2016

Big Data and the Future for Privacy

Neil M. Richards
Washington University in St. Louis School of Law, nrichards@wustl.edu

Jonathan H. King

Big Data and the Future For Privacy

Neil M. Richards
Jonathan H. King

Abstract

In our inevitable big data future, critics and skeptics argue that privacy will have no place. We disagree. When properly understood, privacy rules will be an essential and valuable part of our digital future, especially if we wish to retain the human values on which our political, social, and economic institutions have been built. In this paper, we make three simple points. First, we need to think differently about "privacy." Privacy is not merely about keeping secrets, but about the rules we use to regulate information, which is and always has been in intermediate states between totally secret and known to all. Privacy rules are information rules, and in an information society, information rules are inevitable. Second, human values rather than privacy for privacy's sake should animate our information rules. These must include protections for identity, equality, security, and trust. Third, we argue that privacy in our big data future can and must be secured in a variety of ways. Formal legal regulation will be necessary, but so too will "soft" regulation by entities like the Federal Trade Commission, and by the development of richer notions of big data ethics.

Word Count: 7,718 words, incl. footnotes

Big Data and the Future For Privacy

Neil M. Richards*
Jonathan H. King**

Big data is our future, but what place will privacy have in that future? Many technologists and futurists predict a digital future in which privacy has no place (Barnett, 2010; Ferenstein, 2013). Others argue that the benefits of open data and data science mean that certain kinds of privacy rules (like limitations on collection or deletion requirements) make privacy either an obstacle to progress or something highly impractical to enforce in our ubiquitous digital future (Mundie, 2014; Toobin, 2014). Data scientists consider privacy to be an obstacle to the kind of innovative work they want to do, while the leading manuals for data warehouse engineers largely ignore considerations of privacy altogether (Birnhack et al., 2015). At the level of theory, then, privacy is an anachronism hostile to progress, while at the level of practice, it is just impractical and gets in the way of doing things.

Such accounts are common, but their dismissal of privacy as a foolish anachronism is belied by both common sense and a small but growing scholarly and public literature about the importance of privacy for the kind of sustainable, humanist society we should want to build. In this paper, we attempt to lay out some of the legal and ethical principles we should build into that future society.

* Professor of Law, Washington University.

** Visiting Scholar, Washington University; Head of Cloud Strategy for Ericsson. The views and opinions expressed by the author are not necessarily the views of his employer. For helpful comments and perspective we would like to thank Lucas Carlson, Andrew Clyne, Andrew Higginbotham, Woody Buckner, Jason Hoffman, Brian Hughes, Matthew Johnson, Gavin McMurdo, Alex Williams and Jared Wray. We would also like to thank Ujjayini Bose, Matthew Cin, and Carolina Foglia for their very helpful research assistance.

Our claim is that the shiny future asserted as inevitable by the pro-innovation anti-privacy rhetoric turns out to be both shallow and unappealing on closer analysis. We believe there is an alternative future where we can have the benefits of data science while at the same time preserving meaningful legal and ethical protections for data subjects. We are not alone in this view. Over the past decade, scholars and commentators writing in fields as diverse as law, economics, sociology, and even computer science have argued that our digital future must include meaningful protections for values that we care deeply about such as privacy, equality and identity. Moreover, there is good evidence not only that privacy can be a market differentiator for businesses in the digital economy, but also that security and data integrity are essential for any entity that handles personal data (Tysiac, 2015). Similarly, we suggest that the digital future we are building is not inevitable – whether and how much it protects privacy and security are dependent upon many individual decisions by regulators, businesses, engineers, and users. Our digital future is not ordained to take a single, shiny, privacy-denying form. Instead, it will be a human creation. As such it must include the human values we have come to cherish – values which are advanced by privacy rules. In this chapter we lay out a vision for the role that privacy can and should play in protecting values in this digital future.

Our argument develops in three simple steps. In Part I, we show how the traditional way of thinking about privacy as secret, unobserved information is incomplete. We argue instead that the best way to think about the control of digital information is not by reference to secrecy, but by thinking of privacy rules as information rules. This has close analogues to the long European scholarly and regulatory tradition of Data Protection. Privacy regulation, in our view, however, is not an end in itself but rather a means to other, more important ends. Thinking about privacy in this way reveals that privacy rules of some sort are inevitable, that information is usually in some intermediate state between wholly private and wholly public, and that existing legal privacy tools like data

protection and confidentiality have an essential part to play in our digital future. In Part II, we illustrate the values that privacy rules should protect. We identify four of these values as particularly salient at the present time: identity, equality, security and trust. And in Part III, mindful of the complexity of our digital society, we suggest how meaningful privacy protection can be achieved in a big data future, through a combination of traditional regulation, "soft" regulation, and the development of big data ethics.

We have long had technology futurists, and in many respects they have done their job. We propose now that privacy futurists should join the debate in earnest. We hope that our cross-disciplinary examination of a digital future for privacy will help stimulate this discussion, with the goal of helping to build a digital future in which humans as well as data will want to live and thrive.

I. HOW TO THINK ABOUT PRIVACY

For a concept that has taken on such importance in our modern, digital society, "privacy" is notoriously slippery and hard to define. Even the leading academic treatments of privacy are reluctant to define it, given its many different usages. Legal scholar Daniel Solove, for example, focuses on "privacy problems" rather than on defining privacy, and takes an observational approach, cataloguing "four general types of privacy problems with sixteen different subgroups" (Solove, 2008). Other legal scholars focus on defining privacy harms in ways that the law can understand and remedy (Calo, 2011); such efforts hearken back to privacy law's origins in tort law, a body of law that by definition remedies civil wrongs to identifiable plaintiffs. Other disciplines take different approaches, but have similarly failed to announce a definition of privacy that can be good for all seasons, or all uses. Information studies scholar Helen Nissenbaum, for example, focuses less on defining privacy than on identifying when a privacy violation is subjectively experienced, by urging us to think about privacy as the

management of information flows in their social and technological contexts (Nissenbaum, 2010).

For better or for worse, "privacy" is the word that Western legal systems have settled on using to deal with problems related to the flows of personal data. This is particularly true in the United States, where the European concept of "data protection" is not a meaningful part of popular discussions about personal data (Richards, 2006). Many American understandings of privacy conceive of it as a secret, under which things that are not known are "private," but things that are shared lose that protection (Solove, 2004). This idea is illustrated perhaps most infamously by Fourth Amendment law's "third party doctrine," under which criminal defendants who share information with third parties can lose a "reasonable expectation of privacy" in that information (Smith v. Maryland, 1979). And in the big data context, this question has taken on added importance with many leading voices (including ostensible privacy advocates) arguing that because the collection of vast amounts of human data is inevitable, the law should abandon regulating the collection of information and instead focus only on its use (Cate et al., 2013; Mundie, 2014; World Economic Forum, 2013).

In this Part, we want to suggest a different way of thinking about privacy, or at least a different way of thinking about the protection of personal information across distributed digital networks and databases. We will argue that most information exists and has always existed in intermediate states between "public" and "private," and that such information should not lose and often has not lost legal protection in such intermediate states. We should think less about whether information is "private" in a metaphysical sense, and more about what sorts of rules should govern our intermediate data, because privacy rules, at bottom are just information rules. In an information society, sustainable information rules of some sort and by some name are inevitable.

A. Information in Intermediate States

Let's start with the idea that privacy is really about secrecy, and that information that is shared ceases to be private. Several good reasons suggest that this is a bad way to understand privacy. First, the idea is problematic even at the level of ordinary verbal usage. If I know something about myself that no one else knows, we might call this a "secret." But if I tell you my secret, the idea doesn't stop being a secret. Odds are that if I am telling you a secret, then there is some kind of informal or formal trust relationship between us. In the government context, secrets and "top secrets" can be known by many people. But even these secrets are protected by a wide variety of legal, technological and operational tools, including criminal and contract law, encryption and other security tools, and the whole trade of spycraft. The same is true of corporations around the world that utilize confidentiality regimes, nondisclosure agreements, and trade secret protection to protect "secret" information.

Our verbal intuitions point up a second observation, which is that information has always existed in intermediate states between wholly public and wholly private. Many still think of privacy as a binary option of "public" or "private," when our everyday experiences remind us that virtually all information that matters exists in intermediate states between these two extremes (Richards, 2015). As Woodrow Hartzog points out in a forthcoming paper, it makes just as little sense to define "privacy" or "private information" in terms of secrets known only to one person as it does to define "public information" as something that every single person knows (Hartzog, 2015). In reality, virtually all information is and has been in intermediate states between these two extreme poles. Faced with such a reality, we should recognize that our law should operate primarily in the vast middle, on information in its intermediate states.

Third, realizing this important fact, our law has in many cases operated on information in intermediate states for a very long

time. As we argued in a prior paper, *even in the big data context, we must recognize that shared private information can remain "confidential"* (Richards & King, 2014). Much of the confusion about privacy law over the past few decades has come from the simplistic idea that privacy is a binary, on-or-off state, and that once information is shared and consent given, it can no longer be protected. Binary notions of privacy are particularly dangerous and can erode trust in our era of Big Data and metadata, in which private information is necessarily shared to some extent in order to be useful. The law has always protected private information in intermediate states, whether through confidentiality rules like the duties lawyers and doctors owe to clients and patients, evidentiary rules like the ones protecting marital communications, or statutory rules like the federal laws protecting health, financial, communications, and intellectual privacies (Solove & Richards, 2014). Neither shared private data (nor metadata) should forfeit their ability to be protected merely because they are held in intermediate states. Understanding that shared private information can remain confidential better helps us see more clearly how to align our expectations of privacy with the rapidly growing secondary uses of big data (Solove & Richards, 2014).

B. Privacy As Information Rules

When we realize that we have been using "privacy" to talk about information in intermediate states, and that our law has long regulated information of that sort, we realize that privacy rules are really information rules. "Privacy," in this broader sense, becomes much more than just keeping secrets, and enters the realm of information governance. We live in an information society, and privacy rules are the rules that govern the information in and out of networks and data sets in this society.

Understanding privacy rules as information rules radically changes the questions we might ask about regulation of personal information in the big data context. If we think about privacy

merely as secrets, the collection of data and its incorporation into a data set would seem to moot any privacy concerns. Information that is collected and shared even to a moderate degree would seem to be "public" or at least "non-private," and therefore beyond any regulation. Such a conclusion would likely meet with the approval of many who wield the powerful tools of big data analytics, but it would leave us with an essentially lawless,anarchic and unsustainable information society. An information society with no rules has no protection against hackers, malicious code, data breaches, revenge porn, child pornography, cybercrime, or any of our other information age maladies.

But by contrast when we ask the privacy question more constructively as what rules should govern the collection of personal information, the question changes entirely, especially when we recognize that most information has always existed in intermediate states. The question becomes not about the metaphysics of "privateness" or "publicness," but rather about what kinds of data uses and which kinds of information regulation support the kind of society we might want to live in and which ones do not. When we do this, we see that the collection and sharing of information need not be the end of the regulatory inquiry, but rather the beginning of it. Unlike much information contained in databases and read by computers, privacy is not binary. Privacy is instead an ethical approach to the management of information flows (Richards & King, 2014). At bottom, then, privacy (and the decisions we make about it) is ultimately a series of human questions that must be informed by human values.

But if the decision to protect or regulate the collection, use, or transfer of personal information is dependent on human values other than the private-ness or public-ness of a data field, we must inquire exactly what those values should be. We'll now take up that question.

II. PRIVACY'S VALUES

Shifting the privacy inquiry from whether something is private to what rules should govern the aggregation and use of personal information begs the question of what values should govern privacy rules. From this perspective, privacy is not a value in and of itself. Privacy is instead a tool that can be used to restrict access to data or to regulate decisions based upon data. (Analogously, the opposite idea of transparency is also a tool that can be used to shed light upon unknown activity.) But the decision about whether or not to impose an information rule, or what sort of rule to impose, must be made upon the basis of values other than privacy itself. Under our account, because privacy is such a vague and notoriously slippery concept, it is not a helpful concept upon which to base policy decisions.

The values that privacy rules can serve, however, are useful. In this Part we offer four such values – identity, equality, security and trust. One can imagine other values that might qualify, but our purpose here is not to be exhaustive in articulating the many values that privacy rules can protect. Our purpose instead is to be illustrative – to sketch out a vision for the kinds of values that would support the imposition of privacy rules in an age of big data. Such an approach is not unusual in the law, at least when we are talking about fundamental human rights or flows of information. First Amendment law, for example, works in a similar way, in which the protection for freedom of speech is instrumental, serving a variety of theories (most notably democratic self-governance and the search for truth), none of which by themselves adequately explains either the technical legal doctrine or the cultural reasons we treat free expression as special (Richards, 2015a). Like privacy law, free speech law is a regulation of information flows.[i] We contend that we should treat privacy the same way.

A. *Identity*

Privacy rules can protect identity – our ability to determine for ourselves who we are, what we value, where we go, what we do and at what times. Big data promises to creates a world of refined

demographic differentiation – the ability to determine what we like and what we might want, even if we don't yet know it ourselves. The prediction engines of Amazon, Netflix, and Spotify, for example, can recommend to us new books, movies, and music based upon our existing preferences that can be highly valuable (Richards & King, 2013).

But such technologies can be used not just to serve existing preferences, but also to shape them (Richards & King, 2013; Leonard, 2013). The system of "targeted" and "behavioral" advertising which support the "free" internet, for example, depend upon highly granular profiles of our intellectual and political preferences based upon surveillance of our reading, web-surfing, associations, and increasingly physical movements (Zuckerman, 2014). Facebook, while apologizing for the reaction over the mood study it conducted on unknowing users, implemented research guidelines that while adding thoughtful governance effectively enshrine their ability to continue such research to package users for advertisers (Wohlsen, 2014). And each American election cycle leads to more refined political dossiers on the political identity of every American voter enabling campaigns to shape elections with voter micro-targeting, allowing personalized messages to nudge the political preferences of individual voters (Rubenstein, 2014).

There is a small but growing scholarship in law and other disciplines suggesting that widespread surveillance affects our ability to form our identities ourselves. When it comes to surveillance of intellectual activities, there is good evidence that surveillance dulls our reading and thinking to the boring, the bland, and the mainstream (Richards, 2015). Philosopher Timothy Macklem has argued that "[t]he isolating shield of privacy enables people to develop and exchange ideas, or to foster and share activities, that the presence or even awareness of other people might stifle. For better and for worse, then, privacy is sponsor and guardian to the creative and subversive" (Macklem, 2008, p. 36). More generally, legal scholar Julie Cohen has argued compellingly that privacy shelters our ability to play, to engage in self-

determination, and to manage our boundaries between our social selves and the world (Cohen, 2013). Privacy of varying sorts – protection from surveillance or interference – is what enables us to define our identities. In turn, free identities enable free market economies and ultimately govern free democracies.

B. *Equality*

Big data allows knowledge, and knowledge is power. But as we have argued elsewhere, big data paradoxically has power effects of its own (Richards & King, 2013). Data science allows firms to better understand their competitors and customers, and permits governments greater transparency over the activities of both non-citizens and citizens. Surveillance of this sort is usually not just to learn about others, but to learn in order to nudge, influence, prevent or control (Calo, 2014). At the same time, scholars have warned about the problem of "unraveling," in which individuals in a favored demographic or statistical category can willingly disclose their favored status, effectively "outing" the remaining silent members of the class as possessing the disfavored characteristic (Barocas & Nissenbaum, 2014a, 2014b; Peppet, 2011).

The work of sociologists writing within the field of surveillance studies is especially helpful in illuminating both the purposes and the consequences of such surveillance. Over two decades ago, Oscar Gandy described the "panoptic sort," in which databases were being used to profile consumers, sort them into categories, and then treat those categories differently, giving different opportunities to each group (Gandy, 1993, p. 15). Gandy was writing about cutting-edge analytic techniques of the early 1990s, but the intervening decades have produced better tools and vastly greater data sets with which to sort consumers and citizens. Subsequent work by David Lyon and other scholars have built on Gandy's foundation to show the ways in which analytic tools are increasingly used to sort citizens and consumers by governments seeking profiles of criminal risk and companies seeking profiles of commercial opportunity (Lyon, 2003; Lyon, 2003a; Gilliom, 2006;

Haggerty & Ericson, 2006). We can see the arrival of Gandy's panoptic sort in data broker labels such as "Rural and Barely Making It", "Retiring on Empty: Singles" or "Credit Crunched" as revealed by a 2013 Senate Commerce Committee report on the data broker industry (Senate Commerce Committee, 2013).

Today we seem to be on the verge of completing the glorious monster of a "free" Internet paid for by advertising targeted on the basis of an unprecedented level of surveillance of human lives (Zuckerman, 2014). And it is data analytics that make the Internet's increasingly specific profiling, targeting, and retargeting possible. There are many possible critiques of our surveillance-based Internet, but we wish to focus here on its effect on economic opportunity. Big data analytics permit efficient "sorting," but the line between "sorting" and "discrimination" is a blurry and dangerous one (Gandy, 2010). We might perhaps be comfortable with certain kinds of economic discrimination; universities and airlines, for example, have engaged in sophisticated price discrimination for many years. But big data analytics hold the promise (or the threat, depending upon one's perspective) of perfect price discrimination, in which the surplus of consumer transactions could potentially be retained entirely by the sophisticated merchants who wield big data's tools to calculate the reserve price of every consumer (Salmon, 2013).

Perhaps even more worryingly, since data analytics rest on correlations, they can be used intentionally or unintentionally to discriminate not on the basis of suspect (and illegal) classifications like race or gender, but on variables or data patterns like residence or shopping habits that correlate with such demographic traits (Crawford & Schultz, 2014). Such uses of data science threaten an end-run around well-established legal principles that forbid government or private entities from engaging in intentional race, gender, or other forms of invidious discrimination. At the same time, the problem of "unraveling" discussed earlier creates potentially pernicious private incentives and behaviors that exacerbate the problem of inequality.

Our goal in this paper is not to demonize big data or data science. Such tools can of course be used to combat discrimination as well, by looking at hidden patterns of bias and denial of opportunity (Polonetsky & Wolf, 2014). But this is precisely our point. Neither the creation of big databases nor the application of data science to them are neutral acts (Dwork & Mulligan, 2013). The design and assembly of a database, the application of algorithmic techniques to that data, and decisions about people based upon the outputs of those algorithms are human decisions made by human beings with goals, incentives, and purposes that can be sharply at odds with the data subjects under analysis. This was one of the most important findings of the White House Big Data and Privacy Working Group. In explaining these findings, its Chair John Podesta explained that:

> The detailed personal profiles held about many consumers, combined with automated, algorithm-driven decision-making, could lead—intentionally or inadvertently—to discriminatory outcomes, or what some are already calling "digital redlining." The federal government's lead civil rights and consumer protection agencies should expand their technical expertise to be able to identify practices and outcomes facilitated by big data analytics that have a discriminatory impact on protected classes, and develop a plan for investigating and resolving violations of law. (Podesta, 2014)

Privacy rules placed upon big data – including restrictions on collection, algorithmic transparency and accountability, and restrictions on the use of analytics to sort and treat people differently – will be an essential element of the future of privacy law in a data science world. The imposition of these kinds of restrictions can ensure that our new analytic tools are not used to sidestep the possibility of civil and economic equality upon which our hard-fought existing laws are based. But the important point is this: without meaningful new procedures and rules protecting

equality, our commitment to equality risks being undermined by big data analytics.

C. *Security*

We have reached a point in our digital society where privacy cannot exist without security and security cannot exist without privacy. Ever more prevalent and powerful computing, networking, and data storage enable the automated and largely costless collection of data concerning nearly everything we do. To protect our privacy against this onslaught of collection, we should expect security and integrity in the systems we use to keep our information private.

Traditional security measures relied principally upon establishing a perimeter and preventing intrusion. Like walls and moats around medieval towns, firewalls, malware detection, password authentication and other tools focused on keeping unwanted intruders out. The problem, summed up by Gus Hunt, former CTO of the CIA, is that "when you get through the outer layers, it is pretty easy to get the goods. The data is soft, often unprotected, once an intruder sneaks through the outer layers." (Rosenbush & Boulton, 2014). The increasing collection of sensitive digital information behind these walls, coupled with increasing vectors of attack have made companies more alluring and vulnerable. Major brands such as Target, Home Depot and Sony, to name just a few, have suffered brand damaging headlines due to security breaches (Sidel, 2014; Barrett, 2014). Apple also suffered an embarrassing blemish when its iCloud backup service was hacked to reveal nude photos of celebrities and others, including even the locations where the nude photos were taken (Hill, 2014).

Rather than relying solely on imperfect perimeter protection, designing for privacy can provide an extra layer of security and integrity within the permeable outer perimeter by protecting the data itself. Hunt and others describe how enabling end-to-end encryption of data at movement and at rest can provide an extra layer of defense and reduce vectors of attack (O'Connor, 2014).

Despite some law enforcement criticism, Apple, Microsoft and others have started to enable two-factor authentication and strong encryption to enhance individual control over the privacy of their data and in turn to promote trust in those providers (Schneier, 2014). Innovative technology startups like Guardtime offer solutions which organizations can use to identify real time changes to data in such a way that the privacy of each change is maintained while the integrity of the change is verified (Ewing, 2014). Such technologies promise to meaningfully add the old fair information practice principle of data quality to our commitment to security. In a world in which digital information can be changed easily and surreptitiously, rules and technologies protecting data quality, privacy, and security simultaneously will be essential.

More generally, by empowering end users with more privacy controls and employing tools which strengthen data integrity, the systems we access are strengthened as we use them. In this manner, privacy and data integrity are becoming a kind of fitness indicator for systems security.

D. *Trust*

Privacy also promotes trust – trust in systems, trust in networks, and trust in the relationships between individual people and the entities that hold their data. Many commentators have argued that privacy rules are somehow bad for business, as if the protection of personal data is a kind of tax on profits or innovation (Birnhack et al., 2015). Arguments of this sort are a staple of critics or skeptics of privacy rules, yet as one of us has argued elsewhere, this suggestion is a myth (Richards, 2015b). More fundamentally, even privacy advocates spend inordinate time on the supposed conflict between privacy and profitability, seeking to demonstrate the "privacy harms" that the failure to protect personal information can produce (Reidenberg et al., 2014). This may be correct, but both the academic and popular literatures about privacy have missed the important insight that privacy rules can promote trust

between users and platforms. And this privacy-backed trust can create value, rather than taxing it.

When users trust that their information will not be misused or abused by their confidantes, they share more information, more freely, and more accurately with those confidantes. The classic example of this is the relationship between doctors and patients. The patient wants to be treated, but may be reluctant to share embarrassing medical details. Nevertheless, because the doctor promises and society enforces confidentiality, the patient shares more completely, and receives better care as a result, secure in the knowledge that sharing with the doctor as a trusted confidant will be for his or her benefit. But note that in this trusted relationship, not only are both parties better off, but also better information is shared. The legally-enforceable promise of privacy (or to be technical, confidentiality) promotes not only the welfare of both parties, but also promotes further trust between the parties, making both of them more likely to deal and share with each other in the future. This is the *information-sharing function of confidentiality* (Richards, 2015).

Privacy rules therefore promote trust, a form of value-creation that negative conceptions of privacy like the privacy versus profits frame of the privacy critics or the privacy harm fixation of the privacy advocates fails to properly comprehend. In a digital environment in which identity can be fluid and everything else is seemingly negotiable and up for grabs, privacy rules create trust, which in turn allows for long-term stable relationships to flourish (Richards & Hartzog, 2015). Such forms of economic and social sustainability will be essential in the digital economy, in order for individuals and corporations to share more information and take advantage of their new digital opportunities over the long term, but only as long as they can trust that their data will not be abused by the other parties in information relationships (Pentland, 2014, p.177).

III. HOW TO PROTECT PRIVACY

Let's sum up our argument so far. We have argued that particularly in an age of big data, privacy should not be understood as a matter of keeping secrets, but rather as a system of rules governing the ethical collection use, and disclosure of personal information. We have also argued that from this perspective, privacy is not itself a value, but it is rather, like its converse transparency, a regulatory tool that should be employed to advance other values. We further suggested several values that the imposition of privacy protections on data could serve, including the protection of identity, equality, security and trust.

But why has such regulation not been successful, and how can and should it be accomplished? For most of the past two decades, the legal regime governing personal data has required little more than notice of data collection and the choice to opt out. In practice, this control-focused approach to personal data management has been a colossal failure in providing people with any meaningful ability to control how data about them is collected and how it is used. People simply lack the time, energy, meaningful choice, technical skills and cognitive bandwidth to manage their data the way they manage their finances, even when they are highly motivated to do so (Richards, 2015b). And as the big data revolution continues apace, with more and more data about people being held in more and more places to make more and more decisions about them, what Daniel Solove calls the "failure of privacy self-management" (Solove, 2013) will only become greater. Something more needs to be done.

In this Part, we suggest several ways in which privacy as information rules can (and is likely) to be effected in a digital networked society in order to supplement the necessary but deminished role that privacy self-management will retain.

A. Regulation

Any meaningful solution to the threats to our cherished values posed by big data will require regulation. The United States

is the only major global democracy without a blanket data protection law or formal data protection agency, and this will have to change. More generally, new laws regulating participation in data-based decisions will have to take two tracks. The first will be procedural, reinforcing transparency of processing, the basis for algorithmic decisions (so-called "algorithmic accountability"), and providing meaningful notice of data practices and actual choice to opt-out of unwanted collection, use, or disclosure. But the second track must be substantive. Certain kinds of data collection, certain kinds of processing, and certain kinds of decisions based upon algorithmic outputs must be taken off the table. In particular, processing that threatens identity, equality, security, data integrity, and trust must be regulated and when necessary forbidden. For example, data use that undermines civil rights laws or which promises a kind of "digital redlining" should be outlawed the way we have outlawed wiretapping or the use of consumer reports for impermissible purposes.[ii] We might also mandate in certain cases what Ryan Calo calls "consumer subject review boards," independent ethical boards assessing the implications of big data for any entity that engages in sensitive consumer analytics at scale (Calo, 2013). Alternatively, the creation of a regulatory commission to oversee the fairness and honesty of big data practices (or the vesting of such authority in an entity like the Federal Trade Commission) might also be a sensible option. Given the problem of "privacy unraveling," we should consider either prohibitions on certain kinds of disclosures that have an unraveling effect, or prohibitions on decisions based upon those criteria. As with other protections of meaningful human equality, procedural rules alone will be insufficient, and substantive prohibitions need to be part of the regulatory equation.

B. "Soft Regulation"

We are pragmatic enough to realize that formal regulation along the lines of statutes or agency rulemaking will not be able to

solve all of the problems we identify and represents at best a limited solution. This is the case not only because of political gridlock in the United States, but more fundamentally because it is possible to over-regulate, and that in a time of rapid technological change, there must at times be a lag between innovation and regulation.

Yet even though formal regulation is an imperfect tool, other kinds of regulation can be used to partially fill the gap. In the United States, in the absence of a data protection agency or data protection statute, other regulatory tools have addressed information policy issues identified as implicating privacy of one sort or another. These forms of "soft regulation" have been unexpected, but have exerted an undeniable regulatory effect on commercial actors in the digital sector (Cao, 2009). Two examples of such soft regulation will suffice.

First, in the absence of a formal data protection agency in the United States, the Federal Trade Commission has partially filled the vacuum under its unfair and deceptive trade practice authority. By investigating alleged breaches of privacy policies and entering into settlement agreements with companies that broke those promises, the FTC has emerged as an important regulator of the data trade (Solove & Hartzog, 2014). These developments suggest that the FTC will not only continue to be an important source of privacy regulation, but will likely become even more important in the future.

Second, even where domestic law is silent, the global nature of the information economy means that actors within the United States will increasingly fall within the regulatory authority of foreign data protection authorities. Early examples of this phenomenon include the attempts by French courts to hold Yahoo! Liable for the sale of Nazi memorabilia in 2000, which led to subsequent litigation in United States courts (Waters, 2005). More recently, the Court of Justice for the European Union held that Google was required to delete search results for a Spanish man who had been adjudged bankrupt in the past (Google Spain Case, 2014;

Kulk & Borgesius, 2014). And in October of 2015, the European Court of Justice invalidated the Safe Harbor Framework between the United States and the European Union finding it provided inadequate protections to EU citizens sharing data with US firms in light of Edward Snowden's disclosure of NSA spying (Schrems Case 2015). Although judgments of these sorts rarely have extraterritorial application as a matter of law, they tend to have extraterritorial application as a matter of effect. For example, Yahoo! ultimately decided not to sell Nazi memorabilia worldwide, the creation of a link removal tool for the European market by Google could well lead Google and other players like Microsoft to introduce similar data removal tools, once created, in the American market as well (White & Benoit, 2014). The recent Safe Harbor invalidation has lead thousands of companies to immediately undertake efforts to continue their ability to transfer personal data of EU citizens to the U.S. while a long term resolution unfolds (Drozdiak & Schechner, 2015). The point here is not whether these controversial foreign judgments are correct, but to observe that precedent changing decisions are occurring and to consider their global effect regardless of their merits as they do.

Third, and perhaps most encouraging, competition around privacy itself can also serve as a form of soft regulation. Microsoft has long been making investments in privacy research and development and their general counsel, Brad Smith, has been an outspoken critic of government encroachment upon privacy (Wingfield, 2014). Apple's rapid response to the iCloud celebrity photo incident shows how seriously they are taking privacy as well. Both of these companies generate their revenue primarily by selling software and hardware-related products and services as opposed to the data-intensive advertising business models of Google and Facebook. As privacy revelations continue to gain headlines, the initial reactions and resulting competitive dynamic offer promise to organically advance privacy. Even Google has started to provide users more insight and control over the data that Google collects about them (Nash, 2012).

Our point here is that regulation in the digital, networked, global society can occur from a variety of perhaps unexpected sources, and that the failure of the U.S. Congress to pass an American data protection law does not stop other regulatory entrepreneurs such as state or foreign governments or unexpected federal regulators like the FTC from stepping into the regulatory vacuum to make or influence regulatory policy. Perhaps the most unexpected and encouraging of all is the emerging industry trend of innovating, advocating and competing for privacy, not just technology.

C. Big Data Ethics

Of course, formal law alone (even from unexpected sources) will not be enough to guarantee effective privacy protection in our digital age. Just as our culture of free speech depends on social norms like a free and critical press, social tolerance for dissent, and technologies of expression from printing presses to encryption to Twitter, so too will any meaningful system of privacy rules rest on technologies and social norms. In prior articles, we argued that because big data creates increased institutional powers of awareness and influence, it is essential that we develop some form of "big data ethics" (Richards & King, 2013; Richards & King, 2014). The precise form of such ethics can vary, but in order to ensure that the critical human values that undergird privacy rules continue to apply in our big data future, we argued that we must have a social conversation beyond merely the compulsions of legal rules in order to ensure that the tools of the new data science are employed in ways that are not merely effective and efficient, but ethical as well (Richards & King, 2014).

Ethical rules need not lag the way that legal doctrine sometimes must. Big data ethics can be embraced into the professional ethos of cross functional leaders on the ground in a way that each can apply their own insights and expertise. By first having a shared vision about what big data ethics mean for their organization, leaders can exert a regulatory effect that emerges

naturally around big data deployments, building trust and responsibly harnessing the full power of big data in the process (Schneider et al, 2015; Hirsch & King, 2015). While the precise application of big data ethics should be left for further resolution, their need is essential. Big data ethics, we contend, should be for everyone: data subjects, data scientists, and data regulators. Big data ethics will be, in large part, the future for privacy.

CONCLUSION

While it might seem that privacy is, at times, a quaint value that has no place in our digitally saturated society, such a view is misguided. Old forms of privacy (say from the 1920s) might seem quaint today. But this is precisely because privacy is the result of our human values filtered through a social conversation about how those values should apply to our society at a particular social and technological moment. The enormous changes that our digital revolution has wrought and that data science promises to make in the future do not excuse us from that conversation. On the contrary, the advent of powerful data science tools that threaten our identity, equality, security and trust in social and digital systems make that conversation essential. When technology leaps forward and destabilizes the social and technological assumptions on which the protection of our values rests, our need to shore up those values is never greater. And when we understand privacy as the rules that govern information flows based upon our values and norms, we can see that rather than being left out of our big data future, an ethical system of privacy rules for the benefit of us all must instead be an essential component of that future.

References

Barnett, E. (2010). Facebook's Mark Zuckerberg says privacy is no longer a 'social norm'. *The Telegraph*. Retrieved from http://www.telegraph.co.uk/technology/facebook/6966628/Facebooks-Mark-Zuckerberg-says-privacy-is-no-longer-a-social-norm.html.

Barocas, S., & Nissenbaum, H. (2014). Big Data's End Run Around Anonymity and Consent. In J. Lane et al. (Eds.), *Privacy, Big Data, and the Public Good: Frameworks for Engagement* (pp. 44-74). Cambridge, GB: Cambridge University Press. Barocas & Nissenbaum, 2014a.

Barocas, S., & Nissenbaum, H. (2014). Big Data's End Run Around Procedural Privacy Protection. *Communications of the ACM*, 57,31-33. Barocas & Nissenbaum, 2014a.

Barrett, P.M. (2014). Forget the Gossip, These Are the Lessons of the Sony Hack. *Bloomberg Businessweek*. Retrieved from http://www.bloomberg.com/bw/articles/2014-12-16/forget-the-gossip-these-are-the-lessons-of-the-sony-hack#p1.

Birnhack, M., et al. (2015). Privacy Mindset, Technological Mindset. *Jurimetrics*, 55. Retrieved from http://papers.srn.com/sol3/papers.cfm?abstract_id=2471415.

Calo, M. R. (2011). The Boundaries of Privacy Harm, *Indiana Law Journal*, 68(1).

Calo, M. R. (2013). Consumer Subject Review Boards: A Thought Experiment. *Stanford Law Review Online*, 66, 97.

Calo, M. R. (2014). Digital Market Manipulation. *George Washington Law Review*, 82, 995.

Cao, Y., & Forrest, E. (2009). Hard vs. Soft Regulation and the Paradoxes of Internet Privacy. *Journal of Global Business Management*, 5(2).

Cate, F. H., et al. (2013). Data Protection Principles for the 21st Century: Revising the 1980 OECD Guidelines. *Microsoft Corporation*. Retrieved from http://www.oii.ox.ac.uk/publications/Data_Protection_Principles_for_the_21st_Century.pdf

Cohen, J. E. (2013). What Privacy is For. *Harvard Law Review*, 126, 1904.

Crawford, K., & Schultz, J. (2014). Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms. *Boston College Law Review*, 55, 93-99.

Drozdiak, N., & Schechner, S., (2015). EU Court Says Data-Transfer Pact With U.S. Violates Privacy. *The Wall Street Journal*. Retrieved from http://www.wsj.com/articles/eu-court-strikes-down-trans-atlantic-safe-harbor-data-transfer-pact-1444121361.

Dwork, C., & Mulligan, D. (2013). It's Not Privacy, and It's Not Fair. *Stanford Law Review Online*, 66, 35.

Ewing, A. (2014). Ericsson Teams With Guardtime in $69 Billion Data-Safety Market. *Bloomberg*. Retrieved from http://www.bloomberg.com/news/2014-09-03/ericsson-teams-with-guardtime-in-69-billion-data-safety-market.html.

Ferenstein, G. (2013). Google's Cerf Says "Privacy May Be An Anomaly". Historically, He's Right. *Tech Crunch*. Retrieved from http://techcrunch.com/2013/11/20/googles-cerf-says-privacy-may-be-an-anomaly-historically-hes-right/.

Gandy, O. (1993). *The Panoptic Sort*. Boulder, CO: Westview.

Gandy, O. (2010). Engaging Rational Discrimination: Exploring Reasons for Placing Regulatory Constraints on Decision Support Systems. *Ethics and Information Technology,* 12, 29.

Gillion, J. (2006). *Overseers of the Poor: Surveillance, Resistance and the Limits of Privacy*. Chicago, IL: University of Chicago Press.

Google Spain Case (2014). Case C-131/12, Google Spain SL and Google Inc. v. Agencia Española de Protección de Datos and Mario Costeja González ECJ.

Haggerty, K., & Ericson, R. V. (2006). The New Politics of Surveillance And Visibility. Buffalo, NY: University of Toronto Press.

Hartzog, W. (2015). The Public Information Problem. Unpublished manuscript on file with authors.

Hill, K. (2014). Stolen Nude Images Reveal Celebs' Location Information. *Forbes.* Retrieved from http://www.forbes.com/sites/kashmirhill/2014/09/02/leaked-nude-images-reveal-celebs-location-information/.

Hirsch, D, & King, J. (2015). Big Data Sustainability, An Environmental Systems Analogy. Unpublished manuscript on file with authors.

Kulk, S., & Borgesius, F. Z. (2014). Google Spain v. González: Did the Court Forget about Freedom of Expression? *European Journal of Risk* Regulation. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2491486.

Leonard, A. (2013). How Netflix Is Turning Viewers into Puppets. *Salon*. Retrieved from http://www.salon.com/2013/02/01/how_netflix_is_turning_viewers_into_puppets.

Lyon, D. (2003). *Surveillance After September 11*. Malden, MA: Polity Press in association with Blackwell Pub. Inc.

Lyon, D. (2003). *Surveillance As A Social Sorting*. London, NY: Routlede. Lyon 2003a.

Macklem, T. (2008). Independence of Mind. Oxford, NY: Oxford University Press.

Mundie, C., Privacy Pragmatism. *Foreign Affairs*, 93(2), 28.

Nash, K. (2012). Privacy: A CIO's Next Competitive Weapon. *CIO.* Retrieved from http://www.cio.com/article/2371662/infrastructure/privacy--a-cio-s-next-competitive-weapon.html.

Nissenbaum, H. (2010). *Privacy In Context*. Stanford, CA: Stanford University Press.

O'Connor, N. (2014). Encryption Makes Us All Safer. *Center For Democracy & Technology Blog.* Retrieved from https://cdt.org/blog/encryption-makes-us-all-safer/.

Pentland, S. (2014). *Social* Physics. New York, NY: The Penguin Press.

Peppet, S. (2011) Unraveling privacy: The personal prospectus and the threat of a full-disclosure future, *Northwestern University Law Review*, 105: 1153-1204. https://www.law.northwestern.edu/lawreview/v105/n3/1153/LR105n3Peppet.pdf

Podesta, J. (2014). Findings of the Big Data and Privacy Working Group Review. *The White House Blog.* Retrieved from http://www.whitehouse.gov/blog/2014/05/01/findings-big-data-and-privacy-working-group-review.

Polonetsky, J., & Wolf, C. (2014). Big Data: A Tool for Fighting Discrimination and Empowering Groups. *Future of Privacy Forum*. Retrieved from http://www.futureofprivacy.org/wp-content/uploads/Big-Data-A-Tool-for-Fighting-Discrimination-and-Empowering-Groups-Report1.pdf.

Reidenberg, J. R., et al. (2014). *Privacy Harms and the Effectiveness of the Notice and Choice Framework.* Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2418247.

Richards, N. M. (2006). The Information Privacy Law Project, *Georgetown Law Journal*, 94, 1087.

Richards, N. M. (2015). *Intellectual Privacy*. New York, NY: Oxford University Press.

Richards, N. M. (2015). Why Data Privacy Law is (Mostly) Constitutional. William & Mary Law Review, 56, 1501. Richards 2015a.

Richards, N. M. (2015). *Four Privacy Myths*: In a world without privacy? Cambridge, MA: Austin Sarat. Richards 2015b.

Richards, N. M., & Hartzog, W. (2015). A Theory of Privacy and Trust. Working paper on file with authors.

Richards, N. M., & King, J. H. (2013), Three Paradoxes of Big Data. *Stanford Law Review Online*, 66, 41-44.

Richards, N. M., & King, J. H. (2014). Big Data Ethics, *Wake Forest Law Review*, 49, 393 (2014).

Rosenbush, S., & Boulton, C. (2014). Data Encryption Will strengthen Privacy Over Long Run, Former CIA Official Says. *The Wall Street Journal*. Retrieved from http://blogs.wsj.com/cio/2014/10/01/data-encryption-will-strengthen-privacy-over-long-run-former-cia-official-says/.

Rubinstein, I. (2014). Voter Privacy in the Age of Big Data. *Wisconsin Law Review* 2014, 861.

Salmon, F. (2013). Why the Internet is Perfect for Price Discrimination. *Reuters. Retrieved from* http://blogs.reuters.com/felix-salmon/2013/09/03/why-the-internet-is-perfect-for-price-discrimination/.

Schneier, B. (2014). Stop the Hysteria over Apple Encryption. Schneier On Security Blog. Retrieved from https://www.schneier.com/essays/archives/2014/10/stop_the_hysteria_ov.html.

Schrems Case (2015). Case C-362/14, Maximillian Schrems v. Data Protection Commissioner ECJ.

Schneider, K. F. et al. (2015). Framing the Big Data Ethics Debate for the Military. *National Defense University Press*. Retrieved from http://ndupress.ndu.edu/Media/News/NewsArticleView/tabid/7849/Article/581865/jfq-77-framing-the-big-data-ethics-debate-for-the-military.aspx.

Senate Commerce Committee on Commerce, Science & Transportation, Office of Oversight & Investigations. (2013). *A Review of the Data Broker Industry: Collection, Use and Sale of Consumer Data for Marketing Purposes*. Retrieved from http://www.commerce.senate.gov/public/?a=Files.Serve&File_id=0d2b3642-6221-4888-a631-08f2f255b577.

Sidel, R. (2014). Home Depot's 56 Million Card Breach Bigger Than Target's. The Wall Street Journal. Retrieved from http://online.wsj.com/articles/home-depot-breach-bigger-than-targets-1411073571.

Solove, D.J. (2004). The Digital Person. New York, NY: New York University Press.

Solove, D. J. (2008). *Understanding Privacy*. Cambridge, MA: Harvard University Press.

Solove, D.J., & Richards, N. M. (2007). Privacy's Other Path: Recovering the Law of Confidentiality. Georgetown Law Journal 96, 123.

Solove, D. J. (2013). Privacy Self-Management and the Consent Dilemma. *Harvard Law* Review, 126, 1880.

Solove, D. J., & Hartzog, W. (2014). The FTC and the New Common Law of Privacy. Columbia Law Review, 114, 583.

Smith v. Maryland, 442 U.S. 735 (1979).

Toobin, J, (2014). The Solace of Oblivion: In Europe, the right to be forgotten trumps the Internet. The New Yorker, Retrieved from http://www.newyorker.com/magazine/2014/09/29/solace-oblivion.

Tysiac, K. (2014), Use data privacy to gain a competitive advantage. Journal of Accountancy. Retrieved from http://www.journalofaccountancy.com/News/20149721.

Waters, M. A. (2005). Mediating Norms and Identity: The Role of Transnational Judicial Dialogue in Creating and Enforcing International Legal Norms. *Georgetown Law Review,* 93, 487.

White, A., & Benoit, A. (2014). 'Google It' Becomes 'Hide It' After Right To Be Forgotten. *Bloomberg*. Retrieved from http://www.bloomberg.com/news/2014-07-09/-google-it-becomes-hide-it-after-right-to-be-forgotten.html.

Wingfeld, N. (2014). Microsoft's Top Lawyer Is The Tech World's Envoy. *The New York Times*. Retrieved from http://www.nytimes.com/2014/07/21/technology/microsofts-top-lawyer-is-the-tech-worlds-envoy.html?_r=0.

Wohlsen, Marcus *Facebook Won't Stop Experimenting on Your. It's Just Too Lucartive*, WIRED, Jan. 2, 2014, available at http://www.wired.com/2014/10/facebook-wont-stop-experimenting-just-lucrative/?mbid=social_twitter.

World Economic Forum. (2013). *Unlocking the Value of Personal Data: From Collection to Usage*. Geneva, CH: World Economic Forum.

Zuckerman, E. (2014). The Internet's Original Sin. *The Atlantic*. Retrieved from http://m.theatlantic.com/technology/archive/2014/08/advertising-is-the-internets-original-sin/376041/?curator=MediaREDEF#.

---

[i] Technically, free speech is a regulation of the regulation of information flows, since it restricts governments from restricting information flows that are "free speech." (Richards 2015a)

[ii] E.g., Electronic Communications Privacy Act, 18 U.S.C. § 2501 et seq.; Fair Credit Reporting Act, 15 U.S.C. § 1681 et seq.