POLICING BY NUMBERS: BIG DATA AND THE FOURTH AMENDMENT

Elizabeth E. Joh*

INTRODUCTION

The age of "big data" has come to policing. In Chicago, police officers are paying particular attention to members of a "heat list": those identified by a risk analysis as most likely to be involved in future violence.[1] In Charlotte, North Carolina, the police have compiled foreclosure data to generate a map of high-risk areas that are likely to be hit by crime.[2] In New York City, the N.Y.P.D. has partnered with Microsoft to employ a "Domain Awareness System" that collects and links information from sources like CCTVs, license plate readers, radiation sensors, and informational databases.[3] In Santa Cruz, California, the police have reported a dramatic reduction in burglaries after relying upon computer algorithms that predict where new burglaries are likely to occur.[4] The Department of Homeland Security has applied computer analytics to Twitter feeds to find words like "pipe bomb," "plume," and "listeria."[5]

---

* Professor of Law, U.C. Davis School of Law (eejoh@ucdavis.edu). Thanks to David Ball, Jack Chin, David Horton, Wayne Logan, Erin Murphy, and Charles Reichmann for comments and suggestions, to the librarians of the Mabie Law Library for research assistance, to the staff of the *Washington Law Review* for the invitation to contribute to the Examining Artificial Intelligence symposium and for their editorial work, and to the U.C. Davis School of Law for institutional support.

1. Jeremy Gorner, *Chicago Police Use 'Heat List' as Strategy to Prevent Violence*, CHI. TRIB. (Aug. 21, 2013), http://articles.chicagotribune.com/2013-08-21/news/ct-met-heat-list-20130821_1_chicago-police-commander-andrew-papachristos-heat-list.

2. Michael Bess, *Assessing the Impact of Home Foreclosures in Charlotte Neighborhoods*, GEOGRAPHY & PUB. SAFETY, Oct. 2008, at 2, 2.

3. Joe Coscarelli, *The NYPD's Domain Awareness System is Watching You*, N.Y. MAG. (Aug. 9, 2012, 5:50 AM), http://nymag.com/daily/intelligencer/2012/08/nypd-domain-awareness-system-microsoft-is-watching-you.html.

4. *See* Erica Goode, *Sending the Police Before There's a Crime*, N.Y. TIMES (Aug. 15, 2011), http://www.nytimes.com/2011/08/16/us/16police.html.

5. Somini Sengupta, *In Hot Pursuit of Numbers to Ward Off Crime*, N.Y. TIMES (June 19, 2013, 10:48 PM), http://bits.blogs.nytimes.com/2013/06/19/in-hot-pursuit-of-numbers-to-ward-off-

Big data has begun to transform government in fields as diverse as public health, transportation management, and scientific research.[6] The analysis of what were once unimaginable quantities of digitized data is likely to introduce dramatic changes to a profession which, as late as 1900, involved little more than an able-bodied man who was given a hickory club, a whistle, and a key to a call box.[7] Real-time access to and analysis of vast quantities of information found in criminal records, police databases, and surveillance data may alter policing[8] in the same way that big data has revolutionized areas as diverse as presidential elections,[9] internet commerce,[10] and language translation.[11] Some have even heralded big data's potential to change our assumptions about social relationships, government, scientific study, and even knowledge itself.[12]

In the private sector, retailers have harnessed big data to produce some seemingly trivial but surprising changes to their practices.[13] A much discussed example stems from Target's extensive use of data analytics to identify certain purchases, such as supplements commonly taken during pregnancy, to know whether a customer is pregnant,

---

crime/?_r=0.

6. *See, e.g.*, TECHAMERICA FOUND., DEMYSTIFYING BIG DATA: A PRACTICAL GUIDE TO TRANSFORMING THE BUSINESS OF GOVERNMENT 12–15 (2012) (describing potential uses of big data in healthcare, transportation, education, fraud detection, cyber security, and weather).

7. *See* Mark H. Haller, *Historical Roots of Police Behavior: Chicago, 1890–1925*, 10 L. & SOC'Y REV. 303, 303 (1976).

8. Beth Pearsall, *Predictive Policing: The Future of Law Enforcement?*, NAT'L INST. JUST. J., June 2010, at 16, 16, *available at* https://www.ncjrs.gov/pdffiles1/nij/230414.pdf (describing its development as having "the potential to transform law enforcement").

9. *See* Michael Scherer, *Inside the Secret World of the Data Crunchers Who Helped Obama Win*, TIME (Nov. 7, 2012), http://swampland.time.com/2012/11/07/inside-the-secret-world-of-quants-and-data-crunchers-who-helped-obama-win/print/ (quoting one Obama campaign official as saying, "We ran the election 66,000 times every night" in computer simulations).

10. *See, e.g.*, VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK 4–5 (2013) (describing development of Farecast, which analyzes data from billions of flight price records to predict airline ticket price variation).

11. *See id.* at 37–39 (describing language translation success of Google using trillion word data set).

12. *See, e.g.*, Adam Frank, *Big Data Is the Steam Engine of Our Time*, NPR (Mar. 12, 2013, 12:28 PM), http://www.npr.org/blogs/13.7/2013/03/12/174028759/big-data-is-the-steam-engine-of-our-time ("Big Data may be the steam engine of our time.").

13. Charles Duhigg, *How Companies Learn Your Secrets*, N.Y. TIMES (Feb. 16, 2012), http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=1 ("Almost every major retailer, from grocery chains to investment banks to the U.S. Postal Service, has a 'predictive analytics' department . . . .").

without the woman disclosing the pregnancy herself.[14] For a retailer, pregnancy is a prime opportunity to target a consumer when shopping habits change and expand. An irate father allegedly complained to Target that his daughter was unfairly targeted as a pregnant woman with coupons only to discover, to his chagrin, that Target was better informed than he was.[15] Similarly, Walmart, through its computerized retail tracking, has discovered that Strawberry Pop-Tarts and beer sell as briskly as flashlights when hurricanes are forecast. These products were quickly shipped to Florida Walmart stores in the path of Hurricane Frances in 2004.[16]

Yet unlike the data crunching performed by Target, Walmart, or Amazon, the introduction of big data to police work raises new and significant challenges to the regulatory framework that governs conventional policing. From one perspective, the Fourth Amendment has proven remarkably flexible over time. Constitutional law has governed ordinary policing whether the crimes involved bootlegging,[17] numbers running,[18] marijuana farming,[19] or cell phones.[20] As the sophistication of criminals has increased, so too have the tools of the police. In the twentieth century, perhaps no two tools have been as revolutionary to modern policing as the two way radio and the patrol car.[21]

In this century, big data—in a variety of forms—may bring the next dramatic change to police investigations. One researcher has concluded that it will soon be technologically possible and affordable for

---

14. *See id.*

15. *See id.*

16. Constance L. Hays, *What Wal-Mart Knows About Customers' Habits*, N.Y. TIMES (Nov. 14, 2004), http://www.nytimes.com/2004/11/14/business/yourmoney/14wal.html.

17. *See, e.g.*, Olmstead v. United States, 277 U.S. 438, 455–58, 465–66 (1928); *see also Photo Gallery*, PBS, http://www.pbs.org/kenburns/prohibition/media_detail/2082733861-olmstead/ (last visited Feb. 17, 2014) (noting that Roy Olmstead was nicknamed the "King of Puget Sound Bootleggers").

18. *See, e.g.*, Katz v. United States, 389 U.S. 347, 348, 358–59 (1967) (illegal wagering).

19. *See, e.g.*, Kyllo v. United States, 533 U.S. 27, 29–33, 40 (2001).

20. In *United States v. Jones*, __U.S.__, 132 S. Ct. 945 (2012), the Supreme Court held that the government's attachment of a GPS tracking device on the defendant's car required a warrant. *Id.* at 954. On remand, the government argued that cell site data could be relied upon without a warrant, which the district court permitted under the good faith exception to the exclusionary rule. *See* United States v. Jones, Crim. Action No. 05-0386 (ESH) (D.D.C. Dec. 14, 2012), *available at* https://ecf.dcd.uscourts.gov/cgi-bin/show_public_doc?2005cr0386-658.

21. *Cf.* Samuel Walker, *"Broken Windows" and Fractured History: The Use and Misuse of History in Recent Police Patrol Analysis*, 1 JUST. Q. 75, 80 (1984) ("The mid-century revolution in American policing involved not just the patrol car but the car in conjunction with the telephone and the two-way radio.").

government to record *everything* anyone says or does.[22] How well will the Fourth Amendment's rules pertaining to unreasonable searches and seizures adapt to the uses of big data? Scholars have widely discussed the shortcomings of applying Fourth Amendment doctrines, once adequate for a world of electronic beepers, physical wiretaps, and binocular surveillance, to rapidly changing technologies.[23] But big data may magnify these concerns considerably.

This article identifies three uses of big data that hint at the future of policing and the questions these tools raise about conventional Fourth Amendment analysis. Two of these examples, predictive policing and mass surveillance systems, have already been adopted by a small number of police departments around the country. A third example—the potential use of DNA databank samples—presents an untapped source of big data analysis. Whether any of these three examples of big data policing attract more widespread adoption by the police is yet unknown, but it likely that the prospect of being able to analyze large amounts of information quickly and cheaply will prove to be attractive. While seemingly quite distinct, these three uses of big data suggest the need to draw new Fourth Amendment lines now that the government has the capability and desire to collect and manipulate large amounts of digitized information.

I. THE RISE OF BIG DATA

What is big data? While not everyone agrees on a single definition of big data, most agree that the term refers to: (1) the application of artificial intelligence (2) to the vast amount of digitized data now available.[24] From this basic definition, a few observations emerge about

---

22. JOHN VILLASENOR, BROOKINGS INSTITUTE, RECORDING EVERYTHING: DIGITAL STORAGE AS AN ENABLER OF AUTHORITARIAN GOVERNMENTS 1 (Dec. 14, 2011), *available at* http://www.brookings.edu/~/media/research/files/papers/2011/12/14%20digital%20storage%20villasenor/1214_digital_storage_villasenor.pdf.

23. A large literature has developed that critiques the limitations of modern search and seizure law as applied to computer software and hardware, the internet, new surveillance technologies, etc. *See, e.g.*, Orin S. Kerr, *The Fourth Amendment and New Technologies: Constitutional Myths and the Case for Caution*, 102 MICH. L. REV. 801, 803 n.7 (2004) (collecting sources espousing this view); Dan Solove, *Digital Dossiers and the Dissipation of Fourth Amendment Privacy*, 75 S. CAL. L. REV. 1083, 1086–87 (2002) (arguing that the current view on Fourth Amendment privacy "is not responsive to life in the modern Information Age").

24. *See, e.g.*, Steve Lohr, *How Big Data Became So Big*, N.Y. TIMES (Aug. 11, 2012), http://www.nytimes.com/2012/08/12/business/how-big-data-became-so-big-unboxed.html?smid=pl-share ("Big Data is a shorthand label that typically means applying the tools of artificial intelligence, like machine learning, to vast new troves of data beyond that captured in standard databases.").

what is distinct and significant about big data.[25]

First, big data alerts us to the sheer *amount* of information that is being produced rapidly every year in *digital* form.[26] The turn towards digitized information has been rapid and dramatic. As recently as the year 2000, only a quarter of the world's stored information was digital; the majority of it was on film, paper, magnetic tapes, and other similar non-digital media.[27] Today, the opposite is true; nearly all of the world's stored information is digital: about 2.7 zettabytes in 2012.[28]

Digital information continues to grow at a rapid pace. According to IBM, ninety percent of the world's data has been generated in the past two years.[29] The Executive Chairman of Google has claimed that we now create as much information in two days as we did from the beginning of human civilization to 2003.[30] Some have suggested that we may run out of ways to quantify numerically the amount of data generated.[31]

Nearly every piece of information today is capable of digitization and storage, including Internet searches, retail purchases, Facebook posts, cellphone calls, highway toll usage, and every last word in books.[32] Cheap, small, and sophisticated sensors and tracking devices have been built into every sort of product and object: smartphones, cars, toll

---

25. Here, too, there are some who dispute whether big data is a new phenomenon at all. *See, e.g.*, Samuel Arbesman, *Five Myths About Big Data*, WASH. POST (Aug. 16, 2013), http://www.washingtonpost.com/opinions/five-myths-about-big-data/2013/08/15/64a0dd0a-e044-11e2-963a-72d740e88c12_story.html (arguing that "big data has been around for a long time").

26. *See* MAYER-SCHÖNBERGER & CUKIER, *supra* note 10, at 8–11 (describing vast quantities of digitized data available today).

27. *See* Kenneth Neil Cukier & Viktor Mayer-Schöenberger, *The Rise of Big Data: How It's Changing the Way We Think About the World*, FOREIGN AFF., May–June 2013, at 28, 28.

28. Albert Pimental, *Big Data: The Hidden Opportunity*, FORBES (May 1, 2012), http://www.forbes.com/sites/ciocentral/2012/05/01/big-data-the-hidden-opportunity/. One zettabyte is 10 to the power of 21 bytes. This is equivalent to the amount of data which could fill 250 billion DVDs. Melody Kramer, *The NSA Data: Where Does It Go?*, NAT'L GEOGRAPHIC (June 12, 2013), http://news.nationalgeographic.com/news/2013/06/130612-nsa-utah-data-center-storage-zettabyte-snowden/.

29. IBM, IBM BIG DATA SUCCESS STORIES 1 (2011), *available at* http://public.dhe.ibm.com/software/data/sw-library/big-data/ibm-big-data-success.pdf.

30. Google, *Eric Schmidt at Technomy*, YOUTUBE (Aug. 4, 2010), http://www.youtube.com/watch?v=UAcCIsrAq70.

31. The largest current recognized number is a yottabye: a digit with twenty-four zeros. *See* Charles Walford, *Information Overload: There Is So Much Data Stored in the World That We May Run Out of Ways to Quantify It*, DAILY MAIL (Dec. 12, 2012), http://www.dailymail.co.uk/sciencetech/article-2247081/There-soon-words-data-stored-world.html.

32. *See* MAYER-SCHÖNBERGER & CUKIER, *supra* note 10, at 83–97 (discussing the "datafication of everything").

transponders, library books, and internet use.[33] The city of Santander, Spain is a prototype of the coming "smart city," with 12,000 sensors buried underground that measure everything from air pollution to free parking spaces.[34] The resulting data doesn't disappear; it ends up in "data barns" that store the ever-growing amount of information.[35] Wal-Mart handles more than a million customer transactions every hour, resulting in databases storing more than 2.5 petabytes of information.[36] In 2008, Facebook boasted storage of 40 billion photos.[37] The Library of Congress decided in 2010 to archive every public "tweet" generated on Twitter: about 170 billion tweets (and counting) in January 2013.[38]

Second, because the term also refers to the artificial intelligence applied to these huge data sets, the big data phenomenon also suggests a change in the way we understand our world. If conventional scientific research begins with a hypothesis or question that then shapes the collection of the relevant data, the big data phenomenon turns such conventions upside down. Because data is being collected and stored all of the time, research questions do not have to shape or limit data collection at all.[39] Researchers need not limit themselves to data sampling, either. Big data permits the study of a phenomenon where the set is nearly everything that is possible to study (another way of stating

---

33. UPS, for example uses telematics sensors in more than 46,000 vehicles; these tell the company about the speed, direction, braking, and drive train performance of their trucks. *See What is Big Data?*, SAS, http://www.sas.com/big-data/ (last visited Feb. 17, 2014).

34. Lauren Frayer, *High-Tech Sensors Help Old Port City Leap Into Smart Future*, NPR (June 4, 2013, 3:27 AM), http://www.npr.org/blogs/parallels/2013/06/04/188370672/Sensors-Transform-Old-Spanish-Port-Into-New-Smart-City.

35. Indeed, a series of investigative reports by the *New York Times* has revealed the relatively little-known environmental costs of huge data centers. *See, e.g.*, James Glanz, *Power, Pollution and the Internet*, N.Y. TIMES (Sept. 22, 2012), http://www.nytimes.com/2012/09/23/technology/data-centers-waste-vast-amounts-of-energy-belying-industry-image.html (quoting an industry executive as describing an "industry dirty secret"); James Glanz, *Data Barns in a Farm Town, Gobbling Power and Flexing Muscle*, N.Y. TIMES (Sept. 23, 2012), http://www.nytimes.com/2012/09/24/technology/data-centers-in-rural-washington-state-gobble-power.html?ref=us (reporting on "sprawling and ubiquitous" "data barns").

36. *See* SAS, BIG DATA MEETS BIG DATA ANALYTICS 1 (2012), *available at* http://www.sas.com/resources/whitepaper/wp_46345.pdf.

37. *See* Doug Beaver, *10 Billion Photos*, FACEBOOK (Oct. 14, 2008, 6:03 PM), http://www.facebook.com/note.php?note_id=30695603919.

38. Erin Allen, *Update on the Twitter Archive at the Library of Congress*, LIBRARY OF CONGRESS BLOG (Jan. 4, 2013), http://blogs.loc.gov/loc/2013/01/update-on-the-twitter-archive-at-the-library-of-congress/.

39. Mayer-Schönberger and Cukier discuss how the combination of cheap and easy data storage with powerful analytic technology makes it possible to constantly store data for purposes that may not be immediately apparent. *See* MAYER-SCHÖNBERGER & CUKIER, *supra* note 10, at 98–106.

that we are approaching n=all).[40] The existence of these massive data sets permits sifting and resifting of the information therein for multiple purposes.[41] Thus, the Library of Congress's continuous collection of "tweets" has interested researchers with questions as diverse as the role of public responses to smoking ads, changes in investor sentiments, and real-time hurricane analysis.[42]

Such massive quantities of information also suggest that the very kinds of questions posed by researchers will be different in the big data context. With so much data available, Viktor Mayer-Schönberger and Kenneth Cukier argue that two conventions of traditional research—a working hypothesis and the search for causality—are no longer necessary given the insights that can be derived from correlations found in big data.[43] The existence of huge amounts of data permits research into correlations that don't require an underlying hypothesis. For instance, Google's mathematical models have identified the forty-five search terms (e.g. "medicine for cough and fever") most strongly identified with historical flu data.[44] The resulting Google Flu Trends has proven to be remarkably accurate in matching the historical surveillance data collected by the Centers for Disease Control.[45] Thus, we can predict new outbreaks of the flu simply by identifying correlations between Google search terms and the spread of seasonal flu.[46] These predictions

---

40. *See id.* at 26 ("In many areas . . . a shift is taking place from collecting some data to gathering as much as possible, and if feasible, getting everything: N = all.").

41. *See id.* at 122 ("The crux of data's worth is its seemingly unlimited potential for reuse: its option value.").

42. Victor Luckerson, *What the Library of Congress Plans to Do with All Your Tweets*, TIME (Feb. 25, 2013), http://business.time.com/2013/02/25/what-the-library-of-congress-plans-to-do-with-all-your-tweets/.

43. *See* MAYER-SCHÖNBERGER & CUKIER, *supra* note 10, at 61 ("In a small-data world, because so little data tended to be available, both causal investigations and correlation analysis began with a hypothesis, which was then tested to be either falsified or verified. . . . Today, with so much data around and more to come, such hypotheses are no longer crucial for correlational analysis.").

44. *See id.* at 2.

45. *See id.*; *see also* Miguel Helft, *Google Uses Searches to Track Flu's Spread*, N.Y. TIMES (Nov. 11, 2008), http://www.nytimes.com/2008/11/12/technology/internet/12flu.html (reporting that Google found "a strong correlation" between five years of its data and the C.D.C.'s reports of flu).

46. *Explore Flu Trends—United States*, GOOGLE.ORG, http://www.google.org/flutrends/us/#US (last visited Feb. 17, 2014). Google has done the same with dengue trends around the world. *Dengue Trends Around the World*, GOOGLE.ORG, http://www.google.org/denguetrends/intl/en_us/ (last visited Feb. 17, 2014). The same approach has been taken with information generated by Twitter. *See* ADAM SADILEK ET AL., ASS'N FOR THE ADVANCEMENT OF ARTIFICIAL INTELLIGENCE, MODELING SPREAD OF DISEASE FROM SOCIAL INTERACTIONS (2012), *available at* http://www.cs.rochester.edu/~sadilek/publications/Sadilek-Kautz-Silenzio_Modeling-Spread-of-Disease-from-Social-Interactions_ICWSM-12.pdf (last visited Feb. 17, 2014). Such big data analysis is not perfect, however. Google's algorithms were grossly inaccurate in winter of 2012–13 and predicted

are useful in their predictive value even though they provide no causal explanation, much in the same way that Amazon's algorithms can predict that you might like a product based on its analysis without caring why.[47] A key contribution of big data is the ability to find useful correlations within data sets not capable of analysis by ordinary human assessment.

II. USE OF BIG DATA IN POLICING

Across the country, some police departments have taken notice that they stand to benefit from big data. While the use of big data in the private sector has raised concerns about consumer privacy, its use by the police raises even bigger questions about the limits of using data to justify surveillance, investigation, and detention by the police. This section discusses three of the most important developments in use of big data by the police: crime prediction, mass surveillance, and DNA databanks.

A. Crime Prediction: Predictive Policing

Perhaps the most visible use of big data by police departments thus far has been predictive policing: the application of computer modeling to historical crime data to predict future criminal activity.[48] While the police have long tried to find patterns of criminal activity on which to focus their resources,[49] predictive policing permits the police to harness thousands of data points to forecast where crime is likely to happen. The most basic models rely on past crimes, but data sources can include factors as variable as payday schedules, seasonal variation, liquor store

---

far more cases than the CDC counted. Part of the problem may be that the flu gained widespread media attention in 2012, which then increased the use of the same search terms that had better predictive value before. *See* Nick Bilton, *Disruption, Data Without Context Tells a Misleading Story*, N.Y. TIMES (Feb. 23, 2013), http://bits.blogs.nytimes.com/2013/02/24/disruptions-google-flu-trends-shows-problems-of-big-data-without-context/; Declan Butler, *When Google Got Flu Wrong*, 494 NATURE 155, 155–56 (2013), *available at* http://www.nature.com/news/when-google-got-flu-wrong-1.12413.

47. *Cf.* MAYER-SCHÖNBERGER & CUKIER, *supra* note 10, at 59 ("The correlations show *what*, not *why*, but . . . knowing *what* is often good enough.") (emphasis in original).

48. *See* JENNIFER BACHNER, PREDICTIVE POLICING: PREVENTING CRIME WITH DATA AND ANALYTICS 14 (2013), *available at* http://www.businessofgovernment.org/sites/default/files/Predictive%20Policing.pdf ("The fundamental notion underlying the theory and practice of predictive policing is that we can make probabilistic inferences about future criminal activity based on existing data.").

49. *See id.* at 7 (observing that "quantitative crime analysis spans centuries").

locations, and potential escape routes.[50]

What is new about predictive policing is not the use of quantitative data.[51] In the 1990s, the N.Y.P.D. ushered in an era of intelligence-based policing.[52] Under Commissioner Bill Bratton, the N.Y.P.D. introduced the now famous CompStat system[53]: weekly meetings at the N.Y.P.D. headquarters at which a revolving group of commanding officers around the city gave accountings of themselves for the recent crime data collected in their precinct.[54] By evaluating performance by the rise or fall of crime data within their precincts, CompStat meetings forced accountability upon commanding officers. Such data-driven policing spread to other departments around the United States when crime rates began to fall dramatically within New York City,[55] a result the police attributed to its reliance upon CompStat, along with the adoption of "broken windows" policing[56] and aggressive stop and frisk tactics.[57]

---

50. *See id.* at 16.

51. *See* WALTER L. PERRY ET AL., PREDICTIVE POLICING: THE ROLE OF CRIME FORECASTING IN LAW ENFORCEMENT OPERATIONS 2 (2013) ("The use of statistical and geospatial analyses to forecast crime levels has been around for decades."); Pearsall, *supra* note 8, at 18 (citing one police chief's skepticism that predictive policing is a break from older trends in intelligence based policing).

52. *See* BACHNER, *supra* note 48, at 6, 9 (noting that "predictive policing is viewed as one pillar of intelligence-led policing").

53. Bratton himself chronicled his tenure as Commissioner in his memoir *The Turnaround*. *See* WILLIAM BRATTON WITH PETER KNOBLER, THE TURNAROUND: HOW AMERICA'S TOP COP REVERSED THE CRIME EPIDEMIC (1998).

54. There are numerous accounts of the perceived innovation and success of the N.Y.P.D. during the 1990s. *See, e.g.*, VINCENT E. HENRY, THE COMPSTAT PARADIGM: MANAGEMENT ACCOUNTABILITY IN POLICING, BUSINESS AND THE PUBLIC SECTOR 17–18 (2003) (describing CompStat meetings as "intensive monthly performance evaluations for every commander of practically every operational unit in the agency"); ELI B. SILVERMAN, NYPD BATTLES CRIME: INNOVATIVE STRATEGIES IN POLICING 97–124 (1999) (describing development of CompStat meetings). In fact, some credit Bratton for thinking of a predictive policing model. *See* PERRY ET AL., *supra* note 51, at 4.

55. *See* BACHNER, *supra* note 48, at 9 (noting CompStat "has been adopted by nearly every law enforcement agency in the country").

56. "Broken windows" policing generally refers to a style of policing that focuses on minor offense enforcement on the assumption that such signs of disorder, if left unchecked, will lead to more serious crimes. *See* James Q. Wilson & George L. Kelling, *Broken Windows: The Police and Neighborhood Safety*, ATLANTIC MONTHLY, Mar. 1982, at 29, *available at* http://www.theatlantic.com/magazine/archive/1982/03/broken-windows/304465/.

57. More recently, however, the credit given to the use of CompStat by the N.Y.P.D. has been criticized. *See, e.g.*, David F. Greenberg, *Studying New York City's Crime Decline: Methodological Issues*, 31 JUST. Q. 154, 182 (2013) (concluding that there is an "absence of evidence pointing to large crime prevention effects in New York from [tactics including] CompStat . . . ."). And the stop and frisk policies of the N.Y.P.D. were eventually held to be unconstitutional. *See* Floyd v. City of New York, No. 08 Civ. 1034(SAS), 2013 WL 4046209 (S.D.N.Y. Aug. 12, 2013).

CompStat and similar programs inspired by it rely on the collection of crime statistics to inform police decision-making.[58]

The innovation of predictive policing is the application of artificial intelligence to such large data sets. CompStat relied heavily on the collection and display of past crime data; predictive policing applies computer analysis to similar information. The identification of future geographic *places* likely to be targeted by criminals has attracted the most attention. These predictive models all rely on well-established observations about the spatial distribution of criminal behavior. Crime is not found randomly across a city, but rather tends to fall within limited, and often very small, areas.[59] (Crime tends to be "lumpy.") For instance, researchers found that over a fourteen year period, about fifty percent of the crime in Seattle was limited to 4.5 percent of the city's street segments.[60] Based upon this connection between crime and place, computer models adopt different approaches towards the prediction of crime.

For instance, the Santa Cruz Police Department uses software that assumes that crime patterns follow a pattern similar to earthquake aftershocks.[61] The software applies a computer algorithm to a database representing five years' worth of crime data (including crime time, location, and type) to assess the likelihood of future crime occurring in the geographic areas within the city, narrowed to squares measured 500 by 500 feet (Figure A).[62] Prior to each shift, Santa Cruz police officers

---

58. Of course, even knowing about these high crime areas might suggest future criminal activity taking place in the same place, but the identification of these areas does not involve prediction specifically. *See* Andrew Guthrie Ferguson, *Predictive Policing and Reasonable Suspicion*, 62 EMORY L.J. 259, 274 (2012).

59. *See*, *e.g.*, PERRY ET AL., *supra* note 51, at 2 ("[C]riminals tend to operate in their comfort zone. That is, they tend to commit the type of crimes that they have committed successfully in the past, generally close to the same time and location.").

60. Anthony A. Braga et al., *The Relevance of Micro Places to Citywide Robbery Trends: A Longitudinal Analysis of Robbery Incidents at Street Corners and Block Faces in Boston*, 48 J. RES. CRIME & DELINQ. 7, 10 (2011) (citing David L. Weisburd et al., *Trajectories of Crime at Places: A Longitudinal Study of Street Segments in the City of Seattle*, 42 CRIMINOLOGY 283 (2004)).

61. *See*, *e.g.*, Martin Kaste, *Can Software That Predicts Crime Pass Constitutional Muster?*, NPR (July 26, 2013, 4:55 PM), http://www.npr.org/2013/07/26/205835674/can-software-that-predicts-crime-pass-constitutional-muster (noting that the software creator "wanted to see if computers could model future crime the same way they model earthquake aftershocks. Turns out they can."). The software, designed by mathematicians and social scientists at UCLA, Santa Clara University, and U.C. Irvine, is called PredPol and is marketed to police departments. *See Policing Meets Big Data*, PREDPOL, http://www.predpol.com/about/ (last visited Feb. 17, 2014).

62. *See* Zach Friend, *Predictive Policing: Using Technology to Reduce Crime*, FBI LAW ENFORCEMENT BULL. (Apr. 9, 2013), http://www.fbi.gov/stats-services/publications/law-enforcement-bulletin/2013/April/predictive-policing-using-technology-to-reduce-crime.

receive information identifying 15 such squares with the highest probability of crime, and are encouraged—though not required—to provide greater attention to these areas.[63] After its experimental introduction in 2011, the Santa Cruz Police Department reported a significant drop in burglaries when compared to a period prior to the adoption of predictive policing.[64] Similar experiments relying upon this software are being conducted by the police in Los Angeles and Seattle.[65]

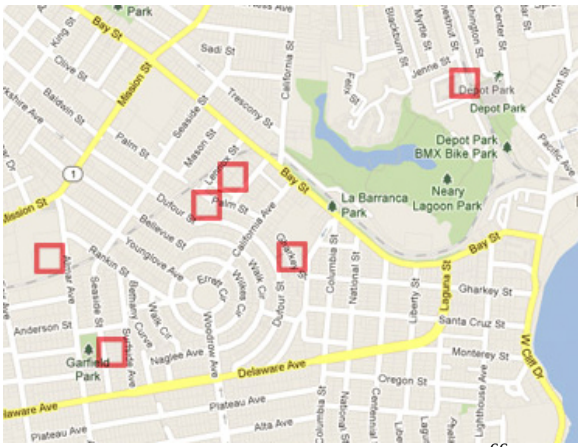

Figure A: Predictive Policing Map[66]

Other approaches may consider additional factors other than the timing and location of past crimes. Risk terrain theory, for example, looks at the social, physical, and behavioral factors that make it more likely certain areas will be targeted by crime.[67] The resulting risk terrain

63. *See* BACHNER, *supra* note 48, at 25.

64. *See id.* at 26. At least one investigative article has raised doubts, however, as to whether PredPol actually delivers on its claims about reducing crime. *See* Darwin Bond-Graham & Ali Winston, *All Tomorrow's Crimes: The Future of Policing Looks a Lot Like Good Branding*, S.F. WEEKLY (Oct. 30, 2013), http://www.sfweekly.com/2013-10-30/news/predpol-sfpd-predictive-policing-compstat-lapd/full/ (suggesting that PredPol's creators have been "most successful [with] its marketing algorithms").

65. *See, e.g.*, David Talbot, *L.A. Cops Embrace Crime-Predicting Algorithm*, MIT TECH. REV. (July 2, 2012), http://www.technologyreview.com/news/428354/la-cops-embrace-crime-predicting-algorithm/ (describing successful use of Predpol software in Foothill precinct of Los Angeles); Sengupta, *supra* note 5 (describing Seattle Police Department's use of PredPol software for property crimes).

66. *Looking Ahead, Not in the Rear View Mirror*, PREDPOL, http://www.predpol.com/technology/ (last visited Feb. 17, 2014).

67. *See, e.g.*, Leslie W. Kennedy et al., *Risk Clusters, Hotspots, and Spatial Intelligence: Risk Terrain Modeling as an Algorithm for Police Resource Allocations*, 27 J. QUANTITATIVE CRIMINOLOGY 339, 342–43 (2011).

map, which gives each factor its own mapping "layer," is a composite map that assigns a risk assessment for all of the factors associated with criminal activity.[68] For example, police in Morris County, New Jersey, use five factors for their risk terrain modeling: (1) past burglaries; (2) the residential location of individuals recently arrested for property crimes; (3) the proximity to major highways; (4) the geographic concentration of young men; and (5) the location of apartment complexes and hotels.[69] Morris County police attribute significant drops in violent and property crime to a reliance on risk terrain modeling.[70]

A second type of predictive technology focuses on the application of algorithms to social media in order to identify likely criminality based on the role an *individual* plays within a social network.[71] This social network analysis[72] begins with the assumption that social networks undergird many crimes: an illegal drug-dealing network may loosely follow the hierarchical structure of a legitimate business, with suppliers, distributors, buyers, and financiers.[73] (Indeed, this type of analysis has its roots in the military study of insurgent groups abroad.[74]) The algorithms used in social network software can help police visualize the density of connections an individual has within a social network. These connections might take the form of exchanges, communications, family ties, participation in crimes, or affiliations with an organization.[75]

---

68. *See id.* at 343.

69. Jeffrey S. Paul & Thomas M. Joiner, *Integration of Centralized Intelligence with Geographic Information Systems: A Countywide Initiative*, GEOGRAPHY & PUB. SAFETY, Oct. 2011, at 5, 7.

70. *See id.* at 7 (noting that since 2007, when the county created an intelligence crime task force, "the total crime index in the county has decreased by 11%, violent crime by 21%, and property crime by 7%").

71. IBM offers, for instance, a social media analytics tool that police departments can use to monitor Facebook and Twitter activity. *See* Somini Sengupta, *Privacy Fears Grow as Cities Increase Surveillance*, N.Y. TIMES (Oct. 13, 2013), http://www.nytimes.com/2013/10/14/technology/privacy-fears-as-surveillance-grows-in-cities.html.

72. Social network analysis should not be confused with police surveillance and infiltration of social media sites such as Facebook and Twitter, which have also proven to be valuable investigative tools.

73. *See* BACHNER, *supra note* 48, at 22–23.

74. *See* Philip Ball, *Unmasking Organised Crime Networks with Data*, BBC (July 9, 2013), http://www.bbc.com/future/story/20130709-unmask-crime-networks-with-data/1.

75. *See* BACHNER, *supra note* 48, at 23.

Figure B: Hypothetical Social Network Analysis[76]

When used by a police department, social network analysis might be used to identify a "central node": a person with a high degree of "connectivity within the network" (Figure B).[77] While traditional police work might easily identify leaders within a criminal organization, social network analysis can identify those with influence or those who transmit information within the group quickly and yet whose roles are not otherwise apparent.[78] The software can even reveal deliberately concealed affiliations. Even if an individual suspected of being part of a criminal organization does not admit his affiliation, social network software can calculate the probability of his membership.[79]

How does such software help police investigations? The identification of a highly "networked" individual could permit the police to infiltrate an organization in the most efficient way, or to identify a hidden source of influence within an organization for further investigation.[80] Also, by revealing hidden relationships among groups, police can disrupt subterfuges by rival criminal organizations. In a gang war, one group likely to retaliate may not do so directly, for fear of being targeted by the

76. Aaron Lester, *Police Clicking into Crimes Using New Software*, BOSTON GLOBE (Mar. 18, 2013), http://www.bostonglobe.com/business/2013/03/17/police-intelligence-one-click-away/DzzDbrwdiNkjNMA1159ybM/"story.html (describing provisional use of new Nucleik software within Springfield, Massachusetts gang unit).

77. *See* BACHNER, *supra* note 48, at 23.

78. *See* Lester, *supra* note 76.

79. *See* Ball, *supra* note 74.

80. *See* BACHNER, *supra* note 48, at 22–24.

police, and instead may enlist an ally gang.[81] Social network analysis can help understand which alliances might require heightened surveillance.[82]

B. Mass Surveillance: Domain Awareness Systems

If predictive policing harnesses data to predict the future, computer surveillance systems help police create a software-enhanced picture of the present, using thousands of data points from multiple sources within a city. As with predictive policing, computer enhanced mass surveillance grows out of other policing techniques.[83] While surveillance has long been an essential tool of the police, what has changed is its supporting technology. Sophisticated yet inexpensive, the surveillance equipment used by the police today produces enormous amounts of information, often too much for the police to use in an efficient way without the help of technology. The N.Y.P.D., for instance, has a database of 16 million license plates captured from its license plate readers, along with the locations of where the plates were photographed.[84]

The N.Y.P.D. has responded to this big data problem by creating a software program with Microsoft. Dubbed a "domain awareness system" ("DAS"),[85] the software collects and analyzes information around the clock within New York City from sources as disparate as the city's 3,000 public surveillance cameras,[86] over 200 automatic license plate readers,[87] more than 2,000 radiation sensors,[88] and information from

---

81. *See generally id.*

82. *See* Ball, *supra* note 74.

83. *Cf.* Jack M. Balkin, *The Constitution in the National Surveillance State*, 93 MINN. L. REV. 1, 3 (2008) ("Government's increasing use of surveillance and data mining is a predictable result of accelerating developments in information technology.").

84. Joseph Goldstein, *Weekly Police Briefing Offers Snapshot of Department and Its Leader*, N.Y. TIMES (Feb. 10, 2013), http://www.nytimes.com/2013/02/11/nyregion/weekly-briefing-provides-lengthy-snapshot-of-kelly-and-nypd.html.

85. Matt Sledge, *NYPD License Plate Readers Will Be Able to Track Every Car Entering Manhattan*, HUFFINGTON POST (Mar. 13, 2013, 5:08 PM), http://www.huffingtonpost.com/2013/03/13/nypd-license-plate-readers_n_2869627.html.

86. Cara Buckley, *New York Plans Surveillance Veil for Downtown*, N.Y. TIMES (July 9, 2007), http://www.nytimes.com/2007/07/09/nyregion/09ring.html; Mark Duell, *The Extraordinary 'Ring of Steel' Around Ground Zero: NYPD Steps Up Dirty Bomb Threat Protection with $200M Project*, DAILY MAIL (July 29, 2011), http://www.dailymail.co.uk/news/article-2020266/NYPD-steps-dirty-bomb-radiation-threat-protection-200m-Manhattan-project.html.

87. Al Baker, *Camera Scans of Car Plates Are Reshaping Police Inquiries*, N.Y. TIMES (Apr. 11, 2011), http://www.nytimes.com/2011/04/12/nyregion/12plates.html?pagewanted=all&_r=0 (observing that in 2011, the N.Y.P.D. had 238 license plate readers, 130 of them mobile). In March 2013, Police Commissioner Raymond Kelly announced plans to install license plate readers in every

police databases.[89] The software's mapping features permit the police to see and understand the information in a way that was not possible before. Located within the N.Y.P.D.'s lower Manhattan Security Initiative Command Center, the Domain Awareness System's operators can quickly use the software to identify potential threats.[90]

This system gives the police real-time access to information that can reveal connections between persons, items, and places in ways that may not be obvious to individual crime analysts. The DAS employs video analytic software designed to detect threats, such as unattended bags.[91] The N.Y.P.D. claims that the DAS can track where a car associated with a suspect is located, and where it has been in the past days, weeks, or months.[92] The DAS can also check license plate numbers, compare them to watch lists, and provide the police with immediate access to any criminal history associated with the car owner.[93] In November 2013, the N.Y.P.D. relied on its DAS to watch nearly every portion of the New York City Marathon route, a potential terrorist target after the Boston Marathon bombings in April 2013.[94]

While New York City has received the most attention for its high-tech approach to surveillance, other cities have shown interest in these mass surveillance systems. Oakland, California, a much smaller city in

---

lane of traffic that serve as exists and entrances to Manhattan, all of which would be linked to the domain awareness system. *See* Sledge, *supra* note 85.

88. Two thousand belt-mounted mobile radiation detectors are carried by N.Y.P.D. officers. *See* Duell, *supra* note 86. The N.Y.P.D. also plans to use very sensitive radiation scanners to detect the presence of concealed weapons on individuals in high crime areas. *See* Slate V Staff, *NYPD Plans Public Radiation Scanners to Detect Guns*, SLATE (Jan. 24, 2013), http://www.slate.com/blogs/trending/2013/01/24/nypd_radiation_scanners_gun_detectors_to_be_set_up_in_public_spaces.html.

89. *See* Sledge, *supra* note 85.

90. Duell, *supra* note 86.

91. *Id.*

92. *See, e.g.*, Press Release, Mayor Bloomberg, Police Commissioner Kelly and Microsoft Unveil New, State-of-the-Art Law Enforcement Technology That Aggregates and Analyzes Existing Public Safety Data in Real Time to Provide a Comprehensive View of Potential Threats and Criminal Activity (Aug. 8, 2012), *available at* http://www.nyc.gov/portal/site/nycgov/menuitem.c0935b9a57bb4ef3daf2f1c701c789a0/index.jsp?pageID=mayor_press_release&catID=1194&doc_name=http%3A%2F%2Fwww.nyc.gov%2Fhtml%2Fom%2Fhtml%2F2012b%2Fpr291-12.html&cc=unused1978&rc=1194&ndi=1.

93. Rocco Paranscandola & Tina Moore, *NYPD Unveils New $40 Million Super Computer System that Uses Data from Network of Cameras, License Plate Readers and Crime Reports*, N.Y. DAILY NEWS (Aug. 8, 2012, 8:50 PM), http://www.nydailynews.com/new-york/nypd-unveils-new-40-million-super-computer-system-data-network-cameras-license-plate-readers-crime-reports-article-1.1132135.

94. *See* Michael Schwirtz, *After Boston Bombings, New York Police Plan Tight Security at Marathon*, N.Y. TIMES (Nov. 1, 2013), http://www.nytimes.com/2013/11/02/sports/video-surveillance-to-be-a-key-component-of-marathon-security.html?_r=0.

comparison but plagued with a high crime rate,[95] has decided to launch a Domain Awareness Center[96] poised to collect and analyze surveillance data "from gunshot-detection sensors in the barrios of East Oakland to license plate readers mounted on police cars patrolling the city's upscale hills."[97] The resulting analysis will be displayed on a bank of giant monitors providing Oakland police with a unified visual representation of the very different sources: police and fire dispatch systems, gunshot detectors, license plate readers, private alarm detection programs, and social media feeds.[98]

C. Genetic Big Data: DNA Databanks

Perhaps less obvious but no less important a big data matter is the collection of DNA for criminal justice databases, which as of June 2013 contained DNA profiles for 10.7 million offenders and 1.7 million arrestees.[99] The United States has used this information to amass the largest DNA database in the world.[100] Police agencies around the country rely on CODIS—the shorthand for the system that links information among the different DNA databases around the country[101]—

95. Forbes named Oakland the third most dangerous city in America in 2012 with a population between 100,000 and 499,000. *See The 10 Most Dangerous U.S. Cities*, FORBES, http://www.forbes.com/pictures/mlj45jggj/3-oakland/ (last visited Feb. 17, 2014) (stating that Oakland's violent crime rate is "1,683 per 100,000 residents").

96. The project was initially sought by the Port of Oakland, but expanded to include the city of Oakland itself. *See* Steven Tavares, *Big Brother in Oakland? There Might Be an App Coming For That*, EBCITIZEN (July 10, 2013), http://www.ebcitizen.com/2013/07/big-brother-in-oakland-there-might-be.html?utm_source=feedburner&utm_medium=email&utm_campaign=Feed%3A+EastbaycitizenBreakingNewsPoliticsLife+%28EastBayCitizen+%7C+Breaking+News+Politics+Life%29.

97. Sengupta, *supra* note 71.

98. Ali Winston, *Oakland Surveillance Center Raises Concerns*, SFGATE (July 17, 2013, 9:46 PM), http://www.sfgate.com/crime/article/Oakland-surveillance-center-raises-concerns-4671708.php. In February 2014, however, Oakland officials delayed voting on a contract to build its DAS after local residents raised concerns about privacy. *See* Associated Press, *Oakland Delays Vote on Surveillance System*, WASH. TIMES (Feb. 19, 2014), http://www.washingtontimes.com/news/2014/feb/19/oakland-delays-vote-on-surveillance-system/.

99. *See CODIS—NDIS Statistics*, FBI, http://www.fbi.gov/about-us/lab/biometric-analysis/codis/ndis-statistics (last visited Feb. 17, 2014).

100. *See* Solomon Moore, *F.B.I. and States Vastly Expand DNA Databases*, N.Y. TIMES (Apr. 18, 2009), http://www.nytimes.com/2009/04/19/us/19DNA.html (noting that CODIS is "the largest [database] in the world"). The U.K., however, has the distinction of having the largest portion of its population—about ten percent—in its DNA database. *See, e.g.*, Jill Lawless, *Spread of DNA Databases Sparks Ethical Concerns*, ASSOCIATED PRESS (July 12, 2013, 8:50 AM), http://bigstory.ap.org/article/spread-dna-databases-sparks-ethical-concerns.

101. Although CODIS specifically refers to the software that links DNA databases around the country for information sharing purposes, it is also used generically as a term to describe the

to match crime scene samples with offender or arrestee DNA profiles.[102] The millions of DNA samples now accessible by the police present another potential use of big data.

The rapid growth of American DNA databases can be attributed to the ever-expanding categories of those deemed eligible for compulsory DNA collection. While the first state DNA databases collected samples only from violent felons or felony sex offenders, today every state collects DNA from all convicted felons.[103] A majority of states collect DNA from those convicted of misdemeanor sex offenses.[104] In 2012, New York became the first "all crimes state."[105] Nearly every person convicted of a crime in New York, regardless of its gravity, will be required to submit a DNA sample for inclusion in the state's DNA database.[106]

The reliance of states upon arrestee DNA collection appears to be following a similar path. In 1997, Louisiana became the first state to authorize the collection of DNA from some categories of arrestees.[107] Today, twenty-eight more states and the federal government have followed Louisiana's lead in requiring some categories of arrestees to provide DNA samples.[108] The Supreme Court's 2013 decision in

---

American DNA database system more generally. *See Frequently Asked Questions (FAQs) on the CODIS Program and the National DNA Index System*, FBI, http://www.fbi.gov/about-us/lab/biometric-analysis/codis/codis-and-ndis-fact-sheet (last visited Feb. 17, 2014).

102. *See, e.g., id.* (describing how CODIS would be used in a hypothetical sexual assault case). CODIS permits states and the federal government to upload and compare DNA profiles on the National DNA Index System (NDIS). *See id.* While state laws specify the types of profiles that can be included in state databases, federal law determines which DNA profiles can be stored and shared at the national level. *See* 42 U.S.C. § 14132(a) (2006). Many resources provide helpful explanatory information on DNA and how it is used in criminal investigations. *E.g.*, SHELDON KRIMSKY & TANIA SIMONCELLI, GENETIC JUSTICE: DNA DATA BANKS, CRIMINAL INVESTIGATIONS, AND CIVIL LIBERTIES 3–27 (2010); *DNA Evidence Basics*, NAT'L INST. JUST. (Aug. 9, 2012), http://www.nij.gov/topics/forensics/evidence/dna/basics/pages/welcome.aspx.

103. *See* NATHAN JAMES, CONG. RESEARCH SERV., R41800, DNA TESTING IN CRIMINAL JUSTICE: BACKGROUND, CURRENT LAW, GRANTS, AND ISSUES 7 (2012).

104. *See id.*

105. *New York DNA Database: Governor Cuomo Signs 'All Crimes' DNA Testing Into Law*, HUFFINGTON POST (Mar. 3, 2012, 10:22 AM), http://www.huffingtonpost.com/2012/03/20/new-york-dna-database-governor-cuomo-all-crimes-dna-testing_n_1366624.html.

106. *See id.* (noting minor exemption "for those convicted of possession of a small amount of marijuana as long as they have no prior convictions").

107. *See* Julie Samuels et al., *Collecting DNA From Arrestees: Implementation Lessons*, NAT'L INST. JUST. J., June 2012, at 18, 19, *available at* https://www.ncjrs.gov/pdffiles1/nij/238484.pdf.

108. *See DNA Arrestee Laws*, NAT'L CONFERENCE OF ST. LEGISLATURES, http://www.ncsl.org/research/civil-and-criminal-justice/dna-arrestee-laws.aspx (last visited Feb. 17, 2014) (reporting that in May 2013, Nevada became the most recent state to require DNA samples from all felony arrestees).

Maryland v. King[109] upholding compulsory arrestee DNA collection[110] will likely mean that the practice will expand to many other states.[111]

How do DNA databases raise big data questions? First, the emerging and controversial use of familial matches is in fact a big data issue. Based on the assumption that close relatives share more genetic information than unrelated individuals, familial searches are searches of DNA databases that look for profiles that only partially match the thirteen STR markers[112] on a DNA profile.[113] (Such a search might take place, for instance, if a CODIS search yields no identical match.[114]) Familial searches take advantage of the big data set that is CODIS: the capability to search millions of individual DNA profiles.[115]Additional testing on DNA samples may be necessary to confirm potential matches.[116]

Similar to other uses of big data, a familial search repurposes (genetic) data collected for another reason (identical matches).[117] Critics of familial searches have focused on issues of privacy and equity, including the concern that familial searches will draw disproportionate

---

109. __U.S.__, 133 S. Ct. 1958 (2013).

110. *Id.* at 1980.

111. The Katie Sepich Enhanced DNA Collection Act provides federal funding for those states that wish to establish arrestee DNA collection programs. *See* Katie Sepich Enhanced DNA Collection Act of 2012, Pub. L. No. 112-253, § 3, 126 Stat. 2407 (2013). For further commentary on the decision, see Elizabeth E. Joh, Maryland v. King: *Policing and Genetic Privacy*, 11 OHIO ST. J. CRIM. L. 281 (2013).

112. The thirteen STR markers refer to the thirteen places on the human chromosome where there is high variability. These identification markers provide law enforcement with a unique identifier for everyone who provides a DNA sample. *See The FBI and DNA*, FBI (Nov. 28, 2011), http://www.fbi.gov/news/stories/2011/november/dna_112811.

113. *See, e.g.*, Erin Murphy, *Relative Doubt: Familial Searches of DNA Databases*, 109 MICH. L. REV. 291, 297–300 (2010) (describing mechanics of familial searches). How those genetically related to the sought after offender turn up depends on a number of factors, including whether a jurisdiction intentionally searches the database for partial matches or whether such matches turn up as the result of a search because the parameters of the search permitted less than an identical match between the sample and the DNA profile. *See id.* at 299.

114. *Id.* at 297–98.

115. *See id.* at 296. In 2008, California became the first state to formally authorize intentional partial matches, or "familial" searches. *Id.* at 293.

116. Matching STRs on the Y chromosomes, in addition to a partial match on the CODIS loci, can show how closely related two men are through their male ancestors. This Y-STR typing can show whether two men share the same genetic father or paternal grandfather. *See, e.g.*, Michael Chamberlain, *Familial DNA Searching: A Proponent's Perspective*, CRIM. JUST., Spring 2012, at 18, *available at* http://www.americanbar.org/content/dam/aba/publications/criminal_justice_magazine/sp12_dna_search_proponents.authcheckdam.pdf (explaining basics of California familial match policy).

117. *See supra* Part I.

attention to racial and ethnic minorities who may be unfairly targeted for genetic surveillance.[118] Yet if we think of familial searches as big data problems, we might also make some useful connections to other areas in which data is being amassed in large quantities for one purpose and later used for another. For instance, policies on familial searches might follow principles of informational privacy used in other database contexts, including limited later analysis to the specific original purpose for which the information was collected.[119]

This potential for repurposing is not limited to familial searches of CODIS, either. A profile in the national DNA database is a string of numbers referring to the thirteen STR locations.[120] Most courts analyzing Fourth Amendment challenges to the compulsory collection of DNA have focused only on the DNA *profile* to deny that their collection and storage by the government raises serious privacy concerns.[121] What is often ignored, however, is that these numbers are generated from biological samples.[122] These samples pose rich data possibilities; information that could be analyzed in different ways for a variety of purposes.[123] Indeed, David Lazer and Viktor Mayer-Schönberger argue

---

118. *See, e.g.*, Murphy, *supra* note 113, at 304 ("Familial searches should be forbidden because they embody the very presumptions that our constitutional and evidentiary rules have long endeavored to counteract: guilt by association, racial discrimination, propensity, and even biological determinism. They are akin to adopting a policy to collect and store the DNA of otherwise database-ineligible persons, solely because they share a blood relation with a convicted person . . . ."); Sonia M. Suter, *All in the Family: Privacy and DNA Familial Searching*, 23 HARV. J.L. & TECH. 309, 368–70 (2010).

119. *See, e.g.*, David Lazer & Viktor Mayer-Schönberger, *Statutory Frameworks for Regulating Information Flows: Drawing Lessons for the DNA Data Banks from Other Government Data Systems*, 34 J.L. MED. & ETHICS 366, 372 (2006).

120. *See, e.g.*, Moore, *supra* note 100 (describing CODIS profile as "numerical sequence").

121. *See, e.g.*, Maryland v. King, __U.S.__, 133 S. Ct. 1958, 1979 (2013) (observing that "the CODIS loci come from noncoding parts of the DNA that do not reveal the genetic traits of the arrestee").

122. *See, e.g.*, Suter, *supra* note 118, at 331 ("Courts often minimize or fail to address the fact that the collection of DNA samples involves two privacy intrusions: the actual collection of biological samples and the retention of samples that contain one's genetic information.").

123. Scholars and judges have expressed a wide range of opinions on whether privacy interests are truly threatened by DNA samples held by the government. *Compare* N. Van Camp & K. Dierickx, *The Retention of Forensic DNA Samples: A Socio-Ethical Evaluation of Current Practices in the EU*, 34 J. MED. ETHICS 606, 606 (2008), *and* Tania Simoncelli & Barry Steinhardt, *California's Proposition 69: A Dangerous Precedent for Criminal DNA Databases*, 33 J.L. MED. & ETHICS 279, 284 (2005) ("While law enforcement authorities would like us to believe that the samples will never be used for anything besides catching criminals, an unlimited span of improper uses remain plausible so long as those samples are retained."), *and* Suter, *supra* note 118, at 335 ("[W]e should be wary of [sample retention] given its substantial threat to privacy and civil liberties."), *with* David H. Kaye, *A Fourth Amendment Theory for Arrestee DNA and Other Biometric Databases*, 15 U. PA. J. CONST. L. 1095, 1155–58 (2013) (expressing extreme skepticism

that the very existence of these DNA samples "*invites* re-purposing at a later stage."[124]

At the moment, however, practical and legal barriers bar this possibility.[125] Although state laws vary with regard to storage, retention, and disclosure requirements,[126] federal law imposes conditions on those samples used to generate profiles for CODIS. For instance, federal law requires that all samples used for CODIS profiles are subject to disclosure only to "criminal justice agencies for law enforcement identification purposes," "in judicial proceedings," "for criminal defense purposes," and for a "population statistics database, identification research and protocol development purposes, or for quality control purposes."[127] At the same time, most state laws contemplate indefinite retention of most DNA samples.[128] Justifications for indefinite DNA sample retention include the need to identify potential sample contamination or mix-ups, to implement changes in the technology used to analyze samples, and to provide lab quality assurance.[129]

---

that DNA samples will ever be used beyond biometric identification).

124. Lazer & Mayer-Schönberger, *supra* note 119, at 372 (emphasis added); *see also* Lawless, *supra* note 100 (quoting one supporter of genetic databases as acknowledging "[t]here is an argument to be made that because that biological sample exists, the government could go back and do other things with it that are not authorized by the law").

125. These legal barriers, however, do not apply to the emerging issue of "offline" or "rogue" DNA databases that are being established by local law enforcement agencies that have no intention of sharing the information with CODIS. *See, e.g.*, Joseph Goldstein, *Police Agencies Are Assembling Records of DNA*, N.Y. TIMES, June 13, 2013, at A1 ("These local databases operate under their own rules, providing the police much more leeway than state and federal regulations.").

126. *See, e.g.*, Sarah B. Berson, *Debating DNA Collection*, NAT'L INST. JUST. J., Nov. 2009, at 9, 11, *available at* https://www.ncjrs.gov/pdffiles1/nij/228383.pdf ("State laws . . . vary with regard to how samples may be used beyond law enforcement and quality control purposes.").

127. *See* 42 U.S.C. § 14132(b)(3) (2006). Some have suggested, however, that a "criminal justice purpose" could be broadly construed to permit law enforcement agencies to analyze samples for a variety of purposes beyond simple matches to crime scene evidence. *See, e.g.*, Suter, *supra* note 118, at 336.

128. *See, e.g.*, JAMES, *supra* note 103, at 5; Mark A. Rothstein & Meghan K. Talbott, *The Expanding Use of DNA in Law Enforcement: What Role for Privacy?*, 34 J.L. MED. & ETHICS 153, 158 (2006) ("There is no national policy on sample retention, but in almost every state the samples are retained indefinitely."). Many of these observations have relied upon a 2005 study by the American Society of Law, Medicine and Ethics. *See* SETH AXELRAD, AM. SOC'Y OF L., MED. & ETHICS, SURVEY OF STATE DNA DATABASE STATUTES, http://www.aslme.org/dna_04/grid/guide.pdf (last visited Feb. 17, 2014).

129. M. Dawn Herkenham, *Retention of Offender DNA Samples Necessary to Ensure and Monitor Quality of Forensic DNA Efforts: Appropriate Safeguards Exist to Protect the DNA Samples from Misuse*, 34 J.L. MED. & ETHICS 380, 381–82 (2006) (Herkenham was the chief of the FBI unit responsible for implementing the NDIS); *see also* JOHN M. BUTLER, FUNDAMENTALS OF FORENSIC DNA TYPING 280–81 (2009) (noting that samples should be preserved for quality control and for "technology advancements in the future" regarding new genetic markers or assays).

For the now, practical barriers also bar comprehensive analysis of the millions of samples collected for criminal justice purposes. While human genome sequencing is vastly cheaper today than it was a few years ago, it likely remains prohibitively costly for states to undertake on a massive scale.[130] These practical and legal impediments may, however, change one day. As technological capabilities change, costs decrease, and a greater understanding of genetic information emerges, the use of DNA databases will raise serious questions for lawmakers about the appropriate balance of big data analysis and privacy protections.[131]

III. HOW BIG DATA CHALLENGES FOURTH AMENDMENT ANALYSIS

These evolving areas raise new questions about how best to regulate the use of big data by the police. In particular, they arise from three characteristics of big data: the use of artificial intelligence, the scale of data storage, and the repurposing of collected data. This section considers some of the difficult questions that judges and lawmakers will face.

A. Human Judgment and Police Suspicion

While popular accounts misleadingly suggest that predictive policing involves police decision-making *controlled* by computers,[132] even partial reliance on artificial intelligence does raise important Fourth Amendment questions. Police are using predictive policing software to direct them to places where they believe there is a high likelihood of criminal activity. Having been directed there by computer analysis, the police must then determine whether any persons located there warrant further investigation. What role should artificial intelligence and human judgment play in Fourth Amendment individualized suspicion?

At a minimum, ordinary investigative detentions by the police require

---

130. In the years since the human genome was first sequenced, the cost of sequencing has fallen dramatically, from nearly $100 million in 2001 to less than $6,000 in 2013. The National Human Genome Research Institute tracks the costs of genome sequencing. *See* K.A. Wetterstrand, *DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP)*, NAT'L HUMAN GENOME RES. INST., http://www.genome.gov/sequencingcosts/ (last visited Feb. 17, 2014).

131. *See, e.g.*, Phil Reilly, *Legal and Public Policy Issues in DNA Forensics*, 2 NATURE REVIEWS GENETICS 313, 317 (2001) (suggesting establishment of "a permanent commission to oversee [DNA databanks], which could review and monitor all requests to use samples for purposes other than forensic identification").

132. *See, e.g.*, PERRY ET AL., *supra* note 51, at 115–16 ("Although much of news coverage promotes the meme that predictive policing is a crystal ball, these algorithms simply predict risks.").

reasonable suspicion[133] based on a totality of the circumstances.[134] The Supreme Court's decisions have permitted the police to formulate reasonable suspicion based not only on their own personal observations, but also on other information, including fellow officers,[135] tips (even anonymous ones),[136] and sometimes even on determinations that particular geographic locations may be labeled as "high crime areas."[137] In particular, tips relied upon by the police must be sufficiently particularized to an individual, in some part predictive of future activity, and corroborated by the observation of the police themselves.[138]

The question here is whether predictive software based on historical crime data is similar to other uses of third party information that have already been held to support a reasonable suspicion determination.[139] Imagine that such software directs the police to a city block to look for property crime, and they observe activity that by itself may not appear obviously suspicious, such as carrying a duffel bag, or peering in windows.[140] A probabilistic determination is not exactly like an informant's tip, particularly since predictive software provides assessments about geographic *areas* and not persons.[141] Nevertheless, a

133. Terry v. Ohio, 392 U.S. 1, 20–21 (1968).

134. Alabama v. White, 496 U.S. 325, 330 (1990) ("Reasonable suspicion . . . is dependent upon both the content of information possessed by police and its degree of reliability. Both factors—quantity and quality—are considered in the 'totality of the circumstances—the whole picture' that must be taken into account when evaluating whether there is reasonable suspicion." (citation omitted)).

135. *Cf.* United States v. Ventresca, 380 U.S. 102, 111 (1965) ("Observations of fellow officers of the Government engaged in a common investigation are plainly a reliable basis for a warrant applied for by one of their number.").

136. Illinois v. Gates, 462 U.S. 213, 244–46 (1983).

137. Illinois v. Wardlow, 528 U.S. 119, 124 (2000).

138. *See, e.g.*, *Gates*, 462 U.S. at 245–46 (noting that anonymous tip "contained a range of details relating . . . to future actions of third parties ordinarily not easily predicted"); Florida v. J.L., 529 U.S. 266, 271 (2000) (noting that the tip in the case "provided no predictive information and therefore left the police without means to test the informant's knowledge or credibility"); 2 WAYNE R. LAFAVE, SEARCH & SEIZURE § 3.3f (5th ed. 2012) ("[I]t seems wise in light of subsequent events to read *Gates* to mean that corroboration of part of an anonymous informant's information will constitute a sufficient substitute for directly-established veracity and basis of knowledge *only* if the corroborated events are in and of themselves quite suspicious.").

139. Professor Andrew Ferguson was among the first to recognize the Fourth Amendment challenges raised by the adoption of predictive policing programs. *See* Ferguson, *supra* note 58, at 305–12 (discussing these analogies).

140. *See id.* at 309 (citing example).

141. Because such programs only make predictions about areas where crime is likely to happen, it would seem more difficult to justify probable cause in the predictive policing analysis, although certainly many stops can ripen into full blown arrests once more information about the suspect is made known to the police during the course of a stop.

court might analogize computerized predictions to informant-based predictions about specific places—such as drug houses and hourly motels—to add to the reasonable suspicion assessment.[142] While likely not sufficient on its own to provide justification for a stop (because of its lack of specificity with regard to persons), such predictions could form the basis of police observation and corroboration.[143]

So long as predictive software is not the sole justification used by police, courts are likely to accept its place within the reasonable suspicion analysis. If, for instance, courts were to borrow assessments of credibility and veracity from the informant context,[144] predictive software may provide more justification than an anonymous informant. The assumptions and inputs of such software, after all, are capable of verification.[145] Indeed, to the extent that the Supreme Court has emphasized that the reasonable suspicion determination is to be objective,[146] reliance on a computer analysis of crime data is arguably more objective than an inference made by an officer or a tip provided by a third party. Software with a demonstrated history of successfully predicting high crime areas based on verifiable crime data is likely to be a highly persuasive factor in the reasonable suspicion formulation.

Indeed, predictive software may remove some of the problems raised by the types of information used. Informants, particularly anonymous ones, can have questionable motivations in aiding the police.[147] In

142. *See* Ferguson, *supra* note 58, at 306–07.

143. *See id.* at 310 (observing that "a common theme in the Fourth Amendment" analysis of reasonable suspicion is "[c]orroboration of individual actions").

144. While the Court in *Illinois v. Gates* adopted a totality of the circumstances tests for probable cause, it nevertheless reaffirmed that these factors continued to be "all highly relevant" and "should be understood [to] illuminate the commonsense, practical question" of probable cause. *See* 462 U.S. 213, 230 (1983). The same can be said of the reasonable suspicion standard as well. *See* Alabama v. White, 496 U.S. 325, 330 (1990) ("Reasonable suspicion is a less demanding standard than probable cause not only in the sense that reasonable suspicion can be established with information that is different in quantity or content than that required to establish probable cause, but also in the sense that reasonable suspicion can arise from information that is less reliable than that required to show probable cause.").

145. *Cf.* Ferguson, *supra* note 58, at 307 (noting that "an objective, well-functioning computer program seems more reliable than your typical police informant").

146. *See, e.g.*, United States v. Cortez, 449 U.S. 411, 417–18 (1981) ("[In an investigative detention] officers must have a particularized and objective basis for suspecting the particular person stopped of criminal activity.").

147. The tip that led to the investigation of Lance and Sue Gates was allegedly given by Sue Gates's hairdresser, annoyed with Sue's boasting. Thomas Y. Davies, *The Supreme Court Giveth and the Supreme Court Taketh Away: The Century of Fourth Amendment 'Search and Seizure' Doctrine*, 100 J. CRIM. L. & CRIMINOLOGY 933, 1005 n.383 (2010). On the problems raised by the "informant institution," see generally Alexandra Natapoff, *Snitching: The Institutional and Communal Consequences*, 73 U. CIN. L. REV. 645 (2004).

addition, most courts are highly deferential to generalized police judgments of what constitutes a "high-crime area."[148] Software that eliminates undesirable biases and requires quantitative precision can introduce more fairness into the police decision-making process.[149]

Some caveats remain, however. First, no predictive policing program is entirely objective. The basic building blocks of a predictive software program necessarily involve human discretion.[150] The assumptions underlying any method of crime prediction rely upon the decision to choose one model of risk prediction over another. The data used to build the models will depend on discretionary judgments about the types of crimes used for prediction, and the type of information used to predict those crimes. Should a police department focus on burglaries; and if so, how are burglaries to be measured? For example, reliance on arrest rates is surely problematic[151] because arrests themselves are discretionary decisions that, if used as the basis to justify more attention, may simply reinforce unjustified police stereotypes that certain neighborhoods need heavier police attention.[152]

Second, prediction models might nudge police judgments in favor of investigative detention in borderline cases because the police rely too

148. *See, e.g.*, Andrew Guthrie Ferguson & Damien Bernache, *The "High-Crime Area" Question: Requiring Verifiable and Quantifiable Evidence for Fourth Amendment Reasonable Suspicion Analysis*, 57 AM. U. L. REV. 1587, 1607 (2008) ("[T]he majority of jurisdictions . . . have relied on an officer's testimony that an area is a 'high-crime area' without much analysis as to the basis of that conclusion.").

149. I have argued elsewhere that an automated traffic enforcement system made possible by a federal intelligent highway initiative could improve fairness and reduce or eliminate racial profiling in traffic stops. *See* Elizabeth E. Joh, *Discretionless Policing: Technology and the Fourth Amendment*, 95 CALIF. L. REV. 199 (2007).

150. Not only are there decisions about which model to use, each model itself involves discretionary judgments about the type and amount of data to use, as well as how to display it. *See, e.g.*, BACHNER, *supra* note 48, at 21 ("Just as with the other clustering methods, the final map is sensitive to analyst judgment.").

151. Measures of crime based on arrest rates—and particularly arrests in minor offenses—are problematic because they represent the greatest exercise of police discretion. *See* Wayne A. Logan, *Policing Identity*, 92 B.U. L. REV. 1561, 1590 (2012). As a result, the resulting data may often reflect racial biases in policing. *See, e.g.*, Simon A. Cole, *Fingerprint Identification and the Criminal Justice System: Historical Lessons for the DNA Debate*, *in* DNA AND THE CRIMINAL JUSTICE SYSTEM: THE TECHNOLOGY OF JUSTICE 63, 82 (David Lazer ed., 2004) (observing that criminal histories "appear to be pure, objective information, when in fact they may reflect the prejudices of police or judicial practitioners").

152. *See, e.g.*, *Predictive Policing: Don't Even Think About It*, ECONOMIST (July 20, 2013), http://www.economist.com/news/briefing/21582042-it-getting-easier-foresee-wrongdoing-and-spot-likely-wrongdoers-dont-even-think-about-it ("It matters . . . whether software crunches reports of crimes or arrests; if the latter, police activity risks creating a vicious circle.").

heavily on probabilistic information.[153] If, for example, a predictive model directs the police to look at a particular block for burglaries, then it may encourage the police to "see" suspicious behavior when there may be none.[154] The danger here is that an overreliance on the objectivity of prediction—which is in fact an informed probabilistic guess—will be determinative, rather than a supplement to independent assessments by the police.

B. Privacy and Surveillance Big Data

What we do in public can be seen by anyone and therefore we generally cannot claim those activities are private. That intuition is embodied in the *Katz* reasonable expectation of privacy test to determine whether the Fourth Amendment applies to police activity at all.[155] But does assuming the risk of police surveillance mean something different when the police have mass surveillance capabilities?

Computer enhanced mass surveillance systems would seem to be the latest example of the increasing sophistication of police technologies to monitor public activity. Decades of police reliance upon CCTV cameras, electronic beepers, listening devices, surveillance aircraft, and other similar sense enhancements have prompted concerns that these measures have significantly eroded any social sense of privacy individuals have in public.[156] Indeed, the Supreme Court has emphasized in a number of cases that our public activities, movements, and even our literal physical characteristics visible to the public lack Fourth Amendment protection.[157]

---

153. The predictive software may drive the officer to use personal observation to *confirm* the potentially suspicious behavior rather than independently assess whether it is truly suspicious. *Cf.* Andrew E. Taslitz, *Police Are People Too: Cognitive Obstacles to, and Opportunities for, Police Getting the Individualized Suspicion Judgment Right*, 8 OHIO ST. J. CRIM. L. 7, 29–30 (2010) (discussing "continuum model" in which observer uses further assessment to confirm initial judgments rather than challenging them).

154. A court may see the issue characterized as a kind of reliable tip—albeit from a computer—that requires less police corroboration precisely because of its reliability. *See, e.g.*, Ferguson, *supra* note 58, at 308 (making this observation).

155. Katz v. United States, 389 U.S. 347, 361 (1967) (Harlan, J., concurring).

156. In fact, the problems of a mass surveillance system like the total domain awareness program were anticipated years before such programs actually existed. *See, e.g.*, Robert H. Thornburg, Comment, *Face Recognition Technology: The Potential Orwellian Implications and Constitutionality of Current Uses Under the Fourth Amendment*, 20 J. MARSHALL J. COMPUTER & INFO. L. 321, 343 (2002) (noting in 2002 that "a networked system could identify an individual in one location on a specific date, and identify that same person at a different location afterwards").

157. *See, e.g.*, United States v. Dionisio, 410 U.S. 1, 14 (1973) ("The physical characteristics of a person's voice, its tone and manner, as opposed to the content of a specific conversation, are

Moreover, a line of Supreme Court cases suggests that any "scientific enhancement" of the senses used by the police to watch activity falls outside of the Fourth Amendment's protections if the activity takes place in public.[158] Thus, the Supreme Court concluded in *United States v. Knotts* that police use of an electronic beeper to follow a suspect surreptitiously did not constitute a Fourth Amendment search.[159] The premise underlying such a conclusion is that if the police could themselves pursue a suspect over the same public roads, then so too could an electronic beeper concealed within a container given to the unwitting suspect.[160]

The surveillance capacities of the police today, however, far exceed even what armies of police officers could accomplish without access to big data.[161] That difference should alter the absence of Fourth Amendment protections. Indeed, several Justices have recently indicated concerns about the big data surveillance capacities of the police.[162] For example, in *United States v. Jones*[163] (regarding the twenty-eight day GPS tracking of a single suspect[164]), five Justices expressed concerns that long-term police surveillance, even of a person's public movements, might constitute a Fourth Amendment search.[165] The premise here, sometimes referred to as the "mosaic theory," is that the danger to Fourth Amendment privacy lies in the aggregation of discrete bits of data, even if each piece standing alone would not be subjected to constitutional protections.[166] Indeed, the majority in *Knotts*

---

constantly exposed to the public. . . . No person can have a reasonable expectation that others will not know the sound of his voice, any more than he can reasonably expect that his face will be a mystery to the world."); Davis v. Mississippi, 394 U.S. 721, 727 (1969) ("Fingerprinting involves none of the probing into an individual's private life and thoughts that marks an interrogation or search."). Christopher Slobogin has convincingly argued, however, that a right to anonymity—even in public—should be protected by the Fourth Amendment. *See* Christopher Slobogin, *Public Privacy: Camera Surveillance of Public Places and the Right to Anonymity*, 72 MISS. L.J. 213 (2002).

158. *E.g.*, United States v. Knotts, 460 U.S. 276, 285 (1983).

159. *Id.*

160. *Id.* ("A police car following [the defendant] at a distance throughout his journey could have observed him leaving the public highway and arriving at the cabin owned by respondent, with the drum of chloroform still in the car.").

161. *See, e.g.*, People v. Weaver, 909 N.E.2d 1195, 1199 (N.Y. 2009) ("The potential for a similar capture [to GPS technology] of information or 'seeing' by law enforcement would require, at a minimum, millions of additional police officers and cameras on every street lamp.").

162. Certainly a number of lower court judges have expressed these concerns as well.

163. __U.S.__, 132 S. Ct. 945 (2012).

164. *Id.* at 948.

165. *See id.* at 955 (Sotomayor, J., concurring); *id.* at 964 (Alito, J., concurring in judgment).

166. The "mosaic theory" is generally attributed to the decision in *United States v. Maynard*, 615

acknowledged that "dragnet-type law enforcement practices," such as "twenty-four hour surveillance of any citizen of this country," might raise a Fourth Amendment problem while the use of a beeper did not.[167]

Not only is the quantity of information collected in the big data context far greater, the very nature of surveillance itself is different. If conventional surveillance involves the intentional tracking of one or a few suspects by actual police officers, what happens when a person "emerges" as a surveillance target as a result of a computer analysis? In the traditional surveillance context, the police have not been constrained by the Fourth Amendment so long as their investigations neither interfered with an individual's movements, nor ranged beyond public spaces.[168] As the Supreme Court has observed, there is no constitutional right to be free from police investigation.[169]

But this *surveillance discretion* may mean something different in the big data context. The intentional surveillance of targeted individuals is not equivalent to the perpetual "indiscriminate data collection"[170] of entire populations. While both approaches involve watching by the government, a program like the N.Y.P.D.'s "total domain awareness" system differs from traditional surveillance enough to warrant a different approach.[171] The very quality of public life may be different when government watches everyone—surreptitiously—and stores all of the resulting information.[172]

F.3d 544, 562 (D.C. Cir. 2010) ("Prolonged surveillance reveals types of information not revealed by short-term surveillance, such as what a person does repeatedly, what he does not do, and what he does ensemble. These types of information can each reveal more about a person than does any individual trip viewed in isolation."). *See, e.g.*, Orin S. Kerr, *The Mosaic Theory of the Fourth Amendment*, 111 MICH. L. REV. 311, 326 (2012) ("[F]ive justices wrote or joined opinions that . . . suggest that a majority of the Court is ready to embrace some form of the D.C. Circuit's mosaic theory.").

167. United States v. Knotts, 460 U.S. 276, 283–85 (1983).

168. *See, e.g.*, Terry v. Ohio, 392 U.S. 1, 19 n.16 (1968) ("Obviously, not all personal intercourse between policemen and citizens involves 'seizures' of persons. Only when the officer, by means of physical force or show of authority, has in some way restrained the liberty of a citizen may we conclude that a 'seizure' has occurred.").

169. Michigan v. Chesternut, 486 U.S. 567, 576 (1988) ("The police [are] not required to have 'a particularized and objective basis for suspecting [respondent] of criminal activity,' in order to pursue him." (quoting United States v. Cortez, 449 U.S. 411, 417–18 (1981))); *cf.* Oyler v. Boles, 368 U.S. 448, 456 (1962) ("[T]he conscious exercise of some selectivity in [law] enforcement is not in itself a federal constitutional violation.").

170. Stephen Rushin, *The Judicial Response to Mass Police Surveillance*, 2011 U. ILL. J.L. TECH. & POL'Y 281, 286.

171. *See* 1 LAFAVE, *supra* note 138, § 2.7(g) (raising similar concerns).

172. *See* United States v. Jones, __U.S.__, 132 S. Ct. 945, 956 (2012) (Sotomayor, J., concurring) ("Awareness that the Government may be watching chills associational and expressive freedoms.").

Practical barriers have long served to protect individual privacy by forcing the police to selectively apply their resources and interests,[173] but those barriers have now largely eroded.[174] The ability of government to record, store, and analyze nearly everything we do is now becoming technologically possible and affordable.[175] By 2015, it will cost just two cents to store all of the audio data generated by the average person in one year; storing a year's worth of a person's movements generated by their cellphone will cost next to nothing.[176] These expanded capabilities raise the possibility of a "surveillance time machine": the capacity of the government to identify a person of interest and then search retrospectively through all of the data that has been stored and collected about that person.[177] While some people have already changed their personal habits to avoid this mass surveillance, many likely have not.[178]

The longstanding doctrines declaring that we lack any Fourth Amendment protections in the public sphere should not hold its traditional force once the police deploy the tools of big data.[179] "Knowing exposure" suggests a degree of control over one's information that is lacking when the government is capable of recording and storing every small detail in perpetuity.[180] Thus the traditional assumptions about Fourth Amendment protections in public spaces,

---

173. *See, e.g.*, *id.* at 963 (Alito, J., concurring in judgment) ("In the pre-computer age, the greatest protections of privacy were neither constitutional nor statutory, but practical. Traditional surveillance for any extended period of time was difficult and costly and therefore rarely undertaken.").

174. *See* Scott Shane, *Data Storage Could Expand Reach of Surveillance*, N.Y. TIMES (Aug. 14, 2012), http://thecaucus.blogs.nytimes.com/2012/08/14/advances-in-data-storage-have-implications-for-government-surveillance/?_php=true&_type=blogs&_r=0.

175. *See id.*

176. *See id.*

177. *See* VILLASENOR, *supra* note 22, at 1. Wayne Logan has persuasively argued that such a capacity has exposed the need to distinguish between identity evidence used strictly for identity verification and that used for forensic investigation. *See* Logan, *supra* note 151, at 1581–93.

178. On the various ways in which people might protest the growing surveillance capacities of the government, see Elizabeth E. Joh, *Privacy Protests: Surveillance Evasion and Fourth Amendment Suspicion*, 55 ARIZ. L. REV. 997 (2013); *see also* VILLASENOR, *supra* note 22, at 7 (observing that the use of encryption, for instance, might attract *greater* government attention).

179. Kevin Bankston and Ashkan Soltani have demonstrated the enormous differences in cost between traditional and new surveillance methods. They estimate that the cost of a using a traditional covert five police car surveillance operation over 28 days—the days the government followed Antoine Jones—is "nearly *775 times more expensive* than the cost of using GPS." Kevin S. Bankston & Ashkan Soltani, *Tiny Constables and the Cost of Surveillance: Making Cents out of* United States v. Jones, 123 YALE L.J. ONLINE 334, 335 (2014), http://yalelawjournal.org/2014/1/9/bankston-soltani.html.

180. Katz v. United States, 389 U.S. 347, 351 (1967) ("What a person knowingly exposes to the public, even in his own home or office, is not a subject of Fourth Amendment protection.").

absent statutory protections from Congress, call out for reexamination and doctrinal adaptation.

C. Repurposing Information

Google's reuse of search terms to identify flu outbreaks represents an upending of a core research convention: formulate a hypothesis first, and then collect the appropriate data.[181] With big data, we can collect (nearly all) the data first, and apply the questions later.[182] Indeed, the data can be analyzed in multiple ways at multiple times.[183] It is this repurposing or resifting of data that has led to some of big data's unexpected insights, like Google's flu analysis.

When it is the police who sift through the data, however, the Fourth Amendment is ill-suited to this particular relationship of data collection and analysis. The Fourth Amendment is primarily interested in the legitimacy of *how* information is acquired.[184] If the acquisition is permissible, how the police use that information thereafter is generally not subject to an additional Fourth Amendment challenge.[185] This suggests that once legitimately within the government's possession, information can be repurposed and reanalyzed without any additional Fourth Amendment justification.[186] In the case of genetic information, courts have been generally dismissive of claims that individuals have any Fourth Amendment claims to DNA samples once lawfully acquired

---

181. *See supra* text accompanying notes 43–47.

182. *See supra* text accompanying notes 43–47.

183. *See* MAYER-SCHÖNBERGER & CUKIER, *supra* note 10, at 104 ("In the big-data age, data is like a magical diamond mine that keeps on giving long after its principal value has been tapped.").

184. *See, e.g.*, Russell D. Covey, *Pervasive Surveillance and the Future of the Fourth Amendment*, 80 MISS. L.J. 1289, 1294–95 (2011) ("Fourth Amendment law . . . has proved singularly inept at dealing with the technological revolution. . . . [This is because it] has purported to regulate and control the non-consensual governmental acquisition of information from individuals in the name of privacy protection.").

185. *See* Erin Murphy, *Back to the Future: The Curious Case of* United States v. Jones, 10 OHIO ST. J. CRIM. L. 325, 330–31 (2012) ("Current Fourth Amendment law emphasizes acquisition: how did the police acquire the DNA sample or financial record or biometric image? It cares little for what happens next—to what use that information is put.").

186. Thus in a case from the 1990s, the New York Court of Appeals rejected a Fourth Amendment challenge to the use of a DNA analysis to connect a defendant to a rape, although the warrant for the blood sample was approved with regard to a different case. People v. King, 663 N.Y.S.2d 610, 614 (N.Y. App. Div. 1997) ("It is also clear that once a person's blood sample has been obtained lawfully, he can no longer assert either privacy claims or unreasonable search and seizure arguments with respect to the use of that sample. Privacy concerns are no longer relevant once the sample has already lawfully been removed from the body, and the scientific analysis of a sample does not involve any further search and seizure of a defendant's person.").

by the police, but used for investigative purposes unrelated to the original justification for the sample's collection.[187]

Is a secondary analysis of an individual's DNA sample to find a familial match sufficiently similar to an analysis to find whether that same individual is responsible for another crime? Are there other sorts of information to be derived from DNA samples that ought to require distinct Fourth Amendment justifications? Repurposing a DNA sample to look for information regarding *someone other than the source of the sample* raises sufficient privacy concerns that some further government justifications may be necessary.[188] Such a search does more than "identify" again the source of the DNA sample in a subsequent police investigation.[189]

The government's ability to reanalyze information—of any sort—in the age of big data calls out for a new approach. What courts could do is shift the focus of the Fourth Amendment from data collection to a more rigorous scrutiny of its intended uses by the government.[190] Indeed, Harold Krent proposed nearly twenty years ago that the repurposing of information by the government obtained at an earlier time could be deemed unreasonable for Fourth Amendment purposes.[191] Professor Krent suggested, for instance, that courts might consider whether the seizure of a person's information would have been reasonable had the government articulated the later use initially.[192] The closer the

187. *See, e.g.*, State v. Hauge, 79 P.3d 131, 144 (Haw. 2003) ("[T]he appellate courts of several states have ruled that expectations of privacy in lawfully obtained blood samples . . . are not objectively reasonable by 'society's' standards. Specifically, a number of jurisdictions have held on analogous facts that once a blood sample and DNA profile is lawfully procured from a defendant, no privacy interest persists in either the sample or the profile."); State v. Emerson, 981 N.E.2d 787, 792–93 (Ohio 2012) (rejecting defendant's claims of privacy in subsequent uses of DNA *profile*); Smith v. State, 734 N.E.2d 706, 710 (Ind. Ct. App. 2000) ("[L]aw enforcement agencies may retain validly obtained DNA samples for use in subsequent unrelated criminal investigations . . . ."), *aff'd*, 744 N.E.2d 437 (Ind. 2001).

188. Kelly Lowenberg argues that subsequent searches of DNA samples that yield new information should require further government justification and a consideration of the reasonableness of that additional search. In the familial search context, Lowenberg suggests that Y-STR typing of a DNA sample of a convicted offender would be permissible without a warrant, while the same analysis conducted on another type of sample (e.g. a volunteer sample) would not. *See* Kelly Lowenberg, *Applying the Fourth Amendment When DNA Collected for One Purpose is Tested for Another*, 79 U. CIN. L. REV. 1289, 1319–23 (2011).

189. *Cf.* Logan, *supra* note 151, at 1586 (distinguishing evidence showing "one's identity (*who* one is), [from the] entirely different question . . . presented by identifying information (revealing *what* one might have done or perhaps will do") (emphasis in original) (footnote omitted)).

190. *See* Covey, *supra* note 184, at 1302.

191. Harold J. Krent, *Of Diaries and Data Banks: Use Restrictions Under the Fourth Amendment*, 74 TEX. L. REV. 49, 60–63 (1995).

192. *See id.* at 80–81.

government's secondary purpose is to its original purpose at the time of acquisition, the more likely it should be that the government could use the data without further justification.

The case for Fourth Amendment protections regarding repurposed information is stronger still should the government one day be interested in gleaning information from DNA samples other than matching profiles to crime scene samples. Here, the Supreme Court has hinted at a willingness to reassess the balance of privacy and government utility at some future date. In *Maryland v. King*,[193] in which the Court upheld the compulsory collection of DNA from arrestees,[194] Justice Kennedy suggested that "[i]f in the future police analyze [DNA] samples to determine [other information], that case would present additional privacy concerns not present here."[195] The resolution by the Court regarding such a dispute may well turn, however, on the purposes claimed by the government to mine that information. In *King*, the Court was willing to permit defendant's cheek swab, and the subsequently generated DNA profile, without individualized suspicion because the police were permitted to find out King's "identity": a term broad enough to encompass any other crimes King had committed.[196]

For now, however, the Court has left open the possibility that it may give greater scrutiny to some sorts of repurposing. That, plus the existing statutory protections on access and disclosure, may allay the concerns of many.[197] Yet it would be overly optimistic to ignore two developments in the other direction: the trend of Fourth Amendment law away from protection in these secondary searches, and the Court's recent expansion of what the government may do for purposes of "identification" when it comes to genetic information.[198]

---

193. __U.S.__, 133 S. Ct. 1958 (2013).

194. *Id.* at 1980.

195. *Id.* at 1979.

196. *Id.* at 1980. Justice Scalia's dissent in *King* was much less sanguine about the threats to privacy in the case, and strongly disputed that the government's interest in the case could be justified as one of "identification." *See id.* at 1988 (Scalia, J., dissenting) (noting "it may one day be possible to design a program that uses DNA for a purpose other than crime-solving").

197. *Cf.* United States v. Jones, __U.S.__, 132 S. Ct. 945, 964 (2012) (Alito, J., concurring in judgment) ("A legislative body is well situated to gauge changing public attitudes, to draw detailed lines, and to balance privacy and public safety in a comprehensive way.").

198. In the *King* case, the majority comfortably found that arrestee DNA profiles could be used to link the defendant to a crime unrelated to the crime of arrest, *King*, 133 S. Ct. at 1965, 1970–80, a definition of "identification" to which Justice Scalia dissented, dramatically. *See id.* at 1982–90 (Scalia, J., dissenting). For further discussion of this issue, see Joh, *supra* note 111.

D. Beyond the Fourth Amendment

Apart from the Fourth Amendment challenges raised by big data policing, an uncritical embrace of these new technologies raises other concerns beyond regulating the police. Whether practical or abstract, these concerns will be easily swept aside by departments eager to be part of the next technological wave in policing.

First, many of these new technologies have been developed by private companies whose motivations and concerns may not always be consonant with those of a public police department. For instance, IBM has spent billions acquiring data analytics companies in order to develop and market predictive tools to the police.[199] Although PredPol was initially developed by academics, it is now a for-profit company.[200] Similarly, Microsoft—and the N.Y.P.D.—will profit from every new police department that adopts a total domain awareness system.[201] Future interest in the further analysis of DNA samples will also benefit some private laboratories.

Second, the introduction of new big data technologies requires attention not only to appropriate regulation, but also to questions about how well these privately developed tools actually help to reduce crime. New technologies possess understandable appeal for departments seeking innovative crime fighting strategies. New strategies lend themselves toward positive media attention in a way that "a poorly attended community meeting in a church basement" does not.[202] Yet, for-profit purveyors of big data products may not provide the best objective assessment of their products. The desirability of these new technologies should not steer attention away from questions about how well they reduce crime and conserve limited public resources compared to traditional methods.

A final concern is much more fundamental. The reliance on big data

---

199. *See* Sengupta, *supra* note 5. Indeed, to the extent that these companies may market both to public police departments and private corporations interested in reducing crime privately, special attention must be paid to claims of public benefit. For further discussion on how private interests can distort public police goals, see Elizabeth E. Joh, *The Forgotten Threat: Private Policing and the State*, 13 IND. J. GLOBAL LEGAL STUD. 357, 384–88 (2006).

200. *See, e.g.*, Bond-Graham & Winston, *supra* note 64 (noting that PredPol incorporated in January 2012 and "has emerged early to dominate the [predictive policing] market").

201. The N.Y.P.D. is said to receive thirty percent of gross revenues from sales of the system to other departments. Sam Roberts, *Police Surveillance May Earn Money for City*, N.Y. TIMES, Apr. 4, 2013, at A23.

202. David Alan Sklansky, *The Persistent Pull of Police Professionalism*, NEW PERSPECTIVES IN POLICING (Harvard Kennedy Sch., Cambridge, Mass. & Nat'l Inst. of Just.), Mar. 2011, at 9, *available at* https://www.ncjrs.gov/pdffiles1/nij/232676.pdf.

by the police also poses the risk that the very definition of policing may be changing. The promise of big data is a vision of policing that is driven and assessed by quantitative measurements. Indeed, those police chiefs that have already embraced big data tout the potential to rely on numbers when budgets for police departments are shrinking.[203] The problem, however, is that a technocratic solution to crime is not the only objective of democratic policing.[204]

Reducing crime is not the only job of the police. Policing as an institution has never been amenable to a single objective,[205] and indeed over time its aims have shifted.[206] What is clear, however, is that democratic policing aims at more than mere crime control and, at its core, relies on skills that do not always lend themselves to statistical analysis. No amount of data-driven policing is likely to assuage communities soured by long histories of tension with the police. Nor will demonstrations of little red boxes on a smartphone necessarily justify to a community the need for a heavy-handed police presence.

CONCLUSION

The use of big data is likely to become an ordinary aspect of policing. The application of artificial intelligence to crime data promises immediate and tangible benefits. We can gain some real insights about how to direct police resources efficiently and effectively in ways that intuition, tradition, and limited information have been unavailing. At the same time, the reliance upon artificial intelligence and the collection of vast amounts of information poses some special challenges in the policing context. Courts and legislatures will need to think of Fourth

---

203. *See, e.g.*, Charlie Beck & Colleen McCue, *Predictive Policing: What Can We Learn from Wal-Mart and Amazon about Fighting Crime in a Recession?*, POLICE CHIEF (Nov. 2009), http://www.policechiefmagazine.org/magazine/index.cfm?fuseaction=display_arch&article_id=1942&issue_id=112009 (arguing that "predictive policing represents an opportunity to prevent crime and respond more effectively, while optimizing increasingly scarce or limited resources, including personnel") (Charlie Beck is the Chief of Detectives for the L.A.P.D.).

204. *Cf.* Sklansky, *supra* note 202, at 9–10 ("A fixation on technology can distract attention from the harder and more important parts of [policing], the parts that rely on imagination and judgment.").

205. Perhaps the ambiguities of policing was best stated by sociologist Egon Bittner, who described the job of policing as: "a mechanism for the distribution of non-negotiably coercive force employed in accordance with the dictates of an intuitive grasp of situational exigencies." *See* EGON BITTNER, THE FUNCTIONS OF THE POLICE IN MODERN SOCIETY: A REVIEW OF BACKGROUND FACTORS, CURRENT PRACTICES, AND POSSIBLE ROLE MODELS 46 (1970).

206. *See, e.g.*, Eric H. Monkkonen, *History of Urban Police*, 15 CRIME & JUST. 547, 555 (1992) (observing that early in American policing history the police were expected to dole out social services to the city's needy).

Amendment issues in new ways to adequately protect notions of individual privacy.