A Few Criminal Justice Big Data Rules

Stephen E. Henderson*

As with most new things, the big data revolution in criminal justice has historic antecedents—indeed, a 1965 Presidential Commission called for some of the same data analysis that police departments and courts are today developing and implementing. But there is no doubt we are on the precipice of a criminal justice data revolution, and it is a good time to take stock and to begin developing guidelines so that, as much as possible, criminal justice systems might reap the benefits and avoid the pitfalls of this newly data-centric world. In that spirit, I propose ten high-level rules to guide criminal justice big data implementations.

* Judge Haskell A. Holloman Professor of Law, The University of Oklahoma. J.D. Yale Law School, 1999; B.S. in Electrical Engineering U.C. Davis, 1995. I would like to thank Ric Simmons, Dennis Hirsch, and their team at Ohio State for organizing this terrific event (The Ohio State University Moritz College of Law Round Table on Big Data and Criminal Law); my fellow participants for their insights and company; and my good friend Joseph Thai for always defending a robust First Amendment.

INTRODUCTION

In 1965, President Lyndon Johnson gave a special message to Congress on Law Enforcement and the Administration of Justice.[1] Decrying the crime rate and its fiscal and human costs, the President gave a far-reaching address in which he announced the appointment of a Presidential Commission to "probe . . . fully and deeply into the problems of crime in our nation."[2] When that Commission, which included future Supreme Court Justice Lewis F. Powell, Jr., issued its 340-page report in 1967, it said this about law enforcement's use of technology:

> The Scientific and Technological revolution that has so radically changed most of American society during the past few decades has had surprisingly little impact upon the criminal justice system. In an age when many executives in government and industry . . . ask the scientific and technical community for independent suggestions on possible alternatives and for objective analyses of possible consequences of their actions, the public officials responsible for establishing and administering the criminal law—the legislators, police, prosecutors, lawyers, judges, and corrections officials—have almost no communication with the scientific and technical community.
>
> Even small businesses employ modern technological devices and systems, but the Nation's courts are almost as close to the quill pen era as they are to the age of electronic data processing. The police, with crime laboratories and radio networks, made early use of technology, but most police departments could have been equipped 30 or 40 years ago as well as they are today.[3]

As is typically the case, a bit of perspective makes seemingly novel changes appear less so, and it is hard to fault current police departments and court systems

---

[1] Lyndon B. Johnson, Special Message to the Congress on Law Enforcement and the Administration of Justice (Mar. 8, 1965) (transcript available at http://www.presidency.ucsb.edu/ws/?pid=26800 [https://perma.cc/BBQ8-AQKA]).

[2] *Id.*; *see also* Exec. Order No. 11,236, 3 C.F.R. 329 (1964–1965), http://www.presidency.ucsb.edu/ws/?pid=105658 [https://perma.cc/6U7P-94SY].

[3] PRESIDENT'S COMM'N ON LAW ENF'T & ADMIN. OF JUSTICE, THE CHALLENGE OF CRIME IN A FREE SOCIETY 245 (1967), https://www.ncjrs.gov/pdffiles1/nij/42.pdf [https://perma.cc/L7VF-PWRT]. Among its more than two hundred recommendations, the Commission was the impetus for the 911 emergency telephone system. *Id.* at 250–51, 291; SEASKATE, INC., THE EVOLUTION AND DEVELOPMENT OF POLICE TECHNOLOGY 2–3 (1998), https://www.justnet.org/pdf/PoliceTech.pdf [https://perma.cc/WT4F-JZ5D]. The Seaskate report includes a police technology timeline running from the 1850s through 1996. SEASKATE, INC., *supra*, at 22–23. For articles on recent developments in policing technology, see *Police Technology: A Curated Collection of Links*, MARSHALL PROJECT, https://www.themarshallproject.org/records/247-police-technology [https://perma.cc/E5HE-V8E9] (last updated Jan. 27, 2018, 11:44 AM).

for seeking to employ the modern techniques of data analysis when history shows that they would otherwise be criticized for ignoring them.[4] The 1965 Commission established a Science and Technology Task Force which sought, among other things, "[t]o identify and describe crime control problems in a form susceptible to quantitative analysis," and "[t]o suggest organizational formats within which technological devices and systems can be developed, field tested, and rendered useful."[5] The Commission had high hopes for technology, since, in its absence, "[v]irtually all the efforts of the Commission [were] hampered by the pervasive lack of adequate objective information about crime and the possible effects of various techniques for crime control."[6] That criticism applied to then-current sentencing practice: "Each year, judges in this country pass roughly 2 million sentences. Almost all sentencing decisions are made with little or no information on the likely effect of the sentence on future criminal behavior."[7] That criticism applied to then-current policing practice: "About 200,000 policemen spend half of their time on 'preventive' patrol. Yet, no police chief can obtain even a rough estimate of how much crime is thereby 'prevented.'"[8] The contemporary 2018 answer to both problems would be the analytics of big data.

This is not to say that the Commission believed technology a panacea. Instead, while it was confident in technology's ability—"[t]echnology can . . . fill most reasonable requests and can thereby provide considerable help to law enforcement"—the Commission explicitly recognized technology's costs: "We must still decide what devices we want relative to the price we are willing to pay in dollars, invasion of privacy, and other social costs."[9] As we today contemplate a

---

[4] As Andrew Ferguson has pointed out, "[a]t some level, most decision-making systems involve prediction [and] [t]he criminal justice system is no exception." Andrew Guthrie Ferguson, *Policing Predictive Policing*, 94 WASH. U. L. REV. 1113, 1120 (2017). Ric Simmons has similarly noted that "[t]he criminal justice system has always been concerned with predictions." Ric Simmons, *Quantifying Criminal Procedure: How to Unlock the Potential of Big Data in Our Criminal Justice System*, 2016 MICH. ST. L. REV. 947, 948 (2016).

[5] PRESIDENT'S COMM'N ON LAW ENF'T & ADMIN. OF JUSTICE, *supra* note 3, at 245.

[6] *Id.* at 247.

[7] *Id.* This is not to say that algorithmic solutions had not been previously proposed. Parole boards began to use mathematical algorithms at least as early as the 1920s. *See* Jason Tashea, *Calculating Crime: Attorneys Are Challenging the Use of Algorithms to Help Determine Bail, Sentencing and Parole Decisions*, 103 A.B.A. J. 54, 57 (2017); OFFICE OF PROB. & PRETRIAL SERVS., ADMIN. OFFICE OF THE U.S. COURTS, AN OVERVIEW OF THE FEDERAL POST CONVICTION RISK ASSESSMENT 4 (2011), http://www.uscourts.gov/sites/default/files/pcra_sep_2011_0.pdf [https://perma.cc/SFU8-6QWA]. And such efforts have continued after the 1965 Commission, including during the 1980s when the original United States Sentencing Commission analyzed 10,000 records "to determine which distinctions were important in pre-guidelines [sentencing] practice." U.S. SENTENCING COMM'N, GUIDELINES MANUAL 5 (2016). It has only been in the past decades, however, that we have witnessed a very significant increase in the use of "evidence-based correctional practices." Cecelia Klingele, *The Promises and Perils of Evidence-Based Corrections*, 91 NOTRE DAME L. REV. 537, 551–52 (2015); *see also* Ferguson, *supra* note 4, at 1121.

[8] PRESIDENT'S COMM'N ON LAW ENF'T & ADMIN. OF JUSTICE, *supra* note 3, at 247.

[9] *Id.* at 246.

record everything world of "Fourth Amendment time machines,"[10] even the Commission's specific examples seem remarkably timely:

> It is technically feasible, for example, to cut auto theft drastically by putting a radio transmitter in every car in America and tracking all cars continuously. But this might cost a billion dollars and, even more important, create an intolerable environment of unending surveillance. Science can provide the capability, but the public as a whole must participate in the value discussion of whether or not the capability is worth its financial and social costs.[11]

The same crime-control/privacy balance is demonstrated in the Commission's recommendations for adjudication and policing, which include the following:

> The Commission recommends[] [that] [s]tatistical aids for helping in sentencing and selection of proper treatment of individuals under correctional supervision should be developed.[12]

> Criminal justice could benefit dramatically from computer-based information systems [that] . . . can aid in the following functions:
> [E]nabling a police officer to check rapidly the identification of people and property . . . .
> [A]ltering police deployment in response to changing patterns of crime on an hourly, daily, seasonal or emergency basis.
> [P]roviding a collection of anonymous criminal histories to find out how best to interrupt a developing criminal career and to achieve a better understanding of how to control crime.[13]

Again, the Commission did not ignore the risks in these developments, including those to information privacy.[14] It thus recommended decentralization of some records, audit controls, encryption, and automated purging of stale data.[15]

---

[10] *See* Stephen E. Henderson, *Fourth Amendment Time Machines (And What They Might Say About Police Body Cameras)*, 18 U. PA. J. CONST. L. 933, 937 (2016).

[11] PRESIDENT'S COMM'N ON LAW ENF'T & ADMIN. OF JUSTICE, *supra* note 3, at 246. *Cf.* United States v. Jones, 565 U.S. 400 (2012) (considering the constitutionality of longer term, single vehicle GPS surveillance).

[12] PRESIDENT'S COMM'N ON LAW ENF'T & ADMIN. OF JUSTICE, *supra* note 3, at 260.

[13] *Id.* at 266–67.

[14] The Commission recognized that:

> [W]henever government records contain derogatory personal information, they create serious public policy problems: [1] The record may contain incomplete or incorrect information. [2] The information may fall into the wrong hands and be used to intimidate or embarrass. [3] The information may be retained long after it has lost its usefulness and

So, big data analytics are relatively novel, including in their application to matters of criminal justice. But as a fifty-year-old Commission report teaches, applying developing technologies to policing and criminal adjudication is not so novel, and "[b]ecause of the enormous range of technological possibilities, it is essential to begin not with technology but with problems."[16] In other words, we ought not to consider in a vacuum what technology might do for criminal justice, but instead examine the specifics of our criminal justice system, identify problems and weaknesses therein (such problems are of course ample), and then ask what technology (and other tools) can do to improve them. Among other strengths, such a systems analysis will prevent the perfect from being the enemy of the good, allowing imperfect improvements to better our systems of criminal justice.

In that spirit, this roundtable symposium seems like an excellent step, bringing together experts in anthropology, sociology, data analysis, and law, and community stakeholders in legislators, interest groups, prosecutors, defense attorneys, judges, and police. And it is no surprise that it has been convened by Ric Simmons, who has written some of the pioneering work on the subject.[17] Moving forward, most difficult, perhaps, but critical, will be to include meaningful representation of 'the people,' here meaning those living on the streets that will be policed and adjudicated. In time, we can together formulate best practices for using modern data analysis in criminal justice.[18] Looking forward to—and hoping to be a part of—that solution, I propose some tentative rules (or, perhaps more technically, standards) that can then be improved through a continuing collaborative process.

> serves only to harass ex-offenders, or its mere existence may diminish an offender's belief in the possibility of redemption.

Id. at 268.

[15] *Id.* at 268–69. "A witness at congressional hearings claimed that 'the Christian notion of the possibility of redemption is incomprehensible to the computer.' By a policy of early purging of the files, computers permit restoring the notion of redemption to the existing manual files." *Id.* at 269.

[16] *Id.* at 246.

[17] *See generally* Simmons, *supra* note 4.

[18] This is not to naively assert that actuarial justice is new to our criminal justice system; as Andrew Ferguson has developed, quite the contrary is the case, and there is some rich literature on the subject. *See* Ferguson, *supra* note 4, at 1123–26. But it seems fair to say that the big data revolution is still upon us, and sometimes it is helpful to take a step back to first principles. Thus, in the words of Ric Simmons, "the rise of big data, with its vast amounts of information and vastly powerful methods of processing that data, brings the promise (or the threat) of a true revolution in the sophistication and the proliferation of [statistical prediction] tools." Simmons, *supra* note 4, at 949 n.2. Elizabeth Joh has similarly recognized that "[i]n this century, big data—in a variety of forms—may bring the next dramatic change to police investigations." Elizabeth E. Joh, *Policing by Numbers: Big Data and the Fourth Amendment*, 89 WASH. L. REV. 35, 37 (2014).

I. RULE ONE: BRING TECHNOLOGY TO THE PROBLEM

In traditional criminal investigations, police "move data to the question."[19] They may want to know, say, who killed *X* or who stole from *Y*. So, they gather evidence, moving data to the specific question at issue. Big data analysis does the opposite: store everything, and then "move the question to the data."[20] Which of these is the right approach for the meta-question—or higher-level question—of how to improve the criminal justice system?

Perhaps someday, data analytics will discover problems in our systems of criminal justice that we knew nothing of. But for the realistic future, I propose the better approach remains that of the 1965 Presidential Commission: "begin not with technology but with problems."[21] Thus, I do not believe the first topic of conversation is *what can big data analysis do for criminal justice?*, but instead *what are the contemporary problems in criminal justice?*. Otherwise, I fear significant time will be spent arguing fascinating problems of philosophy when those questions may become practically important only in a future day when we are seeking to replace quite a good system. The reality is, unfortunately, that our contemporary criminal justice implementation is likely a rather unjust, racially-biased one, in which case we ought not allow the perfect to be the enemy of the good. If criminal justice sentencing is currently a proprietary, too often racially- or otherwise-biased black box—in the form of a single judge's largely unexplained decision[22]—it is *this* to which we should compare any alternative. If *Terry*-stop policing is currently a proprietary, too often racially- or otherwise-biased black box—in the form of a single officer's ex-post boilerplate rationalization[23]—it is *this* to which we should compare any alternative.[24]

Likewise, a theoretically perfect technological solution might be implemented in a manner or system that entirely negates its utility. In other words, when it

---

[19] *See* Margaret Hu, *Orwell's* 1984 *and a Fourth Amendment Cybersurveillance Nonintrusion Test*, 92 WASH. L. REV. 1819, 1845 (2017) (quoting CIA Chief Technology Officer Ira Hunt); *see also* Joh, *supra* note 18, at 40–41 (making the same point for big data more generally).

[20] Hu, *supra* note 19, at 1845. Thus, while a study found that Los Angeles police use Automated License Plate Readers (ALPRs) for a variety of purposes, "the most common use of ALPRs is simply to store data for potential use during a future investigation." Sarah Brayne, *Big Data Surveillance: The Case of Policing*, 82 AM. SOC. REV. 977, 993 (2017).

[21] PRESIDENT'S COMM'N ON LAW ENF'T & ADMIN. OF JUSTICE, *supra* note 3, at 246.

[22] *See* Williams v. New York, 337 U.S. 241 (1949) (permitting largely unfettered judicial discretion in sentencing); *see also* Dean v. United States, 137 S. Ct. 1170, 1175 (2017) ("Sentencing courts have long enjoyed discretion in the sort of information they may consider when setting an appropriate sentence.").

[23] *See* Terry v. Ohio, 392 U.S. 1 (1968) (permitting temporary detentions upon reasonable suspicion).

[24] *See* Simmons, *supra* note 4, at 960–65 (arguing that algorithmic policing can improve *Terry* stops); *see also id.* at 978–79 (demonstrating the potential difference between actual and stated justifications for a stop).

comes to improving criminal justice, it is especially important for those of us working in the ivory tower to work with those in the trenches. We need to keep it real.

II. RULE TWO: BRING ONLY CREDIBLE TECHNOLOGY TO THE PROBLEM

This symposium has us off to a good start by including not only lawyers but also scientists. The criminal justice system has a sad history of folklore masquerading as science, perhaps best demonstrated by decades of arson investigation,[25] but also occurring with bite marks, hair comparisons, and other techniques.[26] When it comes to big data analysis, we instead want to follow the model of DNA, in which genuine, tested science was brought into the courtroom.

Of course, this does not mean that no difficult problems will be encountered—they always are when even a 'gold standard' science is put to novel use.[27] To state an obvious and minimal example, that machine learning can effectively deduce correlations says next to nothing about whether a particular algorithm is accurately predicting future dangerousness. Thus, the criminal justice system will have to work closely with well credentialed, independent data scientists, meaning scientists not employed by a vendor trying to sell a product or solution.

III. RULE THREE: THE DECIDER SHOULD BE HUMAN

There is increasingly robust science in automated decision making and artificial intelligence, and these technologies will surely continue to develop, perhaps even at a rapid pace.[28] But at least for now (and possibly forever) when it comes to criminal justice, it seems the ultimate decision should *always* be a human one. I say this not because I believe a human decision is more likely to be the *right*

---

[25] *See* Mark Hansen, *Badly Burned: Long-held Beliefs About Arson Science Have Been Debunked After Decades of Misuse and Scores of Wrongful Convictions*, 101 A.B.A. J. 37 (2015).

[26] *See* EXEC. OFFICE OF THE PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., FORENSIC SCIENCE IN CRIMINAL COURTS: ENSURING SCIENTIFIC VALIDITY OF FEATURE-COMPARISON METHODS (2016), https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf [https://perma.cc/X26P-XBVB].

[27] *See* Matthew Shaer, *The False Promise of DNA Testing*, ATLANTIC (June 2016), https://www.theatlantic.com/magazine/archive/2016/06/a-reasonable-doubt/480747/ [https://perma.cc/6R4E-R6UX] (chronicling some of the problems with crime laboratories and mixed DNA samples).

[28] *See* MAX TEGMARK, LIFE 3.0: BEING HUMAN IN THE AGE OF ARTIFICIAL INTELLIGENCE 40–42 (2017) (describing timeline disagreement among experts). For an argument that even current capabilities call for an increased use of computers in sentencing, see Mirko Bagaric & Gabrielle Wolf, *Sentencing by Computer: Enhancing Sentencing Transparency and Predictability, and (Possibly) Bridging the Gap Between Sentencing Knowledge and Practice*, 25 GEO. MASON L. REV. (forthcoming 2018).

decision, if 'right' means accurately predicting or discerning what is intended to be predicted or discerned. Indeed, I suspect quite often the human decision is currently rather appalling in this regard. I instead urge this rule simply because, on some intuitive level, it seems important to our humanity.[29]

The 1965 Presidential Commission noted that "[a] witness at congressional hearings claimed that 'the Christian notion of the possibility of redemption is incomprehensible to the computer.'"[30] Without delving into 'Christian' notions or those of any other particular faith, I suspect that most all of them share an intuition there being expressed: that it is preferable to be judged by a flawed human (or, better yet, a superhuman god) than by a computational machine. In the words of Kiel Brennan-Marquez:

> If the answer comes back to error-reduction—if the point of judicial oversight is simply to maximize the overall number of accurate decisions—machines could theoretically do the job as well as, if not better than, humans. But if the answer involves normative goals beyond error-reduction, automated tools—no matter their power—will remain, at best, partial substitutes for judicial scrutiny.[31]

I agree with Brennan-Marquez that the latter is the case,[32] and thus, even though I have not defended the view but instead merely appealed to intuition, I believe our criminal justice system should require that a human decision-maker decide how to act on any computationally-suggested result.[33]

---

[29] It is also possible that retaining a human decision-maker will increase the perceived legitimacy of the system and thus further procedural justice. On the importance of such perceived legitimacy, see Ric Simmons, *Big Data and Procedural Justice: Legitimizing Algorithms in the Criminal Justice System*, 15 OHIO ST. J. CRIM. L. 573 (2018).

[30] PRESIDENT'S COMM'N ON LAW ENF'T & ADMIN. OF JUSTICE, *supra* note 3, at 269.

[31] Kiel Brennan-Marquez, *"Plausible Cause": Explanatory Standards in the Age of Powerful Machines*, 70 VAND. L. REV. 1249, 1250 (2017).

[32] *Id. Cf.* Simmons, *supra* note 4, at 1009–16 (arguing for at least some entirely mechanical determinations of reasonable suspicion, but also positing how an officer might incorporate her intuitions using Bayesian inference). For some very interesting arguments regarding the use of artificial intelligence in administrative decision-making more generally, see Mariano-Florentino Cuéllar, *Cyberdelegation and the Administrative State* (Stanford Pub. Law, Working Paper No. 2754385, 2018, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2754385 [https://perma.cc/9Z3T-BAP7]).

[33] *But see* Andrew D. Selbst, Response, *A Mild Defense of Our New Machine Overlords*, 70 VAND. L. REV. EN BANC 87 (2017) (thoughtfully challenging Brennan-Marquez's specific hypothetical and more generally his assumptions about both human and machine decision-making). To Selbst, human decision-making need not be inherently different from a machine's: "The human's role . . . is to detect when input information might be missing from the model, and then supply it," a "sanity check." *Id.* at 101. Selbst agrees that humans "must remain involved in the process in case the machines fail to take into account certain contextual facts that are important, in case the results implicate values unconsidered by the data miners, or in case the results just make no sense," but he is

IV. RULE FOUR: THE CODE AND DECISION ALGORITHM SHOULD BE ACCESSIBLE (THOUGH NOT NECESSARILY PUBLIC) BUT THE DECISION ALGORITHM NEED NOT BE EXPLAINABLE

We have constitutional norms of public trial[34] and we should have the same for mechanisms of policing.[35] Thus, any underlying algorithm used in adjudication or policing should be publicly available, or—at the very least—available for and subject to inspection by independent authorities.[36] If Amazon prefers not to share the code or ultimate algorithm it uses to better pitch products, so be it. But if a company prefers not to share the code or ultimate algorithm the State of Oklahoma uses in bail, diversion, sentencing, or parole decisions, society should have a ready answer: the state cannot do business with you. The societal cost in lack of trust is simply too high when our criminal justice system runs on knowable but nonetheless-kept-secret algorithms.

By contrast, I do not believe we ought to reject all machine-learned decision algorithms that cannot be meaningfully explained, sometimes termed 'opaque AI' or contrasted with 'explainable AI.'[37] So long as the underlying algorithm that *lead to* that decision algorithm is accessible (this rule), so long as the decision

---

willing to limit that human role because people often step in only to screw up the logic of the process. *Id.* at 101–02.

[34] The Sixth Amendment requires that "[i]n all criminal prosecutions, the accused shall enjoy the right to a speedy and public trial." U.S. CONST. amend. VI. In the words of Justice Harlan:

> Essentially, the public-trial guarantee embodies a view of human nature, true as a general rule, that judges, lawyers, witnesses, and jurors will perform their respective functions more responsibly in an open court than in secret proceedings. A fair trial is the objective, and "public trial" is an institutional safeguard for attaining it.

Estes v. Texas, 381 U.S. 532, 588 (1965) (Harlan, J., concurring) (internal citation omitted) (quoted by the Court in Waller v. Georgia, 467 U.S. 39, 46 n.4 (1984)).

[35] *See* Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. (forthcoming 2018) (arguing against pervasive use of trade secrets in criminal justice); Elizabeth E. Joh, *The Undue Influence of Surveillance Technology Companies on Policing*, 92 N.Y.U. L. REV. 101 (2017) (warning of the role that technology vendors play in contemporary policing).

[36] *See* Selbst, *supra* note 33, at 91, 97 (calling for disclosure and testing); Simmons, *supra* note 4, at 994–99 (same). While I would prefer full public disclosure, some argue that there are longstanding pressures discouraging private companies from developing policing (and perhaps adjudicatory) technologies. *See* SEASKATE, INC., *supra* note 3, at 7–8. Depending upon the relevance of these factors in this context, some measure of proprietary protection might be unavoidable.

[37] *See* Selbst, *supra* note 33, at 90–91 (explaining machine learning, including that "advanced versions of machine learning can key in on variables that have no semiotic value to humans, and thus we cannot truly comprehend them even as approximation."); David Gunning, *Explainable Artificial Intelligence (XAI)*, DEF. ADVANCED RES. PROJECTS AGENCY, https://www.darpa.mil/program/explainable-artificial-intelligence [https://perma.cc/2PDY-SH2M] (last visited Apr. 7, 2018) (introducing explainable artificial intelligence).

algorithm is available for extensive testing for bias and accuracy (this rule),[38] so long as nothing is artificially held back as confidential (this rule), and so long as the ultimate decision maker is human (rule three), I believe there is a place in our criminal justice system for machine learned, non-humanly-explainable decision algorithms. I would, however, reject practices like that in Wisconsin, where sentencing partially relies upon a proprietary code available for independent testing, but for which the algorithms are kept strictly secret.[39] There is an important difference between an algorithm a private company chooses not to share (unacceptable) and a machine-learned algorithm of such complexity that humans can test but not articulate it (acceptable).

V. RULE FIVE: ANY TECHNOLOGY IMPLEMENTATION SHOULD BALANCE COSTS AND BENEFITS

The Fourth Amendment has operated for hundreds of years—and is equipped to operate for hundreds of years more—because it requires balance: searches and seizures must be *reasonable*,[40] which typically requires, says the Court, balancing the government need against the intrusion into security.[41] In other words, how compelling is the government need, how well do these means further that need, and how do these means impact privacy and liberty (a concern that in the big data context includes any decrease in the traditional practical obscurity of records)?

The 1965 Presidential Commission recognized that, as a matter of good policy, this same balance should be considered for changes to our criminal justice

---

[38] *See, e.g.*, Anupam Datta et al., *Algorithmic Transparency via Quantitative Input Influence*, *in* TRANSPARENT DATA MINING FOR BIG AND SMALL DATA 71–94 (Tania Cerquitelli et al., eds., 2017) (proposing a method of such testing).

[39] *See* State v. Loomis, 881 N.W.2d 749, 757, 760–64 (Wis. 2016), *cert. denied*, 137 S. Ct. 2290 (2017). "Northpointe, Inc., the developer of COMPAS, considers COMPAS a proprietary instrument and a trade secret. Accordingly, it does not disclose how the risk scores are determined or how the factors are weighed." *Id.* at 761. Thus, neither party could even determine how the algorithm does or does not account for offender sex. *Id.* at 765. The Wisconsin Supreme Court allowed limited use of the tool at sentencing but required this caution, among others: "The proprietary nature of COMPAS has been invoked to prevent disclosure of information relating to how factors are weighed or how risk scores are determined." *Id.* at 769.

[40] The Fourth Amendment requires that "[t]he right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated . . . ." U.S. CONST. amend. IV.

[41] *See* Riley v. California, 134 S. Ct. 2473, 2484 (2014). The Court explained:

> Absent more precise guidance from the founding era, we generally determine whether to exempt a given type of search from the warrant requirement by assessing, on the one hand, the degree to which it intrudes upon an individual's privacy and, on the other, the degree to which it is needed for the promotion of legitimate governmental interests. Such a balancing of interests supported the [rule in a previous case].

Id. (internal quotation marks and citation omitted).

process even when they are not required by our Constitution,[42] a norm that should apply to both policing and adjudication. So, as part of considering whether to adopt any new technology, it should be asked how it would affect accuracy, fairness (broadly conceived), limited government, and efficiency.[43] This might seem sufficiently obvious as to not merit articulation, but the reality is that even the most obvious criterion—accuracy—is too often ignored in the excitement of the moment. As Andrew Ferguson has chronicled, the history of predictive analytic policing "has been invention first, then adoption, and finally assessment only after the fact."[44] We cannot be unrealistic in our demands, as of course every technology was once novel and sometimes we can only learn as we go. But nor should we ignore the basic norms of legitimacy by adopting inadequately considered technologies.

Moreover, in considering these norms we must remember—and not shy away from—the fact that limited government norms, such as the Fourth Amendment, tend to be *anti*-accuracy norms. Sometimes we can achieve a win-win, which is of course best of all: perhaps police get more reliable information on stopped individuals by scanning a biometric marker, but in return learn only what they need to know, such as whether a driver has a currently valid license or outstanding warrant but not (yet) revealing the details thereof. Other times, however, a win-win is not possible and we might choose a less accurate, fair, or efficient system because it is necessary to keep the government at bay.[45] Other times, we might choose less privacy because it is more fair.[46] The idea is not to prejudge the outcome, but rather to require the honest consideration.

---

[42] The Commission of course recognized the importance of constitutional rules (see, e.g., PRESIDENT'S COMM'N ON LAW ENF'T & ADMIN. OF JUSTICE, *supra* note 3, at 93–95), but it ultimately called for societal normative judgments consistent with those rules (see, e.g., *id.* at 95, 246). *See also supra* notes 9–15 and accompanying text.

[43] *See* JOSHUA DRESSLER & GEORGE C. THOMAS III, CRIMINAL PROCEDURE: PRINCIPLES, POLICIES AND PERSPECTIVES 34–41 (5th ed. 2013) (identifying these as four norms of our criminal process). For an extensive analysis of the many considerations that should factor into the legitimacy of predictive policing, see Ferguson, *supra* note 4, at 1148–94.

[44] Ferguson, *supra* note 4, at 1194.

[45] *See, e.g.*, PRESIDENT'S COMM'N ON LAW ENF'T & ADMIN. OF JUSTICE, *supra* note 3, at 268 (calling for forced inefficiencies in records storage and thus access).

[46] *See* I. Bennett Capers, *Race, Policing, and Technology*, 95 N.C. L. REV. 1241 (2017) (arguing that we ought to increase policing for some—via policing everyone—in order to decrease racial injustice for others).

VI. RULE SIX: ANY PROPOSAL FOR PRIVACY-BASED RESTRICTION ON GOVERNMENT TECHNOLOGY SHOULD CONSIDER NON-GOVERNMENT USE

A longstanding Fourth Amendment principle is that law enforcement need not *alone* shield its eyes from what other persons freely observe.[47] I have defended this as a legal and normative view,[48] and believe it should also have play when it comes to big data and criminal justice. Any time the reason for restricting criminal justice use of big data analytics is for reasons of privacy harm, we should be sure that precisely the same harm is not already occurring in the non-government sector. In general, it does not make sense to handcuff *solely* our criminal justice actors, especially when doing so has accuracy, fairness, or efficiency costs that are not meaningfully offset by any privacy gain. Of course, there might sometimes be special reason to limit only the government,[49] or the proper solution might be to, when possible, legislatively restrict private actors.[50] But if non-government actors are already routinely making use of certain data in a certain way, the burden should be on anyone arguing the government should not be permitted to do the same.

VII. RULE SEVEN: PRIVACY PROTECTIONS SHOULD NOT BE LIMITED TO ACQUISITION RESTRAINTS—THEY SHOULD CONSIDER ROBUST USE RESTRICTIONS

Traditionally, law enforcement restraints—including the judicially declared restraints of the Fourth Amendment—have focused almost entirely upon the initial acquisition of information: there is significant restraint, say, upon law enforcement entering a home,[51] but once information therein is lawfully seized, its analysis is largely independent of any continued constitutional constraint. This system is no longer feasible in a world of massive over-collection of digital data and mass

---

[47] *See* California v. Ciraolo, 476 U.S. 207, 213 (1986) ("The Fourth Amendment protection of the home has never been extended to require law enforcement officers to shield their eyes when passing by a home on public thoroughfares.").

[48] *See* Mark Jonathan Blitz, James Grimsley, Stephen E. Henderson & Joseph Thai, *Regulating Drones Under the First and Fourth Amendments*, 57 WM. & MARY L. REV. 49, 68–72, 74–77 (2015).

[49] *See* 18 U.S.C. § 2702 (2015) (restricting certain company disclosures only if to the government); CHRISTOPHER SLOBOGIN, PRIVACY AT RISK: THE NEW GOVERNMENT SURVEILLANCE AND THE FOURTH AMENDMENT 98–108 (2007) (arguing that there are freedom of speech and association, freedom of movement and repose, right to privacy, and Fourth Amendment reasons to constitutionally restrict government public-camera surveillance).

[50] This may be impossible if, for example, the First Amendment protects that conduct. *See* Rule Seven, *infra* Part VII.

[51] *See* Payton v. New York, 445 U.S. 573, 590 (1980) ("In terms that apply equally to seizures of property and to seizures of persons, the Fourth Amendment has drawn a firm line at the entrance to the house. Absent exigent circumstances, that threshold may not reasonably be crossed without a warrant.").

surveillance, and thus privacy rules should often constrain the later access, use, and dissemination of lawfully collected data.[52]

VIII. RULE EIGHT: ANY CLAIM TO FIRST AMENDMENT RIGHTS IN CONSUMER DATA, IN CRIMINAL JUSTICE ALGORITHMS, OR IN ALGORITHMIC RESULTS SHOULD BE OPPOSED

In both the Apple encryption controversy and the lesser scuffle over law enforcement access to data conveyed through Amazon's 'Echo,' the companies asserted First Amendment rights. For Apple, it was a First Amendment right against compelled speech through forced computer code and use of its encryption keys.[53] For Amazon, it was a First Amendment right in customer voice data and Alexa's responses thereto.[54] While both claims can be seen as consumer-friendly, each also has serious risk. If companies are able to claim First Amendment rights in consumer data, and perhaps even if companies can merely *assert* such rights, it could become difficult to legislatively restrict what companies do with that data, meaning it could become difficult to legislatively protect privacy. Similarly, if companies are able to assert First Amendment rights against needed code, it could become difficult to control algorithms that become embedded in our daily lives.

As a privacy scholar, I am broadly concerned about these types of First Amendment claims. I recognize, however, that these are difficult and contested issues. Courts have held that computer code can be protected by the First Amendment,[55] that corporations have at least some First Amendment rights,[56] that there is a First Amendment right to some information gathering,[57] and that certain

---

[52] *See generally* Henderson, *supra* note 10 (explaining and arguing for use restrictions); *see also* Kiel Brennan-Marquez & Stephen E. Henderson, *Fourth Amendment Anxiety*, 55 AM. CRIM. L. REV. 1 (2018) (arguing that the Supreme Court has implicitly authorized such Fourth Amendment restraints).

[53] Apple Inc.'s Motion to Vacate Order Compelling Apple Inc. to Assist Agents in Search, and Opposition to Government's Motion to Compel Apple Inc.'s Assistance at 32–34, In the Matter of the Search of an Apple iPhone Seized During the Execution of a Search Warrant on a Black Lexus IS300, Cal. License Plate 35KGD203, No. CM 16-10 (C.D. Cal. Feb. 25, 2016), https://assets.documentcloud.org/documents/2722457/Apple-s-Motion-Opposing-iPhone-Order.pdf [https://perma.cc/BZH2-Q78N].

[54] Memorandum of Law in Support of Amazon's Motion to Quash Search Warrant at 9–12, Arkansas v. Bates, No. CR-2016-370-2 (Cir. Ct. Ark. Feb. 17, 2017), https://assets.documentcloud.org/documents/3473747/Amazon-Memorandum-Seeking-to-Quash-Echo-Search.pdf [https://perma.cc/ZA7B-8AJA].

[55] *See* Universal City Studios, Inc. v. Corley, 273 F.3d 429, 445–60 (2d Cir. 2001) (holding computer code can be protected speech but nonetheless permitting a Digital Millennium Copyright Act takedown); 321 Studios v. Metro Goldwyn Mayer Studios, Inc., 307 F. Supp. 2d 1085, 1099–1101 (N.D. Cal. 2004) (same).

[56] *See* Citizens United v. Fed. Election Comm'n, 558 U.S. 310, 365 (2010) (striking down a statute barring corporate expenditures for electioneering communications).

[57] *See, e.g.*, Stanley v. Georgia, 394 U.S. 557, 564 (1969) ("It is now well established that the Constitution protects the right to receive information and ideas."); Fields v. City of Philadelphia, 862

restrictions upon data use and disclosure can violate the First Amendment.[58] But if a criminal justice vendor is unrestricted in using private data in its algorithm, there can be serious privacy cost. And if the vendor can prevent algorithm disclosure via the First Amendment (compelled speech), then the public loses its right to comprehend the criminal justice process. And if a vendor is permitted to shield perhaps flawed results through the First Amendment, or if it refuses to code necessary testing for such flaws where the code has already had criminal justice impact, then the public loses its right to accurate and fair criminal justice.

Thus, when code is distributed for criminal justice use, we require a doctrine of public dedication or waiver that would prohibit the vendor from later claiming First Amendment rights contrary to that criminal justice interest. Governments should therefore contract to prohibit such claims, and perhaps appropriately tailored legislation should require such waiver.

IX. RULE NINE: ANY SIGNIFICANT TECHNOLOGY IMPLEMENTATION SHOULD REQUIRE PUBLIC NOTICE AND COMMENT

As Christopher Slobogin has carefully developed, while police departments are administrative agencies and policing has developed intricate and pervasive forms of wide-scale surveillance, police decision-making has not historically been subjected to the rules applicable to other administrative rulemaking, including mandated notice and comment procedures.[59] I agree with Slobogin that this should change, and I believe we should apply some similarly efficacious framework to structural changes within our systems of criminal adjudication. By requiring some manner of public notice and comment on wide-scale changes—or even by encouraging such voluntary procedures—we would enable significantly better considered and informed decisions in both policing and adjudication, and, in the latter case, make the traditionally least accountable branch slightly more responsive to the will of the people.[60]

F.3d 353, 359 (3d Cir. 2017) ("recording police activity in public falls squarely within the First Amendment right of access to information"); *see also* Blitz et al., *supra* note 48, at 80–81, 85–109 (discussing First Amendment rights to gather and record).

[58] *See, e.g.*, Sorrell v. IMS Health, Inc., 564 U.S. 552 (2011) (striking down a statute that restricted certain disclosures and uses of pharmacy records); United States v. Stevens, 559 U.S. 460 (2010) (striking down a statute that criminalized the commercial creation, sale, or possession of depictions of animal cruelty); Bartnicki v. Vopper, 532 U.S. 514 (2001) (striking down a statute that criminalized intentional disclosure of illegally intercepted communications).

[59] *See generally* Christopher Slobogin, *Policing as Administration*, 165 U. PA. L. REV. 91 (2016).

[60] For some thoughts on how we might benefit from "algorithmic impact statements," see Andrew D. Selbst, *Disparate Impact in Big Data Policing*, 52 GA. L. REV. 109 (2017).

X. RULE TEN: MOST DECISIONS SHOULD BE MADE BY THE STATES

States of course have the general police power, and "[p]erhaps the clearest example of traditional state authority is the punishment of local criminal activity."[61] Thus, big data analysis, like other criminal justice solutions, should primarily be adopted at the state level. As it has often done, however, the federal government can assist through studies, resources, and voluntary standards that will otherwise not be realistically available,[62] as well as to lead by example in the federal courts.

CONCLUSION

The tools of big data analysis are 'dual use,' meaning, just like any other technology, they can be used for good and for ill. Given their increasing ability, there seems little benefit to asking *whether* they have a place in the criminal justice system. Instead, the useful question is *when* they have such a place, and I have proposed—among other preliminary rules—that we first look to the problems in our criminal justice system needing solutions, and then consider big data options along with any others that can be proposed or considered. We should not naively expect our 'big data progress' to be unerringly straight and forward-moving any more than we would expect this of more traditional solutions. It will be, as in all of life, full of zigs, zags, and even some backward falls. If we are too concerned about this, we have not internalized how our criminal justice system currently (mal)functions. At the same time, criminal justice failures are some of society's most upsetting and harmful. Therefore, by beginning to formulate some principles to direct the use of big data analysis, hopefully we can minimize those inevitable errors while beginning to improve our systems of criminal justice.

---

[61] Bond v. United States, 134 S. Ct. 2077, 2089 (2014) (citation omitted).

[62] *See* SEASKATE, INC., *supra* note 3, at 8–19 (describing such role of the National Institute of Justice and other federal agencies).