William & Mary Bill of Rights Journal

Volume *26 (2017-2018)*
Issue 2 *Symposium: Big Data, National Security, and the Fourth Amendment*

Article 3

December 2017

Feeding the Machine: Policing, Crime Data, & Algorithms

Elizabeth E. Joh

FEEDING THE MACHINE: POLICING, CRIME DATA, & ALGORITHMS

Elizabeth E. Joh*

INTRODUCTION

Police departments are increasingly turning to big data tools to answer some familiar questions. Where will the next crime occur? Which person is likely to commit that next crime, and who will be the victim? What threat does that driver sitting in his stopped car pose? Traditionally, the police have answered these questions with a mixture of training, experience, and instinct.[1] Today, police increasingly turn to big data tools—computer algorithms that analyze massive data sets—to provide an answer instead.[2] Police departments typically buy these predictive policing[3] programs

---

* Professor of Law, University of California, Davis, School of Law (King Hall). Thanks to William Isaac, Kristian Lum, Charles Reichmann and the participants in and organizers of the 2017 *Big Data, National Security, and the Fourth Amendment* Symposium sponsored by the *William & Mary Bill of Rights Journal*.

[1] Although unacceptable as the legal justification for a police stop, hunches are nevertheless a reality in police work. *See, e.g.*, Eli B. Silverman, *With a Hunch and a Punch*, 4 J.L. ECON. & POL'Y 133, 134 (2007) ("Despite these warnings, it is important to recognize that police hunches, as integral ingredients of police discretion, are historically ingrained in the very nature of police work.").

[2] This definition of big data comes from VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK 4–5 (2013). There is also a tendency to treat two related concepts—artificial intelligence and machine learning—as synonyms in popular writing. While the two concepts are related, they are distinct. Artificial intelligence refers to the branch of computer science interested in building machines capable of intelligent behavior. Machine learning, a subset of artificial intelligence, refers to the use of algorithms capable of learning from experience. Developments in machine learning have made everyday applications like Facebook tagging, Siri, sophisticated web searching, and movie recommendations possible. *See, e.g.*, Lee Bell, *Machine Learning Versus AI: What's the Difference?*, WIRED UK (Dec. 1, 2016), http://www.wired.co.uk/article/machine-learning-ai-explained [https://perma.cc/QVQ4-QVRG] (noting that machine learning and AI, while related, are distinct concepts). Moreover, "[a]n algorithm is a procedure or set of instructions often used by a computer to solve a problem." Julia Angwin, *Making Algorithms Accountable*, PROPUBLICA (Aug. 1, 2016, 3:21 AM), https://www.propublica.org/article/making-algorithms-accountable [https://perma.cc/RZY6-XM38]. *But see* 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 ("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.").

[3] There is no single definition of predictive policing, but generally predictive policing refers to "the application of analytical techniques—particularly quantitative techniques—to

from surveillance technology company vendors. The same uses of automated analysis seen in shopping, consumer finance, healthcare, and dating have come to policing.[4]

In turn, an emerging body of scholarship and journalism has already begun to question the presumed neutrality, efficiency, and quality of the big data analysis used in policing and other criminal justice institutions. Some have called for greater transparency regarding the "black box" algorithms that can influence decisions about suspicion, bail, sentencing, and parole.[5] Still others have asked whether the private companies responsible for developing these big data programs should be permitted to invoke intellectual property rights to keep some information from defendants, judges, and researchers.[6]

At the same time, courts have shown themselves to be receptive to the use of big data. In 2016, the Wisconsin Supreme Court became one of the first in the nation to uphold the use of a predictive algorithm in sentencing.[7] The trial court judge in the case had used COMPAS, a risk assessment algorithm, to conclude that defendant Eric Loomis presented enough of a "high risk" to the community that he was ineligible for probation.[8]

---

identify likely targets for police intervention and prevent crime or solve past crimes by making statistical predictions." WALTER L. PERRY ET AL., RAND CORP., PREDICTIVE POLICING: THE ROLE OF CRIME FORECASTING IN LAW ENFORCEMENT OPERATIONS 1–2 (2013); *see also* Andrew D. Selbst, *Disparate Impact in Big Data Policing*, 52 GA. L. REV. (forthcoming 2017) (manuscript at 3), https://ssrn.com/abstract=2819182 (defining predictive policing as a method that "uses data mining methods to find correlations between criminal outcomes and various input data they have collected").

[4] *See, e.g.*, Justin Jouvenal, *Police Are Using Software to Predict Crime. Is It a 'Holy Grail' or Biased Against Minorities?*, WASH. POST (Nov. 17, 2016), https://www.washingtonpost.com/local/public-safety/police-are-using-software-to-predict-crime-is-it-a-holy-grail-or-biased-against-minorities/2016/11/17/525a6649-0472-440a-aae1-b283aa8e5de8_story.html?utm_term=.f6975c45dc03 [https://perma.cc/D9B8-F4WR] ("Law enforcement agencies are increasingly trying to forecast where and when crime will occur, or who might be a perpetrator or a victim, using software that relies on algorithms, the same math Amazon uses to recommend books.").

[5] *See, e.g.*, FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 4 (2015) ("Secrecy is approaching critical mass, and we are in the dark about crucial decisions. Greater openness is imperative.").

[6] *See, e.g.*, Elizabeth E. Joh, *The Undue Influence of Surveillance Technology Companies on Policing*, 92 N.Y.U. L. REV. ONLINE 101 (2017); Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. (forthcoming 2018) (manuscript at 5–6), https://ssrn.com/abstract=2920883.

[7] *See* Joe Palazzolo, *Wisconsin Supreme Court to Rule on Predictive Algorithms Used in Sentencing*, WALL ST. J. (June 5, 2016, 5:30 AM), https://www.wsj.com/articles/wisconsin-supreme-court-to-rule-on-predictive-algorithms-used-in-sentencing-1465119008 (noting that the Loomis ruling "would be among the first to speak to the legality of risk assessments as an aid in meting out punishments").

[8] *See* State v. Loomis, 881 N.W.2d 749, 770 (Wis. 2016) ("[I]f used properly with an

The growing conversation about algorithmic decisionmaking in policing, however, often overlooks one issue. Discussions of big data programs used by the police tend to assume the police are merely *end users of big data*.[9] In other words, police departments are consumers and clients of big data, little different than users of Spotify, Netflix, Amazon, or Facebook, and are reliant on the engineers who develop the programs providing the data analysis.

Yet this assumption about predictive policing contains a flaw. Police are not simply end users of big data. They *generate the information* that big data programs rely upon.[10] Crime and disorder are not natural phenomena. These events have to be observed, noticed, acted upon, collected, categorized, and recorded—while other events are not. To be sure, there have already been concerns raised that the inputs for policing algorithms reflect racial biases.[11] Racial bias, however, is just one of the ways crime rates are the product of social processes.

The difference between crime data and actual crime reflects a longstanding observation about the police. Policing is not the passive collection of information, nor the identification of every violation of the law. Every action—or refusal to act—on the part of a police officer, and every similar decision made by a police department, is also a decision about how and whether to generate data.[12] These decisions are inevitable and necessary. Resource constraints mean that "full enforcement" by the police is an impossibility.[13] In practice, though, these choices mean

---

awareness of the limitations and cautions, a circuits [sic] court's consideration of a COMPAS risk assessment at sentencing does not violate a defendant's right to due process.").

[9] *See* discussion *infra* Part I.

[10] *See* discussion *infra* Part II.

[11] The concern about racial bias in algorithmic software is becoming prevalent. *See, e.g.*, Daniel Munro, *The Ethics of Police Using Technology to Predict Future Crimes*, MACLEAN'S (June 18, 2017), http://www.macleans.ca/society/the-ethical-risks-in-police-using-technology-to-predict-future-crimes/ [https://perma.cc/G6NN-FDTB] ("When models draw on flawed or inappropriate data, they may recommend increasing police activity in neighbourhoods with higher proportions of ethnic or racial minorities—not because the risk of crime is higher, but because the input data are biased."); Aaron Shapiro, *Reform Predictive Policing*, 541 NATURE 458, 459 (2017) ("Another concern [apart from the objective of a predictive police program] is the racial bias of crime data.").

[12] Not only do the police produce data about other people, they produce data about themselves. There is a growing interest in whether big data about police could provide insights into predicting which officers might engage in problematic behavior. *See, e.g.*, Rob Arthur, *We Now Have Algorithms to Predict Police Misconduct*, FIVETHIRTYEIGHT (Mar. 9, 2016, 7:32 AM), https://fivethirtyeight.com/features/we-now-have-algorithms-to-predict-police-misconduct/ [https://perma.cc/V35Y-997M] (describing pilot algorithmic prediction system designed to identify officers at risk of misconduct in Charlotte police department).

[13] Joseph Goldstein, *Police Discretion Not to Invoke the Criminal Process: Low-Visibility Decisions in the Administration of Justice*, 69 YALE L.J. 543, 556 (1960) (referring to full

that police decisionmaking shapes the very reality we perceive about crime and law breaking. We know about the crimes the police pay attention to.[14] With others, we often don't.

This Essay explains why predictive policing programs can't be fully understood without an acknowledgment of the role police have in creating their inputs. Their choices, priorities, and even omissions become the inputs algorithms use to forecast crime.[15] The filtered nature of crime data matters because these programs promise cutting edge results, but may deliver analyses with hidden limitations.

I. THE DEBATE OVER BIG DATA POLICING

Many police departments are turning to algorithmic decisionmaking software to help them make predictions about where crimes will happen and who might commit them. Police departments in places like Seattle, Los Angeles, and Atlanta have piloted or adopted predictive policing programs that try to pinpoint geographic locations where crime is likely to occur in the future.[16] The Chicago Police Department (CPD) compiles a "heat list" of persons who are identified by an algorithm as being at high risk of violent crime perpetration or victimization.[17] CPD officials use people from the list to conduct "custom notifications": in person visits where "the police bluntly warn that the person is on the department's radar."[18] The city of Fresno,

---

enforcement as the area in which the police are "not only authorized but expected to enforce fully the law of crimes").

[14] Crime data can be measured in other ways, such as the crime victim survey, but police collected data is dominant. The National Survey on Drug Use and Health (NSDUH) is a nationwide annual survey that asks about 70,000 people selected at random about patterns in the use and abuse of alcohol, tobacco, and illegal substances. *See About the Survey*, NAT'L SURVEY ON DRUG USE & HEALTH, https://nsduhweb.rti.org/respweb/project_description.html [https://perma.cc/YSX6-2WCU] (last visited Dec. 4, 2017).

[15] *See* discussion *infra* Part II.

[16] Nate Berg, *Predicting Crime, LAPD-Style*, GUARDIAN (June 25, 2014, 5:19 AM), https://www.theguardian.com/cities/2014/jun/25/predicting-crime-lapd-los-angeles-police-data-analysis-algorithm-minority-report [https://perma.cc/6D2H-9M53].

[17] The program is formally known as the "Strategic Subject Algorithm," which creates a "risk assessment score known as the Strategic Subject List or 'SSL.' These scores reflect an individual's probability of being involved in a shooting incident either as a victim or an offender. Scores are calculated and placed on a scale ranging from 0 (extremely low risk) to 500 (extremely high risk)." *Strategic Subject List*, CHICAGO DATA PORTAL, https://data.cityofchicago.org/Public-Safety/Strategic-Subject-List/4aki-r3np [https://perma.cc/M95T-VPCT] (last updated Dec. 7, 2017).

[18] Between 2013 and 2016, CPD officers, social workers, and community leaders have visited 1,300 people with "high numbers on the list." Monica Davey, *Chicago Police Try to Predict Who May Shoot or Be Shot*, N.Y. TIMES (May 23, 2016), https://www.nytimes.com

California, briefly considered the use of Beware, a predictive program that assigns a threat score to people the police encounter, based upon various factors including criminal history and social media use.[19]

All of these programs apply algorithms to vast quantities of information to make determinations that traditionally had been made by people. Because today's algorithms are capable of processing massive amounts of data, they can assess much more data more quickly than any individual officer, crime analyst, or department ever could. One predictive policing program, PredPol, has been adopted or tested by dozens of police departments around the country.[20] PredPol analyzes years and sometimes decades of past crime data to identify places where crime is likely to occur in the future.[21] PredPol employs an algorithm that relies on three variables: crime type, date and time, and location; but, other predictive programs can include many other sources of information.[22] Another predictive program, HunchLab, uses machine learning algorithms that incorporate not only public reports of crime, but also weather patterns, moon phases, the location of bars and bus stations, and even the schedules of major sports events.[23]

With the rise of algorithmic decisionmaking in criminal justice, a set of identifiable critiques has emerged.[24] Some have focused on the aims of the algorithms themselves. Others have raised questions about the "black box" nature of algorithms. Finally, some have drawn attention to the "garbage in, garbage out" problem.

/2016/05/24/us/armed-with-data-chicago-police-try-to-predict-who-may-shoot-or-be-shot.html.

[19] 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 [https://perma.cc/G8SP-X84X]. After public criticism, the department withdrew its support for the program. *See* Tim Sheehan, *Fresno Council Halts Purchase of Data Software Wanted by Police*, FRESNO BEE (Mar. 31, 2016, 5:30 PM), http://www.fresnobee.com/news/local/article69337677.html [https://perma.cc/UBQ5-9CDH].

[20] Mara Hvistendahl, *Can 'Predictive Policing' Prevent Crime Before It Happens?*, SCIENCE (Sept. 28, 2016, 9:00 AM), http://www.sciencemag.org/news/2016/09/can-predictive-policing-prevent-crime-it-happens [https://perma.cc/2GXT-F8DT] (noting that PredPol has been adopted by sixty police departments around the country).

[21] *See id.*

[22] Kristian Lum & William Isaac, *To Predict and Serve?*, 13 SIGNIFICANCE 14, 18 (2016) (describing the PredPol algorithm).

[23] *See* Shapiro, *supra* note 11.

[24] What follows is a list of the major criticisms of algorithmic decisionmaking, but certainly not a complete one. One important question not addressed here asks how such algorithmic judgments could be used to justify, at least partially, the necessary reasonable suspicion for an investigative stop for Fourth Amendment purposes. *See, e.g.*, Andrew Guthrie Ferguson, *Predictive Policing and Reasonable Suspicion*, 62 EMORY L.J. 259, 263 (2012) (arguing that "in its idealized form, predictive policing will impact reasonable suspicion analysis and become an important factor in a court's Fourth Amendment calculus").

First, the algorithms themselves may be flawed in some fundamental ways. An algorithm's objective may not be clear.[25] And even if seemingly clear, an algorithm will reflect the perspective and biases of its creators.[26] Its engineers may have given little thought to any negative social effects of its use.[27] At the same time, the very idea of algorithmic decisionmaking may subtly appeal to us as objective because of its mathematical basis.[28] As one data scientist has suggested, however, we should cast aside such "mathwashing": the assumption that algorithmic models don't have subjectivity baked into them because they involve math.[29]

Second, the algorithms that produce life-altering judgments on everything from parole to credit are often themselves "black boxes."[30] An algorithm can be a black box in two senses. In the first sense, an algorithm can be a black box because the calculations used to make a decision may be inscrutable to the person affected by

---

[25] *See, e.g.*, DAVID ROBINSON & LOGAN KOEPKE, UPTURN, STUCK IN A PATTERN: EARLY EVIDENCE ON "PREDICTIVE POLICING" AND CIVIL RIGHTS 3 (2016) ("Published sources do not make clear what these scores [on the Beware predictive system] are intended to measure, much less whether they are accurate in doing so."); Lyria Bennett Moses & Janet Chan, *Algorithmic Prediction in Policing: Assumptions, Evaluation, and Accountability*, POLICING & SOC'Y 1, 2 (2016) ("Predictive policing is also premised on the assumptions that it is possible to use technology to predict crime before it happens, that forecasting tools can predict accurately, and that police will use this knowledge effectively to reduce crime." (internal citation omitted)); Shapiro, *supra* note 11 ("There is no agreement as to what predictive systems should accomplish—whether they should prevent crime or help to catch criminals—nor as to which benchmarks should be used."); *cf.* LEE RAINIE & JANNA ANDERSON, PEW RESEARCH CTR., CODE-DEPENDENT: PROS AND CONS OF THE ALGORITHM AGE 75 (2017), http://www.pewinternet.org/2017/02/08/code-dependent-pros-and-cons-of-the-algorithm-age/ [https://perma.cc/89Y6-8L3V] (quoting Trevor Owens observing that "[a]lgorithms all have their own ideologies"); Kate Crawford & Ryan Calo, *There Is a Blind Spot in AI Research*, 538 NATURE 311, 311–13 (2016) (noting that there is insufficient attention to the question of "whether [an artificially intelligent] system should be built at all").

[26] RAINIE & ANDERSON, *supra* note 25, at 57.

[27] *See, e.g.*, Crawford & Calo, *supra* note 25, at 312 (noting that a social systems approach could "explore whether the use of historical data to predict where crime will happen is driving overpolicing of marginalized communities").

[28] *See* Tyler Woods, *'Mathwashing,' Facebook and the Zeitgeist of Data Worship*, TECHNICAL.LY BROOKLYN (June 8, 2016, 9:18 AM), https://technical.ly/brooklyn/2016/06/08/fred-benenson-mathwashing-facebook-data-worship/ [https://perma.cc/7FBJ-E89C] (calling such use of math terms to hide a subjective reality "mathwashing"); *see also* RAINIE & ANDERSON, *supra* note 25, at 57.

[29] Data scientist Fred Benenson coined the term and states that "[a]lgorithm and data driven products will always reflect the design choices of the humans who built them, and it's irresponsible to assume otherwise." Woods, *supra* note 28.

[30] In a recent white paper, the Pew Research Center identified the growing "need . . . for algorithmic literacy, transparency and oversight" as a major theme in the influence of algorithms in society. *See* RAINIE & ANDERSON, *supra* note 25, at 4.

that decision.[31] And as machine learning algorithms become more complex, they may be inscrutable to the programmers themselves.[32] This has prompted calls to require legal rights for individuals to know the basis of automatic decisionmaking affecting them. For example, new rules contained in the European Union's General Data Protection Regulation, which will come into force in 2018, recognize a right to explanation.[33] A person affected by an algorithmic decision can ask for an explanation of it.[34] Similarly, American scholars have asked whether such automated decisionmaking interferes with due process rights.[35]

An algorithm can also be a black box in another sense; the companies that create them often refuse to divulge information about them.[36] From their developers' perspective, revealing how an algorithm works risks exposing valuable trade secret information to competitors.[37] That justification has been relied upon by judges to reject defendant requests for access to the algorithms that helped convict them, and by police departments to deny requests for predictive policing algorithms.[38]

---

[31] *See, e.g.*, *id.* at 74.

[32] *See, e.g.*, *id.* at 19 (quoting Marc Rotenberg: "Machines have literally become black boxes—even the developers and operators do not fully understand how outputs are produced."); Will Knight, *The Dark Secret at the Heart of AI*, MIT TECH. REV. (Apr. 11, 2017), https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/ [https://perma.cc/TA4B-294L] ("[Machine learning systems that] seem relatively simple on the surface . . . have programmed themselves, and they have done it in ways we cannot understand. Even the engineers who build these apps cannot fully explain their behavior.").

[33] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), art. 13–14, 2016 O.J. (L 119) 1, 40–42, 87.

[34] *See id.* at 40–41 (establishing right to know "meaningful information about the logic involved" in automated decisionmaking "where personal data are collected from the data subject"); *id.* at 41–42 (establishing a similar right to information "where personal data have not been obtained from the data subject").

[35] *See, e.g.*, Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249 (2008) (discussing harms of automation that might be addressed by a due process approach); Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C. L. REV. 93, 121–28 (2014) (calling for procedural due process protections for big data use); Tal Z. Zarsky, *Transparent Predictions*, 2013 U. ILL. L. REV. 1503 (arguing for procedural protections with respect to government use of big data).

[36] *See, e.g.*, Moses & Chan, *supra* note 25, at 3 ("Information on the [predictive policing] tools themselves is often limited and source code is often a trade secret.").

[37] *See, e.g.*, Joh, *supra* note 6, at 125–26 (discussing how TrueAllele protected its source code).

[38] *See, e.g.*, Davey, *supra* note 18 ("The [Chicago] police cited proprietary technology as the reason they would not make public the 10 variables used to create the list . . . .").

This secrecy also thwarts calls for independent audits of algorithms used in policing and sentencing.[39] Without efforts to unlock either sort of black box, we run the risk that "there will be a class of people who can use algorithms and a class used by algorithms."[40]

Finally, algorithmic decisionmaking has been subjected to the "garbage in, garbage out" critique: that any decision is as good or as bad as the data relied upon by the program.[41] For example, when algorithms in the criminal justice system rely upon data that contains racial bias, the machine learning algorithms that use this data to make predictions will inevitably reflect that racial bias.[42] This will be true despite any claims that an algorithm is "race neutral."

A 2016 ProPublica investigation into COMPAS, a widely used sentencing algorithm, highlights some of these issues.[43] The software is designed to help judges make assessments about the kinds of sentences offenders should receive, based on risk scores supplied by the COMPAS algorithm.[44] ProPublica journalists obtained risk scores for more than 7,000 people arrested in Broward County, Florida over a two-year period.[45]

When checked against what actually happened to those arrested in the study, the algorithm's predictions proved "remarkably unreliable" in predicting who was likely to commit violent crime in the future.[46] For instance, only twenty percent of the people identified by the algorithm as likely to commit a violent crime in the future actually did so.[47] Equally significant was the study's discovery that the algorithm's prediction

---

[39] *See, e.g.*, RAINIE & ANDERSON, *supra* note 25, at 79 (quoting Thomas Claburn as saying, "Our algorithms, like our laws, need to be open to public scrutiny, to ensure fairness and accuracy.").

[40] *Id.* at 75 (quoting David Lankes, professor and director of the University of South Carolina School of Library and Information Science).

[41] The 2016 White House Report on Big Data identified algorithm inputs as one of the key challenges in deploying big data and algorithmic systems. *See* EXEC. OFFICE OF THE PRESIDENT, BIG DATA: A REPORT ON ALGORITHMIC SYSTEMS, OPPORTUNITY, AND CIVIL RIGHTS 6–21 (2016).

[42] *See generally* Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [https://perma.cc/9YP3-5DMJ] (discussing racial bias within current software that is designed to predict future criminal activity).

[43] *Id.* ("Northpointe's software is among the most widely used assessment tools in the country.").

[44] *Id.*

[45] *Id.*

[46] *Id.* The Northpointe software uses a benchmark of a new arrest within two years of the current arrest. *Id.*

[47] *Id.*

showed significant racial disparities: black defendants were twice as likely as white defendants to be mislabeled as future criminals.[48]

Most significantly, the ProPublica investigation published its investigation after obtaining risk scores and conducting its own analysis.[49] Claiming protection of its proprietary information, Northpointe, the company that sells COMPAS, did not share with ProPublica the calculations used to produce defendant risks scores.[50]

II. CRIME DATA AND ALGORITHMIC POLICING

When a predictive policing algorithm relies upon crime data, this information is at best a partial representation of crime in the community. Violations of the law are not the same as what is published as crime data—the official recording of crime and crime-related statistics, such as investigative stops. Thus crime data, while representative of some of the crime that occurs, does not represent all of the crime that actually occurs.

The observation that social processes influence the very existence of crime data has interested criminologists since the 1960s. Some scholars went as far as asserting that the crime rate should be considered a "social fact" and thus cannot be considered accurate or inaccurate.[51] In its most extreme (and least convincing) forms, this skepticism questions whether rises in crime rates have any meaning at all.[52]

The more moderate version of this skepticism does clarify the influence of social institutions—particularly the police—on crime data. Crime data does not simply make itself known.[53] Instead, crime rate measurement requires that "crime is (1) uncovered, (2) classified, and (3) recorded."[54] This means that official crime data is the end result

---

48 *Id.*

49 *Id.*

50 *Id.*

51 *See, e.g.*, Donald J. Black, *Production of Crime Rates*, 35 AM. SOC. REV. 733, 734 (1970).

52 *See, e.g.*, Mike Maguire, *Crime Data and Crime Statistics*, *in* THE OXFORD HANDBOOK OF CRIMINOLOGY 241, 249 (Mike Maguire et al. eds., 4th ed. 2007) ("While some criminologists took these ideas in directions that most policy-makers found unconvincing—including arguments that all rises in crime were illusory and in some cases deliberately manufactured by the police, [the more general critique became] widely accepted." (internal citation omitted)).

53 *See, e.g.*, EXEC. OFFICE OF THE PRESIDENT, *supra* note 41, at 22 ("Many criminal-justice data inputs are inherently subjective. Officers use discretion in enforcement decisions . . . just as police officers and prosecutors use discretion in charging . . . . The underlying data reflects these judgment calls.").

54 Wesley G. Skogan, *The Validity of Official Crime Statistics: An Empirical Investigation*, 55 SOC. SCI. Q. 25, 26 (1974). The recorded data also has to be recorded correctly. Mistakes and misclassifications affect algorithmic calculations. *See generally* Wayne A. Logan & Andrew Guthrie Ferguson, *Policing Criminal Justice Data*, 101 MINN. L. REV. 541 (2016) (discussing how mistakes and misclassifications can also lead to innocent persons being detained or arrested).

of many processes and filters that capture some of the crime that actually occurs.[55] Some of these filters involve legislative and prosecutorial decisions,[56] but many of them are attributable to the police.

For instance, scholars of the police have repeatedly pointed out how much enforcement discretion individual officers possess in stopping motorists and pedestrians, or not stopping them at all.[57] That discretion is not only inevitable but useful in the many situations when the public expects the police to "do something" about crime and disorder.[58] "Doing something" doesn't always result in arrest. When police do make arrests, many factors can influence an individual line officer's ultimate decision.

Perhaps the most well-known observation about the decision to arrest is its situational nature.[59] Even when legally authorized to do so, the police may choose not to arrest because of unpredictable exigencies: factors specific to the particular interaction where arrest may be technically justifiable, but practically unappealing.[60] Whether or not a police encounter results in an arrest depends on many factors, including the seriousness of the offense, the wishes of the complainant, the social distance between the suspect and the complainant, and the respect shown to the police.[61]

---

55 *See, e.g.*, Moses & Chan, *supra* note 25, at 5 ("Not only is it impossible to capture every 'crime' that is committed, but such data that are captured will not always be categorised accurately or consistently." (citation omitted)).

56 *See, e.g.*, Mary Breasted, *Carey Signs Marijuana Measure Reducing Penalty for Possession*, N.Y. TIMES (June 30, 1977), http://www.nytimes.com/1977/06/30/archives/carey-signs-marijuana-measure-reducing-penalty-for-possession-carey.html (discussing the 1977 change by the New York legislature that decriminalized possession of small amounts of marijuana not in public view); Stephanie Clifford & Joseph Goldstein, *Brooklyn Prosecutor Limits When He'll Target Marijuana,* N.Y. TIMES (July 8, 2014), https://www.nytimes.com/2014/07/09/nyregion/brooklyn-district-attorney-to-stop-prosecuting-low-level-marijuana-cases.html (reporting former Brooklyn District Attorney's announcement to "stop prosecuting most low-level marijuana cases").

57 *See, e.g.*, GEORGE L. KELLING, NAT'L INST. OF JUSTICE, "BROKEN WINDOWS" AND POLICE DISCRETION 23 (1999) (describing "discovery" of police discretion in 1950s American Bar Foundation study and how "police use discretion throughout their work").

58 *Id.* at 9.

59 *See, e.g.*, Egon Bittner, *The Police on Skid-Row: A Study of Peace Keeping,* 32 AM. SOC. REV. 699, 702–03 (1967) (discussing police decisions not to make arrests for minor offenses).

60 *Id.* at 702–03, 712–13.

61 *See, e.g.*, Black, *supra* note 51, at 736; Donald J. Black, *The Social Organization of Arrest,* 23 STAN. L. REV. 1087, 1104–10 (1971) (discussing similar factors); Donald J. Black & Albert J. Reiss, Jr., *Police Control of Juveniles*, 35 AM. SOC. REV. 63, 69–72 (1970) (discussing the complainant's wishes); Stephen D. Mastrofski et al., *The Helping Hand of the Law: Police Control of Citizens on Request*, 38 CRIMINOLOGY 307, 323, 328 (2000) (discussing police response to the complainant's drunkenness); Irving Piliavin & Scott

Seemingly unrelated issues like workplace pressures can also influence the decision to arrest. Officers with secondary part-time jobs and other personal commitments may purposefully avoid arrests toward the end of their shifts.[62] Labor contract disputes can also result in slowdowns; for the police, this may result in the intentional avoidance of arrests and citations.[63] At the other extreme, "bounty hunter" officers[64] seeking overtime pay may intentionally look for arrests at the end of their shifts in order to take advantage of the time necessary to complete the administrative details of arrest.[65] Furthermore, arrest rates can vary considerably from one officer to the next, depending on factors such as seniority and morale.[66]

Police departments can also provide incentives to individual officers to increase or decrease enforcement activities. American policing is largely a local activity.[67] This means that local factors including geography, populations, and budgets influence what the police do.[68] Funding crises can result in cutting back on service calls for crimes

---

Briar, *Police Encounters with Juveniles*, 70 AM. J. SOC. 206, 210–14 (1964) (discussing the effect on policing of demeanor of juvenile suspects).

[62] In her study of more than 500 NYPD patrol officers, Edith Linn found that one significant factor affecting "non-essential arrests" was personal commitments such as second jobs and dependent family members. *See* EDITH LINN, ARREST DECISIONS: WHAT WORKS FOR THE OFFICER? 125–26 (2009).

[63] *See, e.g.*, Selwyn Raab, *Arrests Drop in New York City*, *Easing Court Delays*, N.Y. TIMES, Oct. 6, 1976, at 41 ("'The cops have no time now for the junk cases,' said one assistant district attorney in Brooklyn. 'They're too preoccupied with getting off work and picketing, rather than hanging around the courts.'").

[64] *See id.* (reporting that "police supervisors, after getting complaints[,] discouraged 'bounty hunters'—that is, police officers who make questionable arrests in order to get overtime pay").

[65] *See* Peter Moskos, *The Better Part of Valor: Court-Overtime Pay as the Main Determinant for Discretionary Police Arrests*, 8 L. ENFORCEMENT EXECUTIVE F. 77, 92 (2008) (discussing research that finds "arrests in high-drug areas [of Baltimore] are primarily the result of an officer's desire for court-overtime pay"). *But see* LINN, *supra* note 62, at 83 (finding in her NYPD study that "[b]ecause of arrest processing difficulties, officers are 'turned off' to arrest processing nearly half the time").

[66] *See* Moskos, *supra* note 65, at 85, 88–90.

[67] The most recent census of state and local law enforcement agencies reports that there are about 765,000 full-time sworn (i.e., with general arrest powers) law enforcement officials around the country. BRIAN A. REAVES, BUREAU OF JUSTICE STATISTICS, U.S. DEP'T OF JUSTICE, CENSUS OF STATE AND LOCAL LAW ENFORCEMENT AGENCIES, 2008, at 1 (2011), https://www.bjs.gov/content/pub/pdf/csllea08.pdf [https://perma.cc/T9XC-GUTM]. There are, by contrast about 120,000 full-time sworn officers employed by the federal government. BRIAN A. REAVES, BUREAU OF JUSTICE STATISTICS, U.S. DEP'T OF JUSTICE, FEDERAL LAW ENFORCEMENT OFFICERS, 2008, at 1 (2012), https://www.bjs.gov/content/pub/pdf/fleo08.pdf [https://perma.cc/L7VP-G9BX].

[68] *See, e.g.*, Leonard Buder, *Police Officers in New York City Grumble, with Some Pride, About 'The Job,'* N.Y. TIMES, June 29, 1980, at 18 ("Many officers also feel that efforts to hold down overtime costs impair their ability to make arrests. 'You get harassed on collars involving overtime,' said an officer in Brownsville, echoing the views of many others. 'It can't help but slow you down.'").

deemed to be low priority.[69] Those crimes have not disappeared, but for official purposes they virtually have.

The decentralized nature of American policing also helps to explain the variation in enforcement priorities—either formal or informal—that departments recognize. Broken windows policing[70]—an approach focused on order maintenance and the enforcement of minor offenses—was widely embraced by many urban police departments in the 1990s.[71] The adoption of that approach also led to higher numbers of arrests for minor offenses.[72]

In recent years, police departments have also succumbed to the pressures of managerial techniques that emphasize quantitative measures of effective policing.[73] If arrests, stops, and citations become gauges of effective policing, increased rates of enforcement follow.[74] The opposite can happen as well. Local alarm at perceived increases in the crime rate can force police departments to discourage aggressive enforcement.[75] That pressure may result in the lack of official recognition of some

---

[69] There are numerous examples of departments making such choices because of budgetary constraints. *See, e.g.*, Matt Bigler, *Cash-Strapped San Jose Police Won't Respond to Low-Priority Calls*, CBS S.F. (Aug. 16, 2011, 11:32 AM), http://sanfrancisco.cbslocal.com/2011/08/16/big-changes-to-cash-strapped-san-jose-department-response-policy/ [https://perma.cc/D9B9-J2FX] (noting a new policy of considering "anonymous noise complaints, burglar alarms and non-injury car crashes . . . under the category of 'non-response'"); Tom DuHain, *Stockton Police to Focus on Violent Crime*, KCRA (June 1, 2012, 7:31 AM), http://www.kcra.com/article/stockton-police-to-focus-on-violent-crime/6397012 [https://perma.cc/2GZY-8QA7] (reporting that "[p]olice in Stockton will no longer respond to low-priority, property crimes in order to free officers for quick response to violent crimes").

[70] George L. Kelling and James Q. Wilson famously outlined the approach in their 1982 *Atlantic* article. *See* George L. Kelling & James Q. Wilson, *Broken Windows: The Police and Neighborhood Safety*, ATLANTIC (Mar. 1982), https://www.theatlantic.com/magazine/archive/1982/03/broken-windows/304465/ [https://perma.cc/R5LJ-A8WF] ("Arresting a single drunk or a single vagrant who has harmed no identifiable person seems unjust, and in a sense it is. But failing to do anything about a score of drunks or a hundred vagrants may destroy an entire community."). Kelling would extend this idea later in GEORGE L. KELLING & CATHERINE M. COLES, FIXING BROKEN WINDOWS: RESTORING ORDER AND REDUCING CRIME IN OUR COMMUNITIES (1996).

[71] *See, e.g.*, KELLING & COLES, *supra* note 70, at 3.

[72] *See, e.g.*, PREETI CHAUHAN ET AL., TRENDS IN MISDEMEANOR ARRESTS IN NEW YORK 76 (2014) ("The residents of New York City . . . have experienced significant increases in the numbers and rates of misdemeanor arrests from 1980 to 2013.").

[73] In this way, predictive policing is conceptually similar to "intelligence-led policing (ILP), data-driven policing, risk-based policing, 'hot spots' policing, evidence-based policing, and pre-emptive [policing]." Moses & Chan, *supra* note 25, at 3 (citations omitted).

[74] *See, e.g.*, Veronica Rocha, *Whittier Police Officers Sue, Say They Were Forced to Meet Quotas*, L.A. TIMES (Mar. 4, 2015, 4:40 PM), http://www.latimes.com/local/lanow/la-me-ln-whittier-ticket-quotas-20150304-story.html [https://perma.cc/P4NR-Q6F6] (reporting lawsuit brought by six Whittier, California, police officers, contending that they faced retaliation for refusing to meet ticket and arrest quotas).

[75] *See, e.g.*, Matthias Gafni, *Pittsburg: Whistleblower Cops Claim Department Falsified*

crimes, and the intentional "downgrading" of serious crimes to minor offenses for official record-keeping.[76]

Not only do police officers and departments wield significant amounts of discretion about what they see and know about crime, so too do individuals and other would-be victims of crime. Banks will reliably call the police if they are robbed; drug dealers will not.[77] When Walmart,[78] the world's largest retailer, decides it will no longer report certain categories of detected theft to local police, those crimes disappear from official view.[79] When it decides it will refer shoplifters to the police at a younger age (from eighteen to sixteen), as it actually did a year after changing its petty theft limit, that has an impact on crime data as well.[80] Crime reporting also varies widely by race, class, and ethnicity.[81] And even those variations can themselves change over time. For example, a perception that local police will help federal officials identify undocumented persons may discourage some of those persons from reporting

---

Reports, Failed to Document Use of Force, EAST BAY TIMES (Aug. 12, 2016, 5:58 PM), http://www.eastbaytimes.com/2016/08/12/pittsburg-whistleblower-cops-claim-department-falsified-reports-failed-to-document-use-of-force/ [https://perma.cc/W27X-JVRV].

[76] *See, e.g.*, *id.* (describing lawsuit by police officers claiming that department policy classified some crimes "as 'suspicious circumstances' rather than felonies to avoid reporting them to the FBI and have them counted as part of the city's crime rate").

[77] *See* Carl B. Klockars, *Some Really Cheap Ways of Measuring What Really Matters*, *in* MEASURING WHAT MATTERS: PROCEEDINGS FROM THE POLICING RESEARCH INSTITUTE MEETINGS 195, 195 (Robert H. Langworthy ed., 1999) ("If I had to select a single type of crime for which its true level—the level at which it is reported—and the police statistics that record it were virtually identical, it would be bank robbery. Those figures are likely to be identical because banks are geared in all sorts of ways . . . to aid in the reporting and recording of robberies and the identification of robbers. And, because most everyone takes bank robbery seriously, both Federal and local police are highly motivated to record such events.").

[78] Walmart can have an outsized influence on crime because of its sheer size. In 2016, violent crimes happened at a rate of about one per day at the company's properties around the country. Shannon Pettypiece & David Voreacos, *Walmart's Out-of-Control Crime Problem Is Driving Police Crazy*, BLOOMBERG BUSINESSWEEK (Aug. 17, 2016), https://www.bloomberg.com/features/2016-walmart-crime/ [https://perma.cc/GS5J-QVK6]. There were also likely "hundreds of thousands" of petty crimes committed on its properties as well in one year. *See id.*

[79] *See* Michael Barbaro, *Wal-Mart Eases Policy on Petty Shoplifters*, N.Y. TIMES (July 13, 2006), https://www.nytimes.com/2006/07/13/business/worldbusiness/13iht-wal-mart.2190898.html?smid=pl-share (reporting corporate policy change where shoplifters caught stealing goods worth less than $25 dollars will not be reported to police).

[80] *See Wal-Mart Lowers Age in Shoplifting Policy*, NBC NEWS (July 11, 2007, 6:47 PM), http://www.nbcnews.com/id/19714916/ns/business-us_business/t/wal-mart-lowers-age-shoplifting-policy/#.WXbdnogrLyR/ [https://perma.cc/73RL-AWY8].

[81] *See, e.g.*, Shapiro, *supra* note 11 ("Norms differ for reporting crime across lines of race, class and ethnicity. Foreign-born citizens . . . are less likely to report crimes than are US-born citizens.").

crimes or agreeing to be witnesses.[82] And even when crimes are reported, the police do not necessarily record them at all.[83]

That all of these factors influence official accounts of crime is well-known to those who study the police.[84] It may be less well-known, let alone acknowledged, by those who develop the algorithms for predictive programs.[85] While algorithmic policing programs vary in the types of information they employ, most at a minimum rely on historical data compiled by the police themselves. This can include both reported crimes and crimes discovered by the police.[86]

Predictive policing systems are as good as the data they possess.[87] By design, machine learning algorithms learn and reproduce the data they are given.[88] If the

---

[82] LAPD Police Chief Charlie Beck said "reports of sexual assault have dropped 25% among the city's Latino population since the beginning of 2017 compared with the same period last year [and added] that reports of domestic violence have fallen by 10%." James Queally, *Latinos Are Reporting Fewer Sexual Assaults Amid a Climate of Fear in Immigrant Communities, LAPD Says*, L.A. TIMES (Mar. 21, 2017, 8:25 PM), http://www.latimes.com/local/lanow/la-me-ln-immigrant-crime-reporting-drops-20170321-story.html [https://perma.cc/46MH-2UCH]. No similar decreases were seen with other ethnic groups in the city. *See id.*; *see also* Rob Arthur, *Latinos in Three Cities Are Reporting Fewer Crimes Since Trump Took Office*, FIVETHIRTYEIGHT (May 18, 2017), https://fivethirtyeight.com/features/latinos-report-fewer-crimes-in-three-cities-amid-fears-of-deportation/ [https://perma.cc/JBS2-NKVZ] (citing analysis of crime reports from Dallas, Denver, and Philadelphia that "supports the notion that immigrants, or Latinos more generally, could be reporting fewer crimes since Trump took office"); Brooke A. Lewis, *HPD Chief Announces Decrease in Hispanics Reporting Rape and Violent Crimes Compared to Last Year*, HOUSTON CHRON. (Apr. 6, 2017, 10:01 AM), http://www.chron.com/news/houston-texas/houston/article/HPD-chief-announces-decrease-in-Hispanics-11053829.php [https://perma.cc/Z442-XQ6R] (noting that Police Chief Acevedo "emphasized the importance of enforcing laws with immigrant communities in a way that does not create fear").

[83] For example, a 2014 report by the United Kingdom's Her Majesty's Inspectorate of Constabulary (HMIC) found that about one in five crimes reported to the police each year were not recorded by the police. *See* HMIC, CRIME-RECORDING: MAKING THE VICTIM COUNT 19 (2014) (calling the failure rate "deplorable").

[84] *See, e.g.*, Klockars, *supra* note 77, at 195 ("It has been known for more than 30 years that, in general, police statistics are poor measures of true levels of crime. This is in part because citizens exercise an extraordinary degree of discretion in deciding what crimes to report to police, and police exercise an extraordinary degree of discretion in deciding what to report as crimes.").

[85] *See Statement of Concern About Predictive Policing by ACLU and 16 Civil Rights Privacy, Racial Justice, and Technology Organizations*, ACLU (Aug. 31, 2016), https://www.aclu.org/other/statement-concern-about-predictive-policing-aclu-and-16-civil-rights-privacy-racial-justice [https://perma.cc/HED8-S3WV] ("Decades of criminology research have shown that crime reports and other statistics gathered by the police primarily document law enforcement's *response* to the reports they receive and situations they encounter, rather than providing a consistent or complete record of all the crimes that occur.").

[86] *See* ROBINSON & KOEPKE, *supra* note 25, at 6–8 (reporting survey results of vendors of predictive police systems).

[87] *See* Lum & Isaac, *supra* note 22, at 15 (explaining that if bias data is used to train models, the models will reflect that bias).

[88] *Id.* at 16.

data police provide to these systems already reflects a variety of priorities, filters, and decisions, then the results will too repeat those choices. And as police rely upon these predictive policing results to deploy their resources, they produce even more data that appear to confirm what the algorithm has predicted. That feedback loop reproduces a pattern of future policing, not future crime.

Data scientists Kristian Lum and William Isaac demonstrated the impact of this feedback loop on patterns of policing in Oakland, California.[89] The study authors used two sources of comparison: public health survey derived estimates about patterns of illegal drug use, and predictions based on the algorithm provided by the PredPol company to predict patterns of drug crimes (which would be used by the police to guide their enforcement activity).[90]

A comparison of these figures "tell[s] dramatically different stories about the pattern of drug use in Oakland."[91] Drug arrests have happened—and more importantly would likely happen—chiefly in poor and non-white neighborhoods.[92] If relied upon by the police, the algorithm would flag areas of the city that are "already over-represented in the historical police data."[93] Sending police to places where they have been before makes it all the more likely to observe new crimes that confirm the predictions.[94] Fed with the new data, the algorithm's model becomes more confident that the predictions about these places were right.[95] Thus, "selection bias meets confirmation bias."[96] Lum and Isaac conclude from their data that "predictive policing of drug crimes results in increasingly disproportionate policing of historically over-policed communities."[97]

Their insights about crime data and policing, however, can be applied more broadly to all of the crime data that the police see or do not see. Any predictive policing model that uses crime data will always already have been filtered through these processes. It is that filtered data, not any hypothetical measure of "real crime," that predictive policing algorithms rely upon. To some degree, then, predictive programs will tend to identify people and places that reflect prior police contacts, and they will not identify those that have been ignored.

---

89 *Id.* at 14–17.

90 The study's authors used National Survey on Drug Use and Health data to create a "synthetic population." *Id.* "A synthetic population is a demographically accurate individual-level representation of a real population . . . [where individuals identified by] sex, household income, age, race, and the geo-coordinates of their home . . . are assigned so that the demographic characteristics in the synthetic population match data from the US Census at the highest geographic resolution possible." *Id.* at 16.

91 *Id.* at 17.

92 *Id.*

93 *Id.* at 18.

94 *Id.* at 16.

95 *Id.*

96 *Id.*

97 *Id.* at 19.

This general pattern—that police will repeat past behavior—is not necessarily a feature of big data programs used by the police. Any human decisions that distribute limited resources can be subject to the same biases and omissions discussed here.[98] Predictive policing systems cannot eliminate these problems, even if their developers lay claim to their objectivity or to use of the scientific method. As long as crime and crime detection are mostly human activities, they will reflect human shortcomings.

Big data programs, however, raise distinct challenges. The use of algorithmic decisionmaking further obscures the human decisionmaking behind the crime data. The discretion and biases inherent in the production of crime data become difficult to challenge by those who may be affected directly by these algorithms, and yet unable to understand or access this black-box decisionmaking. What is more, increasing reliance on these systems "shifts accountability from departmental decision-makers to black-box [algorithms]."[99]

CONCLUSION

The objective of stopping crime before it happens is alluring.[100] That appeal explains why crime prediction has long been a subject of fascination in science fiction.[101] With dozens of police departments adopting or testing predictive policing algorithms, crime forecasting programs will likely become commonplace.[102] And while the adoption of algorithmic decisionmaking may appear to solve issues about resources, efficiency, and discretion, a closer look at the "raw data" fed to these algorithms reveals some familiar problems. Many of these issues will become even more difficult to identify as algorithmic decisionmaking becomes integrated into larger data management systems used by the police. A predictive crime program could be merged, for instance, with GPS data on individual officers and body camera video.[103] Yet as long as policing is fundamentally a set of decisions by people about other people, the data fed to the machine will remain a concern.

---

[98] *Id.* (discussing past processes by which human analysts allocated police resources, allowing police chiefs to justify policing decisions).

[99] *Id.*

[100] *See* PERRY ET AL., *supra* note 3, at 8 ("There is an obvious appeal to being able to prevent crime as opposed to merely apprehending offenders after a crime has been committed.").

[101] The classic story, of course, is Philip K. Dick, *The Minority Report*, *in* FANTASTIC UNIVERSE 4 (Leo Margulies ed., 1956).

[102] *See supra* notes 87–99 and accompanying text.

[103] *See* Shapiro, *supra* note 11 ("Developments to more-integrated systems are likely also to incorporate the locations of individual police officers from Global Positioning System data, as well as footage from body-worn cameras.").