# THE OPEN SOCIETY AND ITS DIGITAL ENEMIES: A REPLY TO PROFESSORS BAMBAUER AND GARRETT

*Erik Luna*† *& Joshua A.T. Fairfield*††



> [I]f we wish to remain human, then there is only one way, the way into the open society.
>
> – Karl Popper[1]
>
> [T]rust, but verify.
>
> – Ronald Reagan[2]

## INTRODUCTION

Don't let our title fool you. Jane Bambauer and Brandon Garrett are friends of an open society, not its enemies. In distinct and distinctly important ways, they have engaged and expounded upon our concept of digital innocence,[3] and in so doing, they have emboldened us to find the courage of our convictions and indeed to sharpen and intensify our sense that an open society demands even more government transparency, especially in data-driven criminal cases. In fact, we agree with their animating principles. Professor Bambauer reminds us that technology must be made to serve the human demand for truth and the drive to achieve it within a decent system of law. For Professor Garrett, the issue is one of recognizing and applying our concept over the full ambit of criminal justice. Their pieces can be construed as responding not only to us but to one another as well (inadvertently, of course). This interactive work is critical, providing precisely the kind of rigorous

† Sydney and Frances Lewis Professor of Law, Washington and Lee University School of Law.

†† Professor of Law, Washington and Lee University School of Law.

1 1 K.R. POPPER, THE OPEN SOCIETY AND ITS ENEMIES 177 (1945).

2 Remarks on Signing the Treaty Eliminating Intermediate-Range and Short-Range Nuclear Missiles, 23 WEEKLY. COMP. PRES. DOC. 1457, 1458 (Dec. 8, 1987).

3 *See* Jane Bambauer, *Collection Anxiety*, 99 CORNELL L. REV. ONLINE 195 (2014); Brandon Garrett, *Big Data and Due Process*, 99 CORNELL L. REV. ONLINE 207 (2014).

challenge that improves legal thought. Relatedly, our thesis for this brief Essay is that the entire system of data-driven prosecution must be susceptible to testing if it is to be considered reliable.

In Part I, we offer brief responses to the separate pieces by Professors Garrett and Bambauer. Part II then details an undertheorized danger from unequal access to data-mining technology, one that could fundamentally destabilize the relationship between citizen and state. We believe that systems that are functionally ex parte and hidden, untested and unchallengeable, operating in an environment of unequal access to data, are the digital enemies of an open society.

## I
## RECIPROCITY VERSUS ACCURACY VERSUS ACCESS

In *Digital Innocence*, we argue that mass cyber-surveillance data must be accessible to exonerate the innocent,[4] appealing to the basic notion of fairness manifested in reciprocity. What the government uses to convict must also be available to exonerate. Professor Garrett's introductory paragraph paints with exquisite clarity a picture of people's "electronic footprints."[5] But as Professor Bambauer incisively argues, digital reciprocity may suggest that gathering even more data is a good thing. After all, the information society is currently building a virtuous cycle of data reciprocity. Shouldn't the cycle be further stimulated by allowing more information to be gathered by both citizen and state? Is this not the opposite of the usual parade of horribles about mass data collection? In response, we wish to make a few points regarding the goal of reciprocity and the relationship between accuracy and access.

Initially, we reiterate what may be a self-evident point: requiring the government to provide information it possesses is not the same as approving the government's acquisition of the information to begin with—and it certainly is not saying the government should now gather more information. By analogy, if prosecutors charge someone after he

---

4 *See* Joshua A.T. Fairfield & Erik Luna, *Digital Innocence*, 99 CORNELL L. REV. 981 (2014).

5 Garrett, *supra* note 3, at 207.

> Electronic traces, left through a smartphone or other device, can be tracked to the scene of a crime, or they can place a person far from a crime scene. Those traces can sometimes be tracked far more reliably than the types of trace evidence traditionally examined at crime scenes, like hairs, fibers, fingerprints, or tool marks. Cases have already come to light in which individuals have cleared their names by using digital evidence, whether a surveillance video, an E-Zpass tag, a cellphone-tower signature, or an e-mail chain, and far more are certain to occur in the future. By the same token, individuals may be falsely implicated due to errors in large government or commercial databases, and evidence of innocence may linger in such archives without ever coming to light.

*Id.* (footnotes omitted).

has been tortured by state agents, the government is obliged to provide him a copy of any recording of his statements[6]—but nobody would say the provision of a transcript serves as approval of torture or, worse yet, an invitation for more of it. Likewise, law enforcement may be required to provide a receipt for the things seized during the execution of a warrant, which, in many jurisdictions, has to be returned to the issuing magistrate with an inventory of the seizure.[7] Neither of these incidents of search-and-seizure law[8] vouchsafes the legality of the underlying warrant itself or the ensuing searches and seizures. Nor does providing information to the court and the affected individuals imply a need for still more government incursions into constitutionally protected areas.

Generally speaking, just because something is required of government because of its actions, and that requirement has some connection to what most people would agree to be an unmitigated good—the exoneration of the innocent, for instance—does not mean that the interests of the public, and those of suspects and defendants in particular, would necessarily be served by government increasing the activity that triggers the aforesaid requirement in the first place. Ultimately, more data-mining does not by itself beget better outcomes in terms of citizen-state reciprocity, or necessarily lead to improvements overall in wrongful accusations and convictions. Professor Bambauer joins Jack Balkin in noting that the government gorges on data yet behaves like an information scrooge.[9] The government seems unlikely to share willingly what it devours and digests. One should also keep in mind that reciprocity can be an essential part of both virtuous and vicious cycles. Think of tit-for-tat games. If I treat you well, and you treat me well, we both benefit. But if one of us treats the other poorly, and the response is poor treatment in kind, then everyone is worse off. Although reciprocity has many merits,[10] it is not the be-all and end-all for us in *Digital Innocence*. Instead, reciprocity is a means to another end—truth-seeking—offering a way to ensure accurate verdicts and fair sentences via a more balanced relationship between citizen and state in criminal proceedings.

Professor Bambauer's concerns about collection anxiety would be entirely correct if citizens had equal ability to gather data about and from the government on their own account. Many of these issues were

---

6 *See, e.g.*, FED. R. CRIM. P. 16(a)(1)(B).

7 *See, e.g.*, 2 WAYNE R. LAFAVE ET AL., CRIMINAL PROCEDURE § 3.4(*l*) (3d ed. 2007).

8 *See, e.g.*, *id.* (noting that, although such conditions come from statute or court rule, "the Supreme Court has intimated that the Fourth Amendment is implicated to some degree in the exhibition/delivery requirement").

9 *See* Jack M. Balkin, *The Constitution in the National Surveillance State*, 93 MINN. L. REV. 1, 17 (2008).

10 One of us has suggested that it is important in citizen-state relations and the creation of trust. *See* Erik Luna, *Transparent Policing*, 85 IOWA L. REV. 1107, 1163 (2000).

broached by David Brin in his groundbreaking manifesto, *The Transparent Society*.[11] In principle and as expressed by Brin, increased data-gathering should inure to the benefit of all. Citizens' increased access to data about government activity ought to provide greater protection of civil liberties than the losses occasioned by street cameras, license-plate readers, and, of particular interest, warrantless bulk collection of call, social network, Internet lookup, and geolocation data. This does not happen in practice, however. Citizens have limited access to the data and are unlikely to obtain more from the government as a matter of grace. So whether one views the question of digital exoneration as being about *access* or *accuracy* can lead to starkly different outcomes in the citizen-state balance of power.

Bambauer is persuasive in arguing that criminal justice decision-making must be more accurate and further, that accuracy will only be achieved if more data is gathered. She is to be lauded for her push to move from digital witchcraft to digital science. But an essential aspect of scientific discovery is testability or even "falsifiability,"[12] as Karl Popper termed it. Putative increases in accuracy are not to be believed if they cannot be tested or, as Popper would require, if they are not capable of being proven false.[13] Such claims should survive the crucible of empiricism, including challenges by those interested in demonstrating error in a particular case. But today, testing of the kind called for by the concept of digital innocence—expressed as a request for reciprocity in sharing the fruits of mass data-gathering—takes place in a less-than-cooperative environment. The era's current standards are personal transparency and state secrecy, which means that a demand for reciprocity now and under these circumstances is really a call for balance—specifically, allowing the surveilled access to the products and mechanisms of state surveillance to prove their innocence.

The balance has both constitutional and constitutive aspects, which

---

11 *See* DAVID BRIN, THE TRANSPARENT SOCIETY 81 (1998) (discussing "reciprocal transparency"); *see also* Fairfield & Luna, *supra* note 4, at 986 n.26 and accompanying text.

12 *See generally* KARL POPPER, THE LOGIC OF SCIENTIFIC DISCOVERY 17 (1959) ("This means that their form must be such that to verify them and to falsify them must both be logically possible.").

13 To be sure, Popper's many critics claim "his account of science is fundamentally flawed." David Papineau, *The Proof Is in the Disproof*, N.Y. TIMES (Nov. 12, 2000), § 7, at 28, *available at* http://www.nytimes.com/2000/11/12/books/the-proof-is-in-the-disproof.html.

> It can seem plausible to view science as a succession of brave conjectures and honest refutations. But few philosophers today think that this explains the worth of science. The whole point of science is to provide a trustworthy guide to the future, not a series of hopeful guesses.

*Id.* Accepting this as true, falsifiability would still be among the principles employed by scientists and others (e.g., policymakers) to determine whether a claim is even scientific, or instead it must compete with other, at times unscientific, normative principles (equality, freedom of expression, protection against unreasonable invasions of privacy, etc.) to decide issues of law and concrete fact-patterns.

have been fundamentally destabilized by the state's largely untrammeled ability to collect and parse data, creating a one-way ratchet in favor of government data-gathering and data-mining. Professor Bambauer recognizes the problem in a concluding caveat:

> [T]he benefits of data collection that I trumpet in this Essay cannot be harnessed without transparency. . . . Like gluttons, the government will collect everything it can; and like misers, it will keep the data and its operations secret. Even if the gluttony is not as bad as we think it is, a lack of transparency will throw all the benefits into doubt.[14]

We agree, in fact, so strongly that we think the argument for increasing the powers of collection is premature until the government's aversion to transparency is addressed in no uncertain terms. It is access to data, and not accuracy, that is currently askew. Citizens' information flows easily to government, but government information does not return to the people, especially when it concerns methods and techniques (e.g., databases and search terms). Otherwise, Edward Snowden would have revealed nothing. Of course, one can imagine an arrangement in which the public would have the same access to the government's information, and maybe to its data-gathering methods too. But this is not our world. The Snowden saga has shown that in order for people everywhere to appreciate online reality—and conceivably move in the direction of the sort of reciprocity that Professor Bambauer posits already exists—a hero/traitor had to risk career, reputation, and freedom.[15]

Professor Bambauer claims not only that "the state is in the best position to . . . ensure that exonerating information is preserved," but that "law enforcement agencies could use the data to exclude the suspect before an arrest or charge is brought in the first place."[16] The argument is forceful, but it does not reflect the lived experience of either Americans or Europeans with government-sponsored data-gathering. Alas, there is the ratchet of data collection, where increasing the scope of information gathered tends to work in one direction only, in favor of government surveillance. As data-gathering increases, more citizens' lives become more vulnerable to more state actions, but the data-gathering measures themselves remain obscure to the public. And there is no reason to believe that law enforcement

---

14 Bambauer, *supra* note 3, at 204–05 (footnotes omitted).

15 In theory, the U.S. Department of Justice could have pursued capital charges against Snowden, though it declined to do so in 2013 in hope of obtaining Snowden's extradition. *See, e.g.*, Adam Gabbatt, *US Will Not Seek Death Penalty for Edward Snowden, Holder Tells Russia*, THE GUARDIAN (July 26, 2013, 2:03 PM), http://www.theguardian.com/world/2013/jul/26/us-no-death-penalty-edward-snowden-russia.

16 Bambauer, *supra* note 3, at 195.

would like it any other way. Society has not known regularized access to digital information in order to exonerate the innocent or, for that matter, just to cast light on government action. Instead, our history of signals collection includes opaque regimes of government surveillance coupled with officials denying the same.[17] We are thus left to speculate whether the predicates could come to fruition, or whether instead a law enforcement agency might use looser search terms until it gets the match it wants (we return to this later).[18]

The size of the government-maintained databases is not the problem, then—it is the lack of access for those embroiled in the criminal justice system, and perhaps others as well (e.g., researchers), who have an interest in testing the validity of the systems. Accurate information is not particularly helpful to the defense when it remains unshared. Gathering more data increases law enforcement power, but only providing access can give the innocent defendant a chance. The concept of digital innocence speaks about both accuracy and access, for sure; but of the two, access is the critical balancing force. And access is precisely what the public has been continually denied as government expands surveillance.

Claims of accuracy depend on access to the algorithms, searches, and results in order to test those claims. Such testing is impeded if not effectively foreclosed by the asymmetry in data access between citizen and state. The upshot of government obfuscation, coupled with people's digital nakedness, can be painfully ironic. While the NSA gathers what it terms "metadata"—the term is absurdly deceptive—involving every electronic contact of every U.S. person, citizens can be arrested for recording conversations to which they themselves are a party when stopped in a public place by a police officer, for instance, as seen in recent incidents in Ferguson.[19] That is, citizens are vulnerable to the state in their most unguarded transactions in private matters, and the state is protected from the public eye even in its most public transactions that are actively circumscribed by constitutional protections.

---

17 *See, e.g.*, Fairfield & Luna, *supra* note 4, at 1012–13, 1028 n.304. Long before it was collecting people's e-mail, the NSA was gathering most of the international telegraphs sent to or from the United States. *See* SELECT COMM. TO STUDY GOVERNMENTAL OPERATIONS WITH RESPECT TO INTELLIGENCE ACTIVITIES, SUPPLEMENTARY DETAILED STAFF REPORTS ON INTELLIGENCE ACTIVITIES AND THE RIGHTS OF AMERICANS, BOOK III, S. REP. NO. 94-755, at 740, 765 (1976), *available at* http://www.intelligence.senate.gov/pdfs94th/94755_III.pdf.

18 *See infra* Part II.

19 *See* Hillel Italie, *Ferguson Arrests Include at Least 10 Journalists*, THE ASSOCIATED PRESS (Aug. 19, 2014, 6:58 PM), http://bigstory.ap.org/article/ferguson-arrest-include-least-10-journalists ("48 American media organizations, including The Associated Press, sent a letter to law enforcement officials in Ferguson, criticizing the treatment of reporters. 'Officers on the ground must understand that gathering news and recording police activities are not crimes,' the letter read.").

Examples abound. Professor Bambauer cites the literature on increased DNA collection, but the basic argument generalizes: if *x* exonerates, why not gather more *x* and share the fruits of that collection with citizens to prove innocence? The difficulty lies not in the theoretical necessity or technological practicability of this happening. Professor Garrett reminds us that "DNA evidence is often also electronic evidence, since when a search is done through the CODIS system of databases, it is a search against a string of numbers entered based on DNA test results."[20] The hitch is in the resistance to sharing *x* (or *y*).[21] "Even for DNA evidence, defense access to the CODIS set of databases can be highly contested," as Garrett notes, and generally speaking, "due process regulation of discovery in criminal cases is not yet well adapted to Big Data or electronic discovery."[22]

We draw strength in our analysis from the nexus between Professor Bambauer's insistence on accuracy and Professor Garrett's argument that accuracy must be tested in order to fit within the general structure of due process. For this to happen, mere access to big data outputs will not suffice. Rather, a defendant's "meaningful inquiry"[23] into the evidence and methods used to convict him requires more than just the raw output data. On this point, Garrett posits the single most important question: "How can a defendant impeach a database?"[24] Currently, the answer is that a defendant cannot impeach or otherwise contest a government database—a posture that must change to prevent the databases' use as a law enforcement tool unchallenged and unchallengeable by adversarial testing. For instance, impeaching an expert witness presenting Big Data findings "may require discovery regarding the reliability of and procedures used to produce the underlying evidence."[25]

---

[20] Garrett, *supra* note 3, at 210. "CODIS is the acronym for the 'Combined DNA Index System' and is the generic term used to describe the FBI's program of support for criminal justice DNA databases as well as the software used to run these databases." *Frequently Asked Questions (FAQs) on the CODIS Program and the National DNA Index System*, FBI, http://www.fbi.gov/about-us/lab/biometric-analysis/codis/codis-and-ndis-fact-sheet (last visited Sept. 10, 2014).

[21] *See, e.g.*, Sarah M. Ruby, *Checking the Math: Government Secrecy and DNA Databases*, 6 I/S: J.L. & POL'Y FOR INFO. SOC'Y 257, 257 (2010) ("[O]ffender DNA databases are secret, government agencies use them to find suspects, and only their own technicians have access to them."); *see also id.* at 306 ("DNA database software does not look only for perfect matches. . . . Whether profiles determined to be non-matches are discoverable is adjudicated on a case-by-case basis."); Robert P. Mosteller, *Protecting the Innocent: Part of the Solution for Inadequate Funding for Defenders, Not a Panacea for Targeting Justice*, 75 MO. L. REV. 931, 940 (2010) ("Except for the defendant and the investigatory leads he or she can provide, the prosecution has greater access to witnesses and to information about the crime than does the defense.").

[22] Garrett, *supra* note 3, at 210.

[23] *See id.* at 213.

[24] *Id.* at 212.

[25] *Id.*

That the "presumptions of regularity in . . . databases may not be warranted" is an understatement.[26] Big Data databanks are buggy, both in terms of input errors by those within the control of law enforcement and the inclusion of false information from outside of law enforcement. As examples, Professor Garrett points to the Supreme Court's *Herring* decision concerning inaccurate entry in a warrant database,[27] and a 2014 audit of DNA databases that revealed dozens of errors due to handwriting gaffes and other mistakes by laboratory technicians.[28] Moreover, post-Snowden revelations demonstrate the government actively misrepresenting the information in large datasets.[29] On various occasions, judges on the nation's spy magistracy—the Foreign Intelligence Surveillance Court (FISC)—have castigated the government for its misstatements to the court.[30] Add to this the fact that outputs can point in multiple directions, many of which may not implicate the defendant.

At a minimum, impeaching a database requires access to and

---

26 *Id.*

27 *See* Herring v. United States, 555 U.S. 135, 137 (2009).

28 *See* Garrett, *supra* note 3, at 212.

29 *See, e.g.*, *[redacted]*, No. *[redacted]*, 2011 WL 10945618, at *5 & n.14, *6, *9 (Foreign Intelligence Surveillance Ct. Oct. 3, 2011) (noting that "for the first time, the government has now advised the Court that the volume and nature of the information it has been collecting is fundamentally different from what the Court had been led to believe," and stating that the court was "troubled that the government's revelations regarding NSA's acquisition of Internet transactions mark the third instance in less than three years in which the government has disclosed a substantial misrepresentation regarding the scope of a major collection program"); *In re* FBI for an Order Requiring the Production of Tangible Things from *[redacted]*, No. BR 9-13, 2009 WL 9150896, at *2 (Foreign Intelligence Surveillance Ct. Sept. 25, 2009) (stating that the court was "deeply troubled" by noncompliance incidents, which occurred shortly after the NSA's completion of "end to end review" of the relevant processes "and its submission of a report intended to assure the Court that NSA had addressed and corrected the issues giving rise to the history of serious and widespread compliance problems"); *In re* Production of Tangible Things from *[redacted]*, No. BR 08-13, 2009 WL 9150913, at *2–9 (Foreign Intelligence Surveillance Ct. Mar. 2, 2009) (describing government misrepresentations and violations of court orders); *In re* Production of Tangible Things from *[redacted]*, No. BR 08-13, 2009 WL 9157881, at *2 (Foreign Intelligence Surveillance Ct. Jan. 28, 2009) ("The Court is exceptionally concerned about what appears to be a flagrant violation of its Order in this matter . . . ."); Memorandum Opinion at 3, *[redacted]*, No. PR/TT *[redacted]* (Foreign Intelligence Surveillance Ct. *[redacted]*) ("NSA exceeded the scope of authorized acquisition continuously during the more than *[redacted]* years of acquisition . . . ."), *available at* http://www.dni.gov/files/documents/1118/CLEANEDPRTT%202.pdf; *see also In re* All Matters Submitted to the Foreign Intelligence Surveillance Court, 218 F. Supp. 2d 611, 620–21 (Foreign Intelligence Surveillance Ct. 2002) (describing government's "misstatements and omissions of material facts" in FISA applications), *rev'd on other grounds*, *In re* Sealed Case, 310 F.3d 717 (Foreign Intelligence Surveillance Ct. Rev. 2002).

30 *See supra* note 29. *But see, e.g.*, Jennifer S. Granick, *FISA Court Rolls Over, Plays Dead*, FORBES (Aug. 28, 2013, 10:50 AM), http://www.forbes.com/sites/jennifergranick/2013/08/28/fisa-court-rolls-over-plays-dead/ (describing how the FISC "failed to respond in any real way to the NSA's defiant, dishonest lawbreaking").

cross-examination of the experts who created and maintained the database. True debugging would further require the source code and the inputs. In the end, there may be no other way to assure accuracy within a system of criminal justice than by providing some form of defense access to the databases, including the possibility of testing those searches the government purports to be probative. In Professor Garrett's words:

> The understanding of exculpatory evidence under *Brady* [*v. Maryland*] may need to be redefined when it is not just the results of a database search that may inculpate or exclude but also the nature of the search terms, the reliability of the database entries, and the manner in which the database is maintained. Each may be crucial information for the defense in order to effectively present a case.[31]

We share Garrett's hope for a future where courts "develop how *Brady* obligations require a meaningful inquiry into the sources and structure of digital evidence, just as a witness must be asked questions about more than just the substance of a formal statement to the police, or just as chain of custody must be documented for trace evidence."[32] Our expectation for the short term is decidedly darker, however. Digital chains of custody have been falsified through the Orwellian practice of "parallel construction,"[33] and, unfortunately, courts continue to reduce the ability of defendants to meaningfully inquire into the sources of digital evidence.[34] But as writers, we are emboldened by Professors Garrett and Bambauer, and suggest an expansion of our earlier argument: digital innocence must mean access not only to the fruits of government data-gathering, but also to its methods, inputs, and source code as well. For meaningful inquiry to occur, the legal and technological interface at the center of data-driven investigation must be thoroughly and adversarially debugged.

## II
## AGAINST TECHNO-TOTALITARIANISM

Digital innocence appeals to the idea of reciprocity not for its own sake but rather for an ultimate purpose of criminal law, truth-seeking, achieved through a more balanced relationship between citizen and state. Increasing the amount of data available to analytics will increase accuracy of those techniques, no doubt, but the increase means little to the accused, and to society at large as well, without some minimal level of transparency. This is not merely a question of what data the government has but also whether in fact the government's methods

---

31 Garrett, *supra* note 3, at 213 (discussing Brady v. Maryland, 373 U.S. 83 (1963)).

32 *Id.*

33 Fairfield & Luna, *supra* note 4, at 1042.

34 *See infra* note 53.

produce accurate or flawed results. Untestable claims of accuracy are rightly considered unscientific[35]—and when asserted by the state, beware: unchallengeable "facts" are among the oldest covers for tyranny, petty or otherwise.

It is here that Karl Popper's work on the philosophy of science—especially his imperative that scientific hypotheses be exposed to refutation through rigorous testing—overlaps with his more popular writing on political theory. In *The Open Society and Its Enemies*,[36] Popper challenged the notion that human history was the product of inexorable laws, arguing instead that human experience was the result of growth in knowledge achieved by an "open society," where state action is transparent, accountable, and responsive, all as the result of a critical approach to public dialogue and the maintenance of liberal democratic processes. By contrast, a closed society adopts the structures of totalitarian governance, which, *inter alia*, centralize decision-making and often quash debate about state action. Today, the opaqueness of government mass surveillance has all the bells and whistles of a more closed digital society.

Let's not mince words: government secrecy and personal nakedness will not end well. The current regime of obscure government surveillance not only sends people to prison, it helps kill people without trial.[37] The relevant systems should be rigorously tested, both scientifically and legally. This, in turn, requires that the systems be open source or otherwise accessible for meaningful testing and use. Without such access, the systems present a clear challenge to an open society that, as a rule, objects to withholding knowledge from the public. An approach that does not strike the balance in favor of personal privacy and government transparency—that is, a system consistent with a closed society—will prove difficult to square with modern conceptions of constitutional democracy.[38] Although discussion of liberal governance is well beyond the scope of this piece, it is always in the back of our minds.

So are concerns about the loss of privacy caused by digital surveillance systems that can damage us individually and as members of

---

35 Here, we take no position on the value of qualitative research that is verifiable but not necessarily falsifiable.

36 *See* 1 & 2 POPPER, *supra* note 1.

37 *See* David Cole, '*We Kill People Based on Metadata*,' NYR BLOG (May 10, 2014, 10:12 AM), http://www.nybooks.com/blogs/nyrblog/2014/may/10/we-kill-people-based-metadata/. Stewart Baker, the former general counsel of the NSA, stated plainly: "[M]etadata absolutely tells you everything about somebody's life. If you have enough metadata, you don't really need content." *Id.* Former NSA director Michael Hayden called Baker's statement "absolutely correct" and added, "We kill people based on metadata." *Id.*

38 For a critique of openness and open society, open source, etc., see Evgeny Morozov, *Open and Closed*, N.Y. TIMES, Mar. 17, 2013, at SR11, *available at* http://www.nytimes.com/2013/03/17/opinion/sunday/morozov-open-and-closed.html.

a community.[39] People who claim they have nothing to hide really mean they have no time to reflect upon the subject and the consequences for themselves and for society at large. Daniel Solove is surely correct on this point: "Even if a person is doing nothing wrong, in a free society, that person shouldn't have to justify every action that government officials might view as suspicious. A key component of freedom is not having to worry about how to explain oneself all the time."[40] Privacy is not only a longstanding right within our legal system,[41] it is also recognized across cultures and may even be a basic human right.[42] The government's modern mass-surveillance programs represent a new kind of threat with the capacity to wreak havoc on most forms of privacy.[43] We do not underestimate the danger but merely

---

[39] *See* Julie E. Cohen, *What Privacy Is For*, 126 HARV. L. REV. 1904, 1927 (2013) ("Privacy rights protect individuals, but to understand privacy simply as an individual right is a mistake. The ability to have, maintain, and manage privacy depends heavily on the attributes of one's social, material, and informational environment."); Neil M. Richards, *The Dangers of Surveillance*, 126 HARV. L. REV. 1934, 1939 (2013) ("Big Data is notable not just because of the amount of personal information that can be processed, but because of the ways data in one area can be linked to other areas and analyzed to produce new inferences and findings.").

[40] Daniel J. Solove, *Five Myths About Privacy*, WASH. POST (June 13, 2013), http://www.washingtonpost.com/opinions/five-myths-about-privacy/2013/06/13/098a5b5c-d370-11e2-b05f-3ea3f0e7bb5a_story.html. *But cf.* Conor Friedersdorf, *This Man Has Nothing to Hide—Not Even His Email Password*, THE ATLANTIC (Aug. 26, 2014, 11:30 AM), http://www.theatlantic.com/politics/archive/2014/08/this-man-has-nothing-to-hide/379041/.

[41] *See, e.g.*, Samuel D. Warren & Louis D. Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193, 193 (1890); Boyd v. United States, 116 U.S. 616, 634–35 (1886).

[42] *See, e.g.*, European Commission, *Comparative Study on the Situation in the 27 Member States as Regards the Law Applicable to Non-Contractual Obligations Arising Out of Violations of Privacy and Rights Relating to Personality*, JLS/2007/C4/028, at 37–40 (Feb. 2009), *available at* http://ec.europa.eu/justice/civil/files/study_privacy_en.pdf (comparing the constitutional treatment of the right to privacy within the European Union); *see also infra* note 55.

[43] People are already feeling the impact of the government's surveillance power, and some have attempted to shield themselves from it. For instance, a recent MIT study of Google search trends indicated that "searches for terms deemed to be sensitive to government or privacy concerns have dropped 'significantly' in the months since Edward Snowden's revelations in July." Alex Pasternack, *In Our Google Searches, Researchers See a Post-Snowden Chilling Effect*, MOTHERBOARD (May 5, 2014, 10:59 AM), http://motherboard.vice.com/read/nsa-chilling-effect (citing Alex Marthews & Catherine Tucker, *Government Surveillance and Internet Search Behavior* (unpublished manuscript) (Aug. 28 2014), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2412564). For their part, Europeans reacted to Snowden's NSA revelations by pushing for stricter data-protection measures. *See* Simon Shuster, *E.U. Pushes for Stricter Data Protection After Snowden's NSA Revelations*, TIME (Oct. 21, 2013), http://world.time.com/2013/10/21/e-u-pushes-for-stricter-data-protection-after-snowden-nsa-revelations/; *see also* Joined Cases C-293/12 & C-594/12, Digital Rights Ireland Ltd v. Minister for Commc'ns, Marine & Natural Res., 2014 E.C.R. (Apr. 8, 2014), *available at* http://curia.europa.eu/juris/document/document.jsf?docid=150642&doclang=EN (judgment of EU's Court of Justice that data retention violated right of privacy). Moreover, recent reports reveal that the NSA is intercepting and recording phone calls in countries not previously thought to be part of its surveillance program, such as the Bahamas, Mexico, the Philippines, and Kenya. *See* Ryan Devereaux et al., *Data Pirates of the Caribbean: The NSA is Recording Every Cell Phone Call in the Bahamas*, THE INTERCEPT (May 19, 2014, 12:37 PM),

note that a full exposition is merited and may be drawn together elsewhere.

Here, our concern is the damage done to a justice system by the misguided belief that government mechanisms used for criminal prosecution ought to go untested by the adversarial process. Among other things, subject-centered searches by hidden machine-learning algorithms lurk in the background. Differential access renders accuracy irrelevant and makes subject-centered searches a true digital enemy of an open society. Government agents who find that a narrow search does not get the job done can simply loosen the search criteria, for instance; as a consequence, the defense will confront information against the accused without the ability to show that the same search criteria could finger untold numbers of people. These algorithms and search criteria should be available, not just to improve the accuracy of government search results, but to contest the accuracy of the process and the probative value of the results.

Most experts consider a closed development process to be anathema to both accuracy and security. The current use of Big Data in law enforcement lacks debugging or "red-teaming," as it is called in software parlance.[44] "Red teams" are groups tasked with testing software for fatal flaws. For security software, a red team might be trying to break in, for instance, while a red team for Big Data analytics might seek to actively poison the information flow as it enters the system in order to find ways in which it can be caused to return inaccurate results. But red-teaming has not proven to be enough, as most companies test their own software and yet miss the bugs that eventually bring down the software. The best companies—those whose security efforts drive the internet—actively seek and support the people who find flaws in the software.[45] Many of these companies have instituted bounty programs and "hackathons" in which the very purpose is to break the software.

The unmistakable parallel in criminal justice is the commitment of

---

https://firstlook.org/theintercept/article/2014/05/19/data-pirates-caribbean-nsa-recording-every-cell-phone-call-bahamas/. The United States is not the sole perpetrator of surveillance, however, as several European government agencies have assisted the NSA's data-collection efforts. *See* Adam Entous & Siobhan Gorman, *Europeans Shared Spy Data with U.S.*, WALL ST. J. (Oct. 29, 2013), http://online.wsj.com/news/articles/SB10001424052702304200804579165653105860502.

44 *See* Lev Grossman, *World War Zero: How Hackers Fight to Steal Your Secrets*, TIME (July 10, 2014), http://time.com/2972317/world-war-zero-how-hackers-fight-to-steal-your-secrets/ (describing companies that are paid to locate vulnerabilities—technical bugs—for businesses); Zack Whittaker, *Google Building Privacy 'Red Team' After FTC Settlement*, ZDNET (Aug. 23, 2012, 7:43 AM), http://www.zdnet.com/google-building-privacy-red-team-after-ftc-settlement-7000002859/ ("A red team typically works independently and internally at a company to shine a bright light on policies, the workforce, or products and services. Consider it like quality control to the next level in order to make the organization work and flow better. Matters of privacy and security are no different.").

45 *See* Grossman, *supra* note 44.

the common law and the Constitution to adversarial testing of evidence.[46] The defense bar serves as the system's red-teamers, with counsel paid (often not enough) to spot and find the bugs in the legal order. They are the ones who are incentivized to find and stop back doors, not install them. But the Big Data systems interfacing with the legal system have been designed through a process that produces a sort of "heartbleed"[47] justice: a bug-riddled, actively dangerous, and, above all, inaccurate technology that feeds the worst in flawed social systems. The accuracy of Big Data systems cannot be improved if they are built to law enforcement's specifications and are actively kept hidden from bug seekers, both technical and legal. For software to get better—and for the law that surrounds its use to be evenhanded—the tools, terms, code, legal process, and results must be adversarially debugged. The failure to do so neglects one of the most important practices of code: debug it, keep debugging it, and debug it with people who benefit from finding the bugs. Today, however, coders have little incentive to find bugs unrelated to law enforcement specifications and goals.

The analysis must also recognize the current legal reality and, in particular, the nation's overcriminalized justice systems. As discussed at length elsewhere, the scope of criminal justice has continually expanded through the creation of new crimes, broader culpability principles, and harsher punishments, to the point that it can be said that everyone is, or could be, a criminal.[48] This is not a new development. Almost three-quarters of a century ago, then–Attorney General Robert Jackson warned of the danger that a law enforcer

> will pick people that he thinks he should get, rather than pick cases that need to be prosecuted. With the law books filled with a great assortment of crimes, a prosecutor stands a fair chance of finding at least a technical violation of some act on the part of almost anyone. In such a case, it is not a question of discovering the commission of a crime and then looking for the man who has committed it, it is a question of picking the man and then searching the law books, or putting

---

46 A perfectly decent, historic method of fact-finding—the inquisitorial approach of continental law—illustrates this parallel as well. *See, e.g.*, Erik Luna, *Prosecutor Kings*, 1 STAN. J. CRIM. L. & POL'Y (forthcoming 2014).

47 *See* Lorenzo Franceschi-Bicchierai, *Report: NSA Knew About Heartbleed Bug for 2 Years and Said Nothing*, MASHABLE (Apr. 11, 2014), http://mashable.com/2014/04/11/nsa-heartbleed-report/.

48 *See, e.g.*, Glenn Harlan Reynolds, *Ham Sandwich Nation: Due Process when Everything is a Crime*, 113 COLUM. L. REV. SIDEBAR 102, 102 (2013); Michelle Alexander, *I'm a Criminal and So Are You*, CNN (May 19, 2010), http://www.cnn.com/2010/OPINION/05/18/alexander.who.am.i/; Alex Kozinski & Misha Tseytlin, *You're (Probably) a Federal Criminal*, *in* IN THE NAME OF JUSTICE 43, 50 (Timothy Lynch ed., 2009); Erik Luna, *The Overcriminalization Phenomenon*, 54 AM. U. L. REV. 703, 716 (2005); GO DIRECTLY TO JAIL: THE CRIMINALIZATION OF ALMOST EVERYTHING (Gene Healy ed., 2004); William J. Stuntz, *The Pathological Politics of Criminal Law*, 100 MICH. L. REV. 505, 511 (2001).

> investigators to work, to pin some offense on him. . . . It is here that law enforcement becomes personal, and the real crime becomes that of being unpopular with the predominant or governing group, being attached to the wrong political views, or being personally obnoxious to . . . [the agent] himself.[49]

Jackson's words are truer now than they were then. Imagine that law enforcement wishes to punish us for writing *Digital Innocence* (may we have that much influence!), so it uses a data-driven system to find some other reason to arrest us. Agents enter a tailored list of search criteria that will surely return one or both of us as a result. Somewhere in the massive, connected sets of arrangements associated by machine-learning algorithm will be a network constellation of variables, which a machine-learning algorithm has associated with something criminal. A list of search terms related to our article would do nicely. The agents then engage in spy-like methods such as parallel construction, actively falsifying the source of information so that the defense (and the judge and jury) cannot track the background and basis for the accusations and related evidence.[50] Maybe the agents leave out the fact that the search was tailored to return the defendant as one of its results, or maybe the legal fact-finder is not informed that the same search terms could have landed, say, half the jurors, the judge, and the bailiff in the dock as well. All of this and more are the products of government obfuscation and the denial of defense access to government databases.

Subject-based searching is just one problematic method that must be scientifically debunked and legally debugged. If such searches become commonplace in an overcriminalized environment—where everyone is a criminal and evidence of such is readily obtainable—all law enforcement has to do is decide whom to arrest. In a sense, this is a high-tech version of the pretextual law enforcement associated with "racial profiling." In one common manifestation, all-encompassing vehicle codes provide law enforcement vast discretion to conduct traffic stops, since few drivers travel even a modest distance without breaking some traffic law.[51] This supplies a pretense for detention, perhaps drawn out by a search, while also hiding any selectivity based on race, ethnicity, class, and so on. The practice of racial profiling was so widespread in minority communities that it was labeled "D.W.B.":

---

49 Robert H. Jackson, *The Federal Prosecutor*, 24 J. AM. JUDICATURE SOC'Y 18, 19 (1940).

50 *See* Fairfield & Luna, *supra* note 4, at 1042 ("[L]aw enforcement is trained to 'recreate' information through a process euphemistically termed 'parallel construction': laundering the information in question by concocting independent sources through field interviews, confidential informants, physical searches and seizures, etc.") (footnotes omitted).

51 *See* David A. Harris, *"Driving While Black" and All Other Traffic Offenses: The Supreme Court and Pretextual Traffic Stops*, 87 J. CRIM. L. & CRIMINOLOGY 544, 557–58 (1997).

"Driving While Black (or Brown)."[52] Given the ubiquity of domain awareness, social radar systems, and similar Big Data–based approaches—all supported with the pretext of algorithmically derived information—selective enforcement by digital evidence might as well be called "Texting While Human."

The threat is not additional cases of false suspicion and wrongful convictions (though one could imagine that happening as well). Instead, it is that accusations will be based on highly sophisticated machine-learning algorithms applied to data gathered through comprehensive dragnet surveillance. The accusations and convictions may be accurate, just as many if not most of the racially profiled were guilty of traffic violations, and others were in fact guilty of the offenses that truly motivated the detaining agents (e.g., drug offenses). But the outcomes will be the product of invidious discrimination or ad hoc and even ad hominem selectivity among a population, more consistent with the rule of men than the rule of law.

## CONCLUSION

Rules made for humans, applied to searches performed by machines, threaten the relationship between citizen and state. One is left to wonder, for instance, whether an open society can even survive an overcriminalized justice system affected by subject-centered Big Data data-mining techniques. On these points, we bet there is common cause with scholars such as Professor Bambauer, whose polestar is the truth, and Professor Garrett, who insists that truth be tested through meaningful inquiry. The area of accord might extend further to challenge laws, policies, and interpretations that are unmindful of the transparency threat identified by Bambauer, or that undercut Garrett's hopes for the development of a jurisprudence of digital due process that will permit defendants to meaningfully challenge adverse evidence. What law at times takes as gospel, science throws out as untestable or nonfalsifiable, and an open society opposes as totalitarian.[53] The current trend must be opposed: with few exceptions, courts have held that state claims of secrecy trump equal access to data;[54] and, of course, the FISC stands for little else, with massive amounts of government data-gathering conducted on the FISC's watch and kept secret in the name

---

52 *See, e.g.*, Erik Luna, *Foreword: The New Face of Racial Profiling*, 2004 UTAH L. REV. 905, 907 (discussing phenomenon).

53 Certainly that is the current and dangerous trend. After our original article went to press, for instance, two courts asked to allow access to law enforcement–sourced data refused to do so for fear of revealing law enforcement sources and methods. *See* United States v. Daoud, 755 F.3d 479 (7th Cir. 2014), *supplemented by* 761 F.3d 678 (7th Cir. 2014); United States v. Mohamud, 3:10-CR-00475-KI-1, 2014 WL 2866749 (D. Or. June 24, 2014).

54 *See, e.g.*, Jewel v. Nat'l Sec. Agency, 965 F. Supp. 2d 1090 (N.D. Cal. 2013) (rare court rejection of state-secrets privilege).

of protecting the intelligence community's sources and methods.[55]

So we concede—and to some extent, actively support—Professor Bambauer's arguments on accuracy and cost. As accuracy rises, certain manifestations of the system's primary cost (i.e., erroneous case decision-making) begin to fall, including wrongful convictions and resources spent pursuing dead ends. For related reasons, we must endorse Professor Garrett's expanded application of those basic legal doctrines, such as *Brady* and its progeny, which support accuracy through transparency and reciprocity.

But our larger fear is neither collection nor inaccuracy, though both concern us a great deal. Instead, the fear is the increasingly uneven distribution of costs and benefits between citizen and state. More concretely, our anxiety is that a growing technocracy will not let people see what search terms it is typing, what machine-learning algorithms it is using, and above all, it will not let anyone test for scientific validity, false positives and negatives, and manipulations that generate tautological results. As we said, a digitally naked public before a secret government does not end well.

---

55 *See generally* ERIK LUNA & WAYNE MCCORMACK, UNDERSTANDING THE LAW OF TERRORISM ch. 4 (forthcoming 2014) (discussing FISC).