEPA. \textit{See supra} Part III.A. However, for the same reasons discussed \textit{infra} Part V.A, the Ninth Circuit should likewise decline to extend the federal defendant rule to bar intervention to defend procedural challenges under other statutes.

\textsuperscript{224} \textit{See} Forest Conservation Council v. United States Forest Serv., 66 F.3d 1489, 1497 (9th Cir. 1995).

\textsuperscript{225} \textit{See}, e.g., Sagebrush Rebellion, Inc. v. Watt, 713 F.2d 525, 527 (9th Cir. 1983).

\textsuperscript{226} \textit{See} Cascade Natural Gas Corp. v. El Paso Natural Gas Co., 386 U.S. 129, 134 n.3 (1967).

\textsuperscript{227} \textit{See} Coalition of Ariz./N.M. Counties v. United States Dep’t of the Interior, 100 F.3d 837, 841–42 (10th Cir. 1996) (concluding that the ESA creates a protectable interest for a proposed intervenor who had petitioned for listing an endangered species).

\textsuperscript{228} \textit{See} Sagebrush Rebellion, 713 F.2d at 527.

\textsuperscript{229} \textit{See} Kleissler v. United States Forest Serv., 157 F.3d 964, 971 (3d Cir. 1998).
Rule 24 Intervention in NEPA Litigation

federal defendant rule contradicts the Ninth Circuit’s more liberal grant of intervention of right to defend laws other than NEPA cases. 230

1. The Ninth Circuit’s Federal Defendant Rule Misapplies Rule 24(a)(2) by Failing To Assess the Practical Effects of Pending Litigation on Protectable Environmental Interests

The federal defendant rule undermines the purpose of Rule 24 because district courts must deny intervention of right to groups that assert protectable interests in the environment even when the outcome of the litigation is likely to directly and profoundly affect these interests. 231 The U.S. Supreme Court has specifically rejected this type of rigid interpretation of the protectable interest requirement, instead directing courts to adopt a pragmatic approach. 232 As noted by the Third Circuit in Kleissler, the federal defendant rule is an unusually narrow and formalistic reading of Rule 24 that ignores the U.S. Supreme Court’s focus on practical effects in evaluating intervention requests. 233

The Kootenai court demonstrated this narrow reasoning by drawing an imperceptibly fine distinction between satisfying the injury-in-fact requirement for Article III standing and the protectable interest requirement for Rule 24(a)(2), concluding that the intervenors satisfied the former but not the latter. 234 Although the purposes of standing and intervention of right differ, the requirements for both are similar. 235 In its analysis of standing, the Kootenai court concluded that the intervenors had suffered an injury in fact, or an invasion of a “legally-protected interest,” to their use and appreciation of the areas to be protected by the Roadless Rule. 236 This injury was caused by the injunction that blocked implementation of the Roadless Rule, and could be redressed by lifting the injunction. 237 When the court applied Rule 24(a)(2)’s test for intervention of right, however, it invoked the federal defendant rule to conclude that the intervenors lacked a “significantly protectable

230. See supra Part III.A.
231. See Kootenai II, 313 F.3d 1094, 1108 (9th Cir. 2002).
232. See FED. R. CIV. P. 24 advisory committee’s notes to 1966 Amendments.
233. See Kleissler, 157 F.3d at 971.
235. See supra notes 102–04 and accompanying text.
236. Kootenai II, 313 F.3d at 1109.
237. Id. at 1110.
interest”238 in the litigation and were not so situated that “disposition of the action may as a practical matter impair or impede” their ability to protect this interest.239 In reaching these apparently contradictory conclusions, the court narrowly distinguished the intervenors’ protectable interest in the implementation of the Roadless Rule from their interest in the rulemaking process itself.240

This distinction fails to account for the practical effect of a court’s conclusion that the government violated NEPA: a significantly increased likelihood that it will enjoin the regulation that absentees seek to defend.241 Although the environmental organizations would not be directly bound by such a judgment, they would face significant practical barriers in future litigation based on principles of stare decisis.242 The Ninth Circuit’s approach in Forest Conservation Council of granting limited intervention of right to an absentee asserting environmental interests, but excluding the intervenor from the liability phase of the trial,243 likewise fails to consider the practical effect of a decision on the merits. A court is not likely to issue an injunction in the remedial phase of a NEPA challenge unless it has already concluded that the federal government violated NEPA in the liability phase.244 In Kleissler, the Third Circuit recognized this concern and rejected the Ninth Circuit’s bifurcated approach of denying intervention in the liability phase of a NEPA trial while allowing intervention in the remedial phase. The court concluded that it could not pragmatically apply the Ninth Circuit’s approach without “unduly attenuating” the absentees’ protectable interests.245

Finally, the Ninth Circuit’s federal defendant rule overlooks the effect on the rulemaking process when the government chooses not to appeal an adverse judgment. The Ninth Circuit in Didrickson recognized that the government’s acquiescence to an adverse judgment is effectively equivalent to the promulgation of a regulation reversing the agency’s

240. Kootenai II, 313 F.3d at 1108.
241. See Kleissler v. United States Forest Serv., 157 F.3d 964, 972 (3d Cir. 1998).
242. See supra notes 91–96 and accompanying text.
243. Forest Conservation Council v. United States Forest Serv., 66 F.3d 1489, 1499 (9th Cir. 1995).
244. See supra notes 91–96 and accompanying text.
245. See supra notes 91–96 and accompanying text.
Rule 24 Intervention in NEPA Litigation

original decision. 246 Had the agency directly issued such a regulation, advocates of the original regulation could have brought suit under the APA to challenge its decision. 247 Yet, under the Kootenai rule, the party that prevailed in the agency’s rulemaking process could be denied the opportunity to join litigation to defend the challenged regulation. 248 Because the Kootenai decision allows only permissive intervention under Rule 24(b)(2), the district court could exercise its discretion to deny intervention to environmental groups seeking to defend the agency’s original decision, 249 regardless of the extent of their participation in the NEPA rulemaking process or the subject of the litigation. If the district court denies permissive intervention, an appellate court is likely to uphold the decision because it reviews such a denial under the deferential abuse of discretion standard. 250 This effectively means that the parties that originally prevailed in the rulemaking process can be shut out of litigation despite their interests in the implementation of the regulation.

2. The Ninth Circuit’s Use of the Federal Defendant Rule To Deny Intervention by Conservation Groups Is Inconsistent with the Environmental Purpose of NEPA

The Kootenai court’s application of the federal defendant rule to deny intervention of right to groups asserting conservation interests conflicts with NEPA’s stated purpose of protecting the environment. 251 The Kootenai rule places environmental groups at a distinct practical disadvantage in future NEPA challenges seeking to overturn conservation-oriented regulations. Under the Ninth Circuit’s federal defendant rule, once the agency makes a decision, only the side that opposed that decision is entitled to participate in the liability phase of NEPA litigation. 252 As a result, the side that successfully persuades an agency to adopt its preferred regulation does not have an equal opportunity to participate in future litigation of the matter. The practical effect of denying intervention to environmental groups is potentially

248. See Kootenai II, 313 F.3d 1094, 1108 (9th Cir. 2002).
249. See FED. R. CIV. P. 24(b)(2) (providing that a court “may” grant intervention).
250. Greene v. United States, 996 F.2d 973, 978 (9th Cir. 1993).
251. See 42 U.S.C. § 4332(2) (2000); see also supra Part I.
252. See Kootenai II, 313 F.3d at 1108.
quite significant: if the federal government chooses not to defend the case, a court may invalidate a regulation regardless of whether the government complied with NEPA.

3. The Ninth Circuit’s Federal Defendant Rule Conflicts with Its Otherwise Liberal Grant of Intervention To Defend Challenges Brought Under Laws Other Than NEPA

The Ninth Circuit’s limited rationale for using the federal defendant rule in NEPA cases is further undermined by the court’s refusal to apply the rule in comparable non-NEPA cases. Notably, the court has not clearly articulated a distinction between NEPA litigation and other cases involving procedural or constitutional challenges. In *Sagebrush Rebellion*, for example, the court granted intervention of right to conservation groups to defend the federal government’s procedures in establishing a bird conservation area because the groups had participated in the administrative rulemaking process. Likewise, in *Didrickson*, the Ninth Circuit permitted environmental groups to independently appeal a decision overturning a wildlife regulation adopted through an APA rulemaking process because they had actively advocated for the regulation. In addition, the Ninth Circuit has repeatedly allowed private parties to intervene of right to defend a variety of constitutional challenges to government actions ranging from ratification procedures for a constitutional amendment to voter initiatives to block the import of radioactive wastes.

As nongovernmental organizations, the intervenors in these cases could not have violated either the constitutional provisions or the procedural requirements of the statutes at issue, but the Ninth Circuit
Rule 24 Intervention in NEPA Litigation

granted them intervention of right to defend the federal government’s actions. These grants of intervention in comparable non-NEPA cases directly contradict the reasoning behind the Ninth Circuit’s federal defendant rule that because only the government can violate the law, only the government can be a defendant. Applying the federal defendant rule in NEPA cases directly conflicts with the Ninth Circuit’s granting of intervention to defend procedural and constitutional challenges when other laws are at issue.

In Sagebrush Rebellion and related cases, the Ninth Circuit correctly applied Rule 24 by focusing on the practical effects of the litigation on the interests of absentees. In these cases, the court did not focus on whether the intervenors could violate the statutes in question, but instead applied Rule 24 to grant intervention based on the absentees’ protectable interests that would be affected by the outcome of the litigation.

B. The Environmental Intervenors Were Entitled To Intervene of Right Because They Asserted Injuries Within NEPA’s Zone of Interest for the Environment and Participated Directly in the Rulemaking Process

Under a proper application of Rule 24, the Kootenai intervenors satisfied the requirements for intervention of right. The court expressly recognized that the intervenors met three of the four requirements of Rule 24(a)(2): the application was timely, their interests could be impaired by a decision in the case, and the government did not provide adequate representation. The court should have recognized that the environmental intervenors met the remaining requirement—that of having a significantly protectable interest—because they asserted conservation interests in the subject of the litigation that are protected by NEPA. In addition, the intervenors had a protectable interest in defending the rulemaking process under the court’s holding in Sagebrush Rebellion because they had participated directly in the adoption of the Roadless Rule.

259. See supra Part III.A.
260. See Kootenai II, 313 F.3d 1094, 1108 (9th Cir. 2002).
261. See supra notes 121–26 and accompanying text.
262. See supra Part III.A.
263. See Kootenai II, 313 F.3d at 1108–11.
264. See supra notes 87–89 and accompanying text.
The *Kootenai* intervenors asserted interests in the Roadless Rule that fell within NEPA’s zone of concern for the environment and thus established a protectable interest sufficient to intervene of right under Rule 24(a)(2). When evaluating intervention motions, the court should consider the purpose of the underlying statute to determine whether it creates a protectable interest. This is particularly appropriate in NEPA litigation because the APA, which provides the cause of action, requires plaintiffs to assert an injury within the zone of interest of the underlying statute. The Ninth Circuit has acknowledged that NEPA’s environmental purpose is relevant to evaluating intervention requests. In *Forest Conservation Council*, the court granted intervention of right to the State of Arizona to defend the remedial phase of a NEPA challenge because the state asserted interests in the environmental health of its adjacent forests that were “concrete, plausible interests within NEPA’s zone of concern for the environment.”

The Ninth Circuit should have recognized that the potential environmental injuries asserted by the *Kootenai* intervenors were likewise within the zone of interest protected by NEPA. Instead, without distinguishing its grant of intervention of right in *Forest Conservation Council*, the *Kootenai* court rejected the proposition that groups asserting environmental interests can intervene of right to defend NEPA cases. The practical effect of denying intervention in this case would have been the invalidation of a conservation rule protecting vast tracts of ecologically significant forestland. Although environmental groups would not have been directly bound by res judicata in future litigation of the issue, stare decisis would pose a significant practical barrier to relief. This change in the regulation would result not from

---

265. See supra notes 87–89 and accompanying text.
269. *Forest Conservation Council v. United States Forest Serv.*, 66 F.3d 1489, 1497 (9th Cir. 1995).
270. *Id.*
271. *See Kootenai II*, 313 F.3d 1094, 1108 (9th Cir. 2002).
272. *Id.*
273. *See id. at 1111* (noting that the federal government declined to defend the Roadless Rule).
274. See supra notes 91–96 and accompanying text.
Rule 24 Intervention in NEPA Litigation

the environmentally-informed, public decisionmaking process mandated by NEPA, but instead from a court’s decision in a case that was not defended. Such an outcome would contravene the fundamental purpose of NEPA.

In addition to having a protectable interest under NEPA, the environmental intervenors had a separate protectable interest in the rulemaking process under Sagebrush Rebellion because they directly and actively promoted the adoption of the Roadless Rule. The Ninth Circuit has repeatedly granted intervention of right to groups to defend procedural or constitutional challenges to laws they helped to adopt. For example, in Sagebrush Rebellion, the court granted intervention of right to environmental groups that had actively supported the federal government’s establishment of a national conservation area during the rulemaking process. Similarly, in Kootenai, the environmental intervenors actively participated in the NEPA rulemaking process, encouraging participation by the public that led to submission of more than 1.15 million comments. Both cases involved a change in presidential administration that resulted in a dramatically different approach to conservation policy after the rules were adopted. The only significant distinction is that Kootenai is a NEPA case and Sagebrush Rebellion is not.

VI. CONCLUSION

The Kootenai decision illustrates that the Ninth Circuit’s federal defendant rule for NEPA cases misapplies Rule 24 by failing to account for the practical effects of litigation on the protectable interests of absentees. These interests include environmental concerns protected by NEPA and participation interests in the administrative rulemaking process. The Ninth Circuit should abandon its blanket federal defendant
rule and instead determine whether the absentee asserts a significantly protectable interest in the subject matter of the litigation that may, as a practical matter, be impaired by the outcome of the litigation. Applying this test, the *Kootenai* intervenors met the requirements to intervene of right under Rule 24(a)(2) because they asserted environmental interests protected by NEPA, these interests would be practically impaired by the outcome of the litigation, and the interests were not adequately represented by the existing parties.
ABROGATION OR REGULATION? HOW ANDERSON V. EVANS DISCARDS THE MAKAH’S TREATY WHALING RIGHT IN THE NAME OF CONSERVATION NECESSITY

Zachary Tomlinson

Abstract: From 1787 to 1871, the federal government and various Indian tribes entered into hundreds of treaties. Under well-established U.S. Supreme Court precedent, the U.S. Congress has plenary authority to abrogate or modify any of these treaties. The U.S. Supreme Court is reluctant to find congressional intent to do so, however, and requires that this intent be clear and plain. States have no such power to qualify treaties, but the Court has allowed states to regulate treaty rights when doing so is necessary for species conservation. While the U.S. Supreme Court has kept these two lines of cases distinct, the U.S. Court of Appeals for the Ninth Circuit has merged the two doctrines in recent years. The Ninth Circuit’s recent decision in Anderson v. Evans, in which the court held that the Marine Mammal Protection Act (MMPA) applied to the Makah Tribe’s treaty whaling rights, dramatically illustrates this practice. This Note argues that the Ninth Circuit’s conflation of federal treaty abrogation principles with state conservation necessity principles is analytically indefensible and in direct contravention to established U.S. Supreme Court precedent. Under correctly applied U.S. Supreme Court precedent, the Makah’s treaty whaling right is not subject to the MMPA. The Anderson decision is a violation of the United States’ treaty obligation to the Makah Tribe and should be overturned.

Long before gray whales were hunted to the brink of extinction in the early 1900s, the animals were central to the cultural, spiritual, and economic existence of the Makah Tribe. Until the 1930s, the Makah hunted gray whales in the waters off the northwest coast of present-day Washington State. Nowhere is the Makah’s connection with the gray whale better illustrated than in the 1855 Treaty of Neah Bay, in which the Tribe ceded most of its land in part for the right to take whales “at usual and accustomed grounds and stations . . . in common with all

1. See United States v. Washington, 384 F. Supp. 312, 363 (W.D. Wash. 1974) (“The Makah Indians, prior to treaty times, were primarily a seafaring people who spent their lives either on the water or close to the shore. Most of their subsistence came from the sea where they fished for salmon, halibut and other fish, and hunted for whale and seal. The excess of what they needed for their own consumption was traded to other tribes for many of the raw materials and some of the finished articles used in the daily and ceremonial life of the village.”), aff’d, 520 F.2d 676 (9th Cir. 1975).
2. See Anderson v. Evans, 314 F.3d 1006, 1009 (9th Cir. 2002); Metcalf v. Daley, 214 F.3d 1135, 1137 (9th Cir. 2000).
citizens of the United States." In fact, the Makah is the only tribe in the United States with an explicit treaty right to hunt whales. However, a recent ruling by the U.S. Court of Appeals for the Ninth Circuit has severely limited the ability of the Makah Tribe to exercise this right. In Anderson v. Evans, the court held that the Marine Mammal Protection Act (MMPA), which generally prohibits the taking of marine mammals, applied to the Makah’s whaling efforts. Without ruling that Congress had abrogated the Makah’s treaty whaling right, the court held that the application of the MMPA to the Tribe’s whaling efforts was necessary to achieve the statute’s “conservation purpose.”

The U.S. Supreme Court has consistently held that Congress has plenary power to abrogate or modify Indian treaties. However, because the federal government and the tribes negotiated the treaties in conditions severely disadvantageous to the Indian tribes, the Court has created canons of Indian treaty construction that favor Indian tribes and protect treaty rights. Foremost among these canons is the principle that congressional intent to abrogate a treaty right must be clear and plain, and will not be easily imputed.

States, on the other hand, have no such power to abrogate or modify Indian treaties. By virtue of the Supremacy Clause, Indian treaties take precedence over any conflicting state laws. Nonetheless, in deference to the police power of the states, the U.S. Supreme Court has held that states can regulate Indian treaty fishing rights when doing so is necessary for conservation. However, courts are reluctant to find such “conservation necessity” in state regulations.
Abrogation or Regulation?

This Note argues that in Anderson v. Evans, the Ninth Circuit erred in using the conservation necessity doctrine to hold that the Makah Tribe is subject to the MMPA, a federal statute. This error is a direct result of the court’s previous conflation of two analytically distinct U.S. Supreme Court doctrines: treaty abrogation principles, properly applied only in cases involving federal abrogation or modification of treaty rights; and the conservation necessity principle, properly applied only in cases involving state regulation of treaty rights. Part I of this Note reviews the development and rationale behind each doctrine. Part II examines the Ninth Circuit’s combination of these two doctrines and notes the rejection of this approach in other circuits. Part III discusses the Ninth Circuit’s recent affirmation and extension of this approach in Anderson. Finally, Part IV argues that in Anderson, the Ninth Circuit improperly analyzed the impact of the MMPA on the Makah’s treaty whaling right. Under a proper abrogation analysis, the Makah are not subject to the provisions of the MMPA because there is no indication that Congress, in passing the statute, intended to modify or abrogate the Makah’s treaty right to hunt whales. Using the conservation necessity doctrine as a substitute for abrogation analysis in the context of federal legislation is doctrinally inexplicable and unprecedented in U.S. Supreme Court jurisprudence. The Ninth Circuit’s Anderson decision is a violation of the United States’ treaty obligation to the Makah, and should be overturned.

I. THE SCOPE OF FEDERAL AND STATE POWER OVER INDIAN TREATY RIGHTS: TWO DIFFERENT RATIONALES, TWO DIFFERENT ANALYSES

The U.S. Supreme Court has held that both the U.S. Congress and the states have some ability to limit the exercise of Indian treaty rights.\(^\text{19}\) The nature of the federal and state powers is fundamentally different, however, and developed in two distinct lines of cases.\(^\text{20}\) Congress has plenary power to modify or abrogate treaty rights; the focus of a judicial inquiry is on whether Congress intended to do so in a specific instance.\(^\text{21}\) In contrast, states have only a limited ability to affect Indian treaty rights may be regulated only upon showing that unregulated treaty fishing would cause “irreparable harm” to fisheries within state).

\(^{19}\) See infra Part I.A–B.
\(^{20}\) See infra Part I.A–B.
\(^{21}\) See infra Part I.A.
Among other restrictions, courts insist that state regulations be reasonable and necessary for conservation purposes.\(^{23}\)

**A. Although Congress Has Plenary Power To Modify or Abrogate Indian Treaties, Its Intent To Do So Must Be Clear and Plain**

As a result of its plenary authority over Indian affairs, Congress has the power to statutorily abrogate or modify Indian treaties.\(^{24}\) Courts strongly presume, however, that federal statutes of general applicability do not automatically affect Indian treaty rights.\(^{25}\) Instead, prior to concluding that a congressional act affects a treaty right, courts must find that Congress considered the Indian treaty right in question and subsequently chose to abrogate or modify the right by passing the statute.\(^{26}\)

The U.S. Supreme Court offered one of its strongest expressions of the presumption against treaty abrogation in *Menominee Tribe of Indians v. United States*.\(^{27}\) The Court held that a 1954 statute terminating the trust relationship with the Menominee Tribe\(^ {28}\) did not extinguish the Tribe’s hunting and fishing treaty rights.\(^ {29}\) While the statute provided that “all statutes of the United States which affect Indians because of their status as Indians shall no longer be applicable to the members of

22. See infra Part I.B.

23. See infra notes 73–83 and accompanying text.

24. See Lone Wolf v. Hitchcock, 187 U.S. 553, 565–66 (1903). However, as later cases have clarified, the exercise of such power would likely subject the United States to a takings claim under the Fifth Amendment. See, e.g., *Menominee Tribe of Indians v. United States*, 391 U.S. 404, 413 (1968).

25. See COHEN, supra note 12, at 222–23. In *Federal Power Commission v. Tuscarora Indian Nation*, 362 U.S. 99 (1960), the U.S. Supreme Court stated that “general Acts of Congress apply to Indians as well as to all others in the absence of a clear expression to the contrary.” Id. at 120. However, Tuscarora was not a treaty rights case, and, as the Ninth Circuit has noted, the Tuscarora rule “does not apply to Indians if the application of the general statute would be in derogation of the Indians’ treaty rights.” Donovan v. Navajo Forest Prods. Indus., 692 F.2d 709, 711 (10th Cir. 1982).


29. Menominee, 391 U.S. at 412–13. The Menominee Tribe’s hunting and fishing rights were not explicitly granted by treaty. See Treaty of Wolf River, May 12, 1854, U.S.-Menominee Tribe, 10 Stat. 1064. However, the Court held that the treaty language reserving for the Tribe a home “to be held as Indian lands are held” included the right to hunt and fish. See Menominee, 391 U.S. at 405–06.
Abrogation or Regulation?

the tribe," the Court held that the use of the word “statutes” was “potent evidence” that treaty rights remained unaffected. Reaffirming the principle that “the intention to abrogate or modify a treaty is not to be lightly imputed to the Congress,” the majority refused to construe the statute as a “backhanded” way of abrogating the Menominee’s hunting and fishing rights.

While the Menominee decision was a strong statement about the importance of treaty rights, it did not establish whether Congress’ intent to abrogate a treaty must be explicit in the language of the statute. The Court addressed this question in United States v. Dion, holding that there must be “clear evidence that Congress actually considered the conflict between its intended action on the one hand and Indian treaty rights on the other, and chose to resolve that conflict by abrogating the treaty.” Congressional intent to abrogate can be found in either the language of the statute or from “clear and reliable evidence in the legislative history of a statute.”

In Dion, the state convicted Dwight Dion, Sr., a member of the Yankton Sioux Tribe, of shooting four bald eagles on the Tribe’s reservation in South Dakota in violation of both the Eagle Protection Act and the Endangered Species Act (ESA). In his defense, Dion asserted his treaty right to hunt eagles for noncommercial purposes. Rejecting Dion’s treaty defense, the Court first reasoned that the language of the Eagle Protection Act strongly suggested congressional intent to abrogate Indian treaty rights to take eagles. Although the

---

32. Id. at 412–13 (citing Pigeon River, Co. v. Charles W. Cox, Ltd., 291 U.S. 138, 160 (1934)).
33. Id. at 413.
34. Id. The Court also felt it unlikely that Congress would have intentionally subjected the federal government to a claim for compensation under the Fifth Amendment. Id.
36. Id. at 739–40.
37. Id. at 739 (quoting COHEN, supra note 12, at 223).
38. 16 U.S.C. §§ 668a–668d (2000). The Eagle Protection Act, as originally enacted, only applied to bald eagles. However, in 1962, the protections were extended to golden eagles as well. See Golden Eagle Protection Act, Pub. L. No. 87-884, 76 Stat. 1246 (1962) (codified at 16 U.S.C. § 668a (2000)).
41. Id. at 745. The Court found that it did not need to consider whether the ESA abrogated Dion’s
Eagle Protection Act contained no language that explicitly mentioned abrogation of Indian treaty rights, the Court noted that the Eagle Protection Act authorized the Secretary of the Interior to issue permits to Indian tribes for the taking of eagles for “religious purposes.” The Court noted that such language was powerful evidence that Congress considered the treaty rights and subsequently chose to abrogate them.

In addition, the Dion Court looked to the legislative history of the Eagle Protection Act. The original version of the statute, passed in 1940, only applied to bald eagles and contained no explicit references to Indians. In 1962, however, Congress extended the statute’s prohibitions to golden eagles and instituted the permitting process described above. The amendment reflected an “unmistakable and explicit legislative policy choice” that unregulated Indian hunting of bald eagles was inconsistent with the need for species preservation. As a whole, the legislative history provided additional evidence that Congress considered the treaty right of all tribes to take eagles and chose to abrogate those rights.

The Court has applied the Dion framework in two subsequent cases. In South Dakota v. Bourland, the Court cited the Dion test with approval in holding that Congress abrogated the Cheyenne River Sioux Tribe’s treaty right to “absolute and undisturbed use and occupation” of former tribal trust lands when it acquired the lands from the Tribe for a dam and flood control project. In acquiring the lands, Congress explicitly reserved the right of the Tribe to hunt and fish around the newly created reservoir “subject to the regulations governing the corresponding use by other citizens of the United States.” Following treaty rights because the Eagle Protection Act had already removed his “treaty shield.”

42. Id. at 740–41.
43. Id. at 740 (quoting Eagle Protection Act, 16 U.S.C. § 668a (2000)).
44. Id.
45. Id. at 740–44.
47. Dion, 476 U.S. at 741.
48. Id. at 745.
49. Id. at 743.
51. Id. at 693.
52. Id. at 690 (citing Cheyenne River Act of Sept. 3, 1954, 68 Stat. 1191, 1193 (1954)).
Abrogation or Regulation?

*Dion*, the Court held that such a limited reservation of rights could not be explained without finding that Congress intended to abrogate the right of the Tribe to regulate hunting and fishing on the land by non-Indians.53 Most recently, in *Minnesota v. Mille Lacs Band of Chippewa Indians*,54 the Court again cited the *Dion* test with approval in holding that there was no “clear evidence” that Congress intended to abrogate Chippewa treaty rights in passing Minnesota’s Enabling Act.55 The Court noted that the Act made no mention of Indian treaty rights, and provided “no clue that Congress considered the reserved rights of the Chippewa and decided to abrogate those rights when it passed the Act.”56 Both *Bourland* and *Mille Lacs* illustrate the continued vitality of the *Dion* framework.

Thus, U.S. Supreme Court precedent weighs heavily against the finding of congressional intent to abrogate treaty rights.57 Under *Dion*, the Court requires evidence that Congress considered the treaty right in question and nonetheless chose to modify or abrogate that right.58 While the intention to abrogate treaty rights need not be explicit on the face of a statute, evidence of abrogation must be “clear and plain” in light of the “fundamental importance” of Indian treaty rights.59

B. States Cannot Modify or Abrogate Indian Treaty Rights, but They Have a Limited Ability To Regulate Off-Reservation Treaty Rights When Necessary for Conservation

By virtue of the Supremacy Clause,60 Indian treaties are superior to conflicting state laws or constitutional provisions.61 As a result, states do not have the ability to qualify the rights guaranteed in Indian treaties.62

53. *Id.* at 693–94. While the Court analyzed the abrogation issue under the *Dion* framework, the Court’s holding of abrogation provoked a strong dissent. *See id.* at 700 (Blackmun, J., dissenting) (stating that the “majority adopts precisely the sort of reasoning-by-implication that [Dion and other cases] reject”).
55. *Id.* at 203.
56. *Id.*
57. *See COHEN, supra* note 12, at 222–23.
59. *Id.*
60. U.S. CONST. art. VI, cl. 2.
61. *See COHEN, supra* note 12, at 62.
62. *See, e.g., Payallup I*, 391 U.S. 392, 398 (1968) (stating that a treaty right to fish at all usual and accustomed places “may, of course, not be qualified by the State”).
However, in recognition of the potential effects of a treaty on a state’s police powers, the U.S. Supreme Court has held that states do have the power to regulate off-reservation Indian fishing rights when necessary for conservation.63

The Court first recognized the ability of states to regulate Indian treaty fishing rights for purposes of conservation in 1968.64 In *Puyallup Tribe v. Department of Game (Puyallup I)*,65 the Court held that off-reservation Indian treaty fishing rights were subject to limited state regulation.66 At issue in the case was the ability of Washington State to enforce regulations banning net fishing of steelhead and salmon by the Puyallup Tribe.67 The Court first reasoned that the Puyallup’s treaty68 guaranteed the Tribe the right to fish at its “usual and accustomed places,” but not the right to fish in its “usual and accustomed manner.”69 Furthermore, the Court noted that the off-reservation fishing right was not an exclusive one, but rather one held in common with all citizens of Washington Territory.70 The Court held that the terms of the treaty preserved “the overriding police power of the State” to enact

63. See id. The regulation of *on-reservation* Indian treaty rights is beyond the scope of this Note as the Makah’s treaty whaling right is an off-reservation right held “in common with the citizens of the territory.” In general, the justification for allowing the regulation of *off-reservation* treaty rights lies primarily in the non-exclusive nature of such rights, indicated by the “in common” language of many Indian treaties. However, in at least one case, the U.S. Supreme Court has applied this same rationale to state regulation of *on-reservation* treaty rights, which are otherwise exclusive rights according to the terms of the treaties. See *Puyallup Tribe v. Dep’t of Game*, 433 U.S. 165, 173–77 (1977).

64. See *Puyallup I*, 391 U.S. at 398. Prior to 1968, the Court had only indirectly addressed whether states could regulate Indian treaty rights. See, e.g., *Tulee v. Washington*, 315 U.S. 681, 684–85 (1942) (holding that the state of Washington could not charge treaty fishermen a fee to exercise “the very right their ancestors intended to reserve,” but noting in dicta that the state could impose “purely regulatory” restrictions on Indian treaty rights to the extent “necessary for the conservation of fish”); *United States v. Winans*, 198 U.S. 371, 384 (1905) (holding that treaty fishermen had an easement to cross and use private land in the exercise of their treaty fishing rights, but noting in dicta that treaty did not “restrain the state unreasonably, if at all, in the regulation of the right”).


66. See id. at 398. The case arose when the Washington Department of Game brought suit against two Indian tribes to prevent them from fishing in violation of a blanket ban on the use of net fishing in Washington territorial waters, applicable to both Indians and non-Indians. *Id.* at 395–96, 400.

67. *Id.* at 395–96.

68. Treaty of Medicine Creek, Dec. 26, 1854, 10 Stat. 1132.


70. *Id.*
Abrogation or Regulation?

nondiscriminatory measures to conserve fish resources. The Court described this new conservation necessity standard, and its limitations, in the following terms:

The right to fish “at all usual and accustomed” places may, of course, not be qualified by the State . . . . But the manner of fishing, the size of the take, the restriction of commercial fishing, and the like may be regulated by the State in the interest of conservation, provided the regulation meets appropriate standards and does not discriminate against the Indians.

The Court has developed these limitations on state regulation in subsequent cases. To show that regulation of treaty rights meets “appropriate standards,” the state must demonstrate that the regulation is a “reasonable and necessary conservation measure, and that its application to the Indians is necessary in the interest of conservation.” Moreover, the principle of non-discrimination also guarantees treaty fishermen an actual share of the harvestable fish, and not just the mere opportunity to fish on the same terms as other citizens. For example, the second Puyallup case to reach the Court, Department of Game v. Puyallup Tribe (Puyallup II), directly addressed whether a state steelhead fishing regulation prohibiting net fishing met the Puyallup I standard. Although the regulation was facially neutral, the Court found that it had the effect of banning all tribal steelhead fishing, and was thus

71. Id. Commentators have suggested that the Puyallup I decision was grounded more on a fear that the salmon would be fished to extinction, rather than on any principled legal basis. See Ralph W. Johnson, The States Versus Indian Off-Reservation Fishing: A United States Supreme Court Error, 47 WASH. L. REV. 207, 208 (1972) (“No valid basis for the existence of such state power can be found . . . . The treaties with the Indians do not provide for state regulation and Congress has never authorized such regulation.”).


76. Id. at 48–49. Under Washington law at the time, the Department of Fisheries regulated salmon fishing, while the Department of Game regulated steelhead fishing. After the Puyallup I decision, the Department of Fisheries changed its blanket prohibition against salmon net fishing to allow Indian net fishing in certain areas of the Puyallup River. Id. at 46. The Department of Game, however, maintained its blanket prohibition of steelhead net fishing in the River, and it was this regulation that the Puyallup Tribe contested in Puyallup II. Id.
discriminatory. The Court refined this line of reasoning in Washington v. Washington State Commercial Passenger Fishing Vessel Ass’n, holding that Indians would have understood the treaty term “taking fish in common with all citizens of the Territory” to guarantee the tribes an actual share of the fish.

Lower courts have interpreted the scope of a state’s conservation necessity power quite narrowly. Some courts have required a showing of imminent species extinction, and all have required proof of substantial species peril. Further, some lower courts have also concluded that states cannot regulate for purposes of conservation when other, less restrictive means are available, such as existing tribal self-regulation. For example, in United States v. Washington, a district court required the state to demonstrate that existing tribal regulation or enforcement was “inadequate to prevent demonstrable harm to the actual conservation of fish,” and that other, less restrictive means or methods could not achieve the conservation goals.

---

77. Id. at 48. The Court found that while the regulations allowed line and hook fishing by all citizens, including Indians, such fishing was “entirely pre-empted by non-Indians,” thus effectively granting the entire steelhead run to non-Indian sports fishermen. Id. Thus, while the treaty fishing rights did not give the Puyallup Tribe a right to “pursue the last living steelhead until it enters [its] nets,” the state did not have the power to regulate so as to deny the Puyallup a fair apportionment of the steelhead in the river. Id.


79. Id. at 675–85. The Court noted that the premise that “each individual Indian would share an ‘equal opportunity’ with thousands of newly arrived individual settlers is totally foreign to the spirit of the [treaty] negotiations.” Id. The rights possessed by the Indians “are admittedly not ‘equal,’ but are to some extent greater than, those afforded other citizens.” Id. at 677. In practice, this has meant a court-imposed allocation of fish between the interested treaty and non-treaty parties. See, e.g., United States v. Washington, 384 F. Supp. 312, 343 (W.D. Wash. 1974) (holding that “in common” off-reservation treaty fishing right entitled the Puyallup Tribe up to fifty percent of harvestable run of fish after allowing for escapement), aff’d, 520 F.2d 676 (9th Cir. 1975).

80. See, e.g., Sohappy v. Smith, 302 F. Supp. 899, 908 (D. Or. 1969) (holding that “conservation necessity” required the state to show that the continued exercise of the unregulated treaty fishing right would imperil the existence of the species).

81. See, e.g., United States v. Oregon, 718 F.2d 299, 304–05 (9th Cir. 1983) (stating that treaty rights may be regulated only where necessary to preserve a “reasonable margin of safety” against the imminence of extinction); United States v. Michigan, 653 F.2d 277, 279 (6th Cir. 1981) (holding that treaty rights may be regulated only upon showing that unregulated treaty fishing would cause “irreparable harm” to fisheries within state); Purse Seine Vessel Owners Ass’n v. State, 92 Wash. App. 381, 392, 966 P.2d 928, 934 (1998) (noting that state regulation of treaty fishing rights must be “essential” to the conservation of fisheries).

82. 384 F. Supp. 312 (W.D. Wash. 1974), aff’d, 520 F.2d 676 (9th Cir. 1975).

83. Id. at 415; see also Mille Lacs Band v. Minnesota, 952 F. Supp. 1362, 1380 (D. Minn.) (noting that state cannot regulate Indian treaty fishing if tribe “can effectively self-regulate and such tribal regulations are sufficient to meet conservation needs”), aff’d, 124 F.3d 905 (8th Cir. 1997),
Abrogation or Regulation?

In sum, states have the ability to regulate tribal treaty rights to the extent necessary for conservation. In practice, however, courts have required that states meet a formidable burden in showing the necessity of such regulation. The state may not regulate to the extent that it denies the treaty tribes an actual share of the harvestable resource—even where conservation regulations are facially neutral. Furthermore, while the state may not have to show that the species is subject to imminent extinction, it must prove substantial species peril. Finally, the state must also show that existing tribal self-regulation, or other, less restrictive means or methods, is inadequate to achieve conservation.

II. THE NINTH CIRCUIT DEVELOPS A NEW TREATY ABROGATION STANDARD, BUT FINDS ITSELF ALONE

As the above cases illustrate, the U.S. Supreme Court has kept its method of analyzing treaty abrogation cases (governed by the Dion standard) distinct from its review of state regulation of treaty rights (governed by the Puyallup line of cases). However, the Ninth Circuit has combined the two methodologies, using state conservation necessity principles to determine the effect of federal statutes on Indian treaty rights. Most other courts have not adopted the Ninth Circuit’s methodology, instead requiring a “clear and plain” showing under Dion that Congress considered the Indian treaty right in question and subsequently chose to abrogate that right.

A. The Ninth Circuit’s Federal Conservation Necessity Standard

The U.S. Supreme Court has never used state conservation necessity principles to analyze whether a federal statute has abrogated Indian treaty rights. The Ninth Circuit, however, has done so on several occasions, in a line of cases originating with United States v. Fryberg.

aff’d, 526 U.S. 172 (1999); Lac Courte Oreilles Band v. Wisconsin, 668 F. Supp. 1233, 1236 (W.D. Wis. 1987) (noting that state regulation of Indian treaty fishing must be the “least restrictive alternative available”).
85. See United States v. Dion, 476 U.S. 734, 739–40 (1986); supra notes 23–57 and accompanying text.
86. See supra notes 64–79 and accompanying text.
87. See infra notes 89–108 and accompanying text.
88. See infra notes 109–15 and accompanying text.
89. 622 F.2d 1010 (9th Cir. 1980).
In *Fryberg*, the court faced the same issue that was later addressed by the U.S. Supreme Court in *Dion*: whether the Eagle Protection Act modified or abrogated treaty rights to take and kill bald eagles.\(^90\) Using reasoning different from that later adopted by the *Dion* Court, the Ninth Circuit held that the application of the Eagle Protection Act to the treaty rights was necessary to achieve the statute’s “conservation purpose.”\(^91\)

The *Fryberg* court noted that the statute did not itself indicate an “unambiguous express intent” to abrogate or modify Indian treaty rights.\(^92\) However, the court held that it was appropriate to look beyond the language of the statute to the Eagle Protection Act’s legislative history and “surrounding circumstances” to help determine congressional intent.\(^93\) In particular, the *Fryberg* court held that such “surrounding circumstances” could include the statute’s inherent “conservation purpose.”\(^94\) The court observed that the U.S. Supreme Court, in the state regulatory context, had “long recognized that reasonable and non-discriminatory conservation statutes implicitly affect treaty rights to the extent necessary to achieve their conservation purpose.”\(^95\) Looking to the *Puyallup* line of cases, the court established the following test:

\[ \text{[R]easonable conservation statutes affect Indian treaty rights when (1) the sovereign exercising its police power to conserve a resource has jurisdiction in the area where the activity occurs; (2) the statute applies in a non-discriminatory manner to both treaty and non-treaty persons; and (3) the application of the statute to treaty rights is necessary to achieve its conservation purpose.} \]

\(^90\) *Id.* at 1014–16. Dean Fryberg, an enrolled member of the Tulalip Indian Tribe, had been charged with the unlawful taking, shooting, and killing of a bald eagle in violation of the Eagle Protection Act. *Id.* Fryberg contested the charges, asserting that he had a general on-reservation right to hunt eagles under the Treaty of Point Elliot. *Id.*

\(^91\) *Id.* at 1016.

\(^92\) *Id.*

\(^93\) *Id.*

\(^94\) *Id.* at 1013–14. The court noted that this was especially true in the case at hand, which did not “involve the termination or diminishment of a reservation, . . . or the total extinguishment of hunting and fishing rights, as in *Menominee*, or even a substantial infringement on fishing and hunting rights.” *Id.* at 1014. Rather, the case involved a “relatively insignificant” restriction on Fryberg’s treaty hunting rights. *Id.* The court noted that there was no evidence that the bald eagle had ever “provided the Indian with any commercial benefit or had any subsistence value.” *Id.*

\(^95\) *Id.* at 1014.

\(^96\) *Id.* at 1015.
Applying this test to the Eagle Protection Act, the Fryberg court held that there was federal jurisdiction, the statute was non-discriminatory, and application of the Eagle Protection Act to treaty rights was necessary to effectuate the conservation purpose of the statute. The court found that the purpose of the Eagle Protection Act was to prevent the extinction of the bald eagle, and not “merely to conserve a resource.” So precarious was the situation of the eagle that “all threats, including takings pursuant to Indian treaty, should be banned to assure the species’ survival.”

Prior to the recent Anderson v. Evans decision, only one Ninth Circuit decision cited Fryberg for its use of the conservation necessity doctrine in the treaty abrogation context. In United States v. Eberhardt, decided prior to the U.S. Supreme Court’s adoption of the “actual consideration and choice” test in Dion, the Ninth Circuit noted in dicta the validity of the Fryberg analysis for interpreting the effect of congressional statutes on Indian treaty rights. However, the Eberhardt court overturned a district court decision that applied the Fryberg analysis.

97. The Eagle Protection Act applied “within the United States or any place subject to the jurisdiction thereof,” including Indian reservations. Id.
98. The “sweeping language” of the Eagle Protection Act banned all threats to the bald eagles’ survival from both treaty and non-treaty persons. Id.
99. Id. at 1013.
100. Id. at 1015.
101. Id.
102. Shortly after the Dion decision, a federal district court in Nevada noted that Fryberg precluded a treaty defense to a civil fine under the ESA. See United States v. Thirty Eight (38) Golden Eagles or Eagle Parts, 649 F. Supp. 269, 280–81 (D. Nev. 1986). However, the court’s mention of Fryberg was dicta as it found that the tribal member had no treaty right to take eagles, making the Fryberg analysis inapplicable. Id. at 281. Moreover, the district court was apparently unaware of the U.S. Supreme Court’s recent Dion decision as it noted only that the Court had granted certiorari in the case. Id. at 280. In addition, the Ninth Circuit has invoked the conservation necessity doctrine at least twice in treaty cases involving prosecutions under the Lacey Act. See United States v. Williams, 898 F.2d 727, 729 (9th Cir. 1990); United States v. Sohappy, 770 F.2d 816, 823 (9th Cir. 1985). However, in both cases, the court analyzed the validity of the underlying state wildlife conservation laws, on which prosecutions under the Lacey Act are based. In both cases, the court did not cite to Fryberg, and instead relied solely on the Puyallup line of cases. See Williams, 898 F.2d at 729; Sohappy, 770 F.2d at 823.
103. 789 F.2d 1354 (9th Cir. 1986).
104. The Eberhardt court consolidated review of two lower court decisions, both of which found the Puyallup line of cases inapplicable when there was no risk of “imminent extinction.” Id. at 1358.
105. Id. at 1358 (overturning United States v. Wilson, 611 F. Supp. 813 (N.D. Cal. 1985) (holding Department of Interior regulations invalid because the general trust statutes contained “no reflection of the congressional intent necessary to abrogate reserved rights as required by [Fryberg]”).

1113
analysis to administrative regulations affecting treaty fishing rights on
the Hoopa Valley Reservation in California. The Eberhardt court held
that Fryberg was inapplicable to regulations promulgated by the
Department of the Interior in its capacity as trustee for the tribes
occupying the reservation. Even though there was no evidence that
Congress considered treaty rights in authorizing the Department of the
Interior to promulgate fishing regulations, the court reasoned that the
regulations were designed to manage the reservation fisheries for the
“benefit of the Indians,” and not to abrogate or modify any treaty
rights.

B. Abrogation Analysis in Other Circuits

In general, other courts have not followed the Ninth Circuit in
incorporating the conservation necessity test into Indian treaty
abrogation jurisprudence. Prior to the U.S. Supreme Court’s decision in
Dion, the Circuits disagreed about whether courts must find
congressional intent to abrogate from the language of the statute, or
instead, like in Fryberg, could infer this intent from the “surrounding
circumstances” of the statute, such as the statute’s “conservation
purpose.” However, in the years since the Dion decision, no court
outside of the Ninth Circuit has cited Fryberg for its methodology.
Rather, these courts have generally relied on the Dion test and required
“clear and plain” evidence that Congress intended to abrogate treaty
rights.

While courts outside the Ninth Circuit have not specifically cited
Fryberg for its reasoning, one federal district court has used the
“conservation necessity” doctrine in analyzing an Indian treaty

106. Id. at 1361–62.
107. Id. at 1362.
108. Id. at 1361–62.
109. Compare, e.g., United States v. White, 508 F.2d 453, 456 (8th Cir. 1974) (holding that
Indian treaty abrogation must be “clearly expressed”) with United States v. Fryberg, 622 F.2d 1010,
1013 (9th Cir. 1980) (holding that treaty abrogation could be inferred from legislative history and
surrounding circumstances, including effectuation of a general conservation purpose).
110. See, e.g., United States v. Gotchnik, 222 F.3d 506, 509–11 (8th Cir. 2000) (acknowledging the
Dion standard, but holding that the Bois Forte Band of Chippewa Indians’ treaty right to hunt
and fish did not include right to use modern transportation methods in a designated wilderness area
while hunting and fishing); Oyler v. Allenbrand, 23 F.3d 292, 296 (10th Cir. 1994) (holding that
there was “clear evidence” under the Dion standard that the Kansas Enabling Act abrogated the
Shawnee’s individual criminal immunity, even though the Shawnee were not one of the four tribes
named in the Act).
Abrogation or Regulation?

Abrogation case. 111 In United States v. Billie,112 the court ruled that the Endangered Species Act abrogated the defendant’s treaty right to hunt panthers. 113 Citing the Payallup line of cases, the court held that Indian treaty rights “do not extend to the point of extinction,” and that “reasonable, nondiscriminatory” conservation measures may affect Indian treaty rights to the extent necessary to ensure the continued existence of a species. 114 Subsequent courts have not followed Billie, and commentators have criticized the case for ignoring the Dion test. 115

III. THE NINTH CIRCUIT REAFFIRMS THE FRYBERG STANDARD IN ANDERSON V. EVANS

In Anderson v. Evans, the Ninth Circuit again used the conservation necessity doctrine to analyze whether a federal statute had abrogated an Indian treaty. In holding that the Makah’s whaling efforts were subject to the Marine Mammal Protection Act,116 the court relied on the conservation necessity standard it adopted more than twenty years earlier in Fryberg.117 While it was unclear whether Fryberg was still valid following the U.S. Supreme Court’s decision in Dion,118 the Anderson court, without citing the Dion decision, reaffirmed and extended Fryberg’s methodology.119

113. Id. at 1489–90.
114. Id.
116. Anderson v. Evans, 314 F.3d 1006, 1029 (9th Cir. 2002).
117. Id. at 1026–29.
118. United States v. Dion, 476 U.S. 734, 739–40 (1986) (holding that it is “essential” that there be “clear evidence that Congress actually considered the conflict between its intended action on the one hand and Indian treaty rights on the other, and chose to resolve that conflict by abrogating the treaty”).
119. Anderson, 314 F.3d at 1006.
Under the Treaty of Neah Bay, the Makah Tribe has an explicit right to hunt whales.120 In fact, while all treaties in present-day Washington and Oregon were negotiated by either Governor Stevens (Washington) or Colonel Palmer (Oregon), the Treaty of Neah Bay is the only one of these treaties that reserves the right to whale for the Tribe.121 However, while they were once prolific whalers, the Makah stopped hunting gray whales entirely by the mid-1930s in the face of plummeting whale populations.122 In the years that followed, the federal government and international regulatory bodies implemented strict regulations on the taking of whales.123 These conservation efforts resulted in a dramatic comeback of the California gray whale and its removal from the endangered species list.124 Following the delisting, the Makah began working in conjunction with the National Oceanic and Atmospheric Administration (NOAA) to obtain a subsistence-whaling quota under the International Convention for the Regulation of Whaling, to which the United States is a signatory.125

In response to these efforts, the plaintiffs in Anderson, most of which were whale advocacy groups, sued the federal government over its decision to support the Makah’s resumed whaling efforts.126 The plaintiffs contended, inter alia,127 that the Makah were subject to the requirements of the federal MMPA.128 Congress passed the MMPA in 1972 in an effort to provide greater protection to marine mammals,

120. Treaty of Neah Bay, supra note 3, art. 4, 12 Stat. at 940a; see Metcalf v. Daley, 214 F.3d 1135, 1140 (9th Cir. 2000).
121. Metcalf, 214 F.3d at 1140.
122. See id. at 1137.
123. See id. While it is beyond the scope of this Note, and was not raised in the Anderson proceedings, it could be argued that, in addition to the MMPA, the International Convention for the Regulation of Whaling, Dec. 2, 1946, 62 Stat. 1716, 161 U.N.T.S. 72, and its domestic implementing statute, the Whaling Convention Act of 1949 (WCA), 16 U.S.C. §§ 916–916l (2000), either modified or abrogated the Makah’s treaty whaling right. However, as with the MMPA, there is no evidence, under the Dion standard, that Congress considered, and chose to abrogate, Indian treaty whaling rights by passing the WCA. See infra Part IV.C (addressing whether Congress intended to abrogate the Makah’s treaty whaling right by passing the MMPA).
125. See Metcalf, 214 F.3d at 1137.
126. Anderson v. Evans, 314 F.3d 1006, 1009 (9th Cir. 2002).
127. The plaintiffs also challenged the federal government’s approval of the Makah’s whaling plans under the National Environmental Policy Act of 1969 (NEPA). See id. The court held that the government violated NEPA by failing to prepare an Environmental Impact Statement before approving a whaling quota for the tribe. See id. at 1023.
128. Id. at 1009.
including the gray whale.\textsuperscript{129} While the MMPA imposes a general moratorium on the taking of marine mammals,\textsuperscript{130} it contains numerous exceptions to this moratorium, including an exemption for subsistence takings by certain Alaskan Natives.\textsuperscript{131} In addition, the MMPA allows the NOAA to issue permits for the taking of marine mammals in accordance with applicable MMPA regulations.\textsuperscript{132}

Neither the statute nor its legislative history mentioned Indian treaty rights until 1994, when Congress amended the MMPA.\textsuperscript{133} Section 14 of the 1994 Amendments provides that “[n]othing in this Act including any amendments to the Marine Mammal Protection Act of 1972 made by this Act alters or is intended to alter any treaty between the United States and one or more Indian Tribes.”\textsuperscript{134} While, on its face, the language of this provision only applies to the 1994 Amendments, the Senate Commerce Committee Report for the amendments notes that “the MMPA does not in any way diminish or abrogate existing protected Indian treaty fishing or hunting rights.”\textsuperscript{135}

The Makah argued that the MMPA was inapplicable to the Tribe because the statute had not abrogated the Makah’s explicit treaty right to hunt whales.\textsuperscript{136} Without addressing this defense, the court ruled that the MMPA must apply to the Makah in order to “effectuate the conservation purpose of the statute.”\textsuperscript{137} Thus, the court concluded that the NOAA had
violated federal law by issuing a whaling permit without complying with the MMPA.\textsuperscript{138}

In holding that the Makah were subject to the MMPA, the Anderson court applied Fryberg’s conservation necessity test rather than the U.S. Supreme Court’s treaty abrogation test established in Dion.\textsuperscript{139} Citing Fryberg, the court concluded that the conservation necessity test applied to federal conservation statutes, even though it had been developed in the state regulatory context.\textsuperscript{140} Unlike the Fryberg court, however, the Anderson court did not apply the conservation necessity test in the context of a larger abrogation analysis.\textsuperscript{141} Instead, it adopted the three-part Fryberg test to hold that the MMPA allows regulation of the Makah’s treaty whaling rights, without considering whether the MMPA abrogated those rights.\textsuperscript{142} Under the jurisdictional prong of the Fryberg test, the court reasoned that because the MMPA extends to “any person subject to the jurisdiction of the United States,”\textsuperscript{143} the statute would apply to the Makah’s whaling efforts off the coast of Washington State and in the Strait of Juan de Fuca.\textsuperscript{144} The statute also satisfied Fryberg’s non-discrimination prong because tribal members and non-treaty people were both subject to identical provisions.\textsuperscript{145} Finally, the court held that the application of the MMPA to the Tribe was necessary to achieve its conservation purpose.\textsuperscript{146} Reasoning that Congress was not merely concerned with the “survival” of marine mammals, but also with maintaining “optimum sustainable population[s],”\textsuperscript{147} the court concluded that the MMPA could not achieve such objectives without subjecting the Makah to its provisions.\textsuperscript{148}

\textsuperscript{138} Id.

\textsuperscript{139} Id. at 1026.

\textsuperscript{140} Id. at 1026 n.21.

\textsuperscript{141} Id. at 1026.

\textsuperscript{142} See id.


\textsuperscript{144} Anderson, 314 F.3d at 1026. The Strait of Juan de Fuca is the channel of water separating northwest Washington State from Vancouver Island, Canada.

\textsuperscript{145} Id.

\textsuperscript{146} Id. at 1026–28.

\textsuperscript{147} Id. at 1026–27.

\textsuperscript{148} Id. The court also rejected contentions by the government and Makah that even if the Fryberg test was the proper framework for analysis, the Fryberg court’s rationale was limited to cases where species preservation was an issue (it was undisputed that the California gray whale population was near “carrying capacity”). The court disagreed, and found that the Fryberg test mandated consideration of the conservation purpose of the statute as a whole, and not a general
In support of this conclusion, the court offered two scenarios where the Makah Tribe’s unrestricted exercise of its treaty right could jeopardize the gray whale population. First, the court noted that the Treaty of Neah Bay contained no explicit limitation on the number of gray whales the Tribe could take. While the Makah had shown “admirable restraint” in limiting its take to a small number of whales, and in seeking the approval of the United States, the court expressed concern that future generations of Makah could decide to take a significantly larger share of gray whales. Second, the court stated that the affirmation of the Makah’s right to whale free of MMPA restraints could trigger other tribes to claim whaling treaty rights under “less specific treaty language.”

In addition to its analysis of the Makah’s treaty right under the Fryberg framework, the court found that regulating the Tribe’s treaty right under the federal conservation necessity doctrine was consistent with “the principles embedded in the Treaty of Neah Bay itself.” The court noted that the Makah’s treaty right to hunt whales was “in common with all citizens of the United States.” According to the court, such language secures a right to take a “fair share” of the available resource for both the treaty and non-treaty parties. Although this logic suggests that the Makah have a right to take a fair share of the available whales, the court noted that the non-Indian citizens of Washington State were unable to exercise their treaty right to take a fair share of the whales because of MMPA restrictions. This apparent inequity rendered the right to take a “fair share” of the resource inapplicable to the Makah’s treaty right. Thus, the court concluded that the Makah’s fair share inquiry into issues of species preservation.  

149. Id. at 1027–28.
150. Id. at 1027.
151. Id.
152. Id.
153. Id. at 1027–28.
154. Id. at 1028.
155. Id.
156. Id. at 1028–29 n.24. In other words, the treaty guarantees Indians not merely the “equal opportunity” to take fish, but instead “secure the Indians’ right to take a share of each run of fish that passes through tribal fishing areas.” See Washington v. Wash. State Commercial Passenger Fishing Vessel Ass’n, 443 U.S. 658, 676–79 (1978).
158. Id.
under their treaty right to hunt whales in common with the citizens of the state was “no[t] unrestricted,” supporting its conclusion that the Tribe must comply with the MMPA’s permit process.159

IV. THE ANDERSON COURT RELIED ON FLAWED NINTH CIRCUIT TREATY ABROGATION JURISPRUDENCE IN HOLDING THAT MAKAH WHALING IS SUBJECT TO THE MMPA

In Anderson, the Ninth Circuit reaffirmed and extended its confused line of Indian treaty jurisprudence by holding that the MMPA applied to the Makah’s whaling efforts.160 This error was a result of the court’s previous conflation, in Fryberg, of two analytically distinct U.S. Supreme Court doctrines: the conservation necessity principle, properly applied only in cases involving state regulation of treaty rights; and treaty abrogation principles, properly applied only in cases involving federal statutes affecting treaty rights.161 The logic of Fryberg is unsound, and was implicitly rejected by the U.S. Supreme Court in Dion, where the Court reached the same result as in Fryberg, but without relying upon the state conservation necessity doctrine.162 In addition, the Anderson court compounded the error made in Fryberg by divorcing the conservation necessity test from the treaty abrogation framework entirely.163

When analyzed under the correct treaty abrogation analysis established in Dion, it becomes clear that the MMPA neither modifies nor abrogates the Makah’s treaty right to whale.164 There is no evidence that Congress considered Indian treaty rights and chose to abrogate those rights through the MMPA.165 Given the fundamental importance of Indian treaty rights, the application of the MMPA to the Makah requires a greater showing than a finding that unrestricted treaty whaling would run counter to the MMPA’s stated purpose.166 The Anderson decision’s logic ignored the U.S. Supreme Court’s repeated admonition that the

159. Id.
160. Id. at 1028.
161. See supra notes 89–101 and accompanying text.
163. See infra notes 188–201 and accompanying text.
164. See infra Part IV.C.
165. See infra notes 218–24 and accompanying text.
166. See infra Part IV.C.
Abrogation or Regulation?

Court will not easily find congressional intent to modify or abrogate Indian treaty rights.\(^{167}\) The Ninth Circuit’s decision is contrary to the treaty abrogation standard established in *Dion* and should be overturned.

A. The Anderson Court Improperly Relied on Fryberg, Which Applied Conservation Necessity Principles to Its Indian Treaty Abrogation Analysis Contrary to U.S. Supreme Court Precedent

The *Anderson* court should not have relied on *Fryberg* in holding that the MMPA applied to Makah whaling efforts. The *Fryberg* court created and relied upon an improper test for Indian treaty abrogation by importing and misapplying principles of the state conservation necessity doctrine. Such an extension of the conservation necessity doctrine serves only to confuse the traditional Indian treaty abrogation analysis\(^{168}\) and is not supported by U.S. Supreme Court precedent.\(^{169}\)

While the *Fryberg* court was properly concerned with determining whether Congress, in passing the Eagle Protection Act, intended to modify or abrogate Indian treaty hunting rights, it improperly applied the state conservation necessity analysis to determine whether such intent existed in a federal law.\(^{170}\) Prior to *Fryberg*, courts correctly applied the conservation necessity doctrine exclusively to determine whether *states* could regulate certain aspects of treaty rights when necessary for conservation purposes.\(^{171}\) Analysis of state regulation of treaty rights differs from that of congressional treaty abrogation because of the dramatically different powers possessed by each sovereign.\(^{172}\) States have only a limited ability to regulate Indian treaty rights when “necessary for conservation,” and no ability to abrogate or otherwise modify these rights.\(^{173}\) In stark contrast to this limited state power, the ability of the U.S. Congress to abrogate or modify Indian treaties is unquestioned.\(^{174}\) While the fundamental question when analyzing state regulation is whether the regulation is necessary for conservation,\(^{175}\) the

\(^{167}\) See infra Part IV.B.

\(^{168}\) See infra notes 177–79 and accompanying text.

\(^{169}\) See infra notes 183–87 and accompanying text.

\(^{170}\) See supra Part II.A.

\(^{171}\) See supra Part I.B.

\(^{172}\) See supra Part I.B.

\(^{173}\) See supra Part I.B.

\(^{174}\) See supra Part I.A.

\(^{175}\) See supra Part I.B.
only question in federal abrogation issues is whether Congress intended to exercise its power.176

When a court applies the Fryberg test and finds congressional intent to abrogate Indian treaty rights implicit in the “general purpose” of a federal statute, the court eviscerates the logic of the abrogation doctrine. The general purpose of a federal statute, conservation or otherwise, does not address the fundamental abrogation question—whether Congress intended the statute to apply to Indian treaty rights. The Fryberg approach contradicts the U.S. Supreme Court’s admonition that congressional intent to modify or abrogate an Indian treaty “is not to be lightly imputed.”177

In citing to the line of cases establishing the conservation necessity doctrine, the Fryberg court focused on the dire conservation threat posed by unregulated treaty hunting.178 Quoting Puyallup II, the court noted:

Rights can be controlled by the need to conserve a species . . . .
The police power of the State is adequate to prevent the steelhead from following the fate of the passenger pigeon; and the Treaty does not give the Indians a federal right to pursue the last living steelhead until it enters their nets.179

However, the U.S. Supreme Court’s logic in Puyallup II, developed exclusively in the context of state regulation, is wholly inapplicable to the abrogation context.180 The question in abrogation cases, unlike state regulation cases, is not whether Congress has the power to affect the treaty rights—if Congress so chooses, it could completely abrogate all Indian treaty fishing rights.181 Rather, the question is whether Congress has chosen to exercise that power. The U.S. Supreme Court has repeatedly stated that it will only find such intent when there is clear evidence that Congress considered the treaty right in question and then chose to abrogate that right.182 The Fryberg court failed to recognize this fundamental distinction.

Most significantly, the U.S. Supreme Court’s Dion decision illustrates the fallacy of the Fryberg court’s analysis.183 The Dion decision, like the

---

176. See supra Part I.A.
178. See United States v. Fryberg, 622 F.2d 1010, 1015 (9th Cir. 1980).
179. Id. (quoting Puyallup II, 414 U.S. 44, 49 (1973)).
180. See supra Part I.A.
181. See supra Part I.A.
183. See supra Part I.A.
Abrogation or Regulation?

Fryberg decision, addressed whether Congress intended the Eagle Protection Act to modify Indian treaty eagle hunting rights. 184 In reaching its conclusion that Congress intended to modify these rights, the Court found that congressional intent to abrogate was both “strongly suggested” by the language of the Eagle Protection Act and evident from the legislative history of the statute. 185 No part of the Dion opinion discusses the Eagle Protection Act’s general “conservation purpose.” Nor does the opinion import the conservation necessity doctrine into its abrogation analysis. 186 On the contrary, the U.S. Supreme Court reaffirmed the importance of Indian treaty rights, holding that treaty abrogation requires “clear evidence that Congress actually considered the conflict between its intended action on the one hand and Indian treaty rights on the other, and chose to resolve that conflict by abrogating the treaty.” 187 Courts can find such evidence in the statute’s language or legislative history; they cannot find it in a generic examination of the statute’s “conservation purpose.”

B. The Anderson Court Ignored the Abrogation Issue, Compounding the Fryberg Court’s Mistaken Analysis, and Offered Little Justification for a Finding of Conservation Necessity

The Anderson court not only relied on the Fryberg court’s flawed methodology, it also compounded the Fryberg court’s error by ignoring the U.S. Supreme Court’s treaty abrogation requirements. 188 In formulating its novel approach to the analysis of the impact of a federal conservation statute on Indian treaty rights, the Anderson court cited the three-part Fryberg test to determine when “reasonable conservation statutes” affect Indian treaty rights. 189 While the Fryberg court misapplied the conservation necessity test, it did so in an attempt to determine whether Congress intended to modify or abrogate Indian

184. See Dion, 476 U.S. at 745.
185. Id. at 740.
186. See supra Part I.A. It is noteworthy that the court did not even cite to Fryberg, in spite of the fact that the case was featured prominently in the underlying Eighth Circuit’s decision, as well as in the briefs submitted to the U.S. Supreme Court. See United States v. Dion, 752 F.2d 1261 1266–70 (8th Cir. 1985); Brief for the United States at 24, 29, United States v. Dion, 476 U.S. 734 (1986) (No. 85-246).
187. Dion, 476 U.S. at 740.
188. See supra Part II.B.
189. Anderson v. Evans, 314 F.3d 1006, 1026 (9th Cir. 2002).
treaty rights.190 The *Anderson* court, by contrast, applied the test to determine whether the MMPA *regulates* Makah whaling.191 Without explanation, the *Anderson* court held that it need not consider whether “the MMPA applies by virtue of treaty abrogation.”192 The practical result of the court’s decision is that the Makah Tribe has been stripped of a treaty right without the court having found congressional intent to do so. Such an analysis is not only inconsistent with established U.S. Supreme Court Indian treaty abrogation precedent, but also with the methodology of *Fryberg*.193

The *Anderson* court’s contention that the conservation necessity test is not limited to cases involving state regulations194 is also erroneous. The two cases the court cited in support of this proposition, *Fryberg* and *Eberhardt*,195 are simply inapposite. As noted above, the *Fryberg* court improperly used the conservation necessity test within the framework of its abrogation analysis.196 Even under its own terms, however, the *Fryberg* decision merely stands for the premise that a court may look to the conservation necessity doctrine in determining whether Congress intended to modify or abrogate Indian treaty rights.197 It does not stand for the proposition that a court may invoke the conservation necessity principle apart from the abrogation context.

The *Anderson* court’s reliance on *Eberhardt* for the proposition that conservation necessity principles are not limited to cases involving state regulation is equally inappropriate. Although the *Eberhardt* court approved of the *Fryberg* methodology in dicta,198 it ultimately determined the *Fryberg* test to be inapplicable to administrative regulations promulgated by the Department of the Interior in its trust capacity.199 While it is doubtful that the court’s distinction between

---

190. See United States v. *Fryberg*, 622 F.2d 1010, 1013–16 (9th Cir. 1980).
191. See *Anderson*, 314 F.3d at 1026–30.
192. *Id.* at 1029–30.
193. See *Fryberg*, 622 F.2d at 1013–16 (analyzing conservation necessity test in context of treaty abrogation analysis).
194. *Anderson*, 314 F.3d at 1026 n.21.
195. *Id.* at 1026–28.
196. See supra Part IV.A.
197. See supra notes 92–101 and accompanying text.
199. See *id*. The *Eberhardt* court specifically found that “the district court erred in analogizing this case to the ban on taking bald eagles in *Fryberg*.” *Id.* at 1361.
Abrogation or Regulation?

administrative regulations and congressional statutes is valid,\textsuperscript{200} \textit{Eberhardt} nonetheless was decided before \textit{Dion}, and its dicta cannot be read as controlling in light of \textit{Dion}’s methodology.\textsuperscript{201} To justify its holding that the Tribe was subject to the MMPA, the \textit{Anderson} court emphasized two possible consequences of unrestricted Makah whaling. The court first noted that, if left unrestricted, the Makah would be free to expand whaling activities, perhaps to the point of “jeopardizing” the gray whale population.\textsuperscript{202} The court ignored, however, the great likelihood that, were the Makah to hunt to the point of endangering the whale population, Congress would exercise its plenary authority to modify or abrogate the Makah’s treaty right. As noted, Congress has plenary power to abrogate treaty rights; the only question is whether Congress has chosen to do so.\textsuperscript{203} Furthermore, the empirical basis for this speculation is highly questionable. In five years of hunting, the Makah have taken one whale.\textsuperscript{204} The assertion that the Makah would (or could) take enough whales to endanger a species that is presently “at or near [the ocean’s] carrying capacity”\textsuperscript{205} is unsupported.\textsuperscript{206}

The \textit{Anderson} court further asserted that less specific treaty language than that found in the Treaty of Neah Bay could support the unrestricted whaling rights by other tribes.\textsuperscript{207} While the treaties of other Pacific Coast tribes do not contain specific treaty language reserving the right to hunt

\begin{footnotesize}
\begin{itemize}
  \item[200.] See, e.g., Timpanogos Tribe v. Conway, 286 F.3d 1195, 1203 (10th Cir. 2002) (“The Department of the Interior cannot under any circumstances abrogate an Indian treaty directly or indirectly. Only Congress can abrogate a treaty, and only by making absolutely clear its intention to do so.”) (citing United States v. Washington, 641 F.2d 1368, 1371 (9th Cir. 1981), aff’d, 520 F.2d 676 (9th Cir. 1975)); see also United States v. Dion, 476 U.S. 734, 739–40 (1986); Wilson v. Omaha Indian Tribe, 442 U.S. 653, 670 (1979).
  \item[201.] See \textit{Dion}, 476 U.S. at 739–40.
  \item[202.] \textit{Anderson} v. Evans, 314 F.3d 1006, 1027 (9th Cir. 2002).
  \item[203.] See supra Part I.A.
  \item[204.] \textit{Anderson}, 314 F.3d at 1014.
  \item[205.] Id. at 1010.
  \item[206.] While it is beyond the scope of this Note, it could be argued that even if the conservation necessity doctrine had been applied to the Makah’s treaty right, the success of the tribal government in self-regulation would preclude regulation by the federal government. See, e.g., United States v. Washington, 384 F. Supp. 312, 415 (W.D. Wash. 1974) (finding that to regulate for purposes of conservation necessity the state must show, among other things, that “existing tribal regulation or enforcement is inadequate to prevent demonstrable harm to the actual conservation of fish”), aff’d, 520 F.2d 676 (9th Cir. 1975).
  \item[207.] \textit{Anderson}, 314 F.3d at 1027–28.
\end{itemize}
\end{footnotesize}
whales, many have reserved “hunting and fishing” rights. The court implied that such “hunting and fishing” rights could, if broadly construed, include the right to take whales. The court reasoned that granting other Pacific Coast tribes unrestricted whaling rights could present a serious threat to the survival of the gray whale.

Such a scenario is unlikely for several reasons. First, as noted above, Congress is free to modify or abrogate any Indian treaty right, and would likely do so should widespread tribal hunting—by the Makah or any other tribe—endanger the existence of the gray whale. Second, two government representatives negotiated virtually all of the treaties with Pacific Coast tribes: Governor Stevens in the Washington Territory, and Colonel Palmer in the Oregon Territory. While many of these treaties reserved the non-exclusive right to “hunt and fish” at usual and accustomed places, only one—the Treaty of Neah Bay—reserved the explicit right to whale. Arguably, if the general right to hunt and fish included the right to whale, there would have been no need for the inclusion of such language in the treaty. While treaty terms are to be construed “in the sense in which they would naturally be understood” by the tribes, the uniqueness of the whaling language provides a strong argument that none of the other treaties should be interpreted to include the reserved right to whale. The concern of the court seems unfounded.

The Anderson court provided no additional justification for the use of the conservation necessity test. That test is not the appropriate method for analyzing the fundamental Indian treaty abrogation issue presented to the court. Not only did the court fail to discuss the issue within the Dion framework, it improperly applied the Fryberg test by completely ignoring the larger abrogation issue. Given the importance of Indian treaty rights, such analysis is insufficient.

208. Id.
209. See id.
210. Id.
211. See supra notes 24–26 and accompanying text.
213. Treaty of Neah Bay, supra note 3, art. 4, 12 Stat. at 940a.
214. See Fishing Vessel, 443 U.S. at 675–76 (citing Jones v. Meehan, 175 U.S. 1, 11 (1898)).
215. This flows from the legal maxim “expressio unius est exclusio alterius” which means that “to express or include one thing implies the exclusion of the other.” BLACK’S LAW DICTIONARY 602 (7th ed. 1999).
Abrogation or Regulation?

C. Under Well-Established U.S. Supreme Court Abrogation Analysis, the Makah Tribe Is Not Subject to the MMPA

Because the Anderson court’s finding that the Makah Tribe was subject to the MMPA in effect abrogates the Tribe’s treaty right to whale, the court should have determined whether, in passing the MMPA, Congress intended to abrogate the Makah’s treaty whaling rights. The U.S. Supreme Court’s Dion test requires a court to evaluate whether “clear evidence” exists that Congress considered the conflict between its intended action on the one hand and Indian treaty rights on the other, and chose to resolve that conflict by abrogating the treaty.216 Such evidence must be “clear and plain,” and found either on the face of the statute or in the statute’s legislative history.217 Under the Dion test, the Ninth Circuit should have concluded that there was no clear evidence that Congress intended the MMPA to apply to the Makah.

There is no indication from either the language of the MMPA218 or the statute’s legislative history219 that Congress considered and chose to abrogate the Makah’s whaling treaty right.220 In fact, until the 1994 Amendments, the MMPA made no mention of any Indian treaty right, including the Makah’s whaling right under the Treaty of Neah Bay.221 Far from abrogating Indian treaty rights, the 1994 Amendments expressly preserved them.222 Section 14 of the 1994 Amendments provides that “nothing in this Act including any amendments to the Marine Mammal Protection Act of 1972 made by this Act alters or is intended to alter any treaty between the United States and one or more Indian Tribes.”223 Congress’ stated intent in enacting this disclaimer was to “reaffirm that the MMPA does not in any way diminish or abrogate protected Indian treaty fishing or hunting rights.”224

217. Id.
219. Dion, 476 U.S. at 739.
220. Treaty of Neah Bay, supra note 3, art. 4, 12 Stat. at 940a.
223. Id.
Arguably, the MMPA’s exemption for Alaska Natives might indicate that Congress intended the MMPA to apply to all Indian groups except for Alaska Natives, evincing congressional intent to abrogate the Makah’s treaty whaling right. Such a reading, however, is tenuous. Alaska Natives possess no treaty rights, so it is logical for the MMPA to contain an explicit exemption for Alaska Natives. Any statute of general applicability passed by Congress would naturally apply to Alaska Natives were there no exemption. Moreover, such a reading of the statute would effect an abrogation of the Makah’s treaty by negative implication. This falls far short of the “clear evidence” standard of Dion.

In sum, there is no evidence in the language and legislative history of the MMPA that suggests that Congress was even aware of the Makah’s whaling treaty right, much less that it considered the right and intended the MMPA to abrogate it. In fact, there is a strong indication that Congress did not wish to abrogate any Indian treaty rights by enacting the MMPA. In the absence of a finding of intention to abrogate, the Dion decision requires that the treaty right be preserved. As the U.S. Supreme Court has noted on many occasions, such treaty rights are too fundamental for the courts to so easily discard.

V. CONCLUSION

Contrary to the Ninth Circuit’s decision in Anderson, the Makah’s right to take whales under the Treaty of Neah Bay is unfettered by the MMPA. The Anderson court misapplied its own faulty precedent in


227. This seems particularly true given that just one year prior to passing the MMPA, Congress had extinguished all claims based on “aboriginal right, title, use, or occupancy of land or water areas in Alaska” by passing the Alaska Native Claims Settlement Act. Pub. L. No. 92-203, 85 Stat. 688 (codified at 43 U.S.C. §§ 1601–1628). At least, it indicates that Congress was well aware of the non-treaty status of Alaska natives. See also Bresette, 761 F. Supp. at 663 (holding that a similar exception for Alaska natives in the Migratory Bird Treaty Act did not abrogate Indian treaty rights and stating that “[t]o treat the consideration of indigenous Alaskans’ rights as the consideration of Native American treaty rights nationwide, for the simple reason that both groups are regarded as Indians, is disingenuous”).


229. See, e.g., id. at 739 (“Indian treaty rights are too fundamental to be easily cast aside.”).
Abrogation or Regulation?

Fryberg in erroneously concluding that Congress intended to regulate Makah whaling through the MMPA. The Fryberg court, implicitly overruled by the U.S. Supreme Court in Dion, mistakenly incorporated the conservation necessity doctrine into treaty abrogation analysis. The conservation necessity doctrine properly applies only to state regulation of Indian treaty fishing. It is not relevant to the question of whether federal statutes affect Indian fishing rights. Furthermore, the Anderson court compounded the Fryberg court’s error by failing even to consider the abrogation issue. Instead, the Anderson court relied solely on the conservation necessity doctrine in holding that the MMPA must apply to the Makah’s whaling efforts.

The correct treaty abrogation analysis is found in Dion: There must be clear and plain evidence that Congress considered the treaty right in question, and subsequently chose to modify or abrogate that right. There is no evidence that Congress, in passing the MMPA, considered and then chose to abrogate the Makah’s treaty right to whale. Thus, under Dion, the Makah’s treaty right remains unaffected by the MMPA. If Congress decides that the MMPA should regulate Makah whaling, it is free to do so through statutory enactment. Until Congress passes such legislation, however, the United States is bound by its treaty obligation to the Makah. The Anderson decision should be overturned.
AUTHOR INDEX

Allen, Timothy E., M.A.
The Foreseeability of Transference: Extending Employer Liability Under Washington Law for Therapist Sexual Exploitation of Patients 78:525

Ambach, Kirsten Lela
Miranda’s Poisoned Fruit Tree: The Admissibility of Physical Evidence Derived from an Unwarned Statement 78:757

Anderson, Nick
Dr. Jekyll’s Waiver of Mr. Hyde’s Right To Refuse Medical Treatment: Washington’s New Law Authorizing Mental Health Care Advance Directives Needs Additional Protections 78:795

Andrews, Rebecca L.
So the Army Hired an Ax-Murderer: The Assault and Battery Exception to the Federal Tort Claims Act Does Not Bar Suits for Negligent Hiring, Retention and Supervision 78:161

Attestatova, Svetlana G.
The Bonds of Joint Tax Liability Should Not Be Stronger than Marriage: Congressional Intent Behind § 6015(c) Separation of Liability Relief 78:831

Black, Natasha Shekdar, M.S.

Burke, Alafair S.
Unpacking New Policing: Confessions of a Former Neighborhood District Attorney 78:985

Colendich, Katie M.
Who Owns “The Law”? The Effect on Copyrights when Privately- Authorized Works Are Adopted or Enacted by Reference into Law 78:589

Dillon, Lainie M.C.
Conundrums with Penumbras: The Right to Privacy Encompasses Non-Gamete Providers Who Create Prenumbras with the Intent To Become Parents 78:625

Egle, Andre V.
Back to Prima Paint Corp. v. Flood & Conklin Manufacturing Co.: To Challenge an Arbitration Agreement You Must Challenge the Arbitration Agreement 78:199

Farrell, Robert C.
Classes, Persons, Equal Protection, and Village of Willowbrook v. Olech 78:367

Firestone, Jeremy Barber, Robert
Fish as Pollutants: Limitations of and Crosscurrents in Law, Science, Management, and Policy 78:693

Green, Derek D.
Does Free Exercise Mean Free State Funding? In Davey v. Locke, the Ninth Circuit Undervalued Washington’s Vision of Religious Liberty 78:653

Harrison, Rebecca E.
When Animals Invade and Occupy: Physical Takings and the Endangered Species Act 78:867

Jansson, Roger L.
Researcher Liability for Negligence in Human Subject Research: Informed Consent and Researcher Malpractice Actions 78:229

Krebs, Jennifer Amanda

Main, Thomas O.
Traditional Equity and Contemporary Procedure 78:429
Matheny, Stephanie D.
Who Can Defend a Federal Regulation?
The Ninth Circuit Misapplies Rule 24 by
Denying Intervention of Right in Kootenai
Tribe of Idaho v. Veneman 78:1067

Nehf, James P.
Recognizing the Societal Value in Information
Privacy 78:1

Reynolds, Laurie
Intergovernmental Cooperation, Metropolitan
Equality, and the New Regionalism 78:93

Richards, Andrew P.
Aboriginal Title or the Paramountcy Doctrine?
Johnson v. McIntosh Flounders in Federal
Waters off Alaska in Native Village of Eyak v.
Trawler Diane Marie, Inc. 78:939

Robertson, Melissa
Is Assent Still a Prerequisite for Contract
Formation in Today’s E-conomy? 78:265

Simmons, Jill E.
Beggars Can’t Be Voters: Why Washington’s
Felon Re-enfranchisement Law Violates the
Equal Protection Clause 78:297

Snoey, Janis
Water, Property, and the Clean Water Act
78:335

Tomlinson, Zachary
Abrogation or Regulation? How Anderson v.
Evans Discards the Makah’s Treaty Whaling
Right in the Name of Conservation Necessity
78:1101

Wolcher, Louis E.
What Is the Rule of Law? Perspectives from
Central Europe and the American Academy
78:515
## Title Index

**TITLE INDEX**

Black, Natasha Shekdar, M.S.  
78:557 |
| --- | --- |
| A | Aboriginal Title or the Paramountcy Doctrine? *Johnson v. McIntosh*  
Flounders in Federal Waters off Alaska in *Native Village of Eyak v. Trawler Diane Marie, Inc.*  
Richards, Andrew P. 78:939 |
| A | Abrogation or Regulation? How *Anderson v. Evans* Discards the Makah’s Treaty Whaling Right in the Name of Conservation Necessity  
Tomlinson, Zachary 78:1101 |
| A | Back to *Prima Paint Corp. v. Flood & Conklin Manufacturing Co.*: To Challenge an Arbitration Agreement You Must Challenge the Arbitration Agreement  
Egle, Andre V. 78:199 |
| A | Beggars Can’t Be Voters: Why Washington’s Felon Re-enfranchisement Law Violates the Equal Protection Clause  
Simmons, Jill E. 78:297 |
| A | Classes, Persons, Equal Protection, and *Village of Willowbrook v. Olech*  
Farrell, Robert C. 78:367 |
| A | Conundrums with Penumbras: The Right to Privacy Encompasses Non-Gamete Providers Who Create Prenembros with the Intent To Become Parents  
Dillon, Lainie M.C. 78:625 |
| A | Does Free Exercise Mean Free State Funding? In *Davey v. Locke*, the Ninth Circuit Undervalued Washington’s Vision of Religious Liberty  
Green, Derek D. 78:653 |
| A | Dr. Jekyll’s Waiver of Mr. Hyde’s Right to Refuse Medical Treatment: Washington’s New Law Authorizing Mental Health Care Advance Directives Needs Additional Protections  
Anderson, Nick 78:795 |
Krebs, Jennifer Amanda 78:901 |
| A | Fish as Pollutants: Limitations of and Crosscurrents in Law, Science, Management, and Policy  
Firestone, Jeremy & Barber, Robert 78:693 |
| A | Intergovernmental Cooperation, Metropolitan Equality, and the New Regionalism  
Reynolds, Laurie 78:93 |
| A | Is Assent Still a Prerequisite for Contract Formation in Today’s Economy?  
Robertson, Melissa 78:265 |
| A | *Miranda’s Poisoned Fruit Tree*: The Admissibility of Physical Evidence Derived from an Unwarned Statement  
Ambach, Kristen Lela 78:757 |
| A | Recognizing the Societal Value in Information Privacy  
Nehf, James P. 78:1 |
| A | Researcher Liability for Negligence in Human Subject Research: Informed Consent and Researcher Malpractice Actions  
Jansson, Roger L. 78:229 |
| A | So the Army Hired an Ax-Murderer: The Assault and Battery Exception to the Federal Tort Claims Act Does |
Not Bar Suits for Negligent Hiring, Retention and Supervision
Andrews, Rebecca L. 78:161

The Bonds of Joint Tax Liability Should Not Be Stronger than Marriage: Congressional Intent Behind § 6015(c) Separation of Liability Relief
Attestatova, Svetlana G. 78:831

The Foreseeability of Transference: Extending Employer Liability Under Washington Law for Therapist Sexual Exploitation of Patients
Allen, Timothy E., M.A. 78:525

Traditional Equity and Contemporary Procedure
Main, Thomas O. 78:429

Unpacking New Policing: Confessions of a Former Neighborhood District Attorney
Burke, Alafair S. 78:985

Water, Property, and the Clean Water Act
Snoey, Janis 78:335

What Is the Rule of Law? Perspectives from Central Europe and the American Academy
Wolcher, Louis E. 78:515

When Animals Invade and Occupy: Physical Takings and the Endangered Species Act
Harrison, Rebecca E. 78:867

Matheny, Stephanie D. 78:1067

Who Owns “The Law”? The Effect on Copyrights when Privately-Authorized Works Are Adopted or Enacted by Reference into Law
Colendich, Katie M. 78:589
SUBJECT INDEX

Advance Directives
Dr. Jekyll’s Waiver of Mr. Hyde’s Right to Refuse Medical Treatment: Washington’s New Law Authorizing Mental Health Care Advance Directives Needs Additional Protections 78:795

Arbitration
Back to Prima Paint Corp. v. Flood & Conklin Manufacturing Co.: To Challenge an Arbitration Agreement You Must Challenge the Arbitration Agreement 78:199

Constitutional Law
Conundrums with Penumbras: The Right to Privacy Encompasses Non-Gamete Providers Who Create Prenecyos with the Intent To Become Parents 78:625

Does Free Exercise Mean Free State Funding? In Davey v. Locke, the Ninth Circuit Undervalued Washington’s Vision of Religious Liberty 78:653

Contracts
Is Assent Still a Prerequisite for Contract Formation in Today’s E-conomy? 78:265

Copyrights
Who Owns “The Law”? The Effect on Copyrights when Privately-Authored Works Are Adopted or Enacted by Reference into Law 78:589

Criminal Procedure

Miranda’s Poisoned Fruit Tree: The Admissibility of Physical Evidence Derived from an Unwarned Statement 78:757

Unpacking New Policing: Confessions of a Former Neighborhood District Attorney 78:985

Employment Law
So the Army Hired an Ax-Murderer: The Assault and Battery Exception to the Federal Tort Claims Act Does Not Bar Suits for Negligent Hiring, Retention and Supervision 78:161

The Foreseeability of Transference: Extending Employer Liability Under Washington Law for Therapist Sexual Exploitation of Patients 78:525

Environmental Law
Fish as Pollutants: Limitations of and Crosscurrents in Law, Science, Management, and Policy 78:693

When Animals Invade and Occupy: Physical Takings and the Endangered Species Act 78:867

Equal Protection Clause
Beggars Can’t Be Voters: Why Washington’s Felon Re-Enfranchisement Law Violates the Equal Protection Clause 78:297

Classes, Persons, Equal Protection, and Village of Willowbrook v. Olech 78:367

Equity
Traditional Equity and Contemporary Procedure 78:429

Federal Rules of Civil Procedure
Traditional Equity and Contemporary Procedure 78:429


Indian Law
Aboriginal Title or the Paramountcy Doctrine? Johnson v. McIntosh Flounders in Federal Waters off Alaska in Native Village of Eyak v. Trawler Diane Marie, Inc. 78:939

Abrogation or Regulation? How Anderson v. Evans Discards the Makah’s Treaty Whaling Right in the Name of Conservation Necessity 78:1101

Intergovernmental Cooperation
Intergovernmental Cooperation, Metropolitan Equality, and the New Regionalism 78:93

International Law
What Is the Rule of Law? Perspectives from Central Europe and the American Academy 78:515

Medicine
Conundrums with Penumbras: The Right to Privacy Encompasses Non-Gamete Providers Who Create Preembryos with the Intent To Become Parents 78:625

Dr. Jekyll’s Waiver of Mr. Hyde’s Right To Refuse Medical Treatment: Washington’s New Law Authorizing Mental Health Care Advance Directives Needs Additional Protections 78:795

Researcher Liability for Negligence in Human Subject Research: Informed Consent and Researcher Malpractice Actions 78:229

Privacy
Recognizing the Societal Value in Information Privacy 78:1

Conundrums with Penumbras: The Right to Privacy Encompasses Non-Gamete Providers Who Create Preembryos with the Intent To Become Parents 78:625

Tax
The Bonds of Joint Tax Liability Should Not Be Stronger than Marriage: Congressional Intent Behind § 6015(c) Separation of Liability Relief 78:831

Telecommunications

Torts
Researcher Liability for Negligence in Human Subject Research: Informed Consent and Researcher Malpractice Actions 78:229
Subject Index

So the Army Hired an Ax-Murderer: The Assault and Battery Exception to the Federal Tort Claims Act Does Not Bar Suits for Negligent Hiring, Retention and Supervision 78:161

Water Law
Fish as Pollutants: Limitations of and Crosscurrents in Law, Science, Management, and Policy 78:693

Water, Property, and the Clean Water Act 78:335

Washington Law

Beggars Can’t Be Voters: Why Washington’s Felon Re-Enfranchisement Law Violates the Equal Protection Clause 78:297

Does Free Exercise Mean Free State Funding? In Davey v. Locke, the Ninth Circuit Undervalued Washington’s Vision of Religious Liberty 78:653

Dr. Jekyll’s Waiver of Mr. Hyde’s Right To Refuse Medical Treatment: Washington’s New Law Authorizing Mental Health Care Advance Directives Needs Additional Protections 78:795

The Foreseeability of Transference: Extending Employer Liability Under Washington Law for Therapist Sexual Exploitation of Patients 78:525