I/S: A JOURNAL OF LAW AND POLICY FOR THE INFORMATION SOCIETY

The Challenges of Prediction: Lessons from Criminal Justice

DAVID G. ROBINSON*

CONTENTS

* Managing Director and co-founder, Upturn; Adjunct Professor of Law, Georgetown University Law Center. The thoughts in this essay were honed through rich exchanges and shared experience with my colleagues at Upturn, the students in my Spring 2017 seminar on Governing Automated Decisions at Georgetown Law, and a constellation of friends, collaborators and mentors too numerous to list here. I am grateful to them all – particularly to Lewis Robinson for his thoughtful and thorough feedback – and take sole responsibility for whatever errors remain.

I. INTRODUCTION

Government authorities at all levels increasingly rely on automated predictions, grounded in statistical patterns, to shape people's lives. In this, the government follows a trend pioneered in the private sector, and reflected across society: a turn toward quantitative, empirical approaches to decision-making. The move is visible in everything from the piecemeal targeting of political campaigns; to the adaptive learning systems fast spreading through school classrooms; to the pricing and marketing of staplers.[1]

A daunting thicket of jargon surrounds these trends,[2] but no matter how one tells the story, digital technology lies at its core. The cost of digital sensors, data storage, and processing has been plummeting exponentially for more than 50 years,[3] while research in

[1] Jennifer Valentino-DeVries et al., *Websites Vary Prices, Deals Based on Users' Information*, WALL ST. J. (Dec. 24, 2012), https://www.wsj.com/articles/SB10001424127887323777204578189391813881534 [https://perma.cc/B7XQ-Q4KW] (A vibrant if unruly garden of newly-coined terms have arisen to point toward these developments, each phrase offering differences of emphasis and tone).

[2] VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK 83-87 (2013) (data at scale, the "quantified society," see *Quantified Society: Examining the Consequences of Algorithmic Decision Making for Open Societies*, OPEN SOC'Y FOUND. (Feb. 27, 2015), https://scholarshipdb.net/jobs-in-United-States/Quantified-Society-Examining-The-Consequences-Of-Algorithmic-Decision-Making-For-Open-Societies-Open-Society-Foundations=UJUUA6G_5BGUNAAlkGUTnw.html [https://perma.cc/2NKJ-33NT], among other frames whose core foci overlap and are traceable to digital technology).

[3] *See* C.A. Mack, *Fifty Years of Moore's Law*, 24 IEEE TRANSACTIONS ON SEMICONDUCTOR MANUFACTURING 202 (2011).

"machine learning" has recently brought rapid advances in computers' capacity to find and act upon patterns in stored data.[4] The excitement sparked by newly developed applications is also contributing to a broad renewal of public and policymaker interest in statistical and empirical methods, including in some longstanding techniques.[5] These technologies can do much more than make predictions: They recognize and translate speech, diagnose disease, generate art, and do much else. But predicting the future—and relying on those predictions to *change* what happens next—is a favorite application of these new methods.

Software that wields government power deserves special attention, particularly when it uses historical data to decide automatically what ought to happen next. Such software already decides which Medicare claims to pay out and which ones to deny as likely fraudulent;[6] whose tax returns to audit;[7] which airport travelers to pluck out of line for secondary screening;[8] and a myriad of other judgments.[9] These are characteristically governmental uses of "predictive analytics;" and they do, as the organizers of this symposium astutely observe, raise "cross-cutting issues."[10]

---

[4] These developments are broadly in the field of machine learning, which is itself a subfield of artificial intelligence, the general project of getting computers to assume tasks and make decisions formerly reserved for humans. *See* MACHINE LEARNING: THE POWER AND PROMISE OF COMPUTERS THAT LEARN BY EXAMPLE, THE ROYAL SOCIETY 18-19 (Apr. 2017) https://royalsociety.org/~/media/policy/projects/machine-learning/publications/machine-learning-report.pdf [https://perma.cc/2HYH-ZAVP].

[5] COMPAS, for example, was introduced in 1996. *See* T. Brennan et al., *Evaluating the Predictive Validity of the Compas Risk and Needs Assessment System*, 36 CRIM. JUST. & BEHAV. 21, 21-23 (2008).

[6] *See* Medicare Fraud Prevention: CMS Has Implemented a Predictive Analytics System, but Needs to Define Measures to Determine Its Effectiveness (U.S. Gov't Accountability Off., Oct. 2012); Joe Eaton, *Glitch in the Machine*, PACIFIC STANDARD (Jan. 11, 2016), https://psmag.com/news/how-to-scam-medicare-in-four-easy-steps [https://perma.cc/5KWV-P3Q4].

[7] Eaton, *supra* note 7.

[8] *Id.*

[9] *See* MICHAEL VEALE, MAX VAN KLEEK & REUBEN BINNS, FAIRNESS AND ACCOUNTABILITY DESIGN NEEDS FOR ALGORITHMIC SUPPORT IN HIGH-STAKES PUBLIC SECTOR DECISION-MAKING 1 (Feb. 2018), https://arxiv.org/abs/1802.01029 [https://perma.cc/U2K8-B84S].

[10] *See* Dennis Hirsch, *Predictive Analytics Law and Policy: A New Field Emerges*, 14 I/S: J.L. & POL'Y FOR INFO. SOC'Y 1, 1 (2018).

The stakes for governmental use of predictive analytics are particularly high in criminal justice, where the public authority being applied or shaped by software extends to physical coercion and sometimes deadly force.

In this article, I draw examples primarily from the domain of criminal justice—and in particular, the intersection of civil rights and criminal justice—to illustrate structural challenges that sweep more broadly across governmental uses of predictive analytics; and that have some risk of arising whenever law or public policy contemplates adopting predictive analytics as a tool.

These challenges take the form of a gap or disconnect: A difference between the circumstances that would be ideal for applying predictive analytics, and realities within which the government operates. These challenges are likely to arise wherever the government uses predictive analytics. Policymakers and the public would do well to take note.

Section II describes three core tensions in predictive analytics; illustrating with examples from the criminal justice domain. Section III responds to these challenges by describing feasible ways to perform prediction more successfully. Section IV concludes.

II. Three Core Tensions in Criminal Justice Prediction

The literature on predictive analytics in criminal justice repeatedly illustrates three core tensions. In the sections that follow, I describe and illustrate each of them:

1) What matters versus what the data measure;
2) Current goals versus historical patterns; and
3) Public authority versus private expertise.

These tensions are likely to arise whenever the government embraces statistical prediction (the first two reach even more broadly, arising in private and public contexts alike).

These challenges do not suggest that predictive analytics lack any rightful place in government. The key insight is simply that each challenge needs to be considered and addressed whenever predictive analytics are used to wield government power.

A. *What Matters Versus What the Data Measure*

There is often a gap between what is legally or politically significant, and what public sector organizations can or do measure.

Such disconnects have always been a challenge for large organizations that seek to measure and manage their own activities, long predating the current wave of interest in predictive analytics. "Campbell's law," articulated by the psychologist Donald Campbell in 1979, observes that "the more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor."[11] Campbell wrote in the context of educational testing in schools; observing that although the tests were intended to measure general educational performance, they inevitably lost explanatory power—and distorted the school's activities—once teachers recognized the tests' importance, and adapted by focusing pedagogy on improving students' scores.[12] James C. Scott has traced this same pattern back centuries earlier, observing that bureaucracies, in their need to make their world "legible," often distort and misunderstand local knowledge.[13] He describes such dynamics across a myriad of contexts; from the emergence of centralized land tenure in Southeast Asia (which replaced and destroyed complex, subtle village-level practices of land and resource sharing) to the failures of central planning in 20th century authoritarian communism.[14]

This problem takes on special urgency today in the context of data-driven criminal justice. The criminal justice system is a primary driver of deprivations of liberty and has lately earned significant, sustained critique. Most of its leaders have neither firsthand knowledge nor readily available staff expertise in statistics, data science, or related fields. Yet the criminal justice system is increasingly in thrall to the cultural prestige of "data;" allowing optimistic presumptions to replace critical scrutiny. The embrace of predictive analytics in criminal justice is, as a result, all too often both naïve and consequential.

The initial choice of what data to measure—and to base predictions on—is vitally important because it will shape what later *counts* as an accurate or successful prediction.

[11] Donald T. Campbell, *Assessing the Impact of Planned Social Change*, 2 EVALUATION & PROGRAM PLAN. 67, 85 (1979).

[12] *Id.*

[13] *See generally* JAMES C. SCOTT, SEEING LIKE A STATE: HOW CERTAIN SCHEMES TO IMPROVE THE HUMAN CONDITION HAVE FAILED (1998).

[14] *Id.*

1. *Stops, Citations, and Arrests are Limited Proxies for Police Performance*

Police wield extensive, daily power over life and liberty, and they are effectively unregulated by many of the constitutional safeguards that constrain other applications of governmental power.[15] This makes it vitally important that they define and measure their performance in ways that encourage balanced attention to all the goals of the communities they serve; including public order, humane and dignified interactions between police and civilians, and a minimization of the collateral consequences both of crime and of law enforcement.

Unfortunately, the things police most often measure – such as arrests, citations and stops – are poor proxies for these balanced outcomes.

"Arrest rates are not, as many seem to believe, measurements of crime. Arrest rates are measurements of a particular type of law enforcement behavior—arresting suspects."[16]

As one eminent criminologist observed, the criminal justice establishment too often ignores these facts: "most of the research on the parameters of a criminal career utilizes arrest data to estimate the underlying behavioral dynamics of criminal activity. We have fallen into bad habits . . . [those arrested] are not a representative sample of offenders in the population and findings from arrest samples should not be generalized to offenders or offenses in the general population."[17]

The extent of the difference between arrest rates and actual crime rates is very hard to figure, but one key observation in the criminology literature is that some types of arrests are more enforcement-driven, and others are relatively more influenced by the actual, true crime rate. For example, it is likely that all (or nearly all) bank robberies are reported to the police.[18] On the other hand, marijuana possession

---

[15] *See generally* BARRY FRIEDMAN, UNWARRANTED: POLICING WITHOUT PERMISSION (2017).

[16] David A. Harris, *The Reality of Racial Disparity in Criminal Justice: The Significance of Data Collection*, 66 LAW & CONTEMP. PROBS. 71, 80 (2003).

[17] DELBERT S. ELLIOTT, LIES, DAMN LIES, AND ARREST STATISTICS, 1-8 (1995).

[18] Carl B. Klockars, *Some Really Cheap Ways of Measuring What Really Matters*, *in* MEASURING WHAT MATTERS: PROC. FROM THE POLICE RES. INST. MEETINGS 195, 195-201 (1999), https://www.ncjrs.gov/pdffiles1/nij/170610.pdf. [https://perma.cc/BRP3-6Z79] ("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

arrests are notoriously biased, with black Americans much more likely to be arrested than whites who use the drug at similar rates.[19]

Yet arrests – and stops and citations, which suffer from similar challenges – play a central role in how police define and measure their success. New York City's CompStat, a tool developed inside the New York Police Department in 1994, was a pioneering system.[20] The original CompStat involved computerized analysis and mapping of the police department's own crime statistics, together with weekly meetings where police leaders were held accountable based on changes in those statistics.[21] Most big city police forces now have some version of a CompStat-style system.[22]

First-generation CompStat-style crime analysis and prediction systems—focusing solely on crime numbers generated by the police themselves[23]—have reshaped policing, and there is now a rich literature about its impact. A description of the impact in Minneapolis is typical: "In the 1990s, it officially established its crime analysis program, which has since grown into a large-scale report management system and predictive policing effort that is integrated with the department's weekly strategic planning."[24] One of the inventors of

---

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 mostly everyone takes bank robbery seriously, both Federal and local police are highly motivated to record such events.").

[19] ACLU, THE WAR ON MARIJUANA IN BLACK AND WHITE: BILLIONS OF DOLLARS WASTED ON RACIALLY BIASED ARRESTS 4 (June 2013) https://www.aclu.org/report/report-war-marijuana-black-and-white [https://perma.cc/NG3Y-P2BX].

[20] POLICE EXECUTIVE RESEARCH FORUM, COMPSTAT: ITS ORIGINS, EVOLUTION, AND FUTURE IN LAW ENFORCEMENT AGENCIES 2, 30 (2014) . https://www.bja.gov/publications/perf-compstat.pdf [https://perma.cc/L65K-A7ND].

[21] *See id.*

[22] In a 2011 survey conducted by the Police Executive Research Forum, 79% of responding U.S. police departments indicated they are now using some version of this system. *Id.* at 36.

[23] It is important to note that CompStat is an umbrella term for an approach, which different departments implement in their own ways. The original CompStat used crime data only, and most discussions of the CompStat approach focus on crime statistics. Nonetheless, some departments may now be using CompStat in ways that do include other data. *Id.* at 23-36.

[24] WALTER L. PERRY ET AL., PREDICTIVE POLICING: THE ROLE OF CRIME FORECASTING IN LAW ENFORCEMENT OPERATIONS 74 (2013),

"broken windows" policing, writing about the NYPD's Giuliani-era reforms in a generally upbeat essay for a conservative think tank, described CompStat as "a new way of *managing police resources and tactics*" and called it "perhaps the single most important organizational/administrative innovation in policing during the latter half of the 20th century."[25]

Such systems can distort both the statistics that a department collects, and the actual policing activities that are shaped by those numbers. The pressure to "produce" arrests amounts to the tail wagging the dog. A 2012 study by John Eterno, a criminology scholar and former NYPD officer who is one of CompStat's leading critics, surveyed 2,000 retired NYPD officers about the integrity of crime statistics and their impact. As summarized by the New York Times, the study found that "pressure on officers to artificially reduce crime rates, while simultaneously increasing summonses and the number of people stopped and often frisked on the street, has intensified in the last decade."[26] As Eterno wrote in an op-ed:

> Eighteen years after the start of the much-vaunted CompStat system of data-driven crime fighting, the Police Department . . . has become a top-down, micromanaged bureaucracy in which precinct commanders are pitted against one another and officers are challenged to match or exceed what they did the previous year, month and week. Words like "productivity" are code for quotas . . . Reducing crime numbers is simple if you disregard basic rights, ignore victims, mandate quotas and manipulate numbers. The lunacy of this police performance culture can end only

https://www.rand.org/content/dam/rand/pubs/research_reports/RR200/RR233/RAND_RR233.pdf [https://perma.cc/ZEM8-K8UN].

[25] George L. Kelling & William H. Sousa, Jr., *Do Police Matter? An Analysis of the Impact of New York City's Police Reforms*, 22 CIVIC REP. 2, 2 (Dec. 2001), https://www.manhattan-institute.org/pdf/cr_22.pdf [https://perma.cc/YU3P-4LQS] (emphasis added).

[26] Wendy Ruderman, *New York Police Department Manipulates Crime Reports, Study Finds*, N.Y. TIMES (Jun. 28, 2012), http://www.nytimes.com/2012/06/29/nyregion/new-york-police-department-manipulates-crime-reports-study-finds.html.

> with stronger leadership, greater transparency and more meaningful relationships with communities.[27]

An expert on police performance management has observed that CompStat "picked up and encouraged . . . the focus on disorder offenses as a way of reducing fear," but at the same time the system effectively de-emphasized the view that "both the legitimacy and effectiveness of police could be increased by reaching out for effective working partnerships with community groups, and focusing attention on community-nominated problems that might or might not include serious crime problems."[28]

Indeed, this disconnect between what is measured and what matters appears to plague many police departments that have followed New York City's example. Many of the big cities that use CompStat are also committed, at least in principle, to "community policing,"[29] a goal typically defined in terms of responding directly to community priorities in the allocation of police resources. Empowering police to address community concerns—partly through increased autonomy for precinct commanders—was a big part of the rhetoric behind CompStat-style systems.[30] But it was crime, not community concerns, that got routinely measured, and evidence from the field seems to show that in the end, it is the numbers that matter. One study of a half-dozen departments that are committed to both community policing and CompStat found that:

> CompStat's contribution to a data-rich environment helped sergeants identify emerging crime (but not community) problems and focus patrol resources, but it

---

[27] John A. Eterno, Opinion, *Policing by the Numbers*, N.Y. TIMES (Jun. 17, 2012), http://www.nytimes.com/2012/06/18/opinion/the-nypds-obsession-with-numbers.html.

[28] MARK HARRISON MOORE ET AL., RECOGNIZING VALUE IN POLICING: THE CHALLENGE OF MEASURING POLICE PERFORMANCE 175 (2002).

[29] James J. Willis, *First-Line Supervision Under CompStat and Community Policing: Lessons From Six Agencies,* CTR. FOR JUST. LEADERSHIP AT GEORGE MASON UNIV. (Mar. 27, 2011), https://ric-zai-inc.com/Publications/cops-p204-pub.pdf [https://perma.cc/5CUQ-H5CE].

[30] *See, e.g.*, Kelling & Sousa Jr, *supra* note 26, at 2-8 ("Bratton invested enormous authority in precinct commanders, devolving both resources and decision-making to the precinct level . . . This administrative mechanism focused the NYPD on substantive community problems").

> had done little to promote innovative responses to those problems. . . *Because none of these departments had implemented similarly sophisticated data systems to support community policing,* sergeants did not mention receiving information that helped them systematically identify community problems, determine priorities, and document results. Consequently, sergeants tended to learn about those issues on a more ad hoc basis.[31]

Meanwhile, other dimensions of police performance that are of vital interest to the communities they serve too often go unmeasured and unaccounted for. The missed opportunities and distortions introduced by this disconnect are further described in section III(A)(2), below.

2. *Risk of Re-arrest is a Poor Proxy for Dangerousness*

Across the country judges and court systems are increasingly using predictive analytics at bail to predict something that most jurisdictions require judges to assess: the risk that the defendant will, if released pending trial, fail to appear or be a danger to the community.[32]

Unfortunately, the outcome variable that these systems typically use as a measure of a defendant's dangerousness is "rearrests within a set period of time," rather than patterns of violence. In other words, these tools do not make statistical generalizations about whether defendants similar to the one at bar generally tend to commit violent crimes during pre-trial release. Instead, most tools fielded today can only make generalizations about whether defendants similar to the one at bar, who were released before trial, were re-arrested. And most re-arrests are for technical, minor infractions; not for new, violent crimes.

The results of this disconnect could be heart-wrenching: defendants may be classed as "high risk," stigmatized as apt to be

---

[31] *See* Willis, *supra* note 30, at 3, 11.

[32] For a fuller exploration of the ideas in this section, see John L. Koepke & David G. Robinson, *Danger Ahead: Risk Assessment and the Future of Bail Reform*, WASH. L. REV. (Forthcoming 2018), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3041622 [https://perma.cc/5ALU-DBN7].

violent, and jailed, when in fact they pose a low risk of violence, and simply have a high probability of being re-arrested for a technicality.

B. *Current Goals Versus Historical Patterns*

Across a wide range of contexts in criminal justice, the available data upon which predictions can be based reflect historical patterns that are widely recognized as unjust and undesirable.

Discussions of predictive tools sometimes assume that the tools themselves merely *describe* and *anticipate* future patterns. The reality is far more challenging: Because the predictions made in criminal justice systems are action-guiding, the very act of predicting frequently changes what happens in the future. For example, if a predictive policing system predicts that a given neighborhood is apt to see more crime, and police respond by deploying more patrols in that neighborhood, they will be better able to detect crime that happens there than they are to detect crime that happens elsewhere; and are likely as a result to find and document more crime in that neighborhood.

Bernard Harcourt has termed this issue a “ratchet effect.” He describes the problem tidily in his book *Against Prediction*, writing:

> [W]hen the police profile higher-offending individuals, they are effectively sampling more from that higher-offending group. The resulting set of successful searches will contain a disproportionate number of those high-offending individuals—disproportionate as compared to their representation in the offending population. This imbalance will get incrementally worse each year if law enforcement departments rely on the evidence of last year’s correctional traces—arrest or conviction rates—in order to set next year’s profiling targets.[33]

The ratchet effect is a domain-specific example of the problem of “hidden feedback loops,” which bedevils modern applications of

33 Bernard E. Harcourt, *A reader's companion to 'Against Protection*' 265, 271 (U. of Chi. Pub. L. & Legal Theory, Working Paper No. 175, 2007), https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=1007&context=public_law_and_legal_theory [https://perma.cc/H7PX-J79M].

predictive analytics generally,[34] and which is recognized across many applications of machine learning and other predictive statistical techniques.

1. *Predictive Policing and Self-Fulfilling Prophecies*

PredPol is a leading vendor of predictive analytics software for police. Its system forecasts where and when various types of crime may happen, relying on police-held administrative data about the time, location and type of previously reported crimes.[35] The company claims that, because its system "uses no personal information about individuals or groups of individuals," the tool does not raise "any personal liberties and profiling concerns."[36] A mathematician and an anthropologist, who were involved in starting the company, have also published a peer-reviewed study that details the mathematics behind the company's approach and reports a field trial in which software using this approach outperformed human crime analysts.[37]

Recently, two statisticians from the Human Rights Data Analysis Group (HRDAG) used PredPol's approach as a case in point, to investigate what might happen if police were to rely on predictive

---

[34] D. SCULLEY ET AL., MACHINE LEARNING: THE HIGH-INTEREST CREDIT CARD OF TECHNICAL DEBT 1, 3 (2014) https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/43146.pdf [https://perma.cc/86EB-3XRP] (the authors use a hypothetical example of a predictive system designed to maximize the click-through rate on news headlines – rather than a predictive policing system – to illustrate the same troubling dynamic:

> Imagine in our news headline [click through rate] prediction system that there is another component of the system in charge of "intelligently" determining the size of the font used for the headline. If this font-size module [uses the click-through rate] as an input signal, and font-size has an effect on user propensity to click, then the inclusion of CTR in font-size adds a new hidden feedback loop. It's easy to imagine a case where such a system would gradually and endlessly increase the size of all headlines).

[35] *About Us*, PREDPOL, http://www.predpol.com/about/ [https://perma.cc/N8LW-NVBD]. ("PredPol uses only three data points in making predictions: past type of crime, place of crime and time of crime").

[36] *Id.*

[37] G.O. Mohler et al., *Randomized Controlled Field Trials of Predictive Policing*, 110 J. AM. STAT. ASS'N 1399 (2015).

analytics in targeting their response to drug crimes.[38] HRDAG is an NGO staffed by data scientists, which has spent the last 25 years applying statistical expertise to document the extent of genocides overseas. More recently, the group has turned its attention to law and order in the United States, with a growing focus on policing.[39]

The HRDAG analysis compared the locations and times of drug-related arrests in Oakland, California (obtained from the Oakland police) with estimates of the actual level of drug use in the city, which they extrapolated from the 2011 National Survey on Drug Use and Health, a study commissioned by the federal Department of Health and Human Services.[40] They found that:

> drug arrests in the police database appear concentrated in . . . two areas with largely non- white and low-income populations. These neighbourhoods experience about 200 times more drug-related arrests than areas outside of these clusters. In contrast, our estimates [based on the federal drug use survey] suggest that drug crimes are much more evenly distributed across the city. This suggests that while drug crimes exist everywhere, drug arrests tend to only occur in very specific locations – the police data appear to disproportionately represent crimes committed in

---

[38] Kristian Lum & William Isaac, *To Predict and Serve?*, SIGNIFICANCE, Oct. 2016, at 14-19 (2016), http://onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2016.00960.x/epdf [https://perma.cc/A25A-SE5M].

[39] Patrick Ball, *Violence in Blue*, GRANTA MAG. (Mar. 4, 2016), https://granta.com/violence-in-blue/ [https://perma.cc/XWV9-Y7LR] (applying list-dependent estimation techniques across several datasets to estimate the number of people killed by police in the U.S.). The Bureau of Justice Statistics commissioned a report that used the same approach to estimate the "arrested-related deaths that are the result of law enforcement homicides," D. BANKS ET AL., RTI INT'L, ARREST-RELATED DEATHS PROGRAM ASSESSMENT: TECHNICAL REPORT 2 (Mar. 2015), https://www.bjs.gov/content/pub/pdf/ardpatr.pdf [https://perma.cc/25BQ-EY7H]. Both the HRDAG and BJS-commissioned analyses found that the number of such deaths captured by official statistics is likely significantly fewer than the true number.

[40] *Population Data/National Survey on Drug Use and Health*, SAMHSA (Aug. 30, 2016), https://www.samhsa.gov/data/population-data-nsduh [https://perma.cc/3VMK-ST2H] (stating, "[t]he National Survey on Drug Use and Health (NSDUH) is the primary source of information on the prevalence, patterns, and consequences of alcohol, tobacco, and illegal drug use and abuse and mental disorders in the U.S. civilian, non-institutionalized population, age 12 and older.").

> areas with higher populations of non-white and low-income residents.[41]

In the end, the authors find that "rather than correcting for the apparent biases in the police data, the model reinforces these biases. The locations that are flagged for targeted policing are those that were, by our estimates, already over-represented in the historical police data."[42]

2. *Reasonable Suspicion Meets Automation Bias*

Legal scholars who consider predictive analytics in the context of criminal justice have focused extensively on the Fourth Amendment risks of predictive policing.[43] Those risks are real and significant, particularly in heavily policed, heavily surveilled communities.

The primary challenge arises in the context of *Terry* stops—where the police lack probable cause to arrest a person, but do, nonetheless, detain the person on "reasonable suspicion" of involvement in criminal activity.[44] The Supreme Court has held that a person's mere presence in a "high-crime area" suffices as one factor out of the requisite minimum of two factors to create constitutionally reasonable suspicion for a stop.[45] However, the Court has never defined what

---

[41] Lum & Isaac, *supra* note 38.

[42] *Id.* at 18.

[43] Rachel A. Harmon, *The Problem of Policing*, 110 MICH. L. REV. 761, 783 (2012) ("[S]cholars take their subject to be Supreme Court cases involving the Constitution rather than the problem of policing . . . They argue that a police activity falls within or outside the scope of the Fourth Amendment" (citations omitted)). *See, e.g.*, Elizabeth E. Joh, *Policing by Numbers: Big Data and the Fourth Amendment*, 89 WASH. L. REV. 35 (2014); Andrew Guthrie Ferguson, *Predictive Policing and Reasonable Suspicion*, 62 EMORY L.J. 259 (2012); Andrew Guthrie Ferguson, *"Predictive Policing" and the Fourth Amendment*, AM. CRIM. L. REV. ONLINE (Nov. 28, 2011), http://www.americancriminallawreview.com/aclr-online/predictive-policing-and-fourth-amendment/ [https://perma.cc/9MW8-67ZJ]; Alexander H. Kipperman, *Frisky Business: Mitigating Predictive Crime Software's Facilitation of Unlawful Stop and Frisks*, 24 TEMP. POL. & CIV. RTS. L. REV. 215 (2014); Kelly K. Koss, *Leveraging Predictive Policing Algorithms to Restore Fourth Amendment Protections in High-Crime Areas in a Post-*Wardlow *World*, 90 CHI.-KENT L. REV. 301 (2015).

[44] *See* Terry v. Ohio, 392 U.S. 1 (1968).

[45] Illinois v. Wardlow, 528 U.S. 119 (2000). The case also held that unprovoked flight from police taking place in a "high-crime area" can count as the second such factor. *Id.*

counts as a "high-crime area," and in practice police apparently have wide latitude to claim that the area where they made a given stop was "high-crime" without having to substantiate that assertion.[46] For people in an area that courts are prepared to consider high-crime, there may thus be just one further factor needed to justify a stop.

Andrew Ferguson, a leading scholar of the Fourth Amendment issues posed by predictive policing and other new technologies, has written: "A police stop based on a predictive policing forecast soon will be in front of a trial court in a motion to suppress evidence."[47] At that time, "the judge or arresting officer will have no way to test the accuracy" of the correlation that the prediction is based upon, before deciding whether to act on it in a particular case.[48]

Regardless of how accurate or inaccurate a predictive policing system may be, officers may well give credence to any indication of concern or suspicion that comes from a computer—and courts are likely to agree that such flags do constitute reasonable suspicion, particularly in high-crime areas where no other factors are needed.

3. *"Well-Calibrated" Bail Predictions Put a Disparate Burden on Black Defendants*

A recent journalistic investigation of predictive analytics in the courtroom provides a vivid illustration of the gap between legal ambiguity and technological precision. A team of journalists at ProPublica investigated the proprietary COMPAS risk assessment algorithm, which is used to evaluate the reappearance risk and

[46] Andrew Guthrie Ferguson & Damien Bernache, *The "High-Crime Area" Question: Requiring Verifiable and Quantifiable Evidence for Fourth Amendment Reasonable Suspicion Analysis*, 57 AM. U. L. REV. 1587, 1590–91 (2008), http://digitalcommons.wcl.american.edu/cgi/viewcontent.cgi?article=1036&context=aulr [https://perma.cc/RGG8-MA5M] ("The Supreme Court has never provided a definition. Lower court decisions are equally imprecise. Yet, as practicing criminal defense lawyers know, the question is highlighted in almost every Fourth Amendment suppression hearing focused on the legitimacy of a police stop. A police officer takes the stand, explains his actions, testifies to his suspicions, adds the magic words—'high-crime area'—and reasonable suspicion is found as a matter of constitutional law. Rarely is there any analysis of why this particular area is a high-crime area, on what objective, verifiable, or empirical data the police officer has based his conclusion, or whether the officer knew this information before he made the stop." (citations omitted)).

[47] Ferguson, *Predictive Policing and Reasonable Suspicion, supra* note 43, at 312.

[48] ANDREW FERGUSON, THE RISE OF BIG DATA POLICING 127 (2017).

dangerousness of pretrial detainees at bail hearings in Broward County, Florida.[49] Using Florida's freedom of information law with painstaking diligence, the journalists "looked at more than 10,000 criminal defendants in Broward County, Florida, and compared their predicted recidivism rates with the rate that actually occurred over a two-year period."[50]

The report's headline described COMPAS as "biased against blacks." Although the system made mistakes roughly equally often in predicting the recidivism of white and of black defendants, the burden of wrongly being labelled "high risk" was borne disproportionately by blacks. That is, among those defendants who in fact did *not* go on to reoffend, the black defendants were far more likely than the white ones to have been labelled "high risk." On the other hand, the system's mistaken predictions about white defendants trended the other way: the system was far more likely to classify as low risk a white defendant who went on to be rearrested, than to classify as low risk a black defendant who went on to be rearrested. On this basis, the journalists described the system as biased against blacks. Northpointe defended itself by saying that its algorithm was "well calibrated" across races, that is, had the same probability of being correct, regardless of whether the defendant was white or black.[51]

A wave of scholarship, triggered by the ProPublica report, illuminated the statistical challenge at the heart of the argument: Given that the underlying "base rate" of re-arrest is higher for blacks than for whites, it is mathematically inevitable that the burden of false positives will fall more heavily on black defendants than on white ones. In other words, given that more black defendants than white defendants actually do have a high risk of being rearrested, a "high risk" label that is correct 70% of the time for both white and black defendants will still mislabel more black than white defendants as high risk. A study titled "Inherent Tradeoffs in the Fair Determination of Risk Scores" proved mathematically that when re-arrest rates are not equal between races, a well-calibrated tool like Northpointe's –

---

[49] 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/2K23-LP6X].

[50] Jeff Larson et al., *How We Analyzed the COMPAS Recidivism Algorithm*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm [https://perma.cc/JD95-ZT87].

[51] *Id.*

that is, a tool that is mistaken equally often about whites and about blacks – will inevitably have more false positives for blacks.[52] As the authors of a second study explained, to equalize the error rates, one would have to make the tool itself race-conscious, and set "multiple, race-specific [risk of re-arrest] thresholds."[53]

In short, an intuitive understanding of equal protection cannot square with the mathematics of predictive risk scoring. Under real conditions, a tool that is equally often mistaken about white and black defendants will more often send blacks to jail by mistake than send whites to jail by mistake. But an explicitly race-conscious risk assessment tool, that predicted scores differently for whites than for blacks, would itself face serious constitutional challenges. An understanding of "equal protection" that would require race-blindness, and simultaneously require that races are burdened equally by prediction errors, simply does not leave room for risk assessment tools to operate.

C. *Public Authority Versus Private Expertise*

The criminal justice system, like other parts of government, often lacks technological expertise, leaving commercial vendors in a privileged position.[54] Vendors can, and often do, shape officials' understanding not only of how technologies work, but also of how the tools' performance should be understood and assessed. Not surprisingly, vendors often practice technical "shock and awe," encouraging strong optimism about what their systems can do and simultaneously discouraging pointed questions about how those systems work.

---

[52] JON KLEINBERG ET AL., INHERENT TRADE-OFFS IN THE FAIR DETERMINATION OF RISK SCORES (Nov. 17, 2016), https://arxiv.org/pdf/1609.05807.pdf [https://perma.cc/J29A-LPMK].

[53] SAM CORBETT-DAVIES ET AL., ALGORITHMIC DECISION MAKING AND THE COST OF FAIRNESS (June 10, 2017), https://arxiv.org/pdf/1701.08230.pdf [https://perma.cc/4ZKW-H7DY]. *See also* GEOFF PLEISS ET AL., ON FAIRNESS AND CALIBRATION (Nov. 3, 2017), https://arxiv.org/pdf/1709.02012.pdf [https://perma.cc/J3HV-JS64] ("[I]nvestigat[ing] the tension between minimizing error disparity across different population groups while maintaining calibrated probability estimates.").

[54] *See generally* Elizabeth E. Joh, *The Undue Influence of Surveillance Technology Companies on Policing*, N.Y.U. L. REV. ONLINE (Sept. 2017), http://www.nyulawreview.org/sites/default/files/Joh-FINAL_0.pdf [https://perma.cc/XPP2-RC6F].

1. *The Cautionary Tale of Beware*

In 2015, police in Fresno, California tested a software system evocatively named "Beware." Beware integrates with the dispatch screens shown to 911 operators, and assigns a "threat score" to community members who initiate 911 calls—that is, to the people who call for help, rather than those who are suspected of wrongdoing.[55] Although purchases of predictive policing systems often go unnoticed by the public,[56] in this case a community group called Faith in the Valley knew about, and provided input into, the police department's decision-making process.[57] When community members' concerns were not well addressed by the department, they brought their concerns to the city council, triggering national press attention.[58]

If predictive policing systems are trying to predict—or are treated as though they *were* trying to predict—the risk that officers will face violence when responding to a call, the results could be deadly. Yet Beware is likely to create confusion on precisely this point.

The physical risks associated with being a police officer are at or near all-time lows,[59] even after accounting for the events of the last

---

[55] David Robinson, *Buyer Beware: A Hard Look at Police 'Threat Scores'*, EQUAL FUTURE (Jan. 14, 2016), https://medium.com/equal-future/buyer-beware-a-hard-look-at-police-threat-scores-961f73b88b10#.6ym7bx9x7 [https://perma.cc/B62H-TBR3]. Portions of the below discussion of Beware were first published in this earlier short article.

[56] David Robinson & Logan Koepke, *Stuck in a Pattern*, UPTURN 10 (Aug. 2016), https://www.teamupturn.org/reports/2016/stuck-in-a-pattern [https://perma.cc/JV55-QLKJ] (observing that, based on a survey of publicly available information about predictive policing at the nation's 50 largest police departments, "an open public debate regarding a police department's potential adoption of a predictive policing system seems to be the exception to the rule.").

[57] *See* Taymah Jahsi, *It's Time to Shine a Light on Police Surveillance in Fresno*, ACLU OF N. CAL. (Sept. 21, 2016), https://www.aclunc.org/blog/its-time-shine-light-police-surveillance-fresno [https://perma.cc/L4HF-9PNA].

[58] Justin Jouvenal, *The New Way Police Are 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?tid=pm_pop_b [https://perma.cc/P3FR-FYYM].

[59] Radley Balko, *Once Again: There is No 'War on Cops'*, WASH. POST: THE WATCH (Sept. 10, 2015), https://www.washingtonpost.com/news/the-watch/wp/2015/09/10/once-again-there-is-no-war-on-cops-and-those-who-claim-otherwise-are-playing-a-dangerous-game/ [https://perma.cc/42SE-GSBP].

several years.[60] Nonetheless, public discussion of a "war on cops" or a "Ferguson effect" creates the false impression of a problem out of control.[61] Beware leverages this anxiety, arguing for example in one of its brochures that "there are no routine calls."[62]

A system that improved officer insight into which situations were high risk – and was deployed in a way that led to more cautious, deliberate responses – might indeed make people safer. But there is no way to be confident that Beware works this way—and ample reason to fear that the system could actually aggravate risks to innocent life.

Responsible voices in the policing community have long been concerned about the low legal standards that attach to officers' use of force, including deadly force. As former Madison, Wisconsin police chief David Couper writes, ever since the Supreme Court case of *Graham v. Connor*,[63] an officer can "legally use deadly force based on whether the officer reasonably believed his or her life was in danger . . . Before this decision, officers were expected to use only the minimum amount of force necessary to overcome resistance."[64] This gives officers extremely wide latitude: "Add to this decision the fear every police officer has that he or she could be disarmed and shot you have a 'perfect storm' of police using deadly force in almost any situation involving resistance."[65]

Predictive policing that tries to predict risks to officers could make this problem worse. As The Atlantic noted in its coverage of the Beware system, it could create "troublingly perverse incentives,

---

[60] Radley Balko, *In the End, 2015 Saw No 'War on Cops' and No 'National Crime Wave'*, WASH. POST: THE WATCH (Dec. 22, 2015), https://www.washingtonpost.com/news/the-watch/wp/2015/12/22/in-the-end-2015-saw-no-war-on-cops-and-no-national-crime-wave/ [https://perma.cc/H6VB-8S3P].

[61] Samuel Sinyangwe, *Stop Pretending the "Ferguson Effect" is Real*, MEDIUM (Oct. 27, 2015), https://medium.com/@samswey/stop-pretending-the-ferguson-effect-is-real-40e3684fae3d#.3z1a4rz4b [https://perma.cc/HE23-YLQS].

[62] INTRADO, BEWARE INCIDENT INTELLIGENCE 13 (2015), http://www.aclunc.org/docs/201512-social_media_monitoring_softare_pra_response.pdf [https://perma.cc/5NUF-YVX8].

[63] Graham v. Connor, 490 U.S. 386 (1989).

[64] David C. Couper, *Police Use of Deadly Force: Time for Discussion*, IMPROVING POLICE (Mar. 18, 2015), https://improvingpolice.wordpress.com/2015/03/18/police-use-of-deadly-force-time-for-discussion/ [https://perma.cc/DC3V-YJGW].

[65] *Id.*

insofar as cops who use deadly force are judged based on what a reasonable officer would've done in the same situation with the same information . . . Will the fact that police were responding to a call relating to a house or individual with a red threat level now be used to argue that subsequent force was relatively more reasonable?"[66]

The company selling Beware (Intrado, a unit of West Corp.) has disclosed little to the public about how its product works. In fact, even the police departments that rely on this product have limited insight into its workings. The system remains unproven by neutral evidence and is marketed in a way that could easily make officers think it predicts risks to their lives.[67]

By piecing together public information—including marketing materials and internal police documents related to a trial deployment in Fresno[68] which the ACLU obtained via public records request;[69] job descriptions and employee reviews,[70] securities filings and other sources—it is possible to build a composite picture of the system's operation. And that picture is concerning.

A basic question – perhaps the *most* basic question – about any application of predictive analytics is what, exactly, the system aims to predict. In other words, here, what do Beware's "threat scores" actually describe? What would it mean to say that the scores are, or are not, accurate?

The public record leaves even this basic question unanswered. Here are three possibilities – listed in order from least to most feasible, which also happens to be from most to least useful:

(1) *Predicting the risk to officers during their response to a call.*

---

[66] Conor Friedersdorf, *A Police Department's Secret Formula for Judging Danger*, THE ATLANTIC (Jan. 13, 2016), http://www.theatlantic.com/politics/archive/2016/01/a-police-departments-secret-formula-for-judging-danger/423642/ [https://perma.cc/VJ6M-YVN5].

[67] *See* INTRADO, *supra* note 62.

[68] *Id.*

[69] Letter from Matthew T. Cagle, Technology & Civil Liberties Policy Attorney & Matthew W. Callahan, Technology & Civil Liberties Fellow, to Jerry Dyer, Police Chief, Fresno Police Dept., Re: Public Records Act Request Regarding Social Media Monitoring Software (Sept. 16, 2015), https://www.aclunc.org/docs/20150916-fresno_social_media_pra.pdf [https://perma.cc/6M94-KTFX].

[70] *See, e.g.*, Anonymous Emp. in Longmont, CO, *West Safety Services: "Software Engineer"*, GLASSDOOR (Dec. 6, 2015) https://www.glassdoor.com/Reviews/Employee-Review-West-Safety-Services-RVW8839318.htm [https://perma.cc/6RHV-82FZ].

Beware's marketing emphasizes risks to officer safety. For example, a company brochure argues that ambushes or surprise attacks, staged by people previously arrested for violent crimes, account for a large fraction of officer fatalities.[71] Did Intrado, or anyone else, assemble the "commercial records" and "website hits" of people who actually have assaulted police officers, finding enough such people to be able to predict, in a statistically responsible way, which out of thousands of consumer purchases or web site visits make a person more likely to assault an officer? Probably not: In 2014, Fresno police answered 408,718 calls for service, and there were, remarkably, a grand total of only 363 assaults on the city's officers. Even if those 363 assailants shared certain traits—perhaps they were more likely to use Twitter, or to buy Cornflakes, than other Fresno residents—such trends are more likely to be spurious correlations than a meaningful statistical signal. Perhaps Beware uses national statistics on assaults or violent acts, which would offer a much larger sample size from which to draw conclusions. That, however, would also discount any variation across different localities.

(2) *Predicting violence against anyone while officers are on the scene—not just against the officers themselves.*

Even then, the numbers are vanishingly small: There were fewer than 3,000 "crimes against persons" in Fresno all year, and in many of those cases, police were dispatched because the crime had already occurred, rather than because it was imminent.

(3) *Predicting violence against anyone by a person near the call, whether during or after the officer's response.*

Out of the three predictive possibilities outlined here, this scenario is the most plausible one to imagine aiming at, because there's a much larger sample size of people who've been arrested for violent crimes. However, this is also the least useful kind of prediction for the officer or dispatcher since it's least focused on the actual situation that they are dealing with.

When the Fresno police briefed the city council about the Beware system, the council president asked a simple question about those color-coded threat levels: "How does a person get to red?" His police force, it turns out, doesn't know, because the vendor won't tell them. As Elizabeth Joh has written, "The relationships between surveillance technology vendors and police departments show the increasing

[71] INTRADO, *supra* note 62.

degree to which private companies can guide, shape, and limit what the public police do."[72]

As the officer delivering the briefing explained: "We don't know what their algorithm is exactly… We don't have any kind of a list that will tell us, this is their criteria, this is what would make a person red, this is what would make a person yellow. It's their own system."[73] Later in that meeting, one of the council members asked the officers to run a live threat assessment on the council member's own home. It came back yellow, apparently indicating a medium threat in that home (though the council member himself came back green).

In the end, Fresno's city council refused the police department's request for funding to use the Beware system beyond its initial trial period.[74]

2. *The Opacity of Evidence Based Sentencing: The Loomis Case*

There is a rich debate in the criminal justice literature about what role, if any, a defendant's potential commission of *future* crimes should play in shaping his sentence.[75] Much of this debate is traceable

---

[72] Elizabeth E. Joh, *The Undue Influence of Surveillance Technology Companies on Policing*, 92 N.Y.U. L. REV. 101, 102 (2017).

[73] Robinson, *Buyer Beware*, *supra* note 55.

[74] Tim Sheehan, *Fresno Council Halts Purchase of Data Software Wanted by Police*, FRESNOBEE (Mar. 31, 2016), http://www.fresnobee.com/news/local/article69337677.html [https://perma.cc/DM2G-5YU5].

[75] *See generally* John Monahan & Jennifer L. Skeem, *Risk Assessment in Criminal Sentencing*, 12 ANN. REV. OF CLINICAL PSYCHOL. 489 (2016); Sonja B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 STAN. L. REV. 803 (2014) [hereinafter Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*]; Richard G. Kopf, *Federal Supervised Release and Actuarial Data (Including Age, Race, and Gender): The Camel's Nose and the Use of Actuarial Data at Sentencing*, 27 FED. SENT'G REP. 207 (2015); Sonja B. Starr, *The New Profiling: Why Punishing Based on Poverty and Identity Is Unconstitutional and Wrong*, 27 FED. SENT'G REP. 229 (2015); Mark H. Bergstrom & Joseph Sabino Mistick, *Danger and Opportunity: Making Public Safety Job One in Pennsylvania's Indeterminate Sentencing System*, 12 JUST. RES. & POL'Y 73 (2010); Claire Botnick, *Evidence-Based Practice and Sentencing in State Courts: A Critique of the Missouri System*, 49 WASH. U.J.L. & POL'Y 159 (2015); PAMELA M. CASEY ET AL., NAT'L CTR. FOR STATE COURTS, USING OFFENDER RISK AND NEEDS ASSESSMENT INFORMATION AT SENTENCING: GUIDANCE FOR COURTS FROM A NATIONAL WORKING GROUP (2011), http://www.ncsc.org/~/media/microsites/files/csi/rna%20guide%20final.ashx [https://perma.cc/ZK2Z-KTRC].

to fundamental and unresolved disagreements about the purposes of criminal sentencing. Contending visions include the retributivist ideal, wherein a defendant's sentence reflects and is controlled by the conduct of which he has been convicted; a rehabilitative model that aims to return the convict to society whole; and utilitarian approaches that define the optimal sentence in terms of its impact upon future conduct by the defendant (incapacitation and "specific" deterrence) or by others ("general" deterrence). Some scholars have aimed to combine these approaches,[76] and the federal sentencing statute ultimately embraces all of them in its list of objectives for the Federal Sentencing Commission.[77]

There is a growing move toward incorporating statistical predictions of recidivism at sentencing. Professor Sonja Starr—who in 2014 found at least twenty states moving to implement statistical prediction at sentencing—has written the leading critique of this approach. She finds a number of practical problems. To wit, risk-based sentencing instruments "provide wildly imprecise individual risk predictions, that there is no compelling evidence that they outperform judges' informal predictions, that less discriminatory alternatives would likely perform as well, and that the instruments do not even address the right question: the effect of a given sentencing decision on recidivism risk."[78]

In short, evidence-based sentencing is widely adopted but its statistical mechanisms—and the limits of those mechanisms—are poorly understood by judges and courts.

Wisconsin's sentencing regime provides a stark example. As described in Section II(B)(1) above, COMPAS is a proprietary, commercially marketed risk assessment system that was designed to assist judges, before trial, in predicting defendants' likelihood of skipping bail or getting rearrested before their cases are resolved. The

---

[76] *See generally* Marc Miller & Norval Morris, *Predictions of Dangerousness: Ethical Concerns and Proposed Limits*, 2 NOTRE DAME J.L. ETHICS & PUB. POL'Y 393 (1985).

[77] *See* 18 U.S.C. § 3553(a)(2) (listing "the need for the sentence imposed . . . (A) to reflect the seriousness of the offense, to promote respect for the law, and to provide just punishment for the offense; (B) to afford adequate deterrence to criminal conduct; (C) to protect the public from further crimes of the defendant; and (D) to provide the defendant with needed educational or vocational training, medical care, or other correctional treatment in the most effective manner.").

[78] Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, *supra* note 75, at 803.

tool was the target of a journalistic investigation suggesting that it exhibited racial bias. The team of investigative journalists at ProPublica who constructed that analysis were not privy to the code by which COMPAS operates.[79]In Wisconsin, despite the fact that COMPAS was designed as a *pretrial* tool, the state now applies the tool at the sentencing phase, to predict the likelihood of re-offense after release from a prison sentence (something that the tool was never designed to measure, and that its training data – about pre-trial outcomes – does not address.) In the case of *State v. Loomis*,[80] the Supreme Court of Wisconsin considered an appeal from Eric Loomis. At Loomis's trial, the presiding judge had cited Loomis's COMPAS score as one factor among many in justifying his sentence. On appeal, the state Supreme Court considered whether the "proprietary nature of COMPAS prevents defendants from challenging the COMPAS assessment's scientific validity," in violation of defendants' due process rights.[81]

The state's Supreme Court, much like stakeholders elsewhere in the criminal justice system, was hard-pressed to understand or evaluate the workings of the COMPAS instrument. As one Justice wrote in her concurrence, "this court's lack of understanding of COMPAS was a significant problem in the instant case. At oral argument, the court repeatedly questioned both the State's and defendant's counsel about how COMPAS works. Few answers were available."[82]

In the end, the court in *Loomis* did not provide crisp guidance on the constitutionality of COMPAS at sentencing. Instead, in its own words, the court "ultimately conclude[d] that a COMPAS risk assessment can be used at sentencing, [but only] by circumscribing its use[.]"[83] The court mandated that sentencing judges be provided with a written list of five cautions, which it spelled out in bullet form—including that COMPAS is proprietary and its inner workings undisclosed, that "some studies . . . have raised questions" about whether the tool is racially biased, and that the tool "was not

---

[79] Angwin et al., *supra* note 40.

[80] State v. Loomis, 2016 WI App 68, 371 Wis. 2d 235, 881 N.W.2d 749.

[81] *Id.* at ¶ 6.

[82] *Id.* at ¶ 133.

[83] *Id.* at ¶ 35.

developed for use at sentencing."[84] "[T]his advisement," the court went on to say, "should be regularly updated as other cautions become more or less relevant as additional data becomes available,"[85] though the court did not specify who would be responsible for such updating or what process they might use. Moreover, the court warned, "a COMPAS risk assessment may not be determinative in deciding whether a defendant may be supervised safely and effectively in the community."[86]

Court and police access to data-related expertise is a challenge that spans well beyond predictive analytics. The justice system also struggles, for example, to understand the value and limits of forensic tools such as software that deciphers DNA mixtures.[87] Trade secrecy—originally a commercial doctrine—has now spread into the criminal justice domain, where it works to inhibit courts and defense counsel from scrutinizing software code that underpins convictions.[88]

III. DOING PREDICTION RIGHT

Each of the gaps described above requires collaborative effort to bridge. Experts across fields need to work more closely together. Everyone is at the periphery of their competence, in one way or another, in facing these challenges. There are a handful of steps that are natural starting points in nearly any application of predictive analytics.

A. *Find or Create Data About What Really Matters*

The most important design choice that shapes any predictive analytic system is the choice of input data. By choosing (or creating)

---

[84] *Id.* at ¶ 100.

[85] *Id.* at ¶ 101.

[86] *Id.* at ¶ 104.

[87] *See* Logan Koepke, *Should secret code help convict?*, EQUAL FUTURE (Mar. 24, 2016), https://medium.com/equal-future/should-secret-code-help-convict-7c864baffe15 [https://perma.cc/CC64-ERBH].

[88] For a thoughtful treatment of this issue and potential ways to resolve it, see Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, STAN. L. REV. (forthcoming 2018).

input data that aligns with true public policy goals, we can improve our chances of achieving those goals.

Government's capacity to monitor and understand its subjects—and its employees—is rapidly growing. Much has been written about the increasingly panoptic vantage point from which government power is exercised; a vantage point informed by pervasive government surveillance and by government access to a far more extensive infrastructure of corporate, commercial surveillance. However, as troubling as this trend may be, it also offers a silver lining: the world is suffused with new data, and newly cheap and easy ways to *generate* further data, that tell on questions that matter in public life.

1. *Measuring Violence Beyond the Police: Victimization Surveys, Health Data, and the Example of the Cardiff Model*

Police statistics leave much to be desired as measures of criminal activity and of police violence, as described above in Section II(A). Such data are inevitably biased, in the statistical sense, toward greater awareness of the areas where officers concentrate their patrols. Officers face incentives to "downcode" some crime reports in order to create a favorable portrait of the department's performance, and a great deal of crime may go unreported.

One vital way to supplement police data is to ask people in the community about their experiences as victims of crime. Victimization is just one of the areas in which useful data could be generated by a systematic, ongoing survey capacity that allows for documentation of community experiences of safety, as described in subsection 2, *infra*. The information gains from such a strategy could be significant: The Department of Justice conducts a National Crime Victimization Survey, which found that from 2006-2010 (the most recent data now available), 52 percent of violent crime victimizations[89] went unreported to police and 60 percent of household property crime victimizations went unreported.[90] Historically, the National Crime

---

[89] "Crime victimization," according to the Department of Justice, includes violent victimization (rape, sexual assault, robbery, aggravated assault, and simple assault) and property victimization (burglary, motor vehicle theft, and property theft) as well as domestic violence and intimate partner violence. JENNIFER L. TRUMAN & LYNN LANGTON, U.S. DEP'T OF JUSTICE, BUREAU OF JUSTICE STATISTICS, CRIMINAL VICTIMIZATION, 2014 (2015), http://www.bjs.gov/index.cfm?ty=pbdetail&iid=5366 [https://perma.cc/X95Z-6LMU].

[90] LYNN LANGTON ET AL., U.S. DEP'T OF JUSTICE, BUREAU OF JUSTICE STATISTICS, SPECIAL REPORT: VICTIMIZATIONS NOT REPORTED TO THE POLICE, 2006-2010, 2 (2012).

Victimization Survey has shown that police are not notified of about half of all rapes, robberies, and aggravated assaults.[91] More local and more frequent victimization surveys could give local communities an ongoing map of safety and other concerns.[92]

A second potentially vital source of data about community violence rests with the second major bureaucracy that responds to violent incidents: the health system. A pioneering effort of this type began in the Welsh city of Cardiff and is now known as the Cardiff Model; it has been extensively studied and is now being piloted in the United States.[93]

In the Cardiff Model, a multi-stakeholder Crime and Disorder Reduction Partnership is established, and the emergency medical services contribute to that body by "sharing, electronically wherever possible, simple anonymized data about precise location of violence, weapon use, assailants and day/time of violence."[94] Such data includes information about incidents that may not have been reported to the police, and naturally includes information about where and when the injury took place, which can help police in their investigations.[95] Moreover, emergency rooms "are the only sources of information about serial (repeat) injury: a recognized precursor to homicide in the home and elsewhere."[96]

This approach is not a silver bullet. No data source is perfect, and the animating goals of medical treatment are different than those of community safety, which will have some impact on the nature and completeness of the data collected. For example, although the time elapsed since a violent injury may be vitally important for the

---

[91] *Id.*

[92] *See* Mark H. Moore & Margaret Poethig, *U.S. Dep't of Justice, Nat'l Inst. of Justice, The Police as an Agency of Municipal Government: Implications for Measuring Police Effectiveness*, *in* MEASURING WHAT MATTERS: PROC. FROM THE POLICE RES. INST. MEETINGS 164 (1999).

[93] *See Piloting the Cardiff Model for Violence Prevention in the United States to provide better local information to address violence problems*, ROBERT WOOD JOHNSON FOUND., https://www.rwjf.org/en/how-we-work/grants-explorer.html#k=Cardiff%20Model [https://perma.cc/3FR4-NUMA].

[94] JONATHAN SHEPHERD, EFFECTIVE NHS CONTRIBUTIONS TO VIOLENCE PREVENTION: THE CARDIFF MODEL 3 (Cardiff U., Oct. 2007).

[95] *Id.*

[96] *Id.*

treatment of a newly admitted patient, the precise location where the injury took place is not of comparable importance to the medical aims of a hospital, and as a result, may be recorded with less care if at all. A description of barriers to the Cardiff model, published by the model's primary advocate, observes that some medical staff believe it is unduly paternalistic to take an interest in circumstances beyond the patient's immediate medical needs, and also that police sometimes unreasonably demand evidentiary documentation from hospitals that would be typical of police work but is atypical for a hospital.[97]

A 2011 study in the British Medical Journal, which looked at the impact of the Cardiff Model over a four year period, found that "[i]nformation sharing and use were associated with a substantial and significant reduction in hospital admissions related to violence."[98] Moreover, the model had "led to a significant reduction in violent injury and was associated with an increase in police recording of minor assaults in Cardiff compared with similar cities in England and Wales where this intervention was not implemented."[99]

Although work on the Cardiff Model has focused on emergency room treatment data, there is also some early work suggesting that ambulance dispatch records may have significant value as well.[100]

In the United States, in addition to efforts expressly patterned on the Cardiff model (including one funded by the CDC[101]), there are also a number of other early efforts to use health data as indicia of community violence. For example, Hospital Violence Intervention Programs (HVIPs) use interviews with hospital patients to target social services. One effort attempted to track shootings specifically *by*

---

[97] *Id.* at 5.

[98] Curtis Florence et al., *Effectiveness of Anonymised Information Sharing and Use in Health Service, Police, and Local Government Partnership for Preventing Violence Related Injury: Experimental Study and Time Series Analysis*, 342 BMJ 1 (2011), https://www.bmj.com/content/342/bmj.d3313 [https://perma.cc/9CX3-BHSD].

[99] *Id.*

[100] Barak Ariel et al., *Can Routinely Collected Ambulance Data about Assaults Contribute to Reduction in Community Violence?*, 32 EMERGENCY MED. J. 308, 312 (2015).

[101] *Cardiff Model Will Help Guide Strategies to Prevent Violence in Southern DeKalb County*, CDC FOUND., https://www.cdcfoundation.org/pr/2017/cardiff-model-will-help-guide-strategies-prevent-violence-southern-dekalb-county [https://perma.cc/QZD7-4CNG].

police through health data.[102] And an emerging area of research uses health data to forecast child abuse and target child protective efforts.[103]

2. *Police Performance: The New Wealth of Alternate Measures*

In most of the nation, we currently measure outcomes and assess performance based on only a subset of the behaviors, costs, and benefits that matter in policing. If predictive policing systems extend this mistake into the big data era, they may well make policing worse. If, however, these new systems expand and enrich the ways that we measure and understand police performance, they may well be a major benefit to civil rights.

"As our ideas of policing have changed . . . we find that we need information that we do not now have. To the extent that we are interested in preventing crimes through means other than arrests, we have to find ways to recognize such activities, and to evaluate their impact."[104] Serious violent crimes may always "remain important, but community concerns frequently center on other issues. Many crimes are not reported, and therefore police would need to use a broader range of data sources—including public health information and victimization surveys—even to be able to see the full range of problems that matter."[105]

As Professor Rachel Harmon has argued, in terms of the cost-benefit analysis by which so many public programs are evaluated, our current approach overlooks the "coercion costs" of policing.[106] This is a profound asymmetry: "In cost-benefit assessments of programs that

---

[102] Joseph B. Richardson et al., *Who Shot Ya? How Emergency Departments Can Collect Reliable Police Shooting Data*, 93 J. OF URB. HEALTH (Suppl. 1) 8 (2016), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4824690/ [https://perma.cc/X5JA-XE8X].

[103] *See* Dyann Daley et al., *Risk Terrain Modeling Predicts Child Maltreatment*, 62 CHILD ABUSE & NEGLECT 29, 29 (2016); Christopher E. Church & Amanda J. Fairchild, *In Search of a Silver Bullet: Child Welfare's Embrace of Predictive Analytics*, 68 JUV. & FAM. CT. J. 67, 67 (2017).

[104] MOORE, *supra* note 28, at 165.

[105] Robinson & Koepke, *supra* note 56, at 6.

[106] Rachel A. Harmon, *Federal Programs and the Real Costs of Policing*, 90 N.Y.U. L. REV. 870, 870 (2015).

influence local policing, they reason expansively with respect to benefits, recognizing a multitude of costs averted by federal programs when the programs prevent crime, such as the pain and suffering avoided when a federally funded officer prevents an offense. But they consider only the budgetary outlay to the federal agency as costs. They do not contemplate any harm—and, therefore, any cost— from policing itself."[107] Such logic applies broadly, not only to the go/no-go decisions made about particular programs and investments, but also to the more specific and tactical choices made by police in the course of their daily activities.

An ideal measurement and prediction system would track, highlight, and reward *excellence*, not just compliance. "It is possible to both remedy shortcomings and thwart the natural tendencies toward defensive responses by viewing the problem not as one of detecting deficiency but of creating measures of good service."[108]

One key element of this approach might be police investment in "a large, continuing capacity to survey citizens."[109] New technologies make survey methods easier and more cost effective to implement than ever before, including via SMS and other widely deployed technologies.[110] Similarly, "a sample of individuals who call the police (or ask officers on the streets or in station houses) for assistance" should be asked about the service they receive.[111] And those who are stopped or arrested – a group that might be called "obligates" – could usefully be asked about their experience as one source of data about how police conduct themselves in exercising their powers.[112]

New yardsticks that could be constructive to use include:

1. *Surveyed levels of crime victimization, community trust, and police legitimacy*: The public should be asked about its view of the police, its experience of criminal victimization, and its fears and priorities. Surveys could "develop a more accurate picture than we now have about the real level of criminal victimization . . . measure

---

[107] *Id.* at 873.

[108] Klockars, *supra* note 18, at 202.

[109] MOORE & POETHIG, *supra* note 92, at 164.

[110] *See* Koepke & Robinson, *Danger Ahead*, *supra* note 32.

[111] MOORE & POETHIG, *supra* note 92, at 164.

[112] *Id.*

levels of fear as well as victimization . . . measure citizen satisfaction with the quality of police service, and . . . discover the level and type of self-defense that is being used to complement police efforts."[113]

2. *Outreach efforts and proactive problem solving*: Malcolm Sparrow, an expert on police performance measurement, calls for written accounts of problems that the department has proactively addressed. "Each project account will describe how the department spotted the problem in the first place, how it analyzed and subsequently understood the problem, what the department and its partners did about the problem, and what happened as a result," he writes. "The more vigilant the department becomes in spotting emerging problems early, the less available significant crime reductions will be. Aggregate levels of crime may remain low but relatively steady, even as the department works hard to spot new threats before they have a chance to grow out of control."

3. *"Coercion costs," including all uses of force and all types of detention*: Mark Harrison Moore, a leading scholar of police performance measurement, has written that "[i]f we reduced crime, but did so by relying on more intrusive investigative techniques, or patrol techniques that were both more assertive and viewed as biased, then the increased use of authority would have to be viewed as a loss to be put against the gain."[114] The intensity of coercion – including, for example, the average number of times residents in a given community are stopped and searched by police – is a key dimension of cost for police to recognize. Optimal policing should see extra stops as a cost to those stopped, most of whom have done no wrong – and predictive analytics should aim to produce safety while imposing a minimum of this and other types of cost.

Data like this would naturally inform a more balanced approach to policing. Just as CompStat's focus on crime measures leads to a policing style that focuses on crime rates, there is also evidence that other systems (based on other measures) can foster a different mindset.

For example, scholars who work on Risk Terrain Modeling—which can involve steps such as using the map of liquor stores and bars in a city to target policing resources there—report that the technique changed how they thought about policing. Focusing on the geography

---

[113] *Id.*

[114] MOORE, *supra* note 28, at 26.

of physical factors that increase the risk of crime, they wrote, "can ground risk-based policing much more into the contexts in which police operate rather than concentrating police on behavior that they are trying to control."[115] However, "while crime and officer activity (e.g. arrest) data is routinely collected and updated by law enforcement, it is currently unclear how often police agencies collect" geographic feature data related to risk, "or to what extent risk variables are considered at all."[116] RTM's geographic and risk-focused approach, the scholars wrote, "encouraged us to think more about using 'risk clusters' instead of crime hotspots to allocate police resources."[117]

In some jurisdictions, the inherent risks of having the police gather data about themselves has led government to establish separate agencies to collect and analyze crime-related data, outside of police departments.[118]

3. *Measure and Understand the True Causes of Failure to Appear*

In the case of pretrial risk assessments, which predict re-arrest and failure to appear, the tools may come across as if they tell the judge who can safely be released, and who cannot. And there is a flourishing literature about how, within this paradigm of deciding whom to release and whom not to release, algorithms can slightly outperform judges. But investment in refining our capacity to predict outcomes in a broken jail system arguably distracts the public and policymakers from some high impact, feasible interventions. It takes

---

[115] Joel M Caplan & Leslie W. Kennedy, *Risk Terrain Modeling Compendium,* RUTGERS CENT. ON PUB. SECURITY 89 (2011), http://www.rutgerscps.org/uploads/2/7/3/7/27370595/riskterrainmodelingcompendium_caplankennedy2011.pdf [https://perma.cc/YAA5-7R7W].

[116] *Id.*

[117] Eric L. Piza et al., *Risk Clusters, Hotspots, and Spatial Intelligence: Risk Terrain Modeling as an Algorithm for Police Resource Allocation Strategies*, RUTGERS CENT. ON PUB. SECURITY 3 (Jun. 21, 2010), http://www.rutgerscps.org/uploads/2/7/3/7/27370595/newarkrtm_casestudy_brief.pdf [https://perma.cc/BJ5M-99RY].

[118] *See, e.g.*, JOHN A. ETERNO & ELI B. SLIVERMAN, THE CRIME NUMBERS GAME 102 (2012) ("New South Wales [the most populous state in Australia] has an independent Bureau of Crime Statistics to assess police crime reports. The Bureau's findings contradicted and corrected official New South Wales police crime statistics and prompted a Deputy Commissioner's resignation.")

place within a framework of conversation where the bail system we have — and even contingent details of how that system operates — are implicitly treated as immovable givens.

For example, there is data showing that failures to appear are actually easy to change — dramatic improvements are possible even from small adjustments in how the system works. In a 2012 study in Colorado, just by calling defendants and reminding them about their court dates, the failure to appear rate was reduced by a factor of three.[119] In Nevada, sending postcards also produce a substantial reduction in failures to appear.[120] A firm called UpTrust now offers counties an automated way to send text message reminders of upcoming court dates, and claims that its software can dramatically reduce failures to appear.[121]

In 2009, the Pretrial Justice Institute reported that only 5% of pretrial services agencies call people to remind them of court dates.[122] That is a troubling sign, even though 5% of agencies is likely far more than 5% of defendants, because the agencies that do have the resources to send reminders are probably the larger ones.[123] These findings suggest that FTA numbers could be dramatically improved by making phone calls, sending text messages, or mailing post cards, instead of building jail beds.

But the growth of predictive analytics may serve to obscure the potential of these promising steps. Most of today's pretrial risk assessment instruments were trained on populations that did not receive the benefit of convenient reminders of their upcoming court dates. As a result, the prediction such tools are making is, how likely is this person to return to court on schedule *without* being reminded to appear. An actuarial instrument's rootedness in the past – its blindness to the innovative practice of reminding defendants about

---

[119] Timothy R. Schnacke et al., *Increasing Court-Appearance Rates and Other Benefits of Live-Caller Telephone Court-Date Reminders: The Jefferson County, Colorado, FTA Pilot Project and Resulting Court Date Notification Program*, 48 CT. REV. 86, 89 (2012).

[120] Alan J. Tomkins et al., *An Experiment in the Law: Studying a Technique to Reduce Failure to Appear in Court*, 48 CT. REV. 96, 100 (2012).

[121] *What We Do*, UPTRUST, http://www.uptrust.co/what-we-do [https://perma.cc/RZ72-XRA7].

[122] PRETRIAL JUSTICE INST., 2009 SURVEY OF PRETRIAL SERVICES PROGRAMS 50 (2009).

[123] *See id.* at 18-20.

their court dates – will lead it to make unduly pessimistic predictions, because the predictions will not reflect the greater likelihood of reappearance with reminders. As a result, even if a jurisdiction does implement reminders (and does, thereby, powerfully increase the true probability that released defendants will appear on schedule), it may continue to blindly predict failure to appear and thus to jail people who, thanks to the reminders, could now be released with low risk. In a forthcoming article, a colleague and I explore this problem – which we term "Zombie predictions" – in the bail context, and suggest specific steps that can be taken to address the issue there.

B. *Leverage New Measures to Define Success Carefully*

The outcome variables that analytics predicts are, inevitably, yardsticks of management. A police department that tracks week-to-week changes in calls for service, or a pretrial services agency that tracks all kinds of re-arrest as a single aggregate category, will tend to understand its performance in terms of causing those numbers to change.

By tracking and analyzing data that does not correspond to what matters most, police and other organizations set themselves up to chase a flawed vision of success. The better police, courts and other organizations can align their quantifiable definition of success with the things that matter most, the more likely it becomes that they will ultimately achieve those objectives.

New measures, such as those described in Section A above, will make available new definitions of success for community safety, and new ways of tracking, and rewarding performance in the criminal justice system.

C. *Open Predictive Systems to Scrutiny, and Build Public Expertise*

In any application of predictive analytics, it is essential that the builders and financial backers of a proposed system not enjoy a monopoly on understanding how the system works or judging its performance. This is a basic matter of good practice for any organization, in order to avoid a lemons market.

In the context of government – where the predictive analytics is publicly funded and is a component input in the exercise of public authority – the need for public understanding is all the more acute. Particularly in criminal justice, where a predictive assessment could lead directly to a deprivation of liberty, it is critical that predictive analytics not only function well at a technical level, but also earn and deserve public confidence.

Public authorities, particularly at the local level, often lack even the expertise to know what questions to ask about a predictive analytics system. As a result, there may be a role for standardized, mandatory disclosures of key information about predictive analytical systems that are meant for public use. Such disclosures might include, among other things:

- A detailed, descriptive inventory of the training data used in the predictive system
- The system's definition of predictive success (what engineers would call its "loss function")
- A description of how the system's performance will be tracked over time, and how its predictive model will be kept up to date in the future to match changing conditions.

Those questions are specific to predictive analytical systems. At the same time, there are many dimensions of openness that are vital for all kinds of software used by government, which have occasioned a rich literature on code transparency.[124]

IV. CONCLUSION

This article uses criminal justice as its illustrative domain, but the challenges described here sweep far more broadly. Whenever government grounds its exercise of power in a statistical model, similar challenges may arise. Goals may become distorted. The patterns that are reflected in the input data will become part of the state's vision of reality. Quantification can all too easily serve as a moral anesthetic. Public authority is not always matched by public expertise.

These challenges may arise in parallel across multiple government domains – from criminal justice to child protection, education, taxation and immigration -- and lessons learned in one domain may well prove valuable in another. In future work, I plan to extend and deepen the analysis that this essay begins, to describe cross-cutting challenges in the government's use of predictive analytics and to

---

[124] *See e.g.*, FRANK PASQUALE, BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 3 (2015); Danielle Keats Citron, *Open Code Governance*, 2008 U. CHI. LEGAL F. 355, 356-58 (2008).

consider opportunities for stronger institutional design to meet those challenges.