Draft – Not for general distribution

2015-16 Faculty Seminar: Democracy, Citizenship, and Constitutionalism
University of Pennsylvania February 18, 2016

Must Privacy Give Way to Use Regulation?

Helen Nissenbaum
Media, Culture, and Communication
New York University

Abstract: Projected benefits of data science and the paradigm of big data are not compatible with the regulation of information collection prompted by concerns over privacy. So goes a compelling and popular argument. I am skeptical not only because it plays suspiciously well with the dominant business model of the commercial information industry but also because it rests on misconceptions and ambiguities of key terms. My paper seeks to unravel the debate between those who continue to see value in protecting privacy and those who would forgo privacy in favor of use regulation instead. There is no denying some of the genuine and unprecedented challenges to privacy posed by data science, but letting it go will undermine a cornerstone of individual freedom.

* * *

1 Background

In the paper, "Big Data's End Run Around Anonymity and Consent,"[1] Solon Barocas and I demonstrated that two mainstays of privacy regulation were fatally challenged by technical capabilities of data science. The claim was not that consent and anonymity no longer perform any useful function, but only that they can no longer can they serve the critical functions that up till the present they had served up till the present – consent as privacy's gatekeeper, anonymity as a boundary for privacy's remit. Aware that our conclusion could appear to lend force to those favoring data use regulation above as opposed to data collection regulation, we explicitly disavowed the connection. It is important not to confuse the means of protecting privacy, namely, consent and anonymity, with privacy itself, understood as appropriate flow. Our article left unaddressed, however, a position steadily gathering momentum in the academy, the information industry, and policy arena that privacy, insofar as it restricts information collection, is no longer tenable, recommending instead, that attention should rest on how information is used. The present article dissects this position – what it means and whether its worldview is inevitable.

* * *

In January 1999, then CEO of Sun Microsystems, Scott McNealy brashly threw down the

[1] Solon Barocas and Helen Nissenbaum, "Big Data's End Run Around Anonymity and Consent," in *Privacy, Big Data and the Public Good: Frameworks for Engagement*, ed. Julia Lane et al. (Cambridge: Cambridge University Press, 2015), 44-75.

gauntlet with, "You have zero privacy anyway. Get over it!"[2] [ref] Repeated countless times ever since then, it was a statement some chose to ignore and ridicule, others to resist and counter. Because the conception of privacy McNealy presumed was so muddled, his statement hardly bore a serious response and while threatened, privacy was far from dead and continued to inspire defenders. Nevertheless, there was no denying that this "bad -boy" stance on privacy was appealing, and has resurfaced repeatedly in various guises and versions. David Brin's popular book, *Transparent Society* (1998),[3] another instance asserts that privacy not only is untenable, given the direction of technological advancement, but that it is also no longer desirable. Total transparency makes much more sense as it allows historically weaker parties, those captured in the webs of surveillance, to turn the tables on stronger parties, holding them accountable for their actions. A decade-and-a-half later, the cultural phenomenon of big data (and associated technical paradigms of data science) has yielded its own version of the bad boy stance: it is pointless to seek restrictions on the collection of information, or data; instead, the focus should be on restricting its uses.[4]

I would have liked to dismiss these as fringe provocations or venal ploys of the information sector including those who stand to benefit, such as Google, Facebook, Amazon, and Twitter as well as actors less known to the public, such as Acxiom, IMS Health, and LexisNexis. Others in the commercial arena whose interests are served by reduced constraints on collection, though not directly offering information products but still able to capture value from consumer information, are telecommunications companies, financial companies, insurance companies, media and publishing companies, and, increasingly, retail merchants.[5] Horses out the barn, a technological infrastructure designed to capture data, an imperative of data-driven institutional bureaucracies all point to the conclusion that it is impossible and imprudent to resist. Outside the commercial realm, too, many actors are eager to collect, record, and hold on to data, without restraint, including governmental agencies, utilities companies, healthcare organizations, educational institutions, as well as a range of not-for-profit public interest organizations. Unlike previous bad boy stances on privacy, the contemporary position supporting the regulation of use, as opposed to collection, has become broadly compelling and mainstream.

This paper argues that the push to regulate use, instead of collection, is problematic and may even be dangerous. Furthermore, because expression of the position is often confused and ambiguous, it presents a fuzzy target to those who aim to evaluate it meaningfully. Accordingly, before tackling the substantive issues, the first order of business is to clarify some of the conceptual problems and, at least in the context of this paper, to establish terminological consistency. After sharpening the position, the paper turns back to substantive evaluation.

2 Versions of advocacy for use over collection

[2] Polly Sprenger, "Sun on Privacy: 'Get Over It'," *WIRED*, January 26, 1999, accessed February 9, 2016, http://archive.wired.com/politics/law/news/1999/01/17538.
[3] David Brin, *Transparent Society: Will Technology Force Us to Choose Between Privacy and Freedom?*, Cambridge: Perseus Books, 1998.
[4] See also Christian Heller, *Post-Privacy: Prima leben ohne Privatsphäre* (*Post-Privacy: Living just Fine Without Privacy*), Munich: C.H. Beck, 2011.
[5] See Joseph Turow, *The Daily You: How the New Advertising Industry is Defining Your Identity and Your Worth* (New Haven: Yale University Press, 2011).

It is a shame one cannot rest an argument on anecdotal observations because it would then be possible to refer simply to the countless panels and presentations at conferences and workshops at which speakers, with a wave of a hand, relegate collection restrictions to the zone of the impossible and privacy as hopelessly passé. Fortunately, however, there do exist written versions that allow us to take a closer look.

Consider, for example, Recommendation 1 of the President's Council of Advisors on Science and Technology (PCAST) Report to the President: "Big Data and Privacy: A Technological Perspective," which asserts that, "Policy attention should focus more on the actual uses of big data and less on its collection and analysis. By actual uses, we mean the specific events where something happens that can cause an adverse consequence or harm to an individual or class of individuals. In the context of big data, these events ('uses') are almost always actions of a computer program or app interacting either with the raw data or with the fruits of analysis of those data. In this formulation, it is not the data themselves that cause the harm nor the program itself (absent any data), but the confluence of the two. These 'use' events (in commerce, by government, or by individuals) embody the necessary specificity to be the subject of regulation. By contrast, PCAST judges that policies focused on the regulation of data collection, storage, retention, a priori limitations on applications, and analysis (absent identifiable actual uses of the data or products of analysis) are unlikely to yield effective strategies for improving privacy. Such policies would be unlikely to scale over time, or be enforceable by other than severe and economically damaging measures."[6]

Michael Seemann's reasons for reducing constraints on access to personal information have a different rationale: "So instead of trying to defend privacy against surveillance, we should be fighting institutionalized punishment. Authoritarian border controls, racist police cohorts, homophobic social structures, inequality in health and welfare systems, and institutional discrimination are the true danger zones in terms of surveillance. Above all, the state itself, with its monopoly on force and its sweeping claims to regulatory authority, is the source of most of the threat scenarios that *do* jeopardize freedom by way of surveillance."[7] He agrees with Jane Yakowitz who argues that privacy is selfish as "open data is a major source of social welfare."[8] According to Seemann, new capabilities call for a new orientation towards data regulation: "In the Old Game, it was often purposeful to enforce data control in order to limit existing powers. … Privacy was intended to shield civilians from the control exerted by institutions. In the New Game, however, this approach no longer works, and in fact, it may produce exactly the opposite effects.… Data protection requirements give platforms reason to shut themselves off, limiting their interoperability, and reinforcing lock-in effects."[9] "So instead of demanding more privacy, we should convince platform operators to open up their data. Because the more open the data becomes, and the more queries can be applied to it, the easier it will

---

[6] PCAST, "Big Data and Privacy: A Technological Perspective," *White House*, May 2014, accesed February 9, 2016, https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf.

[7] Michael Seemann, *Digital Tailspin: Ten Rules for the Internet After Snowden*, Network Notebooks 09 (Amsterdam: Institute of Network Cultures, 2015), 22.

[8] Seemann, *Digital Tailspin*, 28, quoting Jane Yakowitz, "Tragedy of the Data Commons," *Harvard Journal of Law and Technology* 25, no. 1 (Fall 2011): 1-67.

[9] Seemann, *Digital Tailspin*, 49.

be to fence in the power of platforms."[10]

One of the most fully articulated versions is reflected in an essay by Craig Mundie, Senior Adviser to the CEO and former Chief Research and Strategy Officer of Microsoft who writes, "Today, the widespread and perpetual collection and storage of personal data have become practically inevitable. Every day, people knowingly provide enormous amounts of data to a wide array of organizations, including government agencies, Internet service providers, telecommunications companies, and financial firms. Such organizations -- and many other kinds, as well -- also obtain massive quantities of data through 'passive' collection, when people provide data in the act of doing something else: for example, by simply moving from one place to another while carrying a GPS-enabled cell phone. Indeed, there is hardly any part of one's life that does not emit some sort of 'data exhaust' as a byproduct. And it has become virtually impossible for someone to know exactly how much of his data is out there or where it is stored. Meanwhile, ever more powerful processors and servers have made it possible to analyze all this data and to generate new insights and inferences about individual preferences and behavior.

This is the reality of the era of 'big data,' which has rendered obsolete the current approach to protecting individual privacy and civil liberties. Today's laws and regulations focus largely on controlling the collection and retention of personal data, an approach that is becoming impractical for individuals, while also potentially cutting off future uses of data that could benefit society. The time has come for a new approach: shifting the focus from limiting the collection and retention of data to controlling data at the most important point -- the moment when it is used."[11]

Bert-Jaap Koops, an eminent EU legal scholar also steers attention away from collection to use. Addressing the question, "How can data protection meet the challenge of decisions increasingly being taken on the basis of large-scale, complex, and multi-purpose processes of matching and mining enormous amounts of data?" he answers that "the focus in data protection should shift from ex ante regulation of data processing to ex post regulation of decision-making," supporting, "an alternative approach, one that focuses less on data minimisation, user control, and procedural accountability, but instead directs its arrows at the outcome of computation-based decision making: the decision itself."[12]

Reporting on regional discussions held internationally about privacy regulation in light of big data, Viktor Mayer-Schonberger and Fred Cate observe that "one of the most widely discussed alternatives was focusing more attention on the 'use' of personal information rather than on its 'collection,' given the increasingly pervasive nature of data collection and surveillance, inexpensive data storage and sharing, and the development of

---

[10] Seeman,n *Digital Tailspin*, 55.
[11] Craig Mundie, "Privacy Pragmatism: Focus on Data Use, Not Data Collection," *Foreign Affairs*, March/April 2014, accessed December 26, 2015, https://www.foreignaffairs.com/articles/2014-02-12/privacy-pragmatism.
[12] Bert-Jaap Koops, "On Decision Transparency, or How to Enhance Privacy after the Computational Turn," in *Privacy, Due Process and the Computational Turn: The Philosophy of Law Meets the Philosophy of Technology*, eds. Mireille Hildebrandt & Katja De Vries (New York: Routledge, 2013), 197.

valuable new uses for personal data."[13] Although constraints on collection may be necessary in exceptional cases the focus should be on clarifying what "use" covers, and what outcomes to consider when analyzing the associated costs and benefits.

In their article, "Big Data for All: Privacy and User Control in the Age of Analytics," Omer Tene and Jules Polonetsky call for a retrenchment of data minimization, a traditional pillar of privacy regulation aimed at restricting the collection and retention of data to express purposes. They observe, "The big data business model is antithetical to data minimization. It incentivizes collection of more data for longer periods of time. It is aimed precisely at those unanticipated secondary uses, the 'crown jewels' of big data. After all, who could have anticipated that Bing search queries would be used to unearth harmful drug interactions?" They continue, "legal rules collide with technological and business realities. Organizations today collect and retain personal data through multiple channels including the Internet, mobile, biological and industrial sensors, video, e-mail, and social networking tools. Modern organizations amass data collected directly from individuals or third parties, and they harvest private, semi-public (e.g., Facebook), or public (e.g., the electoral roll) sources. Data minimization is simply no longer the market norm."[14]

The common substantive thread embodied in these excerpts, at least superficially, is to divert regulatory effort toward data use and away from data collection in light of the big data phenomenon. Before evaluating it, we need to unravel terminological ambiguities and expose inconsistencies underlying different expressions of this position that ultimately hinge on different arguments.

3 Privacy

Skeptical, "bad boy" positions on privacy with which I opened this paper challenged the value of privacy in light of more important countervailing values served by information and digital technologies. Leaving aside questions concerning their implicit conceptions of privacy and quality of supporting arguments their key claim is that other values should prevail in a trade-off against privacy. Versions of this rhetoric honed to big data similarly assert: "We can have big data or privacy but not both; we must have big data, ergo ..." Michael Seemann, for example, is ready to give up privacy when he presents the new "big data world order" in which privacy works in paradoxical ways to entrench power in the hands of powerful government and commercial actors. But others who support use regulation alone (or mainly) maintain an explicit commitment to privacy, while disassociating it from collection regulation. Roughly, they are saying, "if we really want to protect privacy, we must protect against harmful uses of information (not against collection)." I will label this stance, *big data exceptionalism.* Proponents do not lend their voices to the chorus – either privacy or big data, but not both. Instead, in what might ultimately come down to a definitional issue, they would say when it comes to big data we need to think differently about what privacy protection demands.

As a matter of fact, proponents of big data exceptionalism frequently hail from communities that strongly identify privacy with Fair Information Practices (FIPs),

---

[13] Fred H. Cate and Victor Mayer-Schönberger, "Notice and Consent in a World of Big Data," *International Data Privacy Law* 3, no. 2 (2013), 69.

[14] Omer Tene and Jules Polonetsky, "Big Data for All: Privacy and User Control in the Age of Analytics," *Northwestern Journal of Technology and Intellectual Property* 11, no. 5 (2013): 259-260.

embodied, for example, in traditional OECD Privacy Principles (and numerous other renderings) governing Collection Limitation, Data Quality, Purpose Specification, Use Limitation, Security Safeguards, Openness, Individual Participation, and Accountability. A number of influential works by experts in international privacy regulation have taken up the challenge of reconciling big data, the call to relax collection limitations, and bedrock Fair Information Practice Principles (FIPPs).[15] When I consider that triad, I will say, only that "something's *gotta* give." As of the writing of this paper, proponents are energetically developing revised formulations of FIPs that will allow for relaxation of the troublesome principles requiring advance specification of purpose, collection of data consistent with purpose, requirements of informed consent, etc. – in other words, existing formulations that might obstruct machinations of big data. These efforts seek to mitigate the damage to traditional FIPs with more rigorous use requirements, achieved through cost-benefit analyses taking into consideration potential privacy harms. More will be said on this, and related issues throughout the paper.

The account of privacy as contextual integrity (CI) that I have advanced[16] is less brittle in the face of some of big data's challenges than those relying on FIPs. I say more about this later in the paper, but to anticipate the comparison, a description of the theory is needed, though a brief one will suffice. According to CI, concern for privacy is satisfied when personal information flows appropriately, which means, in turn, that it complies with expectations reflected in informational norms. These norms are specific to social contexts (e.g. education, healthcare, political citizenship, home-life, etc.) and prescribe and prohibit information flows (e.g. collection, disclosure, etc.) according to senders, recipients, and data subjects (acting in context-defined capacities), information types, and transmission principles. The last of these refers to the constraints under which information of a given time passes from actor to actor, for example, "with a warrant," or "with the requirement of confidentiality," as two well-known instances.

The contextual or context-specific informational norms constitute a rigorous model of privacy expectations, prescribing and proscribing information flows in terms of information type, principled constraints (e.g. confidentiality, informed consent, and many others), and the implicated parties -- senders, recipients, and, of course, information subjects.

Whereas FIPs-based accounts consider informed choice to be necessary and sufficient for privacy (except, arguably, in the few areas where we have statutory protections), CI considers it to be merely one among countless transmission principles, defensible only with the actors and information types appropriately instantiated. (I have painstakingly

---

[15] See Fred H. Cate, Peter Cullen and Victor Mayer-Schönberger, "Data Protection Principles for the 21st Century Revising the 1980 OECD Guidelines," *Oxford Internet Institute*, March 2014, accessed December 26, 2015, http://www.oii.ox.ac.uk/publications/Data_Protection_Principles_for_the_21st_Century.pdf; Ira Rubinstein, "Big Data: The End of Privacy or a New Beginning?" *International Data Privacy Law* 3, no. 2 (2013): 74-87; Tene and Polonetsky, "Big Data for All."

[16] See Helen Nissenbaum, *Privacy in Context: Technology, Policy, and the Integrity of Social Life*, Palo Alto: Stanford University Press, 2010; Helen Nissenbaum, "Respect for Context as a Benchmark for Privacy Online: What it is and isn't," in *Social Dimensions of Privacy: Interdisciplinary Perspectivs*, eds. B. Roessler, D. Mokrosinska (Cambridge: Cambridge University Press, 2015), 278-302.

defended this claim in other work; to do so here would be to a detour too far.)[17] Unlike accounts of privacy defined in terms of FIPs, CI would not render impossible many promising applications of big data technologies, as the substantive prescriptions CI supports may be compatible with these applications even in the absence of informed choice.

When information practices involve information flows that comply with informational norms, contextual integrity is maintained; but norm transgressions are flagged as requiring further analysis. Often it is the deployment of new information technologies (loosely defined) that lead to disruptions in terms of recipients, information types, and transmission principles, and trigger questions about why entrenched practices *should* be respected and disruptions resisted. These normative questions call into play an evaluative dimension of CI, which asserts that information flows deserve to be protected (whether through law or any other type of regulation) based on the balance of interests they serve, the extent to which they promote ethical and political values (e.g. autonomy, fairness, social justice, security), and how well they serve contextual ends and purposes. Although the last of these three are, in fact, invoked in many arguments over privacy, CI is unique in giving it conceptual prominence in the theoretical model.

For the rest of the paper, when I speak of privacy, unless I specify otherwise, I mean privacy understood as contextual integrity. Because collection always involves flow, one may always question whether it complies with contextual integrity or not. Whether and how use is relevant to contextual integrity will depend on how one defines it. Here lies a source of ambiguity that will be revealed and discussed in the next section, before we take up the substantive questions at issue.

4 Collection versus Use: Definitional Conundrum

Foundational to the thesis that use should be regulated while collection, for the most part, be allowed to proceed freely is the dichotomy of collection and use. So, the lack of agreement and clarity surrounding these critical terms creates difficulties for evaluation. For some, *collection* covers only the initial moment of interface between data subjects and data collecting entities (variably labeled data processors, data recipients, first parties, etc.) when data (or information) is first recorded (requested, acquired, observed, sensed, created, documented, constructed, etc.). *Use* is everything else, including various applications of data in order to classify, predict, and make decisions about data subjects, a range of other practices, including aggregation, storage, and analysis, as well as disclosure, distribution, and dissemination to other parties. In other words, anything beyond or in addition to initial uptake counts as use. This narrow definition of collection is consistent with the much aired worry that FIPs consent-based restrictions would cramp the usefulness of the big data apparatus.

Another view conceives of *collection* as including all information flows from initial uptake to storage, onward distribution to other parties, and the extraction of new information based on inferences drawn from the initial set. *Use*, under this view, covers a narrower range, including what data holders *do* with the data, and how data guides prediction, decision-making, and action in relation to data subjects and possibly also other people. While the PCAST report, for example, seems committed to a narrow definition of collection as initial uptake and maintains that everything following it should be closely

[17] Ibid.

regulated, other supporters of big data exceptionalism seem unaware of the need to draw clear lines or are more inclined to limit *use* to the narrower set.

Where one draws the line, obviously, will significantly affect the scope of big data exceptionalism. This is no small quibble. The second view would have a radical impact on privacy conceived in terms of appropriate information flows. Even those who care little about privacy's fate should nevertheless remain interested in such matters of interpretation. Why? Because never mind the ultimate success or failure of consent as the linchpin of collection regulation, a broad interpretation of collection not only opens the door to new-fangled collection practices that are unfamiliar to most ordinary people, but exposes unimaginable volumes of data already collected to the laxer constraints this interpretation allows.

Such anxieties might steer those seeking a reasonable balance toward the narrowest conception of collection, but a bright line might, in fact, be difficult to draw. To begin, consider what we mean by "initial uptake" as the definition of collection in a world where much of the data in question is intermediated. If you communicate with me via Gmail, Google is first in the line of distribution and I only receive the message as a secondary recipient; hence, this flow would count as a *use* by Google of the information. This strikes me as wrong. If receiving the message counts as a point of collection, by the same token, data brokers might argue that they too, are merely collectors of data (and subject to lowered regulation) whether in collecting from government databases or from other first-points of collection.

Concluding this section, I do not see how narrowly defining collection as uptake at initial point of contact with data subjects will not exclude many of the cases that fueled the appeal of big data exceptionalism in the first place, such as allowing medical researchers freer access to primary health data. At the same time, I can see no line drawing that is inclusive of these without going down a slippery slope that would allow in data brokers and other such secondary collectors. The upshot is that those who would reduce or jettison restrictions on collection will be held accountable for most information flows currently under privacy's purview.

5 Big Data Exceptionalism: Focus Regulation on Use

The proposal to lift restrictions on collection would be a significant departure from data protection, or privacy, discourses and hence demands close scrutiny. Although applications of data science (often used interchangeably with "big data") raise important and sometimes distinctive challenges from traditional information practices, and although there is widespread support for adjusting privacy expectations in accordance with these, I find the supporting observations and arguments unconvincing. This paper, therefore, undertakes two tasks. One is to give expression to the thesis, which I argue is not unified but comprises several different strands. The second is to locate and evaluate reasons and arguments supporting these different strands. I suspect it will be surprising both to supporters and critics to discover how incoherent big data exceptionalism is in light of the enthusiasm it seems to have garnered. What exactly does it say?

6 Descriptive versus Normative

One point of departure is whether exceptionalism is a descriptive or normative claim. Craig Mundie seems to believe both: it is simply impossible to apply privacy regulations

to collection and even if it were possible, it would not be desirable. Let us consider each in turn.

6.1 Two Impossibility arguments: the technical and the institutional

Many versions of the impossibility argument are rooted in a commitment to FIPS with the strange result that those most deeply committed to them are ready to throw the towel when it comes to big data because of its dependency on consent as a gatekeeper. In this respect, online tracking was a precursor to a burgeoning array of activity capture, including clickstream monitoring, networked sensors, metadata capture, and geo-location tagging; and passive demographic data a precursor to location within complex social graphs. Common to all is their awkward fit within dominant regulatory approaches.[18] The alternative based on articulation of substantive rules, which I have supported,[19] is not immune because adjudicating whether novel information flows are acceptable requires an assessment of likely outcomes in terms of interests, rights and values, and contextual ends and purposes. Boosters of data science demur: we cannot anticipate outcomes until after enough of the data has been collected – chicken and egg.

A variation of the impossibility argument points to the systems of computational and communications media that we cannot avoid and have thoroughly embraced; it is inherent to the functioning of these systems and devices that information about us leaves indelible traces – voice and text communications as well as a myriad field of metadata that inexorably comes with them, such as biometric markers, dates, times, frequencies, and more. Devices, operating systems, platforms, and networks must absorb these data streams ("exhaust") in their very functioning. Data is collected; therefore, it must be collected. This has been particularly powerful for the swath of actors (many commercial) we might call information or data intermediaries. Web searches and calls, locations visited, and videos streamed all leave indelibly marked trails. Data capture is a technological imperative, therefore, proponents say, it is irresistible.

These arguments belie the contingency of technology and its design. Design differences between, say, the domains of Internet and mobile networks have significant consequences for what is easy and difficult to achieve. The existing designs involving centralized servers enabling the capture and flows of information that many accept fatalistically form part of merely one among other possible architectures.[20] Details aside, the reference to flows determined by entrenched technology design is naïve if not disingenuous. One only needs to recall early Web cookie protocols designed to keep out third parties in light of workarounds of the present day that have made cookies one of

---

[18] Notwithstanding admirable work seeking to improve notice and consent by privacy scholars such as Ryan Calo, "Against Notice Skepticism in Privacy (and Elsewhere)," *Notre Dame Law Review* 87, no. 3 (2013): 1027-1072; Joel R. Reidenberg et al., "Privacy Harms and the Effectiveness of the Notice and Choice Framework," Fordham Law Legal Studies Research Paper No. 2418247, available at SSRN, March 29, 2014, accessed February 9, 2016 http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2418247.

[19] See Helen Nissenbaum, "A Contextual Approach to Privacy Online," *Daedalus* 140, no. 4 (Fall 2011): 32-48; Solon Barocas and Helen Nissenbaum, "Big Data's End Run Around Anonymity and Consent."

20 See Vincent Toubiana et al. "Adnostic: Privacy Preserving Targeted Advertising," *Proceedings Network and Distributed System Symposium*, March 2010.

the greatest sources of data leakage to appreciate design's contingencies. When incentives are aligned, systems have been developed that are able to carefully channel flows based on fine-grained distinctions among recipients, attributes, and purposes – what information is needed for core functionality, what for advertising, or other purposes, and so forth. The immensely complex ecology of targeted advertising depends on such capabilities; to policy makers or academics with limited technical savvy, they may seem overwhelmingly abstruse. Further, in cases that are too difficult to manage technologically, policy may be imposed on top of design. All of this is to say that blunt assertions of impossibility to regulate are often disingenuous.

Perhaps is it not only technological imperative that drives a sense of the impossible concerning data capture because, as difficult as it may be to impose regulations on a range of material flows, including ephemera of data and information, this has not stopped us from attempting to do so. It has been excruciatingly difficult to staunch the influx of illegal narcotics, but we have not ceased in our quest to do so. Financial markets require incredible vigilance, particularly in the confluence of money and data; yet we persist. And, in relation to the holdings of digitized content and information held in the data repositories of private commercial and governmental holders, we do not admit to a sense of helplessness. In a historical moment, we successfully regulated what telecommunications providers ("common carriers") recorded and stored despite their service as data intermediaries, but we balk at doing so for their descendants.

We are not seeing the NSA, the IRS, Google, or Facebook -- not even Diebold when elections are in question -- caving to the irresistible tendency for data to flow. Metaphysics and architecture have little to offer as justifications for big data exceptionalism. If there is an impossibility argument to which supporters of use-restriction may refer it is not something essential to data. I would hazard a guess that the sense of the impossible has more to do with the institutional landscape and the respective parties whose data are and are not being syphoned. When we consider the societal forces arrayed against collection-restriction, and the radical upheaval it would mean for the practices and privileges of the collectors who have amassed not only national but global influence, we conclude it is impossible to regulate; the ten-ton truck is barreling down the road and nothing can stop it. It is institutional inertia, not metaphysical impossibility that confronts us.

Before turning to consider normative perspectives, one final remark related to earlier "bad boy" claims: In David Brin's fantastical world, not unlike Michael Seemann's post-apocalyptic state, clever citizens have opted for full transparency to keep tabs on powerful actors – governmental and commercial – unlike the alternative universe in which we foolishly demand secrecy which not only fails to protect us against those powerful enough to obtain information but, in turn, allows these actors to operate in obscurity.

But if we are resigned to unstoppable collection, the source of hope that any information regulation will be possible is unclear. If restricting collection is impossible because the collection and flow that saturates all domains of life is the lifeblood of public and private institutional incumbents, and the systems that support them, we should expect the same resistance to use-restriction. Mundie's pragmatic vision for holding these actors accountable for usage is to enfold or tag data in metadata and construct a system of verifiable identities that will allow us to express, limit, and monitor what holders of data can do with it. I daresay, with such mechanisms in place even the task of regulating

collection would be greatly eased. In my view, neither the idealistic nor pragmatic visions will work because they imply significant burdens on the very class of actors who have worked determinedly -- and seem to have succeeded -- to shake off meaningful restraints on collection. In short, those whose collection activities defy close monitoring and regulation are unlikely to offer an easier target for use regulation. We are not here talking about Russian mobsters setting up botnets but also mainstream actors who seek immunity from watchdog organizations and public interest vigilantes. They actively use technological as well as legal means, such as, nondisclosure clauses, to obscure problematic information flows. (Nowhere is this more evident than in the mobile arena.) In other words, if the reason for giving up on collection restrictions is that the barn door is open, the cat out of the bag, there's no going back, it will be no easier to regulate use.

6.2 Normative arguments

Many supporters of big data exceptionalism not only make observations of how difficult or impossible it would be to stop or turnaround rampant capture of information (which is just as well because, as we have shown, these reasons are insufficient). They also claim that it *ought* not be restricted. One strain of this argument points out that, as a matter of fact, effective constraints on collection may conflict with political and ethical rights and values. Another is that constraints on collection for the sake of privacy can no longer be defended because such constraints would foreclose important benefits to individuals and societies. The remainder of this paper will focus on the second argument because it is not only the one we hear most often, but also the one that is the most complex and compelling. Nevertheless, I will briefly address the first argument, which is a familiar refrain in the literature on privacy.

6.2.1 Collection restrictions and constitutional rights

According to this argument, effective constraints conflict with ethical and political rights and values to which liberal democracies are committed as a matter of principle or explicit law (or both). Were it the case that collection merely involved furthering the interests of collectors, the assertion of privacy interests could have prevailed. Data collection involves no immediate harm, however, and more importantly is akin to pursuits that are constitutionally defended, such as security, intellectual property, and free speech (and associated intellectual freedoms).[21] To those who have followed privacy debates over the course of time – pick any of its significant milestones: the 1890s landmark article by Warren and Brandeis,[22] the 1960s proposed Federal Data Center,[23] the 2000s social

[21] See e.g. Eugene Volokh, "Freedom of Speech, Information Privacy, and the Troubling Implications of a Right to Stop People from Speaking About You," *Stanford Law Review* 52, no. 5 (2000): 1049-1124.

[22] Samuel D. Warren and Louis D. Brandeis, "The Right to Privacy," *Harvard Law Review* 4, no. 5 (1890): 193-220.

[23] Proposed, in 1965, by the Social Science Research Council of the American Economic Association after a three-year study revealing that neither scholars nor other agencies were able to make use of public data because of decentralization. (Communications on the Preservation and Use of Economic Data, Report to the Social Science Research Council of the American Economic Association 1965, reprinted in House Hearings 195.) The SSRC urged the creation of a federal data center to make basic statistical data from all Federal agencies available to non-Governmental users and other Federal agencies.

media, or this decade's Snowden revelations – these rejoinders will be familiar. Privacy claims are rarely uncontested; usually they challenge another party or other parties' interests, preferences, or even rights (for example, surveillance of media usage through Digital Rights Management systems (DRMs) and Technological Protection Measures (TPMs), or surveillance of populations for the sake of safety and security, where they have been framed as challenges to collective or public interests[24]).

It would be tangential to pursue this line of discussion any further. I point it out for the sake of completeness in identifying the types of reasons defenders of big data exceptionalism invoke. The contemporary move shifts the focus of older arguments to that of collection, mostly by relying on counter-claims associated with First Amendment rights, such as those invoked in the highly disputed *Sorrell v. IMS Health* (2011) Supreme Court case, and more generally, on our rights to pursue knowledge by collecting information as well as to draw inferences, whether with the brute force of human capacities to reason or enhanced by machine learning and statistical techniques.

In my view, these arguments go too far because they rely on simplistic conceptions of privacy as secrecy or as control over information. For them, the question is all-or-nothing. Contextual integrity, which prescribes the modulation of information flows according to several parameters and set within a complex teleological frame, would allow flows that could respect privacy and constitutional values at the same time. Although conflicts may still arise, a refinement of our understanding of privacy would allow many prima facie conflicts to be dissolved. (More on this later.)

6.2.2 Collection restrictions and foreclosure of benefits

The potential of big data to deliver benefits is so great that imposing restrictions on collection and accumulation of data is to withhold the lifeblood of this promising enterprise. Although this prevalent line of reasoning is compelling, I will argue that it claims more than it can prove both in understating the grounds for resisting "open season" on data collection and in overstating the overall value of big data's benefits. Big data and data science do provoke the need for new approaches to articulating and achieving appropriate flows and uses of information, not by categorically dispensing with past approaches, but by understanding and carrying forward their core ethical and political insights.

Let us begin by drilling deeper into the foreclosure-of-benefits reasoning as rendered in the work of many enthusiasts in academia, government, and public life.[25] The confluence of mathematical and computational sciences and technologies, networked sensors and digital networks of fixed and mobile devices, has given us prodigious capacities to collect, amass, store, and analyze data. Although there are no radical discontinuities in

[24] For an overview see e.g. Nissenbaum*, Privacy in Context*, 2010; especially Part II.
[25] See e.g. Kenneth Neil Cukier and Viktor Mayer-Schönberger, "The Rise of Big Data: How It's Changing the Way We Think About the World," *Foreign Affairs* 92, no. 3 (May/June 2013): 28-40; Mundie, "Privacy Pragmatism"; PCAST, "Big Data and Privacy"; Tene and Polonetsky, "Big Data for All"; Yakowitz, "Tragedy of the Data Commons"; White House, "Big Data: Seizing Opportunities, Preserving Values," *Executive Office of the President* (May 2014): 1-79; and a host of newly minted trade books.

these capabilities, no quantum leaps, or scientific revolutions,[26] there are epistemological paradigm shifts[27] that have resulted from the expansion of scale and scope. Interested readers can enrich their knowledge of specific developments through the books and articles I have cited, and many more that I have missed.

There are, however, a few points worth highlighting. One is the range of sensory modalities that can be captured and entered into storable repositories and, related to this, the growth in the capacity to express information in digital formats that allow for quantitative and statistical analysis. The technologies we are using today allow for sensing and recording bits of data – imagine them as data particles – that were not accessible to capture before (e.g. more sensitive and sophisticated sensor apparatuses). Along with traditional sources of information and data, the media and devices that are part of everyday life in technologically advanced societies, including the Internet, mobile phones, and increasingly the *Internet of Things*, wearables, self-tracking devices of the quantified self-movement, etc. contribute to this massive stock. Furthermore, aspects of experience and forms of expression and production that were previously not grist for the data science mill, such as affect, sound, text, image, video, type and strength of relationships, biological characteristics ("biometrics"), "brain waves" indicative of thought patterns and sensory experiences, to name a few, now very much are.[28] (Many also celebrate the capacity of database technologies that have given us unprecedented analytical powers over so-called *unstructured data.)*

Capacities to create interpretable data repositories through cleaning, selection, and aggregation allow for inventive applications of supervised and unsupervised data mining, knowledge discovery and machine learning, in turn revealing complex patterns, profiles, and regularities, which help us understand the past, predict the future, and make better decisions. Although it may be utterly self-evident to experts in the field, it may be worth pointing out that the contributions and promise, as well as the insidious threats of big data, lie in the meshing of data accrual with data analytics. This is most clearly seen in cases of data points that individually may have no meaning or significance, for example, the click or motion of a mouse, or data points that when too sparse fail to meet levels of statistical significance and offer little analytical insight. In such cases, it is the combination of scale, amassed density, and enrichment with contextual data, subjected to analytical mechanisms, that yields knowledge, insight, and meaning, and the grounds for prediction and rational, algorithmic decision-making. Suddenly, the data we have on rare diseases will be significant, we may find we can distinguish innocuous communications from dangerous plots, and discover more or less productive learning patterns. **Although what we learn from such exercises might be the result of hypotheses and hunches, it might also be surprising, counterintuitive, and**

---

[26] See Thomas S. Kuhn, *The Structure of Scientific Revolutions*, 3rd ed., Chicago and London: The University of Chicago Press, 1996.
[27] See Paul Feyerabend, *Against Method*, 4th ed., New Ýork: Verso, 2010.
[28] Some have referred to this as data-fication. See e.g. Katherine J. Strandburg, "Monitoring, Datafication, and Consent: Legal Approaches to Privacy in the Big Data Context," in *Privacy, Big Data and the Public Good: Frameworks for Engagement*, ed. Julia Lane et al. (Cambridge: Cambridge University Press, 2015), 5-43; Victor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, New York: Houghton, Mifflin, Harcourt Publishing Company, 2013.

unanticipated. This distinctive characteristic of big data, in my view, constitutes the most powerful challenge to prior constraints on collection of information.

The literature on big data and data science is filled with dazzling instances – correlating Web search trends with unanticipated drug interactions[29] and flu trends,[30] automated fraud detection, the impact of sentiment manipulation in Newsfeeds with sentiment of expressed only postings,[31] the creation of a winning Major League Baseball Team.[32] Less dramatic but arguably more powerful are reports and forecasts of successful and sustained applications in host of societal domains (even limiting our purview to applications involving data about people) from finance; public health; public safety; evidence-based medicine; national security; commerce; marketing; romantic love; employment; law; cultural creation; personalized, automated information services; recommendation and ranking, algorithmic systems; and advertising.[33] The refrain is the same: 1) we need masses of data in order to extract great and lustrous utility from it; and 2) we do not always know, therefore cannot say, in advance the specific uses to which they will be put and the particular value they will yield.

---

[29] See Ryen W. White et al., "Web-scale Pharmacovigilance: Listening to Signals from the Crowd," Journal of the American Medical Informatics Association 20, no. 3 (2013): 404-408.

[30] But which failed to be replicated; see David Lazar et al., "The Parable of Google Flu: Traps in Big Data Analysis," *Science* 343, no. 6176 (2014): 1203-1205.

[31] See Gregory S. McNeal, "Facebook Manipulated User Feeds to Create Emotional Responses," *Forbes*, June 28, 2014, accessed February 9, 2016, http://www.forbes.com/sites/gregorymcneal/2014/06/28/facebook-manipulated-user-news-feeds-to-create-emotional-contagion/#2715e4857a0b11dcc1245fd8.

[32] See Michael Lewis, *Moneyball: The Art of Winning an Unfair Game*, New York, London: W.W.Norton Company, 2004.

[33] See e.g. "IBM Watson Health," IBM Think, accessed February 9, 2016, http://www.ibm.com/smarterplanet/us/en/think/watson-health/; Rachel Willcox, "Big-Data Analytics: the Power of Prediction," *Public Finance*, January 27, 2016, accessed February 9, 2016, http://www.publicfinance.co.uk/feature/2016/01/big-data-analytics-power-prediction; Paul Wormelli, "The Promise of Big Data in Public Safety and Justice: Making Data Easier to Digest for More Law Enforcement Users," *Government Technology*, September 10, 2012, accessed February 9, 2016, http://www.govtech.com/public-safety/The-Promise-of-Big-Data-in-Public-Safety-and-Justice.html; Kalorama Information, "Evidence-based Medicine: Bringing Big Data to Healthcare Consumers," *Scientific Computing*, November 26, 2014, accessed February 9, 2016, http://www.scientificcomputing.com/news/2014/11/evidence-based-medicine-bringing-big-data-healthcare-consumers; Babak Akhgar et al., *Application of Big Data for National Security: A Practitioner's Guide to Emerging Technologies*, Oxford: Elsevier, 2015; Greg Satell, "The Future Of Marketing Combines Big Data With Human Intuition," *Forbes*, October 12, 2014, accessed February 9, 2016, http://www.forbes.com/sites/gregsatell/2014/10/12/the-future-of-marketing-combines-big-data-with-human-intuition/#2715e4857a0b7fd34974331d; Paul Rubens, "Is Big Data Dating the Key to Long-Lasting Romance?" *BBC*, March 25, 2014, accessed February 9, 2016, http://www.bbc.com/news/business-26613909; Erik Brynjolfsson and Andrew McAfee, *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*, New York, London: W.W. Norton & Company, 2014.

Proponents of big data exceptionalism conclude that collection (or flow, more broadly) should, for the most part, proceed unchallenged. The best of the arguments do not dispute that this constitutes a significant departure from existing privacy regimes. Unlike earlier "bad boy" privacy skeptics, they do not dispute the important role strong privacy regulation can and has served to protect individuals (and societies) against a host of privacy or informational harms regularly cited by advocates. They say, however, that we can no longer sustain privacy insofar as it constrains information collection, storage, and analysis; instead we must focus directly on the informational harms we know about that may result from uses of information. In other words, aware of the vulnerabilities exposed by stripping away the protective shield of privacy (as constraints on flow) these proponents shift the burden of proof to the uses of information. In the past, a privacy skeptic might have insisted on cost-benefit analyses for information collection and flow; these proponents give up on those restraints and shift the location of cost-benefit to use. They propose that the host of uses to which data and data science are being put should be subjected to cost-benefit analyses, with uses allowed only when tradeoffs are justifiable.[34] [35] Responding to this need, a new field of inquiry and practice, labeled data ethics, has emerged. I should qualify: those championing the field of data ethics do not necessarily frame the emergence of this field in relation to the data exceptionalist position.[36]

A burgeoning literature has already identified many of the ethical issues raised by uses and applications of data science. It is worth noting that many of these issues overlap considerably with those raised in the literature on privacy protection. Although a full account is obviously not possible here, I will provide an overview of some of the leading concerns. From my own observations, the most commonly discussed relate, generally, to social justice. Focusing on educational, employment, advertising, and finance, critics have explored the uses of data science to cluster and profile individuals for purposes of predictive analytics, ultimately basing decisions about individuals, critical to quality of life, based on these analyses. These critiques point to error, unfair discrimination, historical prejudice, inequitable allotment of resources and opportunities as potential and sometimes inevitable consequences of automated, algorithmic, prediction and decision making based on machine learning on big data.[37] Even the data itself cannot be assumed to be objective and impartial.[38] Whether a person is stopped at the border,

---

[34] See e.g. Tene and Polonetsky, "Big Data for All"; White House, "Big Data: Seizing Opportunities, Preserving Values".

[35] This move is reminiscent of public debate in the 1980s over computer matching of federal databases to extract useful knowledge, resulting in the 1986 Computer Matching and Privacy Protection Act, which placed restraints on the matching of disparate databases. The details do not matter; what matters is that the path taken was to set protocols in place that required an articulation of benefits while allowing stakeholders (or their representatives) to identify potential harms.

[36] See e.g. danah boyd and Kate Crawford, "Critical Questions for Big Data: Provocations for a Cultural, Technological and Scholarly Phenomenon," *Information, Communication and Society* 15, no. 5 (2012): 662-679; Neil M. Richards and Jonathan H. King, "Big Data Ethics," *Wake Forest Law Review* 49, no. 2 (Summer 2014): 393-432.

[37] See e.g. Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information*, Boston: Harvard University Press, 2015; Solon Barocas and Andrew D. Selbst, "Big Data's Disparate Impact," *California Law Review* 104 (forthcoming).

[38] See boyd and Crawford, "Critical Questions for Big Data".

offered employment, acceptance at a prestigious university, an apartment, or favorable rates for health and life and a mortgage, what prices they are charged for merchandise, what ads and offers they receive, comes down to the results of automated decision systems, which may be biased. Further, because there are no constraints on what data can be applied to which decisions, people can no longer rely on being judged only on the basis of relevant information. Any information that may affect clustering and predictive accuracy may enter into critical decision-making, affecting not only fairness but autonomy.

Closely related is a concern that you may be subject to important decisions as a consequence of processes that are utterly opaque, not only to you but even to those responsible for operating the systems. The models that emerge from statistical learning may map well onto the vast data points, they may offer statistically respectable predictions, but they may defy sense-making for the human mind. Decisions affecting your prospects and well-being, accordingly, may seem as arbitrary as the toss of a coin. With these concerns in mind, critics have raised considerations of due process[39] and urged transparency of key steps involved in automated operations, from an account of the data and algorithms used to criteria invoked in the application of findings to decisions (e.g. thresholds).

Critics concerned with autonomy have also raised questions about targeted manipulation of individuals to serve the interests of third parties. They cite the vast machinations of behavioral advertising and marketing that attempt to shape preferences and actions on the basis of personal information. They resist the presumption of data processors that their ways of identifying individuals, through data clusters and profiles, for example, can take preeminence over our own ways. They point to the chilling effect on speech and association as individuals are aware that the friends they accept online, the opinions they post, and the searches they conduct will earmark them as this or that type of person. They warn of threats to democracy based on political messages targeted at particular households or individuals. They warn of the filter bubbles engineered by recommender systems and personalized ranking algorithms.[40] They anticipate oppressive working conditions in which employees' performance and work schedules are optimized for maximum business efficiency.[41]

The question before us is this: Can we look to the work in data ethics as a sufficient mitigation for letting go of the power to restrict the collection of data? With a promise that data use will be scrutinized and stopped, if found not meeting an acceptable cost-benefit

---

[39] See Kate Crawford and Jason Schultz, "Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms," *Boston College Law Review* 55 (2014): 93-128; Danielle K. Citron and Frank Pasquale, "The Scored Society: Due Process for Automated Predictions," *Washington Law Review* 89, no. 1 (2014): 1-34.

[40] See e.g. Ira Rubinstein et al., "Data Mining and Internet Profiling: Emerging Regulatory and Technological Approaches," *University of Chicago Law Review* 75, no. 1 (2008): 261-285.

[41] See e.g. Karen Levy, "The Future of Work: What Isn't Counted Counts," *Pacific Standard*, August 3, 2015, accessed February 9, 2016, http://www.psmag.com/business-economics/the-future-of-work-what-isnt-counted-counts; Jodi Kantor, "Working Anything but 9 to 5," *New York Times*, August 13, 2014, accessed February 9, 2016, http://www.nytimes.com/interactive/2014/08/13/us/starbucks-workers-scheduling-hours.html.

ratio, are there residual problems unresolved, or doubts unattended if under these conditions we fully let go of restrictions on collection, flow, and accumulation? I am unconvinced. Despite efforts to curtail privacy harms at the point of use instead of collection, I argue that interests of data subjects may still be vulnerable to harm and societal values at risk.

6.3 The Reality Factor

There undoubtedly are new and good reasons, in light of big data and data science, for adjusting how we regulate the flow of information, the capture, aggregation, analysis and use of data about people. Accepting contextual integrity as the root of privacy means protecting appropriate information flows, purposely defined to be a moving target. There is so much more we can do with data; we ought not set these prospects against privacy. Instead, we should recalibrate the instantiation of appropriateness so that the role privacy plays in channeling data to promote interests, ethical and political values, and integrity of social contexts may be fulfilled. Yet I remain skeptical that the adjustment we need conforms to a popular position recommending a lifting of restrictions on collection while focusing them on use alone. My reasons fall into two clusters – substantively, they draw on different observations but may both be labeled as “a sprinkling of realism.”

A good dose of realism is important as we consider these lines of argument. I specifically mean realism and not pragmatism, because while pragmatic considerations which might involve saying, “yes, we know what is right and wrong but practically speaking are not able to enforce or get it done,” I want to consider right and wrong in light of assumptions reasonably made about stakeholders and other ethical and political actors in the real world and its contexts.

6.3.1 Who is We?

Recall the argument. We would be foolhardy to restrict collection because the amassing of data is a necessary first step in the operation of big data techniques, which promise individuals and societies enormous benefits. Further, because results of algorithmic learning are not knowable in advance (particularly with unsupervised learning), and because clusters discovered algorithmically do not necessarily map onto concepts that are natural or meaningful to humans, we cannot require purposes to be specified and cost-benefit analyses to be performed before data is collected and analysis is completed.

It is surprising that the logic of such rhetoric has not been more aggressively challenged, most glaringly for the shifting meaning of crucial terms, beginning with "we" in "*We* should not restrict, otherwise *we* have much to lose."[42] As noted earlier, the evidence put forth includes a handful of spectacular instances – drug interactions, Target's pregnancy revelation[43] and flu trends, the latter discredited.[44] There are other laudable efforts in the works (as of writing), such as NIH researchers turning to big data techniques to learn about HIV infection and treatment efficacy,[45] public utilities companies spurring energy conservation through smarter energy grids, about IBM devoting its AI Watson to the task of amassing health and lifestyle data to inform us about actionable correlations, and educational institutions tapping into data trails from online, mediated systems to understand and address individual learning styles, and so on.[46]

Now, let us add a sprinkle of realism. A vast portion of the data available for computational analysis is concentrated in the hands of a few private, global, commercial entities, including some, such as Google, Apple, and Facebook, with which we interface directly in their provision of services, and also indirectly in their capacities as platforms, operating systems, and intermediaries.[47] There is, of course, a class of global, commercial actors with whom individuals have virtually no direct interface and who hold vast data repositories. Other such entities functioning in the background accruing large data lakes include telecommunications providers, medical insurance companies, and financial institutions (such as banks, credit card providers, and insurance companies), by dint of roles as intermediaries and platforms, sometimes accorded to them through legislation.[48]

---

[42] One exception is Rachel Schutt and Cathy O'Neill, *Doing Data Science: Straight Talk from the Frontline* (Sebastopol: O'Reilly Media, 2014), 6.

[43] See Charles Duhigg, "How Companies Learn Your Secrets," *New York Times*, February 16, 2012, accessed January 5, 2016, http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html.

[44] See Lazar et al., "The Parable of Google Flu."

[45] This was recently discussed at the "Harnessing 'Big Data' to Stop HIV" conference, co-hosted by the NIAID Division of AIDS, NIMH Division of AIDS Research, NIH Big Data to Knowledge, and the Bill and Melinda Gates Foundation. See https://www.niaid.nih.gov/about/organization/daids/Pages/big-data.aspx, accessed February 10, 2016.

[46] See e.g. "IBM Watson Health"; Tene and Polonetsky, "Big Data for All," 248; Marie Bienkowski et al., "Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief," *U.S. Department of Education*, October 2012, accessed February 9, 2016, https://tech.ed.gov/wp-content/uploads/2014/03/edm-la-brief.pdf.

[47] See Bernard Marr, "The 7 Most Data-Rich Companies in the World," *Data Science Central*, April 18, 2015, accessed February 9, 2016, http://www.datasciencecentral.com/profiles/blogs/the-7-most-data-rich-companies-in-the-world; see also Lev Manovich, "Trending: The Promises and the Challenges of Big Social Data," April 28, 2011, accessed February 9, 2016, http://manovich.net/content/04-projects/067-trending-the-promises-and-the-challenges-of-big-social-data/64-article-2011.pdf, quoted in boyd and Crawford, "Critical Questions for Big Data," 673.

[48] See e.g. boyd and Crawford, "Critical Questions for Big Data"; Tene and Polonetsky, "Big Data for All."

I question the faith that *we* the collectors (viz. Apple, Medical Information Bureau, Facebook) are sufficiently identified with *we* the people (and our political representatives) to be relied upon to pursue knowledge and underwrite decisions that serve us as individuals or, for that matter, the common good. I impute no ill will or evil doing on the parts of these companies; to the contrary. It is merely that they are, understandably, driven by different imperatives; the data they record and the questions they ask of it are related to these imperatives. The unthinkably large trove of Web use data is optimized for effective targeted advertising; the massive accumulation of medical data accruing to medical insurance companies for assessing premiums; studies Facebook underwrites with its vast stock of networked data shaped by company interests. To be sure, individuals might benefit and societal needs served, but collaterally, not systematically.

These are best-case scenarios: legitimate, competent, largely well-meaning companies producing useful services, sometimes contributing to knowledge and underwriting decisions that happen to be important to the quality of individual lives and societal wellbeing. Interest might align and benefits flow, but nothing compels this. Utilitarian thinkers, including economists should be asking about opportunity costs; not only whether there is benefit from unregulated collection but about how much more there might be were *we* be allowed to frame the questions, with different interests in mind. Here, the "*we"* in question might be government representatives, independent academic researchers, or public interest organizations. [49] Slippage in the referents of "we," though benign, deserves to be noted because it is not always so.

There are also scenarios in which "we" the data collectors and "we" the data subjects -- affected by the data, but with little say – are not quite aligned. The machination of behavioral advertising is a case in point where claims that we are all better off when ads match our interests are disingenuous, at best. Nevertheless, such claims seem to have mollified some regulators perceiving, even supporting, any and all data collection, online and off, within and across platforms, as a contribution to the benefit of humankind.[50] Another claim of big data serving common interests is fraud spotting by credit card companies. Patterns of normal usage extracted from large data sets form the backdrop against which fraudulent use can be seen. Everyone (both "we-s") is happy (except the fraudsters). Although it is true that machine learning over vast databases is the proximal agent of success, the outcome we all enjoy, surely, has as much to do with the forced confluence of interests due in large measure to strategic regulation and industry standards[51] that assigned liability for losses due to fraud to credit card issuers and not individual consumers or merchants. This point is critical.

Consider the claims of big data's successes with risk reduction and mitigation, of which credit card fraud detection is a celebrated instance. In service of realism, it is crucial to acknowledge the role legislation played in aligning the interests of consumers and credit card companies, in hindsight even more brilliant given the birds-eye view these companies have from the aggregated transaction data pool that flows to them. But the serendipitous alignment of incentives is not guaranteed, and respective interests of the

[49] Experts seem unanimous in worrying about the egregious wealth disparities in the USA and around the world. Perhaps data disparities are at root. This seems to be an important issue worth studying

[50] See White House, "Big Data: Seizing Opportunities, Preserving Values," 40-43.

[51] See e.g. the 1974 Fair Credit Billing Act.

we-s may be torn asunder. Data breach notification laws constitute an effort to knit them together but they remain controversial and perhaps too indirect in their enforcement to ensure that the material harms and insecurities befalling individual data subjects constitute risk sufficiently great for data owners to care (enough). The free-for-all in collection continues to create honey pots for evil hackers, mobsters, and their ilk.[52]

But risk shifting, at times, can also be a zero-sum game: I reduce my risk by increasing yours. Data breaches – some quite spectacular – can constitute such instances, when, for example, a company collects and accrues data with an eye to reducing its own costs or risks while exposing data subjects to increased risk. Or data analytics may enable a company to extract a higher price -- good for a seller but bad for a buyer. Data mining and profiling can also place individuals in adversarial relationships with one another, situated within different clusters and given differential treatment. Labeled personalization for a positive spin or discrimination for a negative, benefit (or reduction of risk of bad outcomes) for one party may exaggerate cost or risk for the other, particularly when resources are limited, such as admission to a prestigious college, or discriminatory pricing. Further, probabilistic modeling means that some people inevitably will be misclassified, not a system error, but inherent to such modeling.[53] From the perspective of the individual, however, he is being wrongfully treated, definitively, often with no recourse. Depending on where thresholds are set for false negatives and positives, data processors may shift the risk of erroneous classification toward or away from themselves. The selection of data fields, too, can affect which individuals are blessed with a positive outcome and which a negative.[54] Cost-benefit analyses will not successfully root out risk shifting unless we could count on a world – not our world – where mechanisms were in place to ensure impartial access to the data and to the questions posed to the data.

Whether one believes – as I do – that the FIPs are a poor model for protecting privacy in the contemporary landscape, the concern that shaped them -- to level the playing field for data holders and data subjects – continues to sustain privacy's value.[55] Information

---

[52] For an overview of state legislation regarding security breach notifications, see "Security Breach Notification Laws," *National Conference of State Legislatures*, January 4, 2016, accessed February 9, 2016, http://www.ncsl.org/research/telecommunications-and-information-technology/security-breach-notification-laws.aspx; for two recent high-profile hacking cases see David E. Sanger et al., "Attack Gave Chinese Hackers Privileged Access to U.S. Systems," *New York Times*, June 20, 2015, accessed July 9, 2016, http://www.nytimes.com/2015/06/21/us/attack-gave-chinese-hackers-privileged-access-to-us-systems.html and Jessica Silver-Greenberg et al., "JPMorgan Chase Hacking Affects 76 Million Households," *New York Times*, October 2, 2014, accessed February 9, 2016, http://dealbook.nytimes.com/2014/10/02/jpmorgan-discovers-further-cyber-security-issues/.

[53] See e.g. Federal Trade Commission, "Big Data: A Tool for Inclusion or Exclusion: Undestanding the Issues," FTC Report, January 2016, accessed February 9, 2016, https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf.

[54] See David Robinson et al., "Civil Rights, Big Data and Our Algorithmic Future," *Upturn*, September 2014, accessed February 2, 2016, https://bigdata.fairness.io/.

[55] See U.S. Department of Health, Education and Welfare, "Report of the Secretary's Advisory Committee on Automated Personal Data Systems, Records, Computer, and

flows are strategically constrained in recognition of not only highly differentiated levels of power and wealth but also differentiated societal roles and positions. For criminals, we may not care to level the playing field but for many others, the modulated flow of information provides security for legitimate needs.

6.3.2 We the NSA

For readers who may still believe that we really do care only about use and not collection except insofar as it prevents problematic uses, it may be informative to look to the constitutional roots of our political democracy. Let us further hone in on the Fourth Amendment and, as a thought experiment, imagine critics advocating for its repeal or amendment. They say that it unnecessarily obstructs law enforcement and national security. Whether or not such arguments could have held sway in 1791, they could be quite powerful at present, first, because we have at our disposal so many more forms of surveillance – monitoring communications media and geo-location, capturing visual images with hidden cameras, digitally recording commercial transactions, etc. – and more reasons to be fearful. But, critics calm us with assurances that full attention will be given to preventing misuse (assuming agreement on what counts as misuse) and holding perpetrators to account. Requiring antecedent specification and ratification of purpose, as required by the Fourth Amendment, severely compromises efficacy in catching criminals, exposing dangerous plots, and other serious threats because it is often impossible to know ahead of time the useful patterns the data will yield.[56]

Some might flatly deny that a liberal democracy can do without some version of the Fourth Amendment as an element of due process.[57] Taking up the challenge of the thought experiment, however, defenders point to the downside of dragnets, flagging their invasive character -- universal "stop and frisk," mandatory drug testing, and house-to-house searches. They panic: the *scope* of government's access to citizens' lives *must* be contained if we are to avoid totalitarianism. Even where we allow government agencies access to *aspects of private life* for administrative purposes, such as, for Medicare reimbursements, long-form tax filings to the IRS, the Decennial Census surveys, and welfare applications, these data holdings must be rigorously silo-ed.[58] They continue that even the *feeling* of being monitored can matter as defenders point to chilling effects on activities crucial to human life and civil society, such as association and speech. The critic merely smiles indulgently. Present day dragnets may be cast with enormous discretion – hidden cameras, unobtrusive motion sensors, concealed listening devices, passive capture of signals from mobile devices, black box recorders attached to

---

the Rights of Citizens," July 1973, accessed February 9, 2016, http://www.justice.gov/opcl/docs/rec-com-rights.pdf.

[56] For a discussion of predictive policing, see Sarah Brayne et al., "Predictive Policing," *Data & Civil Rights*, October 27, 2015, accessed February 9, 2016, http://www.datacivilrights.org/pubs/2015-1027/Predictive_Policing.pdf and Cynthia Rudin, "Predictive Policing: Using Machine Learning to Detect Patterns of Crime," *WIRED*, August 2013, accessed February 9, 2016, http://www.wired.com/insights/2013/08/predictive-policing-using-machine-learning-to-detect-patterns-of-crime/.

[57] See e.g. Canadian Charter of Rights and Freedoms, Article 8: Everyone has the right to be secure against unreasonable search or seizure,

[58] The spirit of the 1974 Privacy Act.

broadband cables, and so forth. You will never know. The conclude with one final calming utterance, familiar to all, "if you have nothing to hide, you have nothing to fear."

Yet, nowhere more than in the relationship of private citizens to government is the need for leveling the playing field greater. Lawmakers and even privacy skeptics have supported constraints on information gathering in its name. But what are the reasons for fencing in this power,[59] even as we acknowledge that it may result in government actors performing their functions less effectively or less efficiently? In the field of political philosophy, one answer can be drawn from Philip Pettit's account of Republicanism.[60] According to Pettit, political liberty involves not only the prevention of repressive governmental actions; it also requires protection against government domination, that is, the *power* of government to interfere arbitrarily in our lives. Take note: the concern does not stop with arbitrary interference, but extends to the *power* of arbitrary interference.[61] We protect ourselves against government domination, namely the power to interfere, when we set in place constraints on inappropriate collection and flow of information about private individuals. (On a more mundane note, the saying "knowledge is power," is not qualified by "when you use it.")

Empowering private citizens against government domination, according to this Republican vision, involves hobbling government's *ability* to curtail freedom, not merely preventing curtailments directly. Containing this ability by restricting access to information has been achieved through principles in the Bill of Rights as well as various other acts of legislation.[62] We might still judge this to be fine, in theory, but is it necessary? Here the sprinkling of realism matters. Centuries of recorded history, featuring rulers who have exploited their subjects and governments that have oppressed their citizens, support the need to establish, guard, and maintain protective barriers. Although many of us may have faith in the integrity of today's executive branch, in the NSA, and in our local police, who arguably could make the most convincing case for risk-reduction through noninvasive, dragnet surveillance, we are informed by the realism of past exploits in resisting it.

Big data exceptionalists could weigh against generalizing this argument beyond government and, in so doing, track much privacy law in the United States. In acknowledging the power of data, whether in government or private hands, to impose serious costs as well as great benefits, however, the boosters of big data surely could be persuaded that this is a good time to expose the power in the hands of private institutional actors, such as corporations (in the past, the Church). They may not have armies at their direct disposal but they do have the ability to affect the lives of private

---

[59] Compare the concept of gezeihra as a "fence around the Torah" in Jewish law, see "Halakha: Jewish Law," *Judaism 101*, n.d., accessed February 9, 2016, http://www.jewfaq.org/halakhah.htm.

[60] Philip Pettit, *Republicanism: A Theory of Freedom and Government*, Oxford: Oxford University Press, 1997, cited in Finn Brunton and Helen Nissenbaum, *Obfuscation: A User's Guide for Privacy and Protest* (Cambridge: MIT Press, 2015), 79-80.

[61] Finn Brunton and I have argued that information yields power to the holder, particularly dangerous when that holder is already powerful, see Brunton and Nissenbaum, *Obfuscation*.

[62] E.g. the 1974 Privacy Act.

individuals in critical ways.[63] Further, whereas governments are mostly limited in their exercise of direct power to those within their nation states, present day information companies have global reach – the envy of any single government. Without claiming that nongovernmental actors have the same breadth and level of powers as governmental actors, there is a concentration of power through concentration of money and information that can affect the achievement of a decent life – shelter, employment, nourishment, family, friends, health, education, and security. Big data boosterism celebrates its power to rise above theory, to forecast without explanation. To those who are stymied in these vital spheres of life, the impacts might as well be arbitrary. They may never be able to prove unfairness or manipulation; that, by any name, is domination. If we have determined that the regulation of use is not a sufficient solution for government, we may have good reasons for apply the same in domains beyond.

7 Recommendations

(TK)

8 Summation

This paper has taken on the challenge of big data exceptionalists, who question the value of privacy, understood as constraints imposed on data collection, accrual, and flow. They view such constraints as anachronistic, even unethical. Instead, the big data worldview should be embraced, subject to the constraints of principles of data ethics focusing on use. I have teased out and challenged some of the weaker forms of this positions, such as, "it's just too difficult to monitor, prohibit, or regulate collection!" and have dwelt mainly on the assertion that we dare not (not that we cannot) restrict collection: the methods of data science require lots of data and it is impossible to know, in advance, what data will produce what knowledge. In order to maintain an acceptable cost-benefit ratio, therefore, knowledge should be allowed to flourish while restrictions should apply to use alone.[64]

Although it is unclear whether grandiose forecasts of solutions to the world's direst problems of economy, health, and security will be fulfilled, we have witnessed successful applications in a range of areas to be optimistic. On these grounds, boosters urge us to set aside restrictions on data collection and accrual. I disagree: the realities of our world and political societies do not support their vision, at best naïve, often muddled and hypocritical, but also potentially dangerous. Why?

There is an enormous concentration of data in the hands of a few. These holders are hoarders of data and, to day, there is no thought to the provision of access by those who may seek to use it in socially progressive ways. Silos are maintained, in part through technology and presumably also through enforcement of property rights or contracts. Unless there is a radical revision in our regulation of the large information companies, we cannot support the illusion that the questions they will ask of their data, and problems they will seek to solve, will serve the common good. By the same token, risks identified and mitigated will serve us all only if interests of data powers align with the public

[63] See Philip N. Howard, *Pax Technica: How the Internet of Things May Set Us Free Or Lock Us Up*, New Haven: Yale University Press, 2015.

[64] Inspired by this argument, there has been growth in research focusing on issues of fairness, due process for decision-making, and so forth.

interest. It will take ingenuity and concerted effort from lawmakers to ensure, generally, what they achieved with credit card fraud liability.

The position comprises two interdependent halves: lifting restrictions on collection and imposing restrictions on use. I have argued that merely lifting restrictions will change none of the other conditions limiting fulfillment of the optimistic promise. On the question of use restrictions, past failures of regulators (or their representatives) to gain access to data holdings to audit owners give little hope in the ability to uncover and regulate harmful uses.[65] Big data exceptionalism is a hopeful idea, but not yet for this world.

Bibliography:

Akhgar, Babak, Saathoff, Gregor, Arabnia, Hamid R., Hill, Richard, Staniforth, Andrew, and Petra Saskia Bayerl, *Application of Big Data for National Security: A Practitioner's Guide to Emerging Technologies*, Oxford: Elsevier, 2015.

Barocas, Solon and Helen Nissenbaum. "Big Data's End Run Around Anonymity and Consent." In *Privacy, Big Data and the Public Good: Frameworks for Engagement*. Edited by Julia Lane, Victoria Stodden, Stefan Bender, and Helen Nissenbaum, 44-75. Cambridge: Cambridge University Press, 2015.

Barocas, Solon and Andrew D. Selbst. "Big Data's Disparate Impact." *California Law Review* 104 (forthcoming).

Bienkowski, Marie, Feng, Minyu, and Barbara Means. "Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief." *U.S. Department of Education*, October 2012. Accessed February 9, 2016. https://tech.ed.gov/wp-content/uploads/2014/03/edm-la-brief.pdf.

boyd, danah and Kate Crawford. "Critical Questions for Big Data: Provocations for a Cultural, Technological and Scholarly Phenomenon." *Information, Communication and Society* 15, no. 5 (2012): 662-679.

Brayne, Sarah, Rosenblat, Alex, and danah boyd. "Predictive Policing." *Data & Civil Rights*, October 27, 2015. Accessed February 9, 2016. http://www.datacivilrights.org/pubs/2015-1027/Predictive_Policing.pdf.

Brin, David. *Transparent Society: Will Technology Force Us to Choose Between Privacy and Freedom?* Cambridge: Perseus Books, 1998.

Brunton, Finn and Helen Nissenbaum. *Obfuscation: A User's Guide for Privacy and Protest*. Cambridge: MIT Press, 2015.

Brynjolfsson, Erik and Andrew McAfee. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. New York, London: W.W. Norton & Company, 2014.

Calo, Ryan "Against Notice Skepticism in Privacy (and Elsewhere)." *Notre Dame Law Review* 87, no. 3 (2013): 1027-1072.

Cate, Fred H. and Victor Mayer-Schönberger. "Notice and Consent in a World of Big Data." *International Data Privacy Law* 3, no. 2 (2013), 67-73.

[65] For a non-big data example of this problem, see David E. Sanger, "Prospect of Self-Inspections by Iran Feeds Opposition to Nuclear Deal," *New York Times*, August 21, 2015, accessed February 9, 2016, http://www.nytimes.com/2015/08/22/world/middleeast/prospect-of-self-inspections-by-iran-feeds-opposition-to-nuclear-deal.html.

Cate, Fred H. “The Failure of Fair Information Practice Principles.” in *Consumer Protection in the Age of the Information Economy*, 341-378. Edited by Jane K. Winn. Hampshire and Burlington: Ashgate Publishing.

Cate, Fred H., Cullen, Peter, and Victor Mayer-Schönberger. "Data Protection Principles for the 21st Century: Revising the 1980 OECD Guidelines." *Oxford Internet Institute*, March 2014. Accessed December 26, 2015. http://www.oii.ox.ac.uk/publications/Data_Protection_Principles_for_the_21st_Century.pdf.

Citron, Danielle K. and Frank Pasquale. “The Scored Society: Due Process for Automated Predictions.” *Washington Law Review* 89, no. 1 (2014): 1-34.

Crawford, Kate and Jason Schultz. “Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms.” *Boston College Law Review* 55 (2014): 93-128.

Cukier, Kenneth Neil and Viktor Mayer-Schönberger. “The Rise of Big Data: How It’s Changing the Way We Think About the World.” *Foreign Affairs* 92, no. 3 (May/June 2013): 28-40.

Duhigg, Charles. “How Companies Learn Your Secrets.” *New York Times*, February 16, 2012. Accessed January 5, 2016. http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html.

Federal Trade Commission. “Big Data: A Tool for Inclusion or Exclusion: Undestanding the Issues.” *FTC Report*, January 2016. Accessed February 9, 2016. https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf.

Feyerabend, Paul. *Against Method*. 4th edition. New Ýork: Verso, 2010.

“Halakha: Jewish Law.” *Judaism 101*, n.d. Accessed February 9, 2016. http://www.jewfaq.org/halakhah.htm.

Heller, Christian *Post-Privacy: Prima leben ohne Privatsphäre* (*Post-Privacy: Living just Fine Without Privacy*), Munich: C.H. Beck, 2011.

Howard, Philip N. *Pax Technica: How the Internet of Things May Set Us Free Or Lock Us Up*. New Haven: Yale University Press, 2015.

“IBM Watson Health.” *IBM Think*. Accessed February 9, 2016. http://www.ibm.com/smarterplanet/us/en/think/watson-health/.

Kalorama Information. “Evidence-based Medicine: Bringing Big Data to Healthcare Consumers.” *Scientific Computing*, November 26, 2014. Accessed February 9, 2016, http://www.scientificcomputing.com/news/2014/11/evidence-based-medicine-bringing-big-data-healthcare-consumers.

Kantor, Jodi. “Working Anything but 9 to 5.” *New York Times*, August 13, 2014. Accessed February 9, 2016. http://www.nytimes.com/interactive/2014/08/13/us/starbucks-workers-scheduling-hours.html.

Koops, Bert-Jaap “On Decision Transparency, or How to Enhance Privacy after the Computational Turn,” in *Privacy, Due Process and the Computational Turn: The Philosophy of Law Meets the Philosophy of Technology*, 196-220. Edited by Mireille Hildebrandt & Katja De Vries. New York: Routledge, 2013.

Kuhn, Thomas S. *The Structure of Scientific Revolutions*. 3rd edition. Chicago and London: The University of Chicago Press, 1996.

Lazar, David, Kennedy, Ryan, King, Gary, and Alessandro Vespignani. “The Parable of Google Flu: Traps in Big Data Analysis.” *Science* 343, no. 6176 (2014): 1203-1205.

Levy, Karen. “The Future of Work: What Isn’t Counted Counts.” *Pacific Standard*, August 3, 2015. Accessed February 9, 2016. http://www.psmag.com/business-economics/the-future-of-work-what-isnt-counted-counts.

Lewis, Michael. *Moneyball: The Art of Winning an Unfair Game*. New York, London: W.W.Norton Company, 2004.

Manovich, Lev. “Trending: The Promises and the Challenges of Big Social Data,” *Manovich*, April 28, 2011. Accessed February 9, 2016. http://manovich.net/content/04-projects/067-trending-the-promises-and-the-challenges-of-big-social-data/64-article-2011.pdf,

Marr, Bernard. “The 7 Most Data-Rich Companies in the World.” *Data Science Central*, April 18, 2015. Accessed February 9, 2016. http://www.datasciencecentral.com/profiles/blogs/the-7-most-data-rich-companies-in-the-world.

Mayer-Schönberger, Victor and Kenneth Cukier. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. New York: Houghton, Mifflin, Harcourt Publishing Company, 2013.

McNeal, Gregory S. “Facebook Manipulated User Feeds to Create Emotional Responses.” *Forbes*, June 28, 2014. Accessed February 9, 2016. http://www.forbes.com/sites/gregorymcneal/2014/06/28/facebook-manipulated-user-news-feeds-to-create-emotional-contagion/#2715e4857a0b11dcc1245fd8.

Mundie, Craig "Privacy Pragmatism: Focus on Data Use, Not Data Collection." *Foreign Affairs*, March/April 2014. Accessed December 26, 2015. https://www.foreignaffairs.com/articles/2014-02-12/privacy-pragmatism.

Nissenbaum, Helen. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Palo Alto: Stanford University Press, 2010.

Nissenbaum, Helen. "A Contextual Approach to Privacy Online," *Daedalus* 140, no. 4 (Fall 2011): 32-48.

Nissenbaum, Helen. "Respect for Context as a Benchmark for Privacy Online: What it is and isn't." In *Social Dimensions of Privacy: Interdisciplinary Perspectivs*, 278-302. Edited by Beate Roessler and Dorota Mokrosinska. Cambridge: Cambridge University Press, 2015.

Pasquale, Frank. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Boston: Harvard University Press, 2015.

PCAST, “Big Data and Privacy: A Technological Perspective,” *White House*, May 2014, accesed February 9, 2016, https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf.

Pettit, Philip. *Republicanism: A Theory of Freedom and Government*. Oxford: Oxford University Press, 1997.

Reidenberg, Joel R. Russell, N. Cameron, Callen, Alexander J., Qasir, Sophia, and Thomas B. Norton. “Privacy Harms and the Effectiveness of the Notice and Choice Framework.” Fordham Law Legal Studies Research Paper No. 2418247. Available at SSRN, March 29, 2014. Accessed February 9, 2016 http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2418247.

Reidenberg, Joel R. and Lorrie Faith Cranor, “Can User Agents Accurately Represent Privacy Policies?” Discussion Draft 1.0 available at SSRN, August 30, 2002. Accessed February 9, 2016, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=328860.

Richards, Neil M. and Jonathan H. King. “Big Data Ethics.” *Wake Forest Law Review* 49, no. 2 (Summer 2014): 393-432.

Robinson, David, Yu, Harlan and Aaron Rieke. “Civil Rights, Big Data and Our Algorithmic Future.” *Upturn*, September 2014. Accessed February 2, 2016. https://bigdata.fairness.io/.

Rubens, Paul. “Is Big Data Dating the Key to Long-Lasting Romance?” *BBC*, March 25, 2014. Accessed February 9, 2016. http://www.bbc.com/news/business-26613909.

Rubinstein, Ira. “Big Data: The End of Privacy or a New Beginning?” *International Data Privacy Law* 3, no. 2 (2013): 74-87.

Rubinstein, Ira, Lee, Ronald D., and Paul M. Schwarz, “Data Mining and Internet Profiling: Emerging Regulatory and Technological Approaches,” *University of Chicago Law Review* 75, no. 1 (2008): 261-285.

Rudin, Cynthia. “Predictive Policing: Using Machine Learning to Detect Patterns of Crime.” *WIRED*, August 2013. Accessed February 9, 2016. http://www.wired.com/insights/2013/08/predictive-policing-using-machine-learning-to-detect-patterns-of-crime/.

Sanger, David E. “Prospect of Self-Inspections by Iran Feeds Opposition to Nuclear Deal.” *New York Times*, August 21, 2015. Accessed February 9, 2016. http://www.nytimes.com/2015/08/22/world/middleeast/prospect-of-self-inspections-by-iran-feeds-opposition-to-nuclear-deal.html.

Sanger, David E., Perlroth, Nicole, and Michael D. Shear. “Attack Gave Chinese Hackers Privileged Access to U.S. Systems.” *New York Times*, June 20, 2015. Accessed July 9, 2016. http://www.nytimes.com/2015/06/21/us/attack-gave-chinese-hackers-privileged-access-to-us-systems.html.

Satell, Greg. “The Future Of Marketing Combines Big Data With Human Intuition.” *Forbes*, October 12, 2014. Accessed February 9, 2016. http://www.forbes.com/sites/gregsatell/2014/10/12/the-future-of-marketing-combines-big-data-with-human-intuition/#2715e4857a0b7fd34974331d.

Schutt, Rachel and Cathy O’Neill. *Doing Data Science: Straight Talk from the Frontline*. Sebastopol: O’Reilly Media, 2014.

“Security Breach Notification Laws.” *National Conference of State Legislatures*, January 4, 2016. Accessed February 9, 2016. http://www.ncsl.org/research/telecommunications-and-information-technology/security-breach-notification-laws.aspx.

Seemann, Michael. *Digital Tailspin: Ten Rules for the Internet After Snowden*, Network Notebooks 09. Amsterdam: Institute of Network Cultures, 2015.

Silver-Greenberg, Jessica, Goldstein, Matthew and Nicole Perlroth. “JPMorgan Chase Hacking Affects 76 Million Households.” *New York Times*, October 2, 2014. Accessed February 9, 2016. http://dealbook.nytimes.com/2014/10/02/jpmorgan-discovers-further-cyber-security-issues/.

Sprenger, Polly. “Sun on Privacy: ‘Get Over It’,” *WIRED*, January 26, 1999. Accessed February 9, 2016. http://archive.wired.com/politics/law/news/1999/01/17538.

Strandburg, Katherine J. “Monitoring, Datafication, and Consent: Legal Approaches to Privacy in the Big Data Context.” In *Privacy, Big Data and the Public Good: Frameworks for Engagement*, 5-43. Edited by Julia Lane, Victoria Stodden, Stefan Bender, and Helen Nissenbaum. Cambridge: Cambridge University Press, 2015.

Tene Omer and Jules Polonetsky. “Big Data for All: Privacy and User Control in the Age of Analytics.” *Northwestern Journal of Technology and Intellectual Property* 11, no. 5 (2013): 239-273.

Toubiana, Vincent, Narayanan, Arvind, Boneh, Dan, Nissenbaum, Helen, and Solon Barocas. "Adnostic: Privacy Preserving Targeted Advertising." *Proceedings Network and Distributed System Symposium*, March 2010.

Turow, Joseph. *The Daily You: How the New Advertising Industry is Defining Your Identity and Your Worth*. New Haven: Yale University Press, 2011.

U.S. Department of Health, Education and Welfare. "Report of the Secretary's Advisory Committee on Automated Personal Data Systems, Records, Computer, and the Rights of Citizens." July 1973. Accessed February 9, 2016. http://www.justice.gov/opcl/docs/rec-com-rights.pdf.

Volokh, Eugene. "Freedom of Speech, Information Privacy, and the Troubling Implications of a Right to Stop People from Speaking About You." *Stanford Law Review* 52, no. 5 (2000): 1049-1124.

Warren, Samuel D. and Louis D. Brandeis, "The Right to Privacy," *Harvard Law Review* 4, no. 5 (1890): 193-220.

White House. "Big Data: Seizing Opportunities, Preserving Values." *Executive Office of the President* (May 2014): 1-79.

White Ryen W.,Tatonetti, Nicholas P., Shah, Nigam H., Altman, Russ B., and Eric Horvitz. "Web-scale Pharmacovigilance: Listening to Signals from the Crowd." *Journal of the American Medical Informatics Association* 20, no. 3 (2013): 404-408.

Willcox, Rachel. "Big-Data Analytics: the Power of Prediction." *Public Finance*, January 27, 2016. Accessed February 9, 2016, http://www.publicfinance.co.uk/feature/2016/01/big-data-analytics-power-prediction.

Wormelli, Paul. "The Promise of Big Data in Public Safety and Justice: Making Data Easier to Digest for More Law Enforcement Users." *Government Technology*, September 10, 2012. Accessed February 9, 2016. http://www.govtech.com/public-safety/The-Promise-of-Big-Data-in-Public-Safety-and-Justice.html.

Yakowitz, Jane. "Tragedy of the Data Commons." *Harvard Journal of Law and Technology* 25, no. 1 (Fall 2011): 1-67.