Illuminating Black Data Policing

Andrew Guthrie Ferguson*

INTRODUCTION

The future of policing will be driven by data. Crime, criminals, and patterns of criminal activity will be reduced to data to be studied, crunched, and predicted.[1] Police departments across the United States—like the civilian population—will learn to adapt to ever-shifting technological innovations and efficiencies.[2] The question of adoption is not "if," but "when," and any delay largely will be a function of money and police culture.

The benefits of big data policing involve smarter policing, faster investigation, predictive deterrence, and the ability to visualize crime problems in new ways.[3] Not surprisingly then, police administrators have been seeking out new partnerships with sophisticated private data companies and experimenting with new surveillance technologies.[4] In Chicago, Los Angeles, New York City, Miami, Boston, and other smaller cities and towns, the beginning of a big data policing mindset is developing.[5]

This potential future, however, has a very present limitation. It is a limitation largely ignored by adopting jurisdictions and could, if left unaddressed,

---

* Professor of Law, UDC David A. Clarke School of Law. Thank you to Ric Simmons for the invitation to present in The Ohio State University Moritz College of Law Round Table on Big Data and Criminal Law.

1 *See generally* Andrew Guthrie Ferguson, *Policing Predictive Policing*, 94 WASH. U. L. REV. 1113 (2017); Ric Simmons, *Quantifying Criminal Procedure: How to Unlock the Potential of Big Data in Our Criminal Justice System*, 2016 MICH. ST. L. REV. 947 (2016).

2 *See* Charlie Beck & Colleen McCue, *Predictive Policing: What Can We Learn from Wal-Mart and Amazon About Fighting Crime in a Recession?*, POLICE CHIEF (Mar. 13, 2014), http://www.policechiefmagazine.org/predictive-policing-what-can-we-learn-from-wal-mart-and-amazon-about-fighting-crime-in-a-recession/ [http://perma.cc/D6HM-GA2T].

3 *See generally* Sarah Brayne, *Big Data Surveillance: The Case of Policing*, 82 AM. SOC. REV. 977 (2017); Andrew Guthrie Ferguson, *Big Data and Predictive Reasonable Suspicion*, 163 U. PA. L. REV. 327 (2015); Elizabeth E. Joh, *Policing by Numbers: Big Data and the Fourth Amendment*, 89 WASH. L. REV. 35 (2014); Andrew D. Selbst, *Disparate Impact in Big Data Policing*, 52 GA. L. REV. 109 (2017).

4 *See, e.g.*, Mark Harris, *How Peter Thiel's Secretive Data Company Pushed into Policing*, WIRED (Aug. 9, 2017, 9:40 AM), https://www.wired.com/story/how-peter-thiels-secretive-data-company-pushed-into-policing/ [https://perma.cc/G8NP-N7PL]; Jacques Peretti, *Palantir: The 'Special Ops' Tech Giant that Wields as Much Real-World Power as Google*, GUARDIAN (July 30, 2017, 9:59 AM), https://www.theguardian.com/world/2017/jul/30/palantir-peter-thiel-cia-data-crime-police [https://perma.cc/F493-SE4J].

5 *See* Ferguson, *supra* note 1, at 1115–17.

delegitimize the adoption and use of new data-driven technologies. Simply put: all big data policing technologies have a "black data" problem.[6]

As I have written previously, "black data" connotes three overlapping concerns.[7] First, big data policing is opaque, lacking transparency because most of the magic happens as a result of "black box" proprietary and mathematically complex algorithms.[8] Second, big data policing is racially encoded, colored by the history of real-world policing that disproportionality impacts communities of color.[9] Police data comes from the real world, and all of the long-standing discriminatory impacts of implicit and explicit bias color that data.[10] Black data is black, brown, and marked by disproportionate impacts on communities of color.[11] Finally, big data policing faces legal uncertainty as old constitutional doctrines built on small data principles no longer work in the new big data age.[12] The future path of traditional Fourth Amendment law is uncertain, dark, and distorted. These different types of darkness make it difficult to see the future clearly. Black data must be illuminated so that the positive elements of algorithmic insights and crime prevention can be used without negatively impacting privacy, liberty, or the well-being of citizens subject to new forms of police surveillance.

To work as intended, big data policing technologies must address this lack of transparency, the legacy of racial discrimination, and the constitutional uncertainty arising from application in the real world. This brief essay, part of a symposium on *Big Data and Criminal Law*, seeks to raise the questions that must be resolved to overcome the black data problem. This symposium essay examines: (1) the puzzle of data-driven transparency; (2) the concern of biased data; and (3) the struggle for constitutional clarity in the face of new technologies.

I. THE GROWTH OF BIG DATA POLICING

To ground our discussion, it is important to be clear about the type of technologies which might fall under the big data policing tent.[13] At present, there are three broad categories of technology: predictive technologies, surveillance technologies, and data mining technologies—all are data-driven innovations which

---

[6] ANDREW GUTHRIE FERGUSON, THE RISE OF BIG DATA POLICING: SURVEILLANCE, RACE, AND THE FUTURE OF LAW ENFORCEMENT 3–4, 131 (2017).

[7] *Id.* at 3.

[8] *Id.* at 136–40.

[9] *Id.* at 131–36.

[10] *Id.*

[11] *Id.*

[12] *Id.* at 140–42.

[13] *See* Ferguson, *supra* note 3, at 352–73.

have been adopted by law enforcement to augment police strategies of deterrence, monitoring, and investigation.[14]

First, predictive analytics are being used by law enforcement in the form of place-based predictive policing and person-based targeting.[15] Both of these strategies share the same core approach that algorithms can be developed to identify risk patterns from collected police data.[16] For example, place-based predictive policing might take historic crime data, or environmental risk factors, or some combination thereof, and combine it with other variables (day of the week, time, weather, etc.) to forecast crimes.[17] The data can be fed into a computer model that seeks to identify geographic areas that may be more at risk of crime than other areas, either because of crime patterns, risks in the environment, or some other factor.[18] The outputs—usually maps of micro-areas of forecast risk—can be provided to police officers as they patrol, thus allowing police to effectively respond to crime patterns in a city.[19] Person-based targeting involves using past data about criminal activities (arrests, age at the time of offense, criminal associates, past violence, etc.) and then using that information to predict those individuals most at risk of violence (either as victims or perpetrators).[20] In places like Chicago, Illinois, the Strategic Subjects List (the so-called "heat list") of

---

[14] *See* Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C. L. REV. 93, 104–05 (2014) (predictive analytics); Rachel Levinson-Waldman, *Hiding in Plain Sight: A Fourth Amendment Framework for Analyzing Government Surveillance in Public*, 66 EMORY L.J. 527, 534 (2017) (surveillance); Steven D. Seybold, Note, *Somebody's Watching Me: Civilian Oversight of Data-Collection Technologies*, 93 TEX. L. REV. 1029, 1032 (2015) (surveillance); Tal Z. Zarsky, *Governmental Data Mining and Its Alternatives*, 116 PENN ST. L. REV. 285, 287 (2011) (data mining).

[15] *Predictive Policing*, U.S. DEP'T OF JUSTICE, http://www.nij.gov/topics/law-enforcement/strategies/predictive-policing/Pages/welcome.aspx [https://perma.cc/525K-ZGMX] (last modified June 9, 2014).

[16] *See* Ferguson, *supra* note 1, at 1126–28.

[17] *See* AZAVEA, HUNCHLAB: UNDER THE HOOD (2015), https://cdn.azavea.com/pdfs/hunchlab/HunchLab-Under-the-Hood.pdf [https://perma.cc/RPT4-5JDM]; *How Predictive Policing Works*, PREDPOL, https://www.predpol.com/how-predictive-policing-works/ [https://perma.cc/84TG-G6EF] (last visited Mar. 25, 2018); RUTGERS CTR. ON PUB. SEC., RISK TERRAIN MODELING COMPENDIUM (Joel M. Caplan & Leslie W. Kennedy eds., 2011).

[18] *See* G.O. Mohler et al., *Randomized Controlled Field Trials of Predictive Policing*, 110 J. AM. STAT. ASS'N 1399 (2015); Aaron Mendelson, *Can LAPD Anticipate Crime with 'Predictive Policing'?*, CAL. REP. (Sept. 9, 2013), https://www.scpr.org/programs/take-two/2013/09/09/33630/can-lapd-anticipate-crime-with-predictive-policing/ [https://perma.cc/ZQS2-2ZJ2].

[19] *See* Tessa Stuart, *Santa Cruz's Predictive Policing Experiment*, SANTACRUZ.COM (Feb. 14, 2012), http://www.santacruz.com/news/santa_cruzs_predictive_policing_experiment.html [https://perma.cc/U2YD-VPYC].

[20] Jeremey 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 [https://perma.cc/GKJ7-29LQ].

targeted individuals ranks the people deemed most at-risk in the city.[21] Police and social service interventions can be planned around this list as sort of a public health strategy to prevent future violence.[22] In addition, the numerical risk score is displayed on the dashboard of police computers, advising police officers of the relative risk of the individual they happen to be stopping.[23]

Surveillance technologies in the form of monitoring advancements can watch, hear, smell, sense, and automatically record and respond to things occurring in public.[24] In lower Manhattan, a growing network of 9,000 linked cameras provide real-time video to a centralized command center.[25] Automated programs search for faces, license plates, and other suspicious activity.[26] In Baltimore, traditional surveillance cameras are augmented by aerial cameras and sensors.[27] Entire neighborhoods have been filmed for days by private companies with the surveillance data searchable for particular crimes which later can be turned over to police.[28] In other cities, police body cameras provide days' worth of images

---

[21] Editorial Board, *Who Will Kill or Be Killed in Violence-Plagued Chicago? The Algorithm Knows*, CHI. TRIB. (May 10, 2016, 5:00 PM), http://www.chicagotribune.com/news/opinion/editorials/ct-gangs-police-loury-algorithm-edit-md-20160510-story.html [https://perma.cc/ZMV4-8A89]; Gorner, *supra* note 20; Mark Guarino, *Can Math Stop Murder?*, CHRISTIAN SCI. MONITOR (July 20, 2014), https://www.csmonitor.com/USA/2014/0720/Can-math-stop-murder [https://perma.cc/4XZC-NBLP].

[22] Andrew V. Papachristos, *CPD's Crucial Choice: Treat Its List as Offenders or as Potential Victims?*, CHI. TRIB. (July 29, 2016, 10:00 AM), http://www.chicagotribune.com/news/opinion/commentary/ct-gun-violence-list-chicago-police-murder-perspec-0801-jm-20160729-story.html [https://perma.cc/MTH5-AHE5].

[23] Josh Kaplan, *Predictive Policing and the Long Road to Transparency*, SOUTH SIDE WKLY. (July 12, 2017), https://southsideweekly.com/predictive-policing-long-road-transparency/ [https://perma.cc/XE7F-2G4L].

[24] *See generally* Bennett Capers, *Crime, Surveillance, and Communities*, 40 FORDHAM URB. L.J. 959 (2013); CHRISTOPHER SLOBOGIN, PRIVACY AT RISK: THE NEW GOVERNMENT SURVEILLANCE AND THE FOURTH AMENDMENT (2007); Christopher Slobogin, *Public Privacy: Camera Surveillance of Public Places and the Right to Anonymity*, 72 MISS. L.J. 213 (2002).

[25] Thomas H. Davenport, *How Big Data is Helping the NYPD Solve Crimes Faster*, FORTUNE (July 17, 2016), http://fortune.com/2016/07/17/big-data-nypd-situational-awareness/ [https://perma.cc/4EAP-NLCU]; *see also* 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 [https://perma.cc/78CW-PVJ4].

[26] *See* AOL, *Digisensory Technologies Avista Smart Sensors*, YOUTUBE (Sept. 14, 2012), www.youtube.com/watch?v=JamGobiS5wg [https://perma.cc/NTK7-BGY7]; Associated Press, *NJ City Leading Way in Crime-Fighting Tech*, CBS NEWS (June 19, 2010, 9:30 AM), https://www.cbsnews.com/news/nj-city-leading-way-in-crime-fighting-tech/ [https://perma.cc/GFG4-5YY5].

[27] Monte Reel, *Secret Cameras Record Baltimore's Every Move from Above*, BLOOMBERG BUSINESSWEEK (Aug. 23, 2016), https://www.bloomberg.com/features/2016-baltimore-secret-surveillance/ [https://perma.cc/ZAB4-2CMD].

[28] Craig Timberg, *New Surveillance Technology Can Track Everyone in an Area for Several Hours at a Time*, WASH. POST (Feb. 5, 2014), https://www.washingtonpost.com/business/technology/new-surveillance-technology-can-track-everyone-in-an-area-for-several-hours-at-a-

uploaded to large data warehouses.[29] In Chicago, audio sensors record gunshots.[30] In Washington D.C., chemical sensors record suspicious substances.[31] All across the country, cell signals and digital clues held by third parties can be requested or even directly intercepted by IMSI ("stingray") devices.[32] All across the internet landscape, social media can be watched and monitored for criminal activity.[33] All of this data, big and small, usable and quite revealing, can be monitored by law enforcement.

Third, growing data collection capabilities have provided incentives to create new search technologies to interrogate the information. Federal law enforcement entities have developed massive criminal justice databases filled with personal information.[34] States have developed smaller versions.[35] Biometric databases now include information about DNA, fingerprints, iris scans, tattoos, gait, facial recognition, and other physical markers.[36] These government databases have been

---

time/2014/02/05/82f1556e-876f-11e3-a5bd-844629433ba3_story.html?utm_term=.145394dbdb2b [https://perma.cc/6SJN-HK6P].

[29] Matt Stroud, *The Company That's Livestreaming Police Body Camera Footage Right Now*, MOTHERBOARD (July 27, 2016, 6:00 AM), https://motherboard.vice.com/en_us/article/9a3ddv/visual-labs-police-body-camera-livestream [https://perma.cc/5782-MSGR].

[30] Patrick M. O'Connell, *Chicago Police Announce Expanded Technology to Curb Shootings*, CHI. TRIB. (Jan. 27, 2017, 4:59 PM), http://www.chicagotribune.com/news/local/breaking/ct-chicago-police-shotspotter-technology-met-20170127-story.html [https://perma.cc/TX99-VFCQ].

[31] Michael J. Penders & William L. Thomas, *Ecoterror: Rethinking Environmental Security After September 11*, 16 NAT. RESOURCES & ENV'T 159, 160 (2002).

[32] Stephanie K. Pell & Christopher Soghoian, *A Lot More Than a Pen Register, and Less Than a Wiretap: What the StingRay Teaches Us About How Congress Should Approach the Reform of Law Enforcement Surveillance Authorities*, 16 YALE J.L. & TECH. 134, 142 (2013).

[33] Ben Austen, *Public Enemies: Social Media Is Fueling Gang Wars in Chicago*, WIRED (Sept. 17, 2013, 6:30 AM), http://www.wired.com/underwire/2013/09/gangs-of-social-media/ [https://perma.cc/C2D4-ADXN]; Joseph Goldstein & J. David Goodman, *Seeking Clues to Gangs and Crime, Detectives Monitor Internet Rap Videos*, N.Y. TIMES (Jan. 7, 2014), https://www.nytimes.com/2014/01/08/nyregion/seeking-clues-to-gangs-and-crime-detectives-monitor-internet-rap-videos.html [https://perma.cc/UY78-VTX7]; Heather Kelly, *Police Embrace Social Media as Crime-Fighting Tool*, CNN (Aug. 30, 2012, 5:23 PM), http://www.cnn.com/2012/08/30/tech/social-media/fighting-crime-social-media [https://perma.cc/DAU2-KR5L].

[34] Erin Murphy, *Databases, Doctrine, and Constitutional Criminal Procedure*, 37 FORDHAM URB. L.J. 803, 806–08 (2009); BUREAU OF JUSTICE ASSISTANCE, U.S. DEP'T OF JUSTICE, FUSION CENTER GUIDELINES 2 (2006), https://it.ojp.gov/documents/fusion_center_guidelines_law_enforcement.pdf [https://perma.cc/8U4Q-6KHK].

[35] Stephen Mercer & Jessica Gabel, *Shadow Dwellers: The Underregulated World of State and Local DNA Databases*, 69 N.Y.U. ANN. SURV. AM. L. 639, 681 (2014).

[36] Laura K. Donohue, *Technological Leap, Statutory Gap, and Constitutional Abyss: Remote Biometric Identification Comes of Age*, 97 MINN. L. REV. 407, 436–38 (2012); Ellen Nakashima, *FBI Prepares Vast Database of Biometrics*, WASH. POST (Dec. 22, 2007), http://www.washingtonpost.com/wp-dyn/content/article/2007/12/21/AR2007122102544.html [https://perma.cc/QED3-STBY]; Wayne A. Logan, *Policing Identity*, 92 B.U. L. REV. 1561, 1575 & n.91 (2012); Craig Timberg & Ellen Nakashima, *State Photo-ID Databases Become Troves for Police*, WASH. POST (June 16, 2013), https://www.washingtonpost.com/business/technology/state-photo-id-databases-

augmented on occasion by private data originally collected by for-profit data-broker companies.[37] Like other consumers, law enforcement can purchase the available data to build larger systems.[38] All of this data can be analyzed using powerful computers and algorithms to divine hidden patterns and clues.[39] Crimes can be solved, suspects can be identified, and unseen connections revealed.

All of these technologies rely on data collection and analysis, with varying degrees in the amount of data assembled. For example, place-based predictive policing companies use collected crime statistics, which in large cities involves a significant number of ever-changing data points (depending on the technology used).[40] However, *person*-based predictive policing uses fewer inputs to identify at-risk individuals, usually limited to traditional criminal justice inputs (arrests, convictions, etc.).[41] While data collection may target all of the people arrested in a big city like Chicago, the types of data are not terribly extensive (especially compared to other big data projects). Surveillance technologies can encompass a tremendous amount of information if one thinks about the accumulated digital footage of all police-worn body cameras, or all networked surveillance cameras, or all cell site and other digital surveillance technologies. In addition, data mining usually involves vast amounts of collected data, but again the definition of "vast" compared to other types of datasets in a big data world may not be too

---

become-troves-for-police/2013/06/16/6f014bd4-ced5-11e2-8845-d970ccb04497_story.html?utm_term=.c9c85764f6f8 [https://perma.cc/BK9S-T3YH]; Sara Reardon, *FBI Launches $1 Billion Face Recognition Project*, NEW SCIENTIST (Sept. 7, 2012), https://www.newscientist.com/article/mg21528804-200-fbi-launches-1-billion-face-recognition-project/ [https://perma.cc/5ZSC-W78C].

[37] Chris Jay Hoofnagle, *Big Brother's Little Helpers: How ChoicePoint and Other Commercial Data Brokers Collect and Package Your Data for Law Enforcement*, 29 N.C. J. INT'L L. & COM. REG. 595 (2004); Andrea Peterson, *Your Location History Is Like a Fingerprint. And Cops Can Get It Without a Warrant.*, WASH. POST (July 31, 2013), https://www.washingtonpost.com/news/the-switch/wp/2013/07/31/your-location-history-is-like-a-fingerprint-and-cops-can-get-it-without-a-warrant/?utm_term=.32d84715062a [https://perma.cc/JW46-RLA2].

[38] Bob Sullivan, *Who's Buying Cell Phone Records Online? Cops*, NBC NEWS (June 20, 2006, 11:59 AM), http://www.msnbc.msn.com/id/12534959/ [https://perma.cc/Q5MX-EKMT]; Hoofnagle, *supra* note 37, at 597.

[39] *See* Christopher Slobogin, *Government Data Mining and the Fourth Amendment*, 75 U. CHI. L. REV. 317, 322–23 (2008); Christopher Slobogin, *Transaction Surveillance by the Government*, 75 MISS. L.J. 139, 167–82 (2005).

[40] DAVID ROBINSON & LOGAN KOEPKE, UPTURN, STUCK IN A PATTERN: EARLY EVIDENCE ON "PREDICTIVE POLICING" AND CIVIL RIGHTS 3–5 (2016), https://www.teamupturn.org/reports/2016/stuck-in-a-pattern [https://perma.cc/PH9H-X48T]. However, some companies use only a very limited dataset. For example, PredPol, a leading predictive policing company only uses three data points (crime type, time, and location). PREDPOL, *supra* note 17.

[41] Kaplan, *supra* note 23 (this information might include things like: "1. Age at most recent arrest (the younger the age, the higher the score); 2. Incidents where victim of a shooting; 3. Incidents where victim of an assault or battery; 4. Violent crime arrests; 5. Unlawful use of weapons arrests; 6. Narcotics arrests (Wernick claimed that this is the least impactful variable, and does not seem to matter that much to the model); 7. Trend in criminal activity (essentially whether or not an individual's rate of criminal activity is increasing or decreasing)").

overwhelming.[42] A biometric search of DNA in the FBI CODIS database is a search of about 11.6 million records, which covers a lot of people, but is a relatively small number for some big data systems.[43] More expansive biometric databases are being collected to search everything from iris patterns to tattoos, which increases the variables, but still remains much smaller than other big data projects.[44]

The point of this overview is that big data policing provides a broad definition of new law enforcement technologies which share certain similarities. These technologies are digital with varying (but increasing) capabilities to collect, store, and analyze the information. And, for purposes of this essay, because these technologies emerge from the real world of policing, they all share the same big impediment: the problem of black data. As this essay seeks to offer insights for the future of criminal law, big data, and the promotion of justice, the next sections raise questions that must be answered on the theme of black data.

II. THE BLACK DATA PROBLEM

Big data policing is the future of law enforcement. But, it is a future that has yet to confront the overlapping problems of transparency, racial bias, and constitutional distortion. The solutions to the black data problem are not simple or easy, but are necessary to engage in order to create a system that will be trusted by communities and police alike. This Part seeks to examine the concerns that give rise to the black data problem in the hopes that by exposing the issues, solutions will emerge *before* wide-scale adoption of these new technologies.

A. *Transparency: Black Data Is Opaque*

Big data policing has a transparency problem. By and large, ordinary citizens cannot peer behind the algorithms that now control police patrol routes or human target lists.[45] More complicatedly, in some cases the code itself cannot be interrogated, because of how machine learning and artificial intelligence systems work.[46] These machines learn from past data without human input, and thus cannot simply be examined to reveal the code.[47] Similarly, because of tactical and

---

[42] VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK 9 (2013).

[43] Margaret Hu, *Small Data Surveillance v. Big Data Cybersurveillance*, 42 PEPP. L. REV. 773, 794–99, 821 (2015).

[44] Donohue, *supra* note 36, at 415.

[45] *See* Stuart, *supra* note 19; Nissa Rhee, *Study Casts Doubt on Chicago Police's Secretive "Heat List,"* CHI. MAG. (Aug. 17, 2016), http://www.chicagomag.com/city-life/August-2016/Chicago-Police-Data/ [https://perma.cc/5Z7E-F77E]; Simmons, *supra* note 1, at 994.

[46] Selbst, *supra* note 3.

[47] Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 679–80 (2017).

safety considerations, police are reluctant to reveal the locations and use of surveillance technologies. From a law enforcement perspective, it does little good to expose the existence of surveillance schemes so that anyone can evade them.

These three transparency problems involving (1) technical barriers, (2) technological barriers, and (3) tactical barriers must be overcome for big data policing to be a success. Unless police address the transparency problems, communities will not trust the systems. And, without solutions to the transparency problem, police will have no way to gain that trust.

1. Technical Barriers

At a very basic level, the use of highly-technical systems undermines transparency. Most police administrators and officers are not computer scientists and must remain largely dependent on private companies to provide technical guidance.[48] In practice, this means that police purchase big data technologies without the ability to interrogate them or even understand them. For patrol officers on the street, this means blindly following predictive policing patrols without the ability to challenge the findings or deconstruct its assumptions. For administrators, it means trusting the algorithm based on the theory (or perhaps the result), but without being able to articulate why the system works.[49] This lack of transparency is even more problematic when applied to predictive policing of individuals. A police officer can see the result of the heightened "risk score," but cannot really explain how the score was calculated. As more and more bits of data get inputted into a system, the more complicated it can be to visualize or explain the outputs. For all intents and purposes, the data systems are dark to the end users and the community.

Similar problems exist with automated surveillance monitoring technologies.[50] Video systems pre-programed to alert to suspicious bags, audio sensors, or facial recognition matches are susceptible to the limitations of the programming.[51] While the systems themselves are transparent about the

---

[48] Elizabeth E. Joh, *The Undue Influence of Surveillance Technology Companies on Policing*, 92 N.Y.U. L. REV. 101, 125, 130 (2017).

[49] Some companies have countered this concern by specifically addressing the "why" of why a predictive technology works. Risk Terrain Modeling is one technology that focuses on the why. *See* Samantha Melamed, *Can Atlantic City's Bold Experiment Take Racial Bias out of Predictive Policing?*, PHILA. INQUIRER (Aug. 10, 2017, 5:03 AM), http://www.philly.com/philly/news/crime/atlantic-city-risk-terrain-modeling-rutgers-predictive-policing-joel-caplan-20170810.html [https://perma.cc/6E8Y-7EVH].

[50] Michael L. Rich, *Machine Learning, Automated Suspicion Algorithms, and the Fourth Amendment*, 164 U. PA. L. REV. 871, 898–901 (2016).

[51] For example, Shotspotter technology designed to identify gun shots can also mistakenly identify fireworks. Peter Nickeas et al., *Chicago Police Express Frustration After More Than 100 Shot in Violent Fourth of July Weekend*, CHI. TRIB. (July 6, 2017, 6:30 AM), http://www.chicagotribune.com/news/local/breaking/ct-chicago-july-4-weekend-shootings-violence-20170705-story.html [https://perma.cc/4R3G-CCF8]. Or threat scores that are designed to identify dangerous

information they are providing to police, there is little ability to determine if the automated alert will be accurate. So, for example, in a system that has been programmed to automatically identify "hand-to-hand transactions" on the street (to signify a potential drug deal),[52] the officers may end up responding to an alert without the ability to understand why the computer alerted to a particular place or action. Officers must blindly trust the system without being able to see or interpret the alerts.

Even something as simple as a static database has technical issues in terms of being able to correct or audit the information. As criminal justice databases have grown, there has been a parallel recognition that errors go uncorrected, systems go unexamined, and injustices result.[53] Much of the problem is that ordinary people cannot see within the system, and even experts cannot fix data that is shared among interconnected systems.[54] So, an individual may not know about the database error, but even if they did learn about it, a systems professional may not be able to correct the errors among the proliferating and interconnected systems that have been designed in a largely ad hoc manner. Once data gets corrupted, it is very hard to see or cleanse the error throughout the various systems.[55]

2. Technological Barriers

Beyond technical literacy, the very nature of big data technology thwarts transparency. First, at a most basic level, computer code and algorithms hide the inner workings of the systems.[56] In addition, many predictive policing systems are commercial operations, being privately owned and sold to law enforcement by companies that wish to keep their trade secrets private.[57] Naturally, in order to protect the value of their commercial technology, companies try to keep the

---

households can erroneously flag an innocent house. Justin Jouvenal, *The New Way Police Are Surveilling You: Calculating Your Threat 'Score,'* WASH. POST (Jan. 10, 2016), https://www.washingtonpost.com/local/public-safety/the-new-way-police-are-surveilling-you-calculating-your-threat-score/2016/01/10/e42bccac-8e15-11e5-baf4-bdf37355da0c_story.html?utm_term=.682fa950438e [https://perma.cc/5QQT-XY86]; CLARE GARVIE, ALVARO BEDOYA & JONATHAN FRANKLE, GEO. L. CTR. ON PRIVACY & TECH., THE PERPETUAL LINE-UP: UNREGULATED POLICE FACIAL RECOGNITION IN AMERICA (2016), www.perpetuallineup.org [https://perma.cc/L83J-MULS].

[52] *See* AOL, *supra* note 26.

[53] Wayne A. Logan & Andrew G. Ferguson, *Policing Criminal Justice Data*, 101 MINN. L. REV. 541, 541–44, 568–71 (2016).

[54] Alex R. Hess, Note, Herring v. United States: *Are Errors in Government Databases Preventing Defendants from Receiving Fair Trials?*, 11 J. HIGH TECH. L. 129, 147 (2010).

[55] Logan & Ferguson, *supra* note 53, at 588–91.

[56] *See* Paul Ford, *What* is *Code?*, BLOOMBERG BUSINESSWEEK (June 11, 2015), https://www.bloomberg.com/graphics/2015-paul-ford-what-is-code/ [https://perma.cc/C99E-PN3Y].

[57] *See* Joh, *supra* note 48, at 119; Ellen Huet, *Server and Protect: Predictive Policing Firm PredPol Promises to Map Crime Before It Happen*s, FORBES (Feb. 11, 2015, 6:00 AM), https://www.forbes.com/sites/ellenhuet/2015/02/11/predpol-predictive-policing/#79a6c0e94f9b [https://perma.cc/TTP9-SN67].

systems hidden.[58] Finally, the systems are complex, vacuuming up vast amounts of data every day and trying to process this information for practical use. In the real world of stopping and solving crime, there is less room for reflection or auditing of the data. While, of course, one would hope such accountability checks would be put in place, in practice, the pressures of daily data collection and analysis make such retrospective evaluations less of a priority than stopping the next crime.

In the future, as more systems become more dependent on machine learning and artificial intelligence, transparency will become even more difficult. With newer big data models, the systems will be designed to teach themselves through artificial intelligence, such that the model will constantly evolve based on continuous learning.[59] In such a system, the concept of transparency becomes almost meaningless because one cannot interrogate the past system for future insight, because the past system has already changed.[60] In addition, the creators of the model do not know what the model is doing to get the output because the machine has been taught to learn from itself and not some visible, programed input.

3. Tactical Barriers

Compounding the technical and technological difficulties are the realities of police tactics. In the pursuit of catching bad guys, police prefer not to reveal their proactive investigative strategies. Revealing the placement of surveillance sensors, or how the heat list is created, might provide tactical advantage for those individuals trying to avoid detection.[61] Police, thus, naturally avoid transparency in order to keep this perceived tactical advantage. Such caution sounds in part as a desire to ensure effectiveness and officer safety, but also speaks to a traditional philosophy to keep police practices relatively closed from public view.[62] As many

---

[58] Darwin Bond-Graham & Ali Winston, *All Tomorrow's Crimes: The Future of Policing Looks a Lot Like Good Branding*, S.F. WKLY. (Oct. 30, 2013), http://archives.sfweekly.com/sanfrancisco/all-tomorrows-crimes-the-future-of-policing-looks-a-lot-like-good-branding/Content?oid=2827968&showFullText=true [https://perma.cc/G35D-F543].

[59] Kroll et al., *supra* note 47, at 679–82.

[60] *Id.* at 657–60.

[61] Noah Hurowitz, *NYPD Terrorism Boss Blasts Council Surveillance Oversight Bill as 'Insane,'* DNAINFO (June 14, 2017, 2:46 PM), https://www.dnainfo.com/new-york/20170614/civic-center/surveillance-act-nypd-council-oversight-terrorism-commissioner [https://perma.cc/8KC2-8PCJ] (quoting John Miller, Deputy Commissioner of Intelligence and Counterterrorism, "Terrorists and criminals do their due diligence and they literally study and adapt to evolving security measures.").

[62] Rachel Harmon, *Why Do We (Still) Lack Data on Policing?*, 96 MARQ. L. REV. 1119, 1129 (2013) ("In practice, police chiefs and other local government actors often limit rather than promote information availability.").

police reformers have found, police processes remain decidedly non-transparent, with the adoption of new technologies producing no different result.[63]

4. Consequences

The consequences of these transparency problems create a real barrier to the successful implementation of big data policing technologies. Secrecy leads to community distrust. Complexity leads to outsourcing to third party experts. And, for police and the communities alike, the technical, technological, and tactical barriers make it difficult to understand if the systems work as advertised. The failure to address these transparency issues risks delegitimizing data-driven strategies, even if the technologies improve policing.

The push for transparency (or at least accountability) has created some pressure to change the technologies. New predictive policing companies have advertised themselves as being more transparent, even going so far as to release their basic code and describe their computer models' approach to removing bias.[64] Others have exposed their underlying models to academic scrutiny,[65] and yet others have explicitly attempted to explain why the outputs from the models come out the way they do.[66] Some of this transparency is driven by public political pressure over use, and some is driven by internal concern that the systems need to work before investing police money in the systems. In recent years, there have been more than a few glimmers of light illuminating the transparency and accountability problem.

B. *Race: Black Data is Racially-Encoded*

Big data policing is still policing, and law enforcement has long struggled with concerns of racial bias.[67] Adding in data collection and analysis or new

---

[63] *See id.* at 1133.

[64] Joshua Brustein, *The Ex-Cop at the Center of Controversy over Crime Prediction Tech*, BLOOMBERG TECH. (July 10, 2017, 5:00 AM), https://www.bloomberg.com/news/features/2017-07-10/the-ex-cop-at-the-center-of-controversy-over-crime-prediction-tech [https://perma.cc/R25N-2D3A]; Dave Gershgorn, *Software Used to Predict Crime Can Now Be Scoured for Bias*, QUARTZ (Mar. 22, 2017), https://qz.com/938635/a-predictive-policing-startup-released-all-its-code-so-it-can-be-scoured-for-bias/ [https://perma.cc/U9B3-DFPT].

[65] *See, e.g.*, JIE XU ET AL., RUTGERS CTR. ON PUB. SEC., CRIME GENERATORS FOR SHOOTINGS IN URBAN AREAS: A TEST USING CONDITIONAL LOCATIONAL INTERDEPENDENCE AS AN EXTENSION OF RISK TERRAIN MODELING (2010); Joel M. Caplan et al., *Joint Utility of Event-Dependent and Environmental Crime Analysis Techniques for Violent Crime Forecasting*, 59 CRIME & DELINQ. 243 (2013); Mohler et al., *supra* note 18.

[66] *See, e.g.*, LESLIE KENNEDY ET AL., RUTGERS CTR. ON PUB. SEC., A MULTI-JURISDICTIONAL TEST OF RISK TERRAIN MODELING AND A PLACE-BASED EVALUATION OF ENVIRONMENTAL RISK-BASED PATROL DEPLOYMENT STRATEGIES (2015).

[67] *See* PAUL BUTLER, CHOKEHOLD: POLICING BLACK MEN 59–61 (2017); POLICING THE BLACK MAN: ARREST, PROSECUTION, AND IMPRISONMENT 178–233 (Angela J. Davis ed., 2017).

surveillance systems does not remove the potential threat of racial discrimination.[68] The data collected, the streets watched, and the police themselves do not change just because of new technologies. Big data policing cannot ignore the resulting racial impact, and in order to be successful, must address the complex realities of policing and race in many American cities. How police patrol, how they respond to problem areas, and how they deal with communities without political power all can potentially influence the discriminatory impact of big data policing and thereby expose the problem of black data policing.

1. Police Patterns

In some jurisdictions, race and place are so closely intertwined that policing certain places means a disproportionate impact on communities of color.[69] If the crime data collected from particular areas becomes the only data in the system, then police data systems will mirror police patrols, not necessarily actual crime rates.[70]

In addition, not all crime is reported in what we know as official crime statistics.[71] For example, if police arrests become a data point for crime statistics, a strategic focus on particular communities of color could result in discriminatory outcomes. Other distortions can occur because certain communities may not report crime. One DOJ report claimed that more than 50% of violent crime goes unreported.[72] Other researchers have correctly noted that certain crimes (sexual assault and domestic violence) remain seriously underreported to police.[73] In recent months, due to a crackdown on immigration enforcement, police have seen

---

[68] *See* Ezekiel Edwards, *Predictive Policing Software Is More Accurate at Predicting Policing Than Predicting Crime*, HUFFINGTON POST (Aug. 31, 2016, 2:58 PM), https://www.huffingtonpost.com/entry/predictive-policing-reform_us_57c6ffe0e4b0e60d31dc9120 [http://perma.cc/D79F-6MJV]; Alvaro M. Bedoya, *The Color of Surveillance*, SLATE (Jan. 18, 2016, 5:55 AM), http://www.slate.com/articles/technology/future_tense/2016/01/what_the_fbi_s_surveillance_of_martin_luther_king_says_about_modern_spying.html [https://perma.cc/88UR-FYB9]; Simmons, *supra* note 1, at 980.

[69] *See generally* I. Bennett Capers, *Policing, Race, and Place*, 44 HARV. C.R.-C.L. L. REV. 43, 62–71 (2009).

[70] Jack Smith IV, *Crime-Prediction Tool May Be Reinforcing Discriminatory Policing*, BUS. INSIDER (Oct. 10, 2016, 7:02 PM), http://www.businessinsider.com/predictive-policing-discriminatory-police-crime-2016-10 [https://perma.cc/RF2N-4WN8].

[71] Press Release, Bureau of Justice Statistics, U.S. Dep't of Justice, Nearly 3.4 Million Violent Crimes Per Year Went Unreported to Police from 2006 to 2010 (Aug. 9, 2012, 10:00 AM), https://www.bjs.gov/content/pub/press/vnrp0610pr.cfm [https://perma.cc/PCY6-NMW7].

[72] *Id.*

[73] *Id.*; *See, also e.g.*, Camille Carey & Robert A. Solomon, *Impossible Choices: Balancing Safety and Security in Domestic Violence Representation*, 21 CLINICAL L. REV. 201, 225 (2014); Jeannie Suk, *Criminal Law Comes Home*, 116 YALE L.J. 2 (2006).

a marked decrease in reported crimes in areas with high Latino populations.[74] This decrease is not necessarily the result of less crime, but rather less reporting due to a fear of community members being deported.

Because of this fear of biased or incomplete data, several predictive policing technologies have specifically avoided including any arrest data in their models.[75] Instead, they rely on reported crimes or calls for service, assuming that allegations of crime victimhood (i.e. calling police for assistance) will be more representative than police suspicion in the form of arrests.[76] Not all predictive policing companies follow this limitation and so the impact of patrol patterns and police targeting could impact the final results, depending on the system chosen.

In addition, person-based predictive models may result in seemingly racially discriminatory effects. For example, if the algorithm that identifies people on the heat list includes information about prior arrests, or connections with people who are arrested, then where police are looking for arrests will impact the resulting risk identification system.[77] Obviously, to be arrested police need probable cause to arrest, and a number of violent arrests could be reasonable data points to record and utilize in an algorithm, but there may still be a racial impact if some of the variables involve arrests (even if justified by good police practice). More problematic are gang affiliations, which tend to correlate with racial minorities in many cities and have no standardized method of correctly identifying gang members.[78] Errors proliferate in gang systems, meaning that this data can impact risk scores and therefore police interaction with those individuals.[79]

Complicating the racial impact is the fact that crime does tend to correlate with poverty and in many urban areas those poor areas also correlate with communities of color.[80] The people getting shot in cities like Chicago, New

---

[74] John Burnett, *New Immigration Crackdowns Creating 'Chilling Effect' on Crime Reporting*, NPR (May 25, 2017, 4:54 AM), https://www.npr.org/2017/05/25/529513771/new-immigration-crackdowns-creating-chilling-effect-on-crime-reporting [https://perma.cc/SRD3-4CAZ].

[75] Both PredPol and HunchLab specifically address this potential problem by not including arrest statistics in their models.

[76] *See* Ferguson, *supra* note 1, at 1153–54.

[77] *See* Kaplan, *supra* note 23.

[78] Joshua D. Wright, *The Constitutional Failure of Gang Databases*, 2 STAN. J. C.R. & C.L. 115, 118 (2005).

[79] Associated Press, *Immigrant Sues Chicago Police Alleging Gang Database Error*, ABC NEWS (July 11, 2017, 6:29 PM), http://abcnews.go.com/amp/US/wireStory/immigrant-sues-chicago-police-alleging-gang-database-error-48567964 [https://perma.cc/98UA-S6Y3]; CAL. STATE AUDITOR, THE CALGANG CRIMINAL INTELLIGENCE SYSTEM 3 (2016), https://www.auditor.ca.gov/pdfs/reports/2015-130.pdf [https://perma.cc/C2Y9-5D4N] (finding numerous errors in the California gang database).

[80] *See* Jack Smith IV, *'Minority Report' is Real—And It's Really Reporting Minorities*, MIC (Nov. 9, 2015), http://mic.com/articles/127739/minority-reports-predictive-policing-technology-is-really-reporting-minorities#.zwXVV93jm [https://perma.cc/35GD-56VL].

Orleans, Kansas City, and New York City are disproportionately minority men.[81] Police are using technology to identify people they think are involved in violence, and a review of the individuals who make the top risk scores on the heat list in Chicago show that they are young men of color. Further, many individuals suspected to be in gangs, or who associate with individuals involved in crime, share race-similar friend groups. If associational suspicion from friend groups is part of the analysis, then people of the same social groups will be more likely to be linked in the same network.

This is not to say that predictive policing is intentionally racially discriminatory, but only that, like traditional policing, it suffers from implicit and explicit racial biases, and tracks the structural problems inherent in policing.[82] None of the algorithms use race in their model (and in fact strip it out), but the perception (correctly) is that the technologies end up targeting communities of color.

Some of the challenge has been to overcome media reports and research reports that sensationalize the potential racial issues.[83] For example, a recent report by the Human Rights Data Analysis Group applied arrest and drug data using PredPol's algorithm to conclude that such an algorithm could contribute to racial bias,[84] despite the fact that PredPol does not use any drug crimes or arrest data in its model. The resulting media coverage misleadingly slammed PredPol for being racially discriminatory.[85]

In order for big data policing to be accepted, it will need to address and overcome this perceived race problem. Predictive policing technologies will be looked at with deserved suspicion unless they can assure communities they do not reify racial inequalities. This assurance is difficult when the end result looks discriminatory (or looks to justify a racially disparate effect). The stated concern is that predictive technologies could create a self-fulfilling feedback loop where biased data collection predicts future biased patrols, which in turn creates more

---

[81] *See, e.g.*, *Tracking Homicides in Chicago*, CHI. TRIB., http://homicides.redeyechicago.com/races/ [https://perma.cc/W5GX-LCRC] (last visited Mar. 25, 2018).

[82] Bryan Llenas, *Brave New World of 'Predictive Policing' Raises Specter of High-Tech Racial Profiling*, FOX NEWS (Feb. 25, 2014), http://latino.foxnews.com/latino/news/2014/02/24/brave-new-world-predictive-policing-raises-specter-high-tech-racial-profiling/ [https://perma.cc/N65R-JDND]; Sidney Perkowitz, *Crimes of the Future*, AEON (Oct. 27, 2016), https://aeon.co/essays/should-we-trust-predictive-policing-software-to-cut-crime [https://perma.cc/E295-RDSS].

[83] *See, e.g.*, William Isaac & Andi Dixon, *Why Big Data Analysis of Police Activity Is Inherently Biased*, PBS (May 10, 2017, 2:42 PM), https://www.pbs.org/newshour/nation/column-big-data-analysis-police-activity-inherently-biased [https://perma.cc/BD4D-97D3].

[84] Kristian Lum & William Isaac, *To Predict and Serve*?, SIGNIFICANCE MAG., Oct. 2016, at 15.

[85] *See, e.g.*, Jack Smith IV, *Crime Prediction Tool PredPol Amplifies Racially Biased Policing, Study Shows*, MIC (Oct. 9, 2016), https://mic.com/articles/156286/crime-prediction-tool-pred-pol-only-amplifies-racially-biased-policing-study-shows#.a7kBWFQyI [https://perma.cc/DS5M-C8MJ].

suspicion in these areas.[86] Worse, the predictive scores can be justified because of the "objectivity" of the data, ignoring that all of the data collection comes from human police officers.

2. Power and Powerlessness

Poor communities of color have had a fraught history with police surveillance. In fact, the history of policing is deeply intertwined with racialized police practices.[87] From the post-slavery patrols, to surveillance of the civil rights movement in the 1960s, to monitoring of the Black Lives Matter movement in this decade, African-Americans have felt the negative impact of police surveillance in pointed and pervasive ways.[88] And, during the entire span of those movements, every day police surveillance and harassment has been a point of tension in poor, minority communities.[89]

Racial bias, thus, remains a concern for the implementation of big data policing techniques. While inanimate algorithms or surveillance cameras cannot be racially biased, how the technologies are built, used, and where they are located can have discriminatory impacts. For this reason, some concern has been raised about the implementation of certain big data surveillance techniques. The Baltimore Police and the FBI used aerial surveillance on police brutality protesters after the death of Freddie Gray, as part of a pattern of surveillance of individuals involved with the Movement for Black Lives.[90] Large scale aerial mass surveillance has been conducted on poor minority areas (including Compton, California, and West Baltimore), and use of the other types of high tech surveillance has been directed at minority neighborhoods.[91] For communities

---

[86] 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-crime/ [https://perma.cc/2N75-BRMB].

[87] *See* Paul Butler, *Stop and Frisk and Torture-Lite: Police Terror of Minority Communities*, 12 OHIO ST. J. CRIM. L. 57, 66–69 (2014); R. Richard Banks, *Beyond Profiling: Race, Policing, and the Drug War*, 56 STAN. L. REV. 571 (2003); Tracey Maclin, *Race and the Fourth Amendment*, 51 VAND. L. REV. 333, 386–92 (1998); CHARLES J. OGLETREE, JR. ET AL., BEYOND THE RODNEY KING STORY: AN INVESTIGATION OF POLICE CONDUCT IN MINORITY COMMUNITIES 24, 52–53 (1995).

[88] *See generally* Sandra Bass, *Policing Space, Policing Race: Social Control Imperatives and Police Discretionary Decisions*, 28 SOC. JUST. 156 (2001); Bedoya, *supra* note 68; FERGUSON, *supra* note 6, at 133–34.

[89] *See* BUTLER, *supra* note 67, 59–61; David A. Harris, *The Stories, the Statistics, and the Law: Why "Driving While Black" Matters*, 84 MINN. L. REV. 265, 268–69 (1999).

[90] Ian Duncan, *New Details Released About High-Tech Gear FBI Used on Planes to Monitor Freddie Gray Unrest*, BALT. SUN (Oct. 30, 2015, 7:04 PM), http://www.baltimoresun.com/news/maryland/freddie-gray/bs-md-ci-fbi-surveillance-flights-20151030-story.html [https://perma.cc/34D5-MNC5].

[91] Laura Moy, *Yet Another Way Baltimore Police Unfairly Target Black People*, SLATE (Aug. 18, 2016, 1:19 PM), http://www.slate.com/blogs/future_tense/2016/08/18/baltimore_police_use_surveillance_technology_to_target_black_neighborhoods.html [https://perma.cc/6NCC-LJ53].

conscious of racial inequality in policing, these initial big data policing experiments signal a pattern to use such technologies more on poor communities of color, than non-minority communities.

It is important not to overstate the placement issue because predictive policing technologies have been applied across jurisdictions and surveillance technologies have also been placed in business districts and places of great wealth and power.[92] But, politically weak communities face a greater risk and danger from the implementation. If history is any guide, "color of surveillance"[93] tends to be black and brown and dark.

3. Consequences

From one perspective, surveillance technologies could offer the potential for unbiased policing strategies both by removing race from the computer models and relying on objective surveillance and sensor technologies. In fact, this is one reason why these technologies have been embraced by communities struggling with police-citizen racial tensions (heightened over the last few years).[94] But, if implemented carelessly, or without recognition of the racial and discriminatory impacts (real or perceived) arising from traditional policing practices, these same technologies can become perceived as illegitimate.

Police need to confront these legacy effects of racial discrimination and not pretend that they do not also impact a data-driven policing strategy. In addition, because of this historical (mis)understanding, police must not blind themselves to the reality that some of their data-driven insights may be influenced by biased data. There exists a real concern of creating a self-reinforcing feedback loop which is driven more by police patterns than crime patterns. There also exists a potential to create two surveillance systems—one focused on poor people and another on communities that have the political and economic means to keep the cameras away.

Fortunately, companies have begun prioritizing avoiding racial bias. Because of the media coverage about possible racial bias, and the growing community concern, police and predictive policing companies have sought ways to avoid biased impacts. While certainly not solved, the issue of race is flagged and will likely not be papered over by an appeal to technological objectivity. For all the fancy math, the reality of big data policing is that it still involves human police

---

[92] One of the most surveilled places in America is lower Manhattan by the financial district in New York City. Chris Dolmetsch & Henry Goldman, *New York, Microsoft Unveil Join Crime-Tracking System*, BLOOMBERG TECH. (Aug. 8, 2012, 7:19 PM), https://www.bloomberg.com/news/articles/2012-08-08/new-york-microsoft-unveil-joint-crime-tracking-system [https://perma.cc/NN67-GDKF].

[93] *See* Bedoya, *supra* note 68.

[94] *See* FERGUSON, *supra* note 6, at 28–31.

officers and human criminal suspects and the ability of technologists to adapt to this known human fallibility.

C. *Law: Black Data Is Distorting Constitutional Doctrine*

Big data technologies threaten to distort a Fourth Amendment doctrine that emerged from a small data era.[95] Early court decisions which interpreted probable cause and reasonable suspicion did so in a context where much of the information came from small data sources—namely police observation, analogue police investigation, and local information sharing.[96] These small data rules can become distorted in a big data world, where mass surveillance systems, mass data collection, and data mining radically expand the available inputs.[97] Because there exists more data to sort through, when that information is applied against a small data rule, it can change how police and courts interpret suspicion.[98] In addition, more extensive surveillance technology has already begun to alter privacy expectations in a world of "panvasive" monitoring.[99]

As a general matter, courts have not addressed the legal implications of the rise of big data policing. A few courts have weighed in on the dangers of data error,[100] and more have wrestled with expectations of privacy in an interconnected, digital age[101] (including the Supreme Court which will decide how the Fourth Amendment should apply to cell-site location tracking in *Carpenter v. United States*).[102] But, by in large, courts have not had to think through the potential distortions of new technology on old law. This section briefly summarizes a few of the questions presented by big data policing.

1. Shadows Around Suspicion

Accepting *arguendo* that predictive policing achieves what it promises—the ability to forecast higher risk places and people who might be involved in crime, then it is important to understand how that prediction might impact police officers

---

[95] *See* Ferguson, *supra* note 3, at 336–38.

[96] *See id.* at 337–38.

[97] Stephen Rushin, *The Judicial Response to Mass Surveillance*, 2011 U. ILL. J.L. TECH. & POL'Y 281, 285–86 (2011); Hu, *supra* note 43, at 803–05.

[98] *See* Simmons, *supra* note 1, at 983–97, 999–1006.

[99] Christopher Slobogin, *Rehnquist and Panvasive Searches*, 82 MISS. L.J. 307, 308 (2013).

[100] *See, e.g.*, Florence v. Bd. of Chosen Freeholders, 566 U.S. 318 (2012); Herring v. United States, 555 U.S. 135 (2009); United States v. Esquivel-Rios, 725 F.3d 1231 (10th Cir. 2013).

[101] *See, e.g.*, United States v. Jones, 565 U.S. 400 (2012).

[102] *See* Stephen E. Henderson, Carpenter v. United States *and the Fourth Amendment: The Best Way Forward*, 26 WM. & MARY BILL RTS. J. 495 (2017); Petition for Writ of Certiorari, Carpenter v. United States, 819 F.3d 880 (6th Cir. 2016) (No. 16-402).

on the streets and judges in court. Or, in other words, how does a prediction impact Fourth Amendment reasonable suspicion?

Take, for example, the use of place-based predictive policing technology in a city. Officers are provided a map (paper or digital) that signals the higher risk areas of crime. Police officers are tasked to patrol these areas when time permits within their other duties in the hopes that the additional police presence at the correct places will deter criminal activity. From a police officer's perspective, a computer algorithm has provided some additional information about the likelihood of a particular activity at a particular place and a particular time. Police officers are not told what to do with this information, but one can imagine that it not only does, but probably should, impact how police view the areas in which they patrol. As they drive through a predicted area, police will be on the lookout for individuals who might be likely thieves. The predictive information will color how they view the actions in those areas and may result in tipping the scales in favor of suspicion.[103] After all, the suspicious man loitering near a parked car might just be waiting for a ride, but he might also be waiting to steal one.

The issue becomes even more difficult in court. If the officer does stop the man, and if the man does have implements of car theft (slim jim, screwdriver, high tech lock pick, etc.), there will be a question of whether this evidence should be suppressed under the Fourth Amendment. A judge will face the following legal question: did the officer have reasonable suspicion to stop a man loitering near a parked car, and does the fact that the suspect was in an area of predicted car theft impact the constitutional analysis? I have written previously about how I think the analysis might come out, but the short answer is that we do not know what a judge might do.[104]

Similarly, right now in Chicago the police have incorporated their "heat list" risk score into their police computer dashboard.[105] The result is that this number (1-500+) shows up next to the letters SSL (strategic subjects list) to symbolize a forecast level of risk of violence. This information provides a revealing and perhaps damning context of the individual's potential risk of violence. Whereas before an officer might judge the individual based on actions (what did the individual do to justify police suspicion), now this observational information can be filtered through a data-rich criminal background history. As currently utilized, this heat list has become a "virtual most wanted list" shortcut for individuals suspected of being involved in crime.[106]

---

[103] Andrew Guthrie Ferguson, *Predictive Policing and Reasonable Suspicion*, 62 EMORY L.J. 259, 304 (2012).

[104] *See id.* at 312.

[105] *See* Kaplan, *supra* note 23.

[106] *See* Jessica Saunders et al., *Predictions Put into Practice: A Quasi-Experimental Evaluation of Chicago's Predictive Policing Pilot*, 12 J. EXPERIMENTAL CRIMINOLOGY 347, 363–64 (2016); Matt Stroud, *Chicago's Predictive Policing Tool Just Failed a Major Test: A RAND Report Shows that the 'Strategic Subject List' Doesn't Reduce Homicides*, VERGE (Aug. 19, 2016, 10:28

But, the real ground-breaking shift will be when facial recognition technology can be used to match individuals on the street to risk-scores in a computer.[107] While currently such real-time facial recognition technology does not exist—being limited to stationary video cameras—the technology is being developed to deploy facial recognition on police cameras and even police body-worn cameras.[108] In this new world, the combination of facial recognition and the heat list will mean that each person on the street can be scored and tracked.

Again, in thinking about suspicion, if police observed the suspect loitering by the car in an area of predicted car theft and also were alerted to the fact that the man had three prior convictions for car theft, then the question will be how this information should impact reasonable suspicion. From the police officer's perspective, this additional information adds to the suspicion. Knowing the suspect is a three-time convicted car thief makes it more likely that he might be up to no good this time. On the other hand, even former car thieves have the right to wait on the street without being stopped by police. More troubling, if merely being in the predicted area with a high risk score is enough for suspicion, then the algorithm (not the individual's actions or the police officer's observation) is reducing the level of Fourth Amendment protection. The basic point is that the predicted risk score distorts traditional Fourth Amendment analysis because such information will likely impact the police officer's determination of suspicion.[109]

So far, courts have not weighed in on the subject of how predictive policing impacts the Fourth Amendment. But, clearly predictive analytics has the potential to distort existing Fourth Amendment doctrine. Because there exists more information about each suspect, this background information will alter the judgment of police, and the court's deference to this discretionary decision. Before big data analytics becomes too engrained in police practice, courts and lawyers will need to address the impact of these new technologies on the Fourth Amendment.

2. Shadows Around Surveillance

The rise of new surveillance technologies has already begun to impact the Fourth Amendment's conception of privacy. In two recent cases, the Supreme

---

AM), https://www.theverge.com/2016/8/19/12552384/chicago-heat-list-tool-failed-rand-test [https://perma.cc/JL3B-BLZR].

[107] *See* Ferguson, *supra* note 3, at 365–69.

[108] Sidney Fussell, *The New Tech That Could Turn Police Body Cams into Nightmare Surveillance Tools*, GIZMODO (Mar. 9, 2017, 10:09 AM), https://gizmodo.com/new-ai-could-turn-police-body-cams-into-nightmare-surve-1792224538 [https://perma.cc/8LSD-QH9A]; Elizabeth Joh, *Free Police Body Cameras Come with a Price*, SLATE (Apr. 5, 2017, 4:12 PM), http://www.slate.com/blogs/future_tense/2017/04/05/taser_international_now_axon_offers_police_free_body_cameras.html [https://perma.cc/G7R3-BFE6] ("Axon CEO Rick Smith has said that he expects to have facial recognition technology in his cameras sometime in the near future.").

[109] *See* Ferguson, *supra* note 3, at 387–89.

Court has recognized the revealing nature of smartphones and digital GPS tracking.[110] The 2017–18 *Carpenter* case will address the future of the third party doctrine in a world where almost all digital devices and communications go through third party intermediaries.[111] Law enforcement routinely requests records for cell sites, internet searches, smart devices, transportation services (like Uber and Lyft), and many other types of digital clues. This form of investigation will only increase in the future, thus, creating new legal challenges.[112]

Constitutional law, as it usually does, muddles along applying analogue analogies to digital realities, and doing the best it can to adapt to a new age.[113] But, many of the questions about expectations of privacy are still unanswered. The internet of things is turning our old-fashioned "effects" into trackable technologies.[114] Almost all of the things we used to think as our "papers" now exist in digital (non-paper) form.[115] Our "houses" are becoming "smart," revealing intimate details through smart monitors and devices.[116] Even our "persons" can be augmented with digitally revealing medical devices and sensors. In fact, in an age of ubiquitous connectivity and mass surveillance, the idea of expectations of privacy may need to be rethought.

The problem of black data policing raises all of these complicated questions. What society once understood to be the legal framework for analysis can appear distorted through a big data lens. Generally speaking legislatures have not acted to keep statutory law current with technological threats, so privacy protections remain

---

[110] Riley v. California, 134 S. Ct. 2473, 2491 (2014); United States v. Jones, 565 U.S. 400, 418 (2012) (Alito, J., concurring); *see also* Ric Simmons, *The Missed Opportunities of* Riley v. California, 12 OHIO ST. J. CRIM. L. 253, 262 (2014).

[111] As a disclosure, I co-authored the Brief of Scholars of Criminal Procedure and Privacy as Amici Curiae in Support of Petitioner Timothy Carpenter, Carpenter v. United States, No. 16-402 (Aug. 14, 2017), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3019294 [https://perma.cc/U7A4-L2FT]; *see generally* Stephen E. Henderson, *Beyond the (Current) Fourth Amendment: Protecting Third-Party Information, Third Parties, and the Rest of Us Too*, 34 PEPP. L. REV. 975, 1015 (2007).

[112] *See* Spenser S. Hsu, *Warrantless Requests to Track Cellphones, Internet Use Grew Sevenfold in D.C. in Three Years*, WASH. POST (July 18, 2017), https://www.washingtonpost.com/local/public-safety/court-warrantless-requests-to-track-cellphones-internet-use-grew-sevenfold-in-dc-in-three-years/2017/07/18/b284ac32-6b36-11e7-9c15-177740635e83_story.html?utm_term=.f86602d370e7 [http://perma.cc/W7S9-VAHG].

[113] *See* James J. Tomkovicz, *Technology and the Threshold of the Fourth Amendment: A Tale of Two Futures*, 72 MISS. L.J. 317, 324–26 (2002); Ric Simmons, *From* Katz *to* Kyllo: *A Blueprint for Adapting the Fourth Amendment to Twenty-First Century Technologies*, 53 HASTINGS L.J. 1303, 1321–22 (2002); *see generally* SLOBOGIN, *supra* note 24.

[114] Andrew Guthrie Ferguson, *The Internet of Things and the Fourth Amendment of Effects*, 104 CALIF. L. REV. 805, 823 (2016).

[115] Donald A. Dripps, *"Dearest Property": Digital Evidence and the History of Private "Papers" as Special Objects of Search and Seizure*, 103 J. CRIM. L. & CRIMINOLOGY 49, 51 (2013).

[116] Andrew Guthrie Ferguson, *The "Smart" Fourth Amendment*, 102 CORNELL L. REV. 547, 585 (2017).

fragmented and unclear. New surveillance technologies are eroding privacy, but what the law has to say about the matter is largely unclear.

3. Consequences

The legal system is not unused to adapting to new technologies.[117] The difficulty is that the technologies change much faster than the law. Courts are relatively slow compared to the ever-changing landscape of technology. Big cases will arise with some guidance from the Supreme Court, but most of the development will happen slowly, case by case.

Police will also adapt to new legal realities. As one example of how predictive policing companies might adapt to potential legal challenge, HunchLab initially marketed a predictive policing product that provided risk maps that changed as the officers drove through the area.[118] For example, an officer might drive from a high burglary area to a high theft area and watch the map change colors as she drove.[119] But, recognizing that this knowledge might distort the officer's suspicion or actions, Hunchlab decided to "blind" the officer to the type of crime at issue. In doing so, they tried to address the distortions of suspicion and the potential bias.

> HunchLab's approach to overcoming bias is to provide less information, not more. It doesn't tell police why a box is selected or even whether it's a high- or average-risk zone. . . . HunchLab works off the notion it's better for officers not to know too much.
>
> "This shouldn't be justification to stop someone," [Jeremy] Heffner [of HunchLab] said. "That's why we actually hide a lot from the officers. We don't tell them the likelihood that a robbery will happen, because people get hung up on probabilities. The goal isn't to go into these places and make a bunch of arrests. The goal is to have nothing happen."[120]

Such a change might well avoid the concerns of an algorithm distorting reasonable suspicion. It might also provide a helpful random-control sample to be able to test HunchLab's theory. Of course, the change might prove to be not quite as effective as providing that crime-type information to officers tasked to stop crime. In the current moment, the debate about the appropriate balance between effectiveness

---

[117] Orin S. Kerr, *The Fourth Amendment and New Technologies: Constitutional Myths and the Case for Caution*, 102 MICH. L. REV. 801, 805, 808–10 (2004).

[118] *See* Maurice Chammah, *Policing the Future*, MARSHALL PROJECT (Feb. 3, 2016, 7:15 AM), https://www.themarshallproject.org/2016/02/03/policing-the-future [https://perma.cc/5NKU-S9TB].

[119] *See id.*

[120] Melamed, *supra* note 49 (internal quotations omitted).

and bias-prevention has yet to be worked out as technologies continue to develop and respond to legal and community pressure.

III. CONTEXTUALIZING THE DARKNESS OF BLACK DATA POLICING

The problems of black data—transparency, racial bias, and unclear constitutional protections—are not really new to policing. Traditional policing suffers from many of the same problems. What is interesting is the way digital technologies expose or heighten the risks of these problems in a more visual way. With data, one can see the long-standing problems more clearly.

First, as to this heightening effect, it would be misleading to characterize policing as having been an overly transparent profession. For years, it has been difficult to get any good data on policing practices (who was stopped, why, or how often).[121] Because of the fragmented nature of the policing profession, it has always been hard to see general patterns or study national practices.[122] And, of course, one cannot get inside a police administrator's or police officer's head to see how the decisions about suspicion actually occur. So, it seems somewhat unfair to criticize an opaque algorithm shaping police suspicion as being a new problem.

Yet, the fear of opaqueness—the perceived general lack of accountability surrounding big data policing—is real and has been a powerful argument against predictive policing and other new technologies. It somehow seems more obvious to highlight the black box nature of the computer model than the black box nature of the human mind. Perhaps the difference is that we see the algorithmic models as new and thus changeable, and do not see the same with people. But, the fear is clearly a heightened one that comes from an uncomfortableness with new technology.

Similarly, researchers have solid empirical evidence that racial bias (both explicit and implicit) has affected policing patterns in the past (and the present).[123] So, again it seems a bit unfair to critique big data models for being biased when they are in fact equal, or even less biased than the status quo. Yet, even though we know that bias exists in policing, that human failing somehow feels less threatening than a computer making a similar bad judgment. With a computer program we have a thing that caused the error which might make it easier to cast a skeptical eye, rather than indicting the entire policing system as biased.

---

[121] Harmon, *supra* note 62, at 1129.

[122] Barry Friedman & Maria Ponomarenko, *Democratic Policing*, 90 N.Y.U. L. REV. 1827, 1843 (2015) ("Policing in the United States is a diffuse business. . . . There are just under 18,000 separate police forces in the United States, and some 765,000 sworn officers.").

[123] As but one recent example, REBECCAC C. HETEY ET AL., STAN. SPARQ, DATA FOR CHANGE: A STATISTICAL ANALYSIS OF POLICE STOPS, SEARCHES, HANDCUFFINGS, AND ARRESTS IN OAKLAND, CALIF., 2013–2014 (2016); *see also* STAN. SPARQ, STRATEGIES FOR CHANGE: RESEARCH INITIATIVES AND RECOMMENDATIONS TO IMPROVE POLICE COMMUNITY RELATIONS IN OAKLAND, CALIF. (Jennifer L. Eberhardt ed., 2016).

Finally, most scholars agree that the Fourth Amendment reasonable suspicion standard is not a model of clarity.[124] Most police officers would likely agree as they are constantly second guessed in court and on the streets. The limited training in the police academy about "reasonable suspicion" or "probable cause" provides little concrete guidance on the streets, so frustration can grow about the lack of clear rules. The fact that technology may also distort these unclear rules may thus not reveal a new problem. And, depending on how the information is provided, it might actually refine the process of establishing reasonable suspicion.[125]

Perhaps this heightened focus on big data policing will be a positive thing, because maybe the concern about black data will allow us to address some of the underlying problems of all too human policing. The truth is that it is easier to be vigilant about computer code than people. Code or computer models can be designed to be accountable (if not transparent), to be policed for bias, and to be regulated by legal rules. Unlike people, designers can test the inputs, and test the outputs. Unlike people, we can relatively easily change the system when we have identified the problems. In fact, each of the black data problems of opacity, human judgment, and uncertain rules are routine design problems for most data-driven systems.

The difficulty is that police or technologists have not started from this premise of designing a system that focuses on the black data problems. Instead, much of the technology has been sold as papering over some of those underlying human issues. In a world where the community is concerned about police brutality, data-driven policing sounds like an objective solution. So, the focus has been on selling big data systems, not addressing black data problems. If overcoming black data policing becomes a priority, and the hard questions about design are asked at the outset, many of the problems, and much of the fear of a big data policing future, can be illuminated and addressed.

---

[124] Andrew Guthrie Ferguson, *Personal Curtilage: Fourth Amendment Security in Public*, 55 WM. & MARY L. REV. 1283, 1293–95 (2014).

[125] Simmons, *supra* note 1, at 999–1009.