To appear in *Digital Media and Democracy*, Ed. M. Delli Carpini, Philadelphia: University of Pennsylvania Press (in press).

CHAPTER 10

Must Privacy Give Way to Use Regulation?

Helen Nissenbaum

Prologue

In "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. It was not that consent and anonymity no longer performed any useful function, but no longer could we count on them for the critical functions they had previously performed—consent as privacy's gatekeeper, anonymity as a boundary for privacy's remit. Although we warned against confusing the *means* of protecting privacy, namely, consent and anonymity, with privacy *itself*, understood as appropriate flow, we realized that our article could lend force to a position steadily gathering momentum in the academy, information industries, and public policy. The position is that since privacy, insofar as it restricts information collection, is untenable, attention should focus instead on how information is used. The present chapter dissects this position—what it means and whether its worldview is inevitable in light of data science—and ultimately finds it flawed.

Privacy Skeptics

In January 1999 Scott McNealy, then CEO of Sun Microsystems, brashly threw down the gauntlet, saying, "You have zero privacy anyway. Get over

it!"[2] Repeated countless times since then, the statement hardly bore serious consideration, partly because the conception of privacy McNealy presumed was muddled and partly because while threatened, privacy is far from dead, and continues to inspire defenders. Nevertheless, there was no denying the popular appeal of this "bad-boy" stance, which has resurfaced in various guises and versions. David Brin's popular book *Transparent Society* (1998),[3] another instance, asserts that privacy, in light of technological advancement, is no longer feasible, and is also no longer desirable. Instead, he supports total transparency, arguing that this would advance the cause of weaker parties, those captured in the webs of surveillance. With transparency, the weaker can turn the tables on the stronger by holding them accountable for their actions. Big data and data science has yielded its own bad-boy stance: forget about restricting information collection; focus on restricting its uses instead.[4]

I would have liked to dismiss these pronouncements either as fringe provocations or as venal ploys of the information sector, including obvious beneficiaries such as Google, Facebook, Amazon, and Twitter, and less publicly visible actors such as Acxiom, IMS Health, and LexisNexis. Other commercial actors, though not information product providers, telecommunications companies, financial companies, insurance companies, media and publishing companies, and, increasingly, retail merchants,[5] also stand to benefit from reduced constraints on collection. Outside the commercial realm, too, many actors eagerly collect, record, and hold on to data without restraint, including governmental agencies, utilities companies, healthcare organizations, educational institutions, and a range of not-for-profit public interest organizations. Unlike previous bad-boy stances, the contemporary position has captured mainstream interest. A technological infrastructure designed to capture data, the imperative of data-driven institutional bureaucracies, and a "horses out the barn" stance all point to the futility of resistance.

This chapter argues that the push to deregulate collection is problematic and possibly even dangerous. Before establishing this conclusion, however, the first step is to expose deep conceptual ambiguities in the position statement and to establish terminological consistency.

Introducing Big Data Exceptionalism (BDE)

It is a shame one cannot rest an argument on anecdotal observations, because it would then be possible to refer to the countless conference panels and pre-

sentations at which speakers, with a wave of a hand, relegate collection restrictions to the zone of the impossible and characterize privacy as hopelessly passé. The term I have coined for the claim that the regulation of collection—no longer tenable in light of big data—must be ceded to the regulation of use is *big data exceptionalism*, or BDE. More a convenient label than a precise defnition, BDE refers to a class of generally similar claims, further explicated throughout the rest of the chapter. Although written accounts are less numerous than the anecdotal, those that exist provide a window into the position and its variations.

One such account can be found in the report of the President's Council of Advisors on Science and Technology (PCAST), *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 specifc 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 specifcity 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 identifable 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]

In another account, Michael Seemann offers a different rationale: "So instead of trying to defend privacy against surveillance, we should be fghting 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 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 toward 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] He continues, "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 be to fence in the power of platforms."[10]

One of the clearest expressions is found in an essay by Craig Mundie, senior advisor to the CEO and former chief research and strategy officer of Microsoft:

> 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, 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?," 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 a series of international, regional discussions about privacy and 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 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 should be considered when analyzing the associated costs and benefits.

Finally, in "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 pillar of privacy regulation that restricts collection and retention of data based on expressed 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 thread running through each of these statements seems to be that regulatory effort should attend to data use rather than data collection, in light of big data. What they *really* mean, however, is impossible to establish before unraveling terminological ambiguities and incompatible supporting arguments.

Ambiguities: Privacy

One source of ambiguity is the conception of privacy underlying different accounts of BDE. In Michael Seemann's "big data world order," for instance, privacy, taken to mean the suppression of data, does not empower individuals but entrenches the powers of overbearing government and commercial actors. Pitting big data's benefts against privacy, Seemann implies that we can have big data or privacy but not both; we must have big data, ergo, no privacy. In contrast, other proponents of BDE see no direct conflict with privacy. They seem ready to say, "If we really want to protect privacy, we must protect against harmful uses of information (not against collection)." These accounts do not join the chorus—either privacy or big data, but not both—but urge changes in how we think about privacy protection in light of big data.

These proponents of big data exceptionalism often identify privacy protection with compliance with Fair Information Practices (FIPs) in the influential OECD Privacy Principles, including Collection Limitation, Data Quality, Purpose Specifcation, Use Limitation, Safeguards, Openness, Individual Participation, and Accountability. Like international experts, such as Bert-Jaap Koops, who are troubled by the incongruity of big data practices with traditional FIPs principles,[15] some have suggested revised formulations of FIPs. These revised formulations would qualify or relax one or more of the principles, such as those requiring advance specifcation of purpose and fne-grained informed consent, for any departures from specifed purpose. Relaxing traditional formulations so as not to obstruct machinations of big data, they suggest flling the gaps with strategic cost-benefit analyses to justify noncompliant uses. In sections that follow, I challenge this approach.

In my own view, when considering the triad—privacy, FIPs, big data—I will say only that "something's gotta give," and that "something" is FIPs. As an alternative, the theory of contextual integrity (CI) offers a less brittle con-

ception of privacy in the face of challenges from big data.[16] According to it, privacy is about the appropriate flow of personal information, not, as other theories assert, control or secrecy. Flow is appropriate when it complies with expectations, that is, with social, informational norms specifc to contexts (e.g., education, healthcare, political citizenship, home life, etc.). Contextual informational norms (also called privacy norms) prescribe flows of certain types of information from senders to recipients, about data subjects (acting in context-defned capacities), under certain constraints, called "transmission principles." Thus, when a patient shares health information with his or her physician *in confidence*, the transmission principle is confdentiality, and when police seek incriminating evidence in a suspect's home *with a warrant*, with-a-warrant is the transmission principle. Whereas FIPs-based accounts typically hold informed choice to be necessary and sufficient for privacy (except, arguably, in the few areas covered by statutory protections), CI considers it to be merely one among countless transmission principles.[17] Unlike conceptions of privacy grounded in FIPs, CI need not always require ex ante consent from the data subject, but instead may impose substantive constraints on illegitimate information flows.

Contextual integrity is upheld when information practices comply with informational norms. Norm transgressions are not necessarily condemned but instead are flagged for further analysis. This applies to countless disruptions resulting from deployments of computational and digital systems that shift what information is disseminated to which recipients under what constraints. A presumption favoring norm compliance inevitably triggers questions about why entrenched practices deserve to be favored in this way, whether through law or any other regulatory modality.[18] To this, CI answers that entrenched practices are likely to reflect a settled accommodation of interests and unlikely to infringe on conspicuously ethical and political values (e.g., autonomy, fairness, social justice, security), and may be well calibrated with contextual ends and purposes. If, however, disruptive flows improve on entrenched flows in these ways, they may legitimately replace them.

Taking *privacy* to mean contextual integrity, *collection* refers to the class of flows of information emanating from data subjects into the hands of collecting agents, or *recipients*. Ascertaining whether collection respects privacy means assessing its compliance with preexisting informational norms, or, where not, evaluating its impact on relevant interests, ethical and political values, and contextual purposes and values. As such, privacy as contextual integrity rejects BDE's blanket assertion that this particular class of flows does

not warrant regulation, irrespective of the values of parameters—subject, sender, recipient, information type, and transmission principle. In return, BDE proponents reject CI's assertion that collection itself can be assessed as acceptable or unacceptable, holding that carefully regulating data usage is sufficient. Whether the BDE challenge holds up to scrutiny is the crux of this chapter; the first obstacle to characterizing the substantive meaning of BDE, however, is locating a coherent conceptual line between collection and use. As shown below, this goal is virtually unattainable.

What Is Collection? What Is Use?

A clear understanding of use and collection is critical; otherwise, the fundamental thesis—that only use and not collection should be regulated—eludes comprehension, let alone evaluation. Despite confidently urging differential treatment, proponents have remained silent on the precise meanings of these central concepts, and we are left to surmise them from common usage and intuition.

You may be thinking that not much rides on sharp lines. After all, countless distinctions drawn with natural language concepts have supported ethical, political, and practical deliberations, despite their fuzzy borders. These cases work, it seems, because typical instances are clear, while rare or exceptional cases falling at the border need not undercut the utility of the distinctions. Theoretical jargon may offer greater precision, but rich concepts drawn from natural language can spike imagination and have a broader appeal. It may appear that the case of collection and use fits this model because intuition is strong in certain instances, such as online merchants *collecting* information from consumers when they complete online order forms and *using* information about consumers' purchases when recommending items of interest to them or delivering goods to consumers' addresses. Unlike other politically sensitive distinctions, however, the fuzzy boundary between use and collection ensnares not just a mere handful of exceptional and rare cases, which may be handled on a case-by-case, ad hoc basis.

Collection and use may once have fit the model of other politically useful if fuzzy dichotomies with "use" implying consequential action, causation, and agency and "collection" implying passivity, reaction, and a mere garnering of material lying about. The cultural implication of collection is one of innocuous, often beneficial and legitimate "cleaning up" (e.g., garbage collec-

tion, church collection). Use tends to be ambiguous because it could imply the harmful as readily as the benefcial. As such, collection can be left alone while use is deserving of guidance and oversight.[19] With the emergence of big data technologies, the borderline has become less defned, now capturing a broad swath of vexing challenges, including practices that mark radical departures from the familiar.

The story of big data told by enthusiastic academics, policy makers, activists, and pundits centers on data science and technology and their unprecedented deployment.[20] They point to the confluence of mathematical discovery and computational insights with feats of engineering, the existence of global digital networks that connect fxed and mobile devices, and a layered software infrastructure with prodigious capacities to collect, amass, store, and distribute data. Sheer bulk is but one factor; another is the diversity of input and data capture modalities. Increasingly sensitive and sophisticated sensing apparatus renders digital what its sensors detect and capture, thereby making it available for quantitative and statistical analysis and networked distribution. These novel data sources supply existing stockpiles drawn from traditional data sources as well as from Internet mediated activity, mobile phones, wearables, self-tracking devices (the so-called Internet of Things)—generally, the ephemera of everyday life, which, through device and platform conduits, are permanently imprinted as data. Dimensions of experience and expression, affect, sound, text, image, video, type and strength of relationships (social network graphs), biological characteristics ("biometrics"), even "brain waves" indicative of thought patterns and sensations, to name a few things that are grist for the data mill—are *datafied*, a term invented to label the transformation of phenomena into collectable and usable data.[21]

Techniques for creating databases, preparing data, extracting knowledge and utility from data (even unstructured data), seeing complex patterns, and rendering models through statistical analysis and machine learning are coalescing in the disciplinary feld of data science. Its quest to make sense of data, to draw insight and meaning, to produce new data from data already at hand, reveals a codependency between data analytics and data accrual. This point, which may be obvious to experts, is revelatory for nonexperts (such as myself) and worth emphasizing. Enhanced capacities to create, capture, and collect data mean not only that we have more data with which to work, but also that accrual can transform insignifcant data, already at-hand, into data that counts, that informs and provides insight. With increased scale, density,

and diversity, previously sparse data on, say, rare diseases, meet statistically signifcant thresholds. The click or motion of a mouse takes on meaning when set against other data felds or combined with clicks and motions from dense population data sets. Against the noisy backdrop of normal communication, we may discern a dangerous plot; against learning patterns of whole populations, we may identify those more and less efficacious. In short, extracting meaning from data is not additive; or, according to a well-worn saying, accrual makes the whole more than the sum of parts.

Even if no one can single out a radical discontinuity, quantum leap, or scientifc revolution,[22] the convergence of factors seems to have spawned an epistemological paradigm shift that construes data as knowledge itself.[23] Whether the data supports our hunches, hypotheses, and theories or surprises and vexes us with results that are counterintuitive and unanticipated, it has, for many, become the primary reference point of knowledge.[24]

In the sections that follow, we see how the paradigm of big data expands the fuzzy boundary between use and collection, yielding troubling ambiguity in the very meaning of BDE.

Minimalist, Maximalist, and In Between

A minimalist notion of collection covers only the initial capture by a collecting agent (e.g., data processor, data controller, data recipient, frst party, etc.) of data[25] as it leaves its impression on a given medium; any processing beyond the moment of uptake counts as use. The simplicity of this defnition is misleading, however, and not only because of conceptual perils lurking within it. Rather, it is undone when considering how to apply it, presumably, to quite clear cases, such as information registered in an online form, mouse-click records, images captured on surveillance cameras, and even the digital exhaust (sometimes called "metadata") such as temporal data, geolocation, and more, to which Craig Mundie refers.

Recent critical scholarship on big data[26] rightly resists the idea of data as a raw resource lying about awaiting collection. Unlike a raw resource, data does not preexist our collection of it, a shapeless thing until hewn into something useful for humanity.[27] Instead, the act of collecting, whether registering, logging, recording, acquiring, observing, sensing, or documenting, involves more. A foot leaves an impression in damp beach sand, but we would hardly

say that the sand "collects" the footprint. Wishing to dispel this myth of data being merely collected, critical scholars have sought to replace the notion of data gathered as a raw resource with data *constructed* or *created* from the signals of countless technical devices and systems (cameras, sensors, receivers, servers, networks, etc.). For the impression in sand to be interesting and worth fghting for, it must be conceived *as* a footprint, perhaps even a human footprint, a male footprint; and in so conceiving, the data is created. Similarly, data is interesting and worth fghting for once it has acquired meaning, whether this meaning comes from context of collection, assignment of labels, categorization, interpretation, scientifc discovery (e.g., geospatial coordinates), or merely how it is conceived in natural language.[28]

It is plain to see, in the cases mentioned earlier, differences between mere impressions and data: a mouse click is a digital pulse, but as data it could be the placement of a shoe order, an acknowledgment of terms of service, a place marker in a document, or expressed interest in a particular online ad. Similarly, a pattern of pixels is an image of a suspect on video surveillance footage; a word or phrase is a web search term or an element in a letter, email, or novel; a timed series of location coordinates recorded by a GPS device is the route driven from home to work. To fully capture the layered complexity of data interpretation and classifcation, a single paragraph is clearly insufficient. Nevertheless, to sustain momentum with the central argument, we must leave this task undone, merely acknowledging the considerable interpretive labor that goes into assigning meaning to pixels, clusters of pixels, and their assemblage before fnally, for example, identifying it as the image of a human face—and even more so when this image is recognized as the face of a particular, identifed individual.

If a minimalist defnition of collection assumes no more than that data collected is interpreted or labeled, questions remain about how to characterize the holding of data—that is, storing it, either in the long or short term—and about processing that may be necessary to make this happen. If storage is classifed not as a type of collection but a type of use, it will be snagged in a net of regulation. Although the collection minimalist might be willing to bite the bullet, there surely are BDE proponents who would balk at liberating collected data from regulation only to see these freedoms dissipate as they are granted only to fleeting ephemera with no staying power. They would prefer to see data storage remain within the realms of unregulated collection, (obviously) along with practices required for the creation and management

of databases (relational or nonrelational). Moreover, as anyone with overstuffed, untidy closets knows, it is no use storing data unless it is organized, tagged, labeled, structured, and available for systematic querying. Finally, who is to say where one database ends and another begins? With storage included within the scope of collection, it seems inevitable that concatenation of databases, whether real or virtual (as the ability to query across distributed databases), would come along for the ride. This inclusive conception of collection is compatible with the defnition proffered by Joris van Hoboken. In his insightful comparative analysis of U.S. and EU perspectives on the collection-use distinction, van Hoboken suggests that proponents of collection deregulation mean to cover all activities that provide "access to or control over (personal) data for any potential use."[29]

Beyond data storage, two further access practices bear consideration. One is analysis; the other, flow. Although the former, "data analytics," may appear to be a new phenomenon, in fact, it is a descendent of a 1980s practice known as "computer matching"—the ability to triangulate records across multiple databases.[30] Almost quaint when compared with the sophisticated present-day techniques of data mining, predictive analytics, and machine learning conducted over vast repositories of aggregated data and distributed databases, computer matching was sufficiently worrying that it provoked the passage of the 1988 Computer Matching and Privacy Protection Act. Setting aside ethical and political questions about matching and analytics, we mark as an open question whether inferring new information from data previously collected counts as collection or use. Here, too, the answer—not obvious—would have ramifcations for the scope and power of BDE.

Information *flow*, or dissemination, raises similar questions. Beyond the dyad, collector, and subject, flow involves other parties—intermediaries—beyond the initial collecting agent. It is no secret that data flows prodigiously; it is disclosed, distributed, shared, and disseminated. It flows under a host of transmission principles—sold, bought, freely given, exchanged, required by law, or, one supposes, even stolen. The information landscape is teaming with parties to this flow—frst, second, third parties and so on—including information service providers, ad networks, analytics companies, public health authorities, insurance companies, data brokers, government agencies, and many more. Should flow of data from one party to another be understood as collection or use? Although a collection *minimalist* would surely say flow is use, there is a coherent, contrasting collection *maximalism* that would

incorporate flow into collection, as well as all the other steps discussed above.

Drawing the Line

Drawing a line between collection and use is important because it defnes the scope of BDE. Although some fuzziness at the border is inevitable, the distance between intuitively clear cases of collection on the one hand and use on the other leaves BDE indeterminate in a broad spectrum of contemporary data practices, including many of the most important and controversial. As noted previously, these range from initial impression (or uptake), creation (or interpretation, conceptualization), assembly, storage in databases (or repositories), structuring (or organization), indexing, and query access, to analysis (or so-called data analytics) and flow, to parties beyond the frst. I characterized collection minimalism as a position that categorized only the frst two as collection, and collection maximalism as covering them all. The balance, in respective cases, would count as use. For the maximalist, this is reserved for intuitive cases, such as delivering a package to a given address or denying a loan because of an applicant's unemployment status.[31]

No written accounts of BDE that I have encountered, except the PCAST report, have clarifed, let alone defned, the key terms *collection* and *use*. The PCAST report belongs in the minimalist camp, as it considers everything from storage onward as use, and consequently, subject to regulation. For purposes of this chapter, however, instead of preemptively declaring a line and proceeding with an evaluation of BDE, I have devised two criteria for assessing where a line might be drawn in the hierarchy from impression to analytics and flow, how the placement of the line affects what BDE is, and the extent to which BDE disrupts normative expectations. One criterion is whether a given placement achieves a sufficient departure from business-as-usual to explain the earnest efforts of BDE proponents on its behalf. A corollary is that the greater the extent to which a defnition of collection satisfes this criterion, the more likely is BDE to vex privacy efforts. The second criterion is coherence. By this I mean whether a line that categorizes certain practices as collection and others as use is internally consistent and nonarbitrary. As we move through the practices identifed in the hierarchy we consider how they affect the scope of BDE, and its challenges, according to the two criteria.

Arguably, the least controversial disruptive version of BDE is one that assumes collection to include only the initial uptake when data is captured as an impression on a variety of media. Until such impressions are given meaning, conceptualized in accordance with a given ontology, understood against a background context allowing, for example, the association of a digital blip with a particular person, they are unlikely to excite proponents of BDE. Although the placement of meaning-giving, interpretive activity, labeling, and classifcation outside the concept of *collection* would raise fewer privacy concerns (perhaps none), it would yield a barely interesting BDE. I will assume, therefore, that even those BDE proponents we have called minimalists, who give narrow berth to collection, would include in collection the processing of digital input that results in the ascription of meaning to it.

But what is collection without storage? Extinguishing restrictions that could staunch the lifeblood of big data, as do ex-ante commitments to explicit purposes and consent,[32] is a small victory for BDE without the ability to record and store data for future access; and what is mere storage unless it involves order, organization, classifcation, and a means of fnding, extracting, querying? Some might think this is a good place to draw the boundary between collection and use. Once data is organized to allow for access and query, however, analytics—processing, extraction, inference, the generation of new information—is but a small conceptual step away. If we allow collection to include processing for storage and access, then there seems to be no independent basis for deciding that analytics is data use, and subject to regulation. On grounds of the second, the coherence criterion, it falls within the scope of BDE.

If, fnally, we suggest that a natural place to draw a line between collection and use is at this point, including analytics but excluding flow, it may be that we will have taken one step too far down the slippery slope and run afoul of the coherence criterion. Why? Processes covered under the label of analytics include discovery of new information from information at hand. Said another way, this involves collection of data not directly from data subjects themselves (either from them or generated by them through their activities). From the data collector's perspective, receiving data from a third party is similar to inference because, likewise, it constitutes collection not directly from data subjects. The coherence criterion would impel us to treat the two alike. But if we are to bless analytics and flow with deregulation, the slippery slope transports us to data brokers, whose practices of gathering and supplying data from and to others (governments and commercial frst-party collectors alike)

would, accordingly, earn freedom from regulation. This conclusion, that data broker practices elude regulation under BDE, might stir queasiness among even the most enthusiastic proponents.

The mistake, in my view, originates with setting too much store by the question of whether data is collected directly from a subject or not. Consider, for example, an exchange of email over Google's Gmail. When someone sends me information via email, we may describe this transaction as one in which I have collected this information. The fact that the transmission is mediated via network nodes and ultimately a Gmail server before landing in my account does not make me a third party to it. Something about the intention of the originator, the sender of the email, is more relevant to who is the frst party collector than the service intermediary, which happens to relay the information. For similar reasons, I would argue, the common practice of large information corporations gobbling smaller ones, and each other (e.g., Google and Waze, Facebook and Instagram, Microsoft and LinkedIn, and countless others), ought not, by itself, allow for deregulation of flow that prior to acquisition was subject to regulation.[33]

Finally, since BDE proponents tout the promise of vast caches of data, such as health-related data from primary (i.e. data subjects) and secondary sources, it would not make sense to invite regulation for the former but not the latter. Van Hoboken's defnition of collection, focusing on the capacity to access data for processing, likewise does not distinguish between collection from subject and from other sources. Although it may be worth considering asymmetric regulation for outward flow (i.e. dissemination) versus inward flow (i.e. collection), this idea will not be further developed, here.

To summarize: we have demonstrated how the defnitions of *collection* and *use* markedly affect the scope of BDE, and by implication, determine how radical its departure from present-day practices is. A minimalist account of collection marks a minimal departure, leaving a critical range of practices within the purview of regulation; this conservatism offers solace to some, but gives little leeway to proponents of BDE. The maximalist, by opening the door wide to the full range of data practices typically associated with big data, does not hold any of the parties to account for any of these practices, except at the moment when lives are directly affected—the individual is or is not stopped at the border, does or does not get the job offer, the mortgage, medical insurance, or ads for a high-paying job.[34] No limits placed on what, how, or how much data is collected, or how long it is stored, how it is stored, and where it travels may make BDE proponents happy, but it raises grave privacy concerns.

Where one draws the use-collection line, similarly, is signifcant for most conceptions of privacy; understood as contextual integrity, it challenges the expectations of appropriate flow. More paths of flow escaping regulation means fewer paths subject to normative constraints, and a radical weakening of privacy. The remainder of this chapter will argue that even if BDE reveals consent to be a flawed gatekeeper, it does not justify wholesale surrender of accountability or answerability for all collection.

In order to proceed with evaluation, it is necessary to specify meanings of *use* and *collection* without the benefit of validation by BDE proponents but in ways that stay true to their key thesis. Collection will cover uptake and classifcation; that is, the *creation* of data. It will extend to assembly of data into organized, systematically accessible repositories, or databases, and also cover retention and processing; that is, analysis and inference. Although I see no way to draw a nonarbitrary line between these and flow, I will not press for a defense of these practices from defenders of BDE.

Big Data Exceptionalism: Descriptive or Normative

Clearing the defnitional hurdle still leaves open questions about BDE itself, whether it stands as a factual (i.e., descriptive) assertion or a normative prescription. According to the frst, it is impossible to apply privacy regulation to collection; according to the second, it is undesirable or wrong. Although they are not mutually exclusive—Craig Mundie, for example, defends both—they warrant separate consideration because they rely on different arguments and call for different responses.

Descriptive BDE: Impossibility

When proponents say, "Forget about regulating collection; it is impossible!," we may rightly wonder what they mean by collection, regulation, and impossible. We have already seen that plausible variation in the meaning of collection signifcantly affects the scope of BDE. But it is worthwhile probing the meanings of regulation and impossibility, too.

Regarding regulation, there is some irony: it seems the stronger one's commitment to the synonymy of FIPs with privacy, the more natural the slide

to BDE. As already noted, the burgeoning array of data sources integrated into big data machinations presents a dilemma: embrace the promise of these technologies and practices knowing that core FIPs principles such as use limitation and informed consent must be compromised, or insist on fair information practice principles (FIPPs) and forgo the technologies. Even before the passion for big data took hold, online tracking—arguably, a precursor—revealed fault lines in FIPPs-based regulation, which has been operationalized, ubiquitously, with so-called privacy notices, requiring ostensible consent. Researchers, academics, regulators, and even nonexpert users are well aware of the inefficacy, even failure, of this approach.[35] In short, FIPPs are incompatible with big data because the potential insights from data cannot be anticipated until after enough of the data has been collected—chicken and egg. In response, proponents of BDE have embraced the first horn of the dilemma.

The dilemma, however, could be an artifact of FIPPs-based collection regulation and regulation generally.[36] Following contextual integrity, the substantive regulation of data flow (including collection as one form of flow) offers an alternative that does not demand specification of purpose to data subjects as a condition of each instance of data collection.[37] Contextual integrity is not completely immune from the requirement of teleological justification, specifically, in terms of stakeholder interests, ethical and political rights and values, and contextual ends and purposes, but these deliberations occur as a matter of societal policy and not in a pairwise transaction relegated to data subjects. This still holds collectors to account and thus will not satisfy those who insist on no restriction on collection whatsoever.

Regarding impossibility, what could BDE proponents mean when they say collection cannot be regulated in either of the senses of regulation? Three plausible alternatives come to mind, which, for the sake of convenience, I have labeled technical, institutional, and prior rights, with the caveat that they are not fully independent of one another.

Technical Impossibility: "It Cannot Be Helped!"

Digital impressions are created in the very functioning of the broad class of technologies that provide computational power, communications networks, and information services. Naïve intuition might conceive of digital technologies as mere conduits for messages sent, calculations performed, information

provided, transactions enabled, and the myriad other activities mediated by information technologies, but instead these occur by the transmission of copies and imprints. This means that all activity leaves behind inexorable traces; it is simply how the technology works. Data is collected because it must be collected. For the swath of actors (many commercial) generally called data intermediaries or information service providers, including Web, email, social networks, content (text, video, etc.) usage leaves digital marked trails. A tweet creates content and a shocking trail of metadata far in excess of this.[38] The capture of these trails as data is a technological imperative; it is irresistible.

If BDE is allowed only these observations about irresistible collection, its minimal scope is unlikely to excite proponents, per our discussion in the previous section. Higher value collecting, including interpretation, storage, and analysis, requires the development of complex systems comprising hardware and software, whose architecture and design is far from inevitable. Capture, transformation, channeling, and pooling of data impressions engages creativity, savvy, and scientifc doggedness, resulting in familiar contemporary systems; the Internet is a prime example, Google's search engine is another. Each could have been different, could have yielded different data flows and repositories. A case in point is the protocol regulating Web cookies: contentious at the time, the "winning" protocol allowed third parties to make an end run around restrictions on websites being able to harvest cookies from other sites. Had alternative protocols prevailed, we might have been spared the great privacy disaster of cross-site tracking.

It may be that once system features are established, data creation and flow is inevitable, as when a canal bed is dug, the direction of water flow is inevitable. Those who claim their services cannot but collect data because they have adopted a centralized architecture that affords capture and aggregation of information flows across multiple users and services are hiding the contingency of their system design.[39] Even after they have been settled, technical properties are often malleable and reversible or, as in the case of contemporary digital networks, allow intermediating layers that correct and refne the actions of layers above and below.[40] Such adjustments may allow systems to hide or expose certain data, or to carefully channel its flow according to fne-grained distinctions among recipients, attributes, and purposes. Facebook may claim that it cannot but collect the metadata of its WhatsApp service, but the design choices Signal has made enable not only

encrypted messaging but a backend that maintains minimal metadata.[41] Denying such contingencies belies the carefully calibrated distinctions that manage targeted advertising, meticulously channeling data streams to the myriad enabling parties, pushing content, recording clicks, performing analysis, running real-time auctions, defning which parties are entitled to what data, and so forth. There is nothing inevitable in this, and policy makers and academics with limited technical savvy who may be unable to imagine alternatives are mistaking unwillingness for impossibility.

Institutional Impossibility

To conclude that BDE is the only logical course given immutable technical properties reflects a limited grasp of design contingencies and a readiness to take as given what, in fact, can be questioned. In the struggle over property rights in digitized content, for example, when supposedly inherent capabilities of digital systems threatened commercial interests, stakeholders redoubled their efforts on both technical and regulatory fronts.[42] In a similar manner, the information industry may readily accept technology determined data flows inward, but resists them for outward flow. In the institutional ecology that has sprung up around information and communications services, companies that have realized the value of data about individuals—consumers, customers, users—have emerged as global powers. Incumbents have resisted efforts to alter the technological rubric in unfavorable ways at the same time they have sought to secure the political economy, nationally and globally, that fostered their emergence and sustains their entrenchment and growth.

Attempts to curtail data practices through privacy regulation, which would signifcantly raise the cost of doing business and dampen potential proft margins, have been rebuffed, and the struggle of incumbents to resist regulation has been costly on regulators who are frequently poorer in both resources and technological savvy. In a political economy that accords companies enormous power over their assets and that views data, even data about individuals, as a company's asset, the amalgamation of data holdings may motivate mergers, purchases, and takeovers.[43] One could view these as standard instances of vertical integration of essential services.[44] Yet, these moves allow companies to acquire personal data through strategic purchases that may have been disallowed by privacy rules where the companies in question

are separate entities. A rigorous account of these imbroglios involving global information corporations, national governments, and public interest organizations lies outside the purview of this chapter. Nevertheless, even an outsider's view reveals a bleak picture of what political representatives and regulatory bodies have been able to achieve in the name of their citizens' privacy interests. When considering the national and global influence of corporate forces arrayed in opposition, it is no surprise that commentators see collection regulation as impossible, not as a hard fact of metaphysics but as a consequence of institutional inertia—that is, intransigent key actors creating institutional barriers resistant to regulatory efforts.

Institutional barriers, like technical ones, are surmountable when there is a rationale and a collective will to do so. Mere difficulty is not necessarily a reason to halt attempts; for example, we have not ceased in our quest to staunch the flow of narcotics in the United States, and fnancial markets require relentless vigilance and oversight, which is costly to all members of society, and yet we persist. Furthermore, in times past, we have successfully regulated data intermediaries, imposing limits on what telecommunications providers ("common carriers") can record and share with third parties. Data may want to be free and, like low-hanging fruit, may entice collection, but the National Security Agency (NSA), the Internal Revenue Service (IRS), Google, Facebook, Diebold, and others—even with the integrity of national elections at stake—have utilized technology and regulation to enclose data they consider theirs.

Conflicting Rights and Values

According to this argument, constraints on collection conflict with ethical values and political rights to which liberal democracies are committed as matters of Constitutional principle, explicit law, or both. Citing national security, intellectual property, free speech, and associated intellectual freedoms,[45] this theme has recurred in public debates and signifcant milestones involving privacy, such as the 1890s landmark article by Warren and Brandeis,[46] the 1960s proposal for a federal data center,[47] the 2000s rise of social media, and the 2010s Snowden revelations. Digital rights management systems (DRMs) and technological protection measures (TPMs) have monitored users in the name of intellectual property; governments have surveilled populations in the name of safety and security;[48] data brokers have aggregated data and performed analytics in the name of First Amendment rights.[49]

Finally, overlapping with concerns raised above, having acquired data, organizations claiming property rights over it may cite their right to use it at their discretion, including to analyze and sell the products of this analysis.

But If Collection Cannot Be Regulated, Can Use Be?

In David Brin's fantastical world, not unlike Michael Seemann's postapocalyptic state, citizens "get over" the death of privacy and cleverly opt for full transparency to keep tabs on powerful governmental and commercial actors. These visions contrast with the present situation in which we foolishly demand secrecy, which not only fails to protect us against these overbearing actors who are able to obtain this information anyway but also allows these actors to operate in obscurity. Assuming we share with these visions the belief that public and commercial incumbents will not forgo the lifeblood of their wealth and power, and resign ourselves to the impossibility of meaningfully regulating data collection, it is unclear that faith in the corollary is justifed. If the powers we wish to check are able to resist efforts to constrain their data collection, surely they will resist with equal vigor and determination equivalent efforts to constrain their uses of data, should these, too, prove to be equally proftable, equal in serving will-to-power, and equally surreptitious. In other words, there is little to support the wishful thinking of those holding bad-boy views, or proponents of BDE, that use will be susceptible to regulation if collection is not.

Craig Mundie's pragmatic vision for holding these actors accountable is to enfold or tag data in metadata and construct systems of verifable identities that will allow use restrictions to be expressed and monitored. I daresay, with such mechanisms in place, the task of regulating collection, too, would be greatly eased. Further, since such approaches impose costs and rely on the cooperation of the very class of actors who have worked determinedly to shake off meaningful restraints on collection (and thus far have succeeded), anything short of direct regulation points away from success. Those whose collection activities defy close monitoring and regulation are unlikely to offer an easier target for use regulation. Here I do not refer primarily to, for example, Russian mobsters seeking to evade discovery when setting up botnets, but to mainstream actors seeking immunity from watchdog organizations and public interest vigilantes. Using technological means such as obfuscation, as well as legal means such as nondisclosure clauses, these

mainstream actors have managed to obscure problematic information flows. (Nowhere is this more evident than in the mobile domain.)

In sum, if the reason for giving up on collection restrictions is that the barn door is open, the cat is out of the bag, then there is little reason to believe that regulation of use is likely to succeed.

Normative BDE: Foreclosure of Benefits

From what we learned in the prior section, if metaphysics, architecture, and institutional obduracy make it impossible to regulate collection, then just as surely, they will impede the regulation of use. Although this calls into question key premises of a descriptive account of BDE, it still leaves open a normative account, which asserts that it is undesirable, even wrong, to regulate collection. The case does not rest on the inability to regulate data collection but on the legitimacy of so doing. Normative BDE's underlying rationale, embodied in statements cited earlier, for example, is utilitarian: the potential of big data to deliver benefits to individuals and societies is so great that we dare not staunch its lifeblood. Imposing constraints on data collection would foreclose the benefits of this promising enterprise, particularly since we cannot tell, in advance of collecting, what these might be. Harms that may follow certain big data practices should be minimized by identifying and regulating problematic uses.

As with descriptive BDE, the strength of supporting arguments depends on both premises, namely, (1) that benefits will be foreclosed through regulation of collection, and (2) that harms can effectively be addressed through the regulation of use. In this section I elaborate on these two premises; in the following section I then evaluate them, concluding that at times, BDE proponents have misidentifed and overvalued the global benefit of unregulated collection for big data systems and have missed and undervalued its costs.

The Benefits

On what grounds have big data champions asserted its unprecedented promise? To answer, they point to some already realized in multiple applications and domains as evidence of more to come. Yet, if we are to take seriously one of the core ideas behind BDE, namely, that one cannot say in advance of col-

lection what the likely fndings will be, then the claims must, of necessity, be general ones. A quick survey of the popular literature on big data and data science reveals some of its dazzling feats: Web search patterns that reveal unanticipated drug interactions[50] and flu trends (later discredited);[51] the ability to detect fraud automatically from subtle credit card usage patterns; fndings on the impact of sentiment manipulation in Facebook news feeds on the sentiment of ensuing user commentary;[52] personalized advertising; a winning Major League Baseball team;[53] and Target fguring out which customers were pregnant.[54] There are laudable efforts in health, such as National Institutes of Health (NIH) researchers turning to big data techniques to learn about HIV infection and treatment efficacy;[55] in public utilities companies spurring energy conservation through smarter energy grids; in IBM's Watson amassing health and lifestyle data to reveal actionable correlations; and in educational institutions employing virtual learning platforms (including massive open online courses, or MOOCs) to draw insight about learning styles.[56] These less dazzling but arguably more important advances have been seen, and are foreseen, across the spectrum of social life, in fnance; public health; public safety; medicine; national security; commerce; marketing; romantic love; employment; law; cultural creation; personalized, automated information services; and more reliable recommendation and ranking systems.[57]

A recent wave of interest in machine learning and artifcial intelligence (AI)[58] has publicized mind-boggling achievements accomplished by cleverly exploiting vast data repositories—in machine translation, robotics, and complex games, such as Go and chess. Although this chapter's focus is on data about people, the repositories yielding these important insights are drawn from a wide range of sources and information types.

Ethics of Use

Generally, supporters acknowledge that this is a signifcant departure from existing privacy regimes. Unlike earlier "bad-boy" privacy skeptics, BDE proponents, generally, do not deny the important role strong privacy regulation can and has served in addressing a host of privacy or informational harms. While acknowledging that BDE constitutes a signifcant departure, they hold that directly focusing on harms resulting from *uses* of information can supplant privacy regulation of collection, broadly construed, without foreclosing

the benefts. With an awareness of the vulnerabilities exposed by stripping the protective shield of privacy (as constraints on flow), the burden of concern falls on justifying uses of information. Whereas previously, privacy may have dictated that a cost-beneft analysis support a given intention to collect information, BDE shifts a cost-beneft assessment to the point of use,[59] allowing such uses only if the assessment supports it.[60]

In support of this thinking, an emerging feld of data ethics has attracted interest not necessarily supporting BDE, but compatible with the idea that data use be the linchpin.[61] Although many issues discussed in an already burgeoning literature are echoes of those aired in privacy scholarship, in the context of big data and AI, their reprise has new urgency and sometimes a new twist. Social justice, to date, has been the most preoccupying: as decisions affecting quality of life and even life itself in all social domains—including workplace and employment, advertising and marketing, fnance and healthcare, education and politics—are increasingly informed by big data analytics, commentators point to error, unfair discrimination, historical prejudice, and inequitable allotment of resources and opportunities as potential consequences of automated, algorithmic prediction and decision making.[62] Whether persons are stopped at the border; whether they are offered employment, acceptance at a prestigious university, an apartment, or favorable rates for health and life insurance or a mortgage; what prices they are charged for merchandise and what ads and offers they receive all comes down to the results of automated decision systems, which may be biased. Calls for accountability apply not only to data mining and analytics algorithms but even to the selection data, which cannot be assumed to be objective and impartial.[63]

Threats to autonomy due to manipulation and exploitation constitute another class of issues that have attracted attention. Since any information that increases accuracy in clustering and prediction may be attractive to data holders, processors, and decision makers, people may have little clue about the bases on which they are being judged. Thus, our fates may be sealed by processes that are opaque (to us and even to the processors themselves) and according to information we may deem irrelevant.[64] Models that emerge from statistical learning may map well onto the training data and offer statistically respectable predictions, but they may defy human sense-making and consequently, human explanation.[65] Decisions affecting your prospects and well-being, accordingly, may seem as arbitrary as the toss of a coin. Raising questions about due process,[66] critics have urged transparency in the key operations of automation, from an account of the data and algorithms to

thresholds and criteria affecting the transition from fndings to practical decision making. Fairness is certainly a factor, but autonomy is challenged when seemingly arbitrary decisions interfere with our capacity to achieve important life goals. Veering to the sinister, practices that critics such as Frank Pasquale[67] have called attention to involve ferreting out information from which particular vulnerabilities are inferred. Preying on these vulnerabilities, which individuals may themselves be seeking to overcome, third parties manipulate those individuals through behavioral advertising, targeted marketing, and disadvantageous offers to which they are likely to accede, thereby diminishing their autonomy.[68] More directly, data ethicists anticipate oppressive working conditions in which employees' performance and work schedules are optimized for maximum business efficiency.[69]

Other harms from data uses include chilling effects—on speech and association—as people grow aware that the friends we keep online, the opinions that we and they post, and the searches we conduct may earmark us as people of this or that type.[70] Critics warn of threats to democracy from political messages fnely targeted down to particular individuals and households.[71] They warn of the flter bubbles engineered by recommender systems and personalized ranking algorithms.[72]

The question to which we turn in the next section is whether we can afford to forgo restrictions on collection, confdently assuming that use regulation, guided by data ethics, subject to cost-beneft scrutiny, will protect against privacy and other harms.

Reality Check

Undoubtedly, there are new and good reasons, in light of big data, to recalibrate contextual informational norms and forge new approaches to privacy regulation. Contextual integrity allows for such reassessment, permitting challengers to entrenched practices to replace them if they meet the normative criteria at least as well, or ideally, better—justly serving interests, promoting ethical and political values, and fostering contextual ends and purposes. But BDE goes further. It wants to situate collection entirely outside the remit of political accountability; it recommends a blanket lifting of constraints on data collection (as defned above) with the expressed *faith* that we will be better off if we do and will suffer an opportunity cost if we do not, and that we will be able to address ills by addressing ethical use.

I am skeptical. In my view, this path will leave data subjects vulnerable to privacy harms and hard-fought political values vulnerable to erosion, with no clear path to compensatory benefts. The evaluation I provide in this section does not challenge the logic of BDE; rather, it is informed by what I would call "a dose of realism." I say realism, not pragmatism, because while pragmatists might agree that collection deregulation is wrong, they may believe that resistance is futile and half-measures (i.e., use restrictions) are all we can achieve. By contrast, I argue that deeply engrained realities of the information and data landscape belie the well-meaning beliefs and assumptions making up the justifcation for BDE. In *this* reality, it simply is not rational to expect that unconstrained data collection will optimally serve societal needs and values. Thus, even if one optimistically holds that we can regulate data use reasonably well, the additional risks to data subjects of removing the cushion of collection constraints are not justifed. Finally, I will argue that even if data use could effectively be regulated to minimize informational harms, there are risks and harms inherent to collection itself that must be directly addressed.

Who Is "We"?

BDE asserts that ex ante restrictions on collection are likely to inhibit the tremendous benefts that big data promises to individuals and societies. Because results of algorithmic learning are not knowable in advance (particularly with unsupervised learning) and may not even map easily onto concepts that are natural or meaningful to humans, we cannot perform cost-benefit analyses or require purposes to be specifed *before* data is collected. We need the data frst in order to extract knowledge from it and be guided in actions and decisions. So, as we have seen, goes the argument.

It is surprising that the logic of this argument has not been more aggressively challenged, most glaringly for the shift in meaning of the crucial term *we* in "*We* should not restrict, otherwise *we* have much to lose."[73] Whereas some of the cases we have cited, such as NIH researchers incorporating big data in their studies of HIV infection and treatment[74] and public utilities companies spurring energy conservation through smarter energy grids, a sprinkling of reality dust reveals that all is not as it is claimed to be.

To begin, the sources of the data deluge and the costliness and even futility of regulation are not predominantly patient health records and student

records, or records from innumerable government databases, which in the 1960s had aroused great privacy fears. Yes, these traditional data stores contribute to the deluge, but the sources that have excited BDE proponents are the emanations of technology-mediated behaviors, including intentional activity (online purchases, searches, comments, ratings, etc.) and the data exhaust created and captured alongside it, including social networks, communications metadata, interest profles, and so on. The "we" surely refers to all of us, the data subjects. What of the "we" whose benefts ought not be foreclosed?

It is worth noting that the bulk of this data is concentrated in the hands of a few private, global, commercial entities. These include the familiar ones with which we interact directly, such as Google, Apple, Twitter, LinkedIn, Amazon, Netflix, and Facebook, and indirectly, in their capacities as platforms, operating systems, and intermediaries.[75] Vast repositories are also assembled by those with whom we have not been aware of contact, such as analytics companies and data brokers like Acxiom. Others about whom we think only rarely, including traditional telecommunications providers such as Sprint and AT&T, Internet service providers such as Verizon, medical and other insurance companies such as Medical Information Bureau and Aetna, and fnancial institutions (such as banks and credit card providers), accumulate vast data stores, sometimes because it is necessary for conducting business, sometimes because data retention regulation requires it, and mostly because in their functioning as platforms and intermediaries, data falls into their possession and nothing prevents them from staking claims to it.[76]

In this reality, *we* the benefciaries are not one and the same as *we* the people, the data subjects and those who represent our interests, exhorted to accept deregulated collection. In reality, there are no assurances that opening the floodgates and relieving these dominant data collectors of accountability will result in celebrated knowledge gains and decisional integrity in service of the individual's or, for that matter, the common good. It is not that it will *not* serve them at all, only that it is unlikely to be the primary motivation. In saying so, I impute no ill will or evildoing on the part of these companies; on the contrary, many of them have contributed greatly to quality of life. It is merely that they are, understandably, driven by different imperatives—business and proft—and the data they record and the questions they ask of it are related (arguably, must be related) to these imperatives. The unthinkably large trove of Web use data is optimized for effective targeted advertising; for the massive accumulation of medical data accruing to medical

insurance companies for assessing premiums;[77] for studies Facebook underwrites with its vast stock of networked data shaped by company interests such as attracting advertising dollars and preventing defection to other services.[78] To be sure, individuals have also benefted and societal needs have been served, but collaterally, not systematically.[79]

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 well-being. Although interests might align at least partially, and benefts flow, there is nothing to compel this. Utilitarian thinkers, including economists, should also be asking about opportunity costs—that is, not only whether there is beneft from unregulated collection but also whether the set-up is optimal. The question to ask is whether greater benefts might accrue from a different arrangement of entitlements if *we* were allowed to frame the questions, where the "we" in question could range over government representatives with citizens' interest in mind, independent academic researchers, and public interest organizations.[80] When the referent of "we" slips from one party to others, the distribution of winners and losers may change, no matter what happens to overall gains and losses. As it stands, not only is much of this data outside the grasp of many who might want to put it to use for the public good, but the view into what data there is and what the collectors do with it is utterly blocked to all but a rarefed few,[81] and even they see only a highly circumscribed, measured slice.[82] Trade secrecy and competitive business advantage routinely trump public interest.[83]

In circumstances where interests of *we* the data collectors and *we* the data subjects are more obviously misaligned, it is particularly important to scrutinize the rhetoric of benefts foreclosed. Online behavioral advertising, which has depended on what is, effectively, unregulated online tracking, is a case in point. Those endorsing it claim that we are all better off when ads match our interests. Although persuasive to lawmakers and regulators who have offered little resistance to data collection, online and off, within and across platforms,[84] these claims are inconsistent with surveys that repeatedly show strong opposition to online surveillance and targeting from those who are its subjects.[85] Undoubtedly there are benefciaries of the practice, but there is little published evidence that the benefts are fairly distributed. Although this is not the place for a full-blown discussion of online, targeted advertising, the case clearly illustrates how the interests of *we* the subjects of deregulated

collection diverge from *we* whose benefts would be foreclosed were deregulation resisted.

To hold moral sway, it is insufficient to demonstrate that one set of stakeholders benefts from deregulating collection; we must show common, nonprejudicial beneft. One promising candidate is risk reduction. If a free hand collecting data from and about people can convincingly be linked to reduced risk and increased security for all against, for example, a terrorist attack, the two "we" groups seem to be aligned. Rigorously evaluating this claim requires strong empirical evidence, but equally importantly, the scope and logic of the argument should be sound. Contextual integrity would require that an analysis specify all relevant parameters. Thus, whereas for some parties freedom to collect certain types of information, under certain constraints, might be justifed, the same may not hold for other collecting parties. In the case of the commercial actors we have been discussing, the evidence for overall risk reduction from deregulated collection is simply not present. In fact, one should remain astute to mere shifting of risk from one party to others, at times, even in a zero-sum confguration—I reduce my risk by increasing yours—masquerading as reductions in overall risk. Massive data breaches often reveal such shifting as companies collect and accrue data with an eye to extracting forward value and reducing their own costs, while in the process exposing individuals to greater risks.[86] One of the most spectacular of these instances, announced in September 2016, was a massive breach that had occurred two years earlier of databases held by Yahoo!, which compromised records of an estimated fve hundred million customers containing user names, log-in credentials, birth dates, and zip codes.[87] To date, there appears to be no recourse in the law for exposure to risk for victims of such breaches.[88]

Another type of risk shifting occurs when companies relying on the results of data analysis and profling are able to identify consumers from whom to extract higher prices for their goods and services—thus lowering their risk and increasing it for buyers.[89] Free reign on collection and analysis may place individuals in adversarial relationships not only with companies (or government agencies) but also with one another, as differentiation among individuals, which is advantageous to companies, may unfairly disadvantage some individuals over others. One person's personalization and reduction of risk may be another's discrimination and exaggeration of risk, particularly in competitive situations where resources are limited, such as admission to a prestigious college, discriminatory pricing, and apartment rentals in desirable urban neighborhoods. Probabilistic modeling inevitably means that

some people will be misclassifed, not so much in error but as an inherent property of such modeling.[90] Here, too, this may mitigate risk for the data collector but increase risk of wrongful treatment for the individual. Depending on where thresholds are set for false negatives and positives, data processors may shift the risk of erroneous classifcation toward or away from themselves. The selection of data felds, too, can affect which individuals are blessed with a positive outcome and which a negative.[91] Cost-beneft analyses on use alone will not successfully root out risk shifting unless this world—not our world—includes political mechanisms to ensure impartial access to data sets and democratic guidance on what questions are posed to data and how emergent models are exploited.

To conclude this section, let us consider one of big data's risk-reduction success stories: credit card fraud detection. The story told is that over time, dogged collection of data has enabled credit card companies to detect anomalous card usage based on patterns of normal usage. Fortuitously, everyone (both "we" groups) is happy (except the fraudsters, to be sure). Why highlight this case? Although it is true that machine learning over vast data sets is the proximal agent of mutual beneft, the confluence of interests was due in large measure to strategic legal regulation and the establishment of industry standards[92] that assigned liability for losses due to fraud to credit card issuers, not to individuals or merchants. In service of realism, it is critical to recognize the role of legislation in aligning the interests of consumers and credit card companies. In hindsight, this allocation of liability was even more brilliant given that transaction data naturally accrues to these companies, placing them in the best position to perform these analyses.

The case of credit card fraud is instructive because it steers away from simple connections between unrestricted collection and mutual benefts. Even in this success story of big data, the win-win outcome is as much a product of smart regulation. Data breach notifcation laws strive to a similar achievement by tying the fate of data holders to data subjects. To date, the sting of notifcation seems not to be painful enough to moderate the accumulation of data, which in turn creates honeypots for destructive hackers, mobsters, fraudsters, and their ilk, with ultimate risk shifted to individual data subjects.[93] This should give pause to those who believe that regulating use and misuse alone will or can be effective in mitigating harm to data subjects.

In the contemporary landscape, fair information practice principles may not protect privacy, but one of the fundamental purposes behind them, lev-

eling the playing feld for data holders and data subjects, remains as vital now as it was in the 1960s and 1970s.[94] Strategically imposing constraints on data flows seeks to address enormous disparities of power and wealth and to sustain differentiated societal roles and positions that are important for societal integrity. We may not care to level the playing feld for everyone—criminals, for example—but for others, the modulated collection and use of information provides security for legitimate ends.

In sum, the key aim of this section is to challenge an implicit assumption behind the normative version of BDE that warns against ex ante limits on collection and the risk of foreclosing unanticipated discoveries based on machine learning and other forms of data analytics over large, aggregated data sets. The assumption is that we individuals should support unrestricted collection in order for us collectively to reap the benefts. But scrutiny reveals that the benefciaries are not the same as the contributors; moreover, those controlling and processing big data, in reality, are not obliged to serve the collective good, nor are they restrained in uses that may even cause undeserved harm. In the present-day political economy of data and digital technology, ordinary people—the data subjects—or those representing our interests will never achieve insight and transparency into what data owners and processors are doing. Aside from gross and obvious instances, it will be impossible to regulate use for the variety of subtle harms against which privacy norms, over thousands of years of social life, have evolved to guard.

We the NSA

A reader may agree with the fndings in the previous section but chalk up the problem to misuse, not to deregulated collection. After all, the worrying cases of risk shifting, unbalanced distribution of costs and benefts, and suboptimal extraction of value from data are due to wrongful uses and, except for the case of data breaches, are only indirectly due to unfettered collection. Such readers misunderstand the argument, which is to reverse the burden of proof. If you assert that I have a social obligation to allow unfettered collection, despite its infringement of privacy norms, you must demonstrate the overwhelming social value of so doing. I have shown the opposite: in the present-day political economy and legal landscape there are few, if any, assurances that general social welfare will guide the extraction of value from this data, or mitigate potential costs to data subjects. These observations

undermine the BDE supporter's opportunity-cost worries. Admitting that the likely across-the-board benefts may be exaggerated undermines the BDE supporters' opportunity-cost worries. More importantly, it raises the bar for efficacious use regulation, because with less clear benefts, we need greater assurances of minimal harm; this, given the existing landscape of practice and policy, is impossible to provide.

Here, however, I want to go further. Even, hypothetically, allowing that efficacious regulation of misuse were possible, I want to suggest that collection itself deserves scrutiny and restraint. To explore this proposition, let us consider an equivalent confguration of means and ends where collection is regulated, independently of whether the ultimate target is restraints on use. I refer to one of the constitutional pillars of political democracy in the United States, the Fourth Amendment of the Bill of Rights.[95] Now, let us proceed with a thought experiment: imagine that Fourth Amendment critics are advocating for its repeal on grounds that it unnecessarily obstructs law enforcement and national security. Leaving aside whether such arguments could have held sway in 1791, consider their strength in the present climate. First, the rising incidence of domestic and international terror means there are more reasons to be fearful; and second, improvements in technologies of surveillance—to monitor communications and geolocation, to capture and log visual images and commercial transactions, to aggregate the above data, and to extract useful insight—means the potential fruits of dragnet surveillance are assuredly plentiful.

Advocates of repeal could calm us with assurances that full attention will be given to preventing misuse (assuming agreement on what counts as such) and holding perpetrators to account. Requiring antecedent specifcation of particularized purpose, as required by the Fourth Amendment, severely handicaps efforts to catch criminals and expose dangerous plots and other serious threats, because we cannot always predict what patterns the data reveals and whether they will be useful.[96] Requiring probable cause undermines the efficacy of big data analytics because a backdrop of normal patterns of communications, activity on social networks, and transactional confgurations is crucial to detecting the suspicious, the abnormal, the worthy of note. This is particularly relevant in applications of unsupervised machine learning, where patterns must be discovered in *all* the data, not merely a focused subset.

Respondents insist, however, that a liberal democracy must retain some version of the Fourth Amendment. They warn against dragnets and remind

us of the invasive and demeaning character of random or universal "stop and frisk," mandatory drug testing, and house-to-house searches. They spell out the important work of the Fourth and other amendments, such as the First, in forestalling totalitarianism by restricting the *scope* of government's intrusions into citizens' private endeavors and maintaining certain spheres—home, religion, political association—as off limits. In the clearly demarcated instances where administrative functions allow government agencies into certain areas of private life—for example, to process Medicare reimbursements, long-form tax flings to the IRS, the decennial census surveys, and welfare benefits—data holdings generated through these interactions have been rigorously siloed.[97] The fact of mass monitoring is bad enough, respondents may say, but equally so is the mere *feeling* of it, chilling activities crucial to quality of life and civil society, including, but not limited to, free association and speech.

The repeal advocate smiles indulgently. Present-day dragnets may be cast with enormous discretion, via hidden cameras, unobtrusive motion sensors, concealed listening devices, passive capture of signals from mobile devices, black box recorders attached to broadband cables, government-installed network malware, third-party data aggregators, and so forth. Citizens will neither *feel* invaded nor will they even know. "If you have nothing to hide, you have nothing to fear" is their fnal reassurance; the innocent should not worry because although collection will be unfettered, harmful uses will be curtailed. My guess is that few readers of this chapter will feel reassured, though they may differ on the grounds for their unease. For some, it is the special relation of government to private citizen that calls for special attention, a need that is evident to lawmakers and even privacy skeptics. The furor over what Edward Snowden revealed about NSA practices speaks clearly to this concern.[98]

But what are the reasons for fencing in government power[99] despite obstacles it may create for its administrative function, guardianship of national security, and protection against crime? Why have we refused to open the floodgates? In addressing these questions, an account of republicanism offered by the political philosopher Philip Pettit provides insight.[100] According to it, to achieve political liberty we must do more than thwart repressive governmental actions; we also must contain government domination, meaning the *power* of government to interfere arbitrarily in our lives. The point is worth emphasizing: concern does not stop with arbitrary interference enacted but extends to the *power*, or potential, of arbitrary interference.[101] Adapting

this principle to data, the inappropriate collection of information about private individuals and mass collection about populations provides government with inordinate *powers to* interfere. With the phrase "knowledge is power," the qualifcation "when you use it" is unnecessary because the mere having of knowledge *is* a form of power in itself. Because the gathering and holding of information are empowering to government, the Bill of Rights and other legislative acts that strengthen informational privacy through procedural barriers and prohibitions[102] are critical measures to protect individuals and populations against government domination.

In the 2016 case *Birchfield v. North Dakota*, which decided whether doctrine permits police officers to conduct warrantless breath tests, the U.S. Supreme Court concluded that while breathalyzer blood alcohol content (BAC) readings do not require a warrant, blood tests do, frst because they are invasive (i.e., pricking the skin) and second because they generate a lasting sample: "A blood test, unlike a breath test, places in the hands of law enforcement authorities a sample that can be preserved and from which it is possible to extract information beyond a simple BAC reading. Even if the law enforcement agency is precluded from testing the blood for any purpose other than to measure BAC, the potential remains and may result in anxiety for the person tested."[103] This brief snippet recognizes the mere potential of use as warranting a higher standard—in particular, the requirement of an ex ante rationale for collection of this type of nonephemeral product. Data collection has equivalent properties, as it generates "a sample that can be preserved and from which it is possible to extract information beyond."

Appreciating the Fourth Amendment as a meticulously crafted trade-off, we may, nevertheless, fnd persuasive the enthusiasm of government agencies—from the NSA to law enforcement to the NIH[104]—for the positive potential of big data. The scales may tip against its barriers and prohibitions on a credible showing that mass collection of communication and transactional data, augmented with vast data commercial holdings, would enable more effective mining of suspicious activity and criminal or terrorist networks, or that enhancing traditional medical records with broad swaths of lifestyle records could afford great cross-over understanding in both spheres.[105] It may be time to recalibrate the balance. But if the benefts of vast data repositories have soared, so have its distinctive threats. Big data (including analytics) produces unpredictable insights, and individuals rightly may worry about what could trigger special interest and scrutiny of them—

ordinary things, features of their tax returns, persons with whom they socialize, where they travel, and how they pay. If the machinations of big data cause greater worry to the terrorist bringing a bomb aboard a plane, their unconstrained application also raises concerns for which past precedent does not readily prepare us.

In her concurring opinion in *United States v. Jones*, Justice Sotomayor highlights distinctive threats associated with GPS-enabled tracking, rightly pointing out that the creation of a "precise, comprehensive record of a person's public movements" exceeds the scope of allowable, "plain view" surveillance by a police officer.[106] Similar concerns have engaged the scholarly literature that seeks to characterize the distinctive, incremental threats from amassing data continuously over time or aggregating data across a myriad of sources.[107] These works locate the potential to destabilize the tenuous balance of power between government and other data holders not merely in the additive powers of expanded data sets, but also in the multiplicative powers of analysis and inference. Although these concerns focus on the capture and accrual of information about individuals, one at a time, the power of big data is once more multiplied by capture and accrual across populations. Traditionally, in the context of governance, population or mass surveillance and "dragnet," or bulk collection are reviled, and typically associated with authoritarian and totalitarian regimes and a disregard for civil liberties and due process. By contrast, big data boosters extol the capacity it offers to acquire and analyze data from whole populations, not mere samples.[108] Machine learning can perform its magic ever more dramatically over ever more data; the larger the population, the greater the number of features that can be integrated in its scope. Intelligence gleaned from wide-ranging features (properties, attributes, or types of data) across large populations may include emergent associations, networks, and relationships and predictive accuracy in areas of life in which government has no legitimate business. Individuals must worry not only about what data they may have produced or information about them that may arouse suspicion, but also about others with whom they associate—not an unprecedented concern—and whom they happen to resemble.[109] Whether such resemblances track natural attributes or those concatenated from the mysterious workings of machine learning algorithms, these processes undermine the discretion ordinary individuals have in defning their relationship with government and, further, increase uncertainty over what might be exposed and exploited in this relationship. These, precisely, are the powers to interfere, arbitrarily, that characterize government

domination and understandably provoke the anxiety anticipated by the Court in *Birchfield*.

A favorite taunt of privacy skeptics is that we have more privacy today than in bygone days when everyone in the village knew everyone else's secrets. But even these skeptics draw the line at government, and not only because of how much it knows. Knowledge may give power to your nosey neighbor, but the government has a lot more of it to wield—it can deprive you of liberty and life. When it comes to powers of the state, therefore, even those generally skeptical of privacy-based constraints on big data are measured in what they support. The harsh realities are sobering. Many of us may trust in the restraint and integrity of today's executive branch, the NSA, and our local police to apply their practices of noninvasive, dragnet surveillance to the singular purposes of societal safety and security and efficient administration. But centuries of recorded history featuring rulers who have exploited and tyrannized their subjects, and governments that have oppressed their citizens time and time again, reaffirms the wisdom of protective barriers. In the age of information and big data, insurance against abuses means selectively diminishing access to data, not merely circumscribing certain uses of it.

Is Government an Exception to Big Data Exceptionalism? Another Reality Check

Viewing government as a special case has been justifed by the historical record. The public outcry following Snowden's revelations shows no inclination to surrender civil liberties and concede to government's wish to amass and hold information on all citizens willy-nilly. Indeed, over the past fve decades, as difficult as it has been to hammer out privacy regulation for the private, commercial sector (with a few exceptions), we have clung to limits on prying, surveillance, intrusions, and invasions provided by a combination of the Bill of Rights and legislation, notably the Privacy Act of 1974 and various wiretap statutes. Proponents of BDE could follow the lead of privacy skeptics who have made an exception of government even as they have resisted legal restraints on commercial actors, citing efficiency, free speech (of these actors), free "stuff," innovation, and the fact that people really do not care. Outlawing harmful uses would suffice as a safety net.

Let us view this familiar position through the reality flter. In the previous section, we cited threats to political liberty from government domination as reason to see collection as a contributing factor in its own right. There is compelling reason to extend this scrutiny to private, commercial organizations, which in a relatively short period of time not only have accrued and enclosed vast data holdings about individuals within discrete nations but also have amassed vast powers, globally, to shape national and international policy.[110] They may not have armies at their direct disposal, but they do have the ability to affect the lives of individuals in basic ways—shelter, security, employment—and to exploit the reliance on them for data that governments cannot obtain by dint of either regulation or incapacity.[111] Whereas, for the most part, national borders circumscribe governments' exercise of direct power, such borders have been notoriously porous where global information companies are concerned. No doubt, as global corporations across different industries have found ways to evade national laws in one country by selectively situating questionable practices where they anticipate least resistance, those in the information services and data industries have an advantage from their command of global digital networks and their direct grip on popular engagement.[112]

Even without the mortal powers wielded by governmental actors, corporate actors armed with the power of data can affect the attainment of a decent life—shelter, employment, nourishment, family, friends, health, education, and security. Tyranny and domination come not from the power merely to interfere with people's actions and choices, but to interfere arbitrarily. Celebrating the promise of big data, its boosters have cited predictive capacities that exceed the capacity to explain systematically. Thus, companies may maximize their utility function in various areas—hiring practices, marketing, operational decisions—on the basis of data alone. Yet, what may be expedient for a decision maker might be a decisive blow to the subjects of such decisions whose prospects are stymied in vital spheres of life, particularly those in the margins of actuarial error, without rhyme or reason. Already much discussed, the opacity of decision systems based on machine learning algorithms, compounded by lack of access to data held in private hands, creates a fortress against public inspection. Where a demand for explanation and justifcation goes unanswered, it is impossible to know whether life-critical decisions are fair or unfair, relevant or irrelevant, and to those affected, they might as well be arbitrary. By any name, this is domination. In

a free society, cumbersome as it may be for private, commercial actors, the burden of accountability falls on them for their collection practices, just as it does on governments.

Summation

In a departure from received principles of privacy regulation, the thesis I have called big data exceptionalism (BDE) supports the deregulation of data collection. Its key assertions are that (1) characteristics inherent to digital technologies make collection inevitable and unavoidable; (2) inherent characteristics of big data make it impossible to anticipate in advance what knowledge may be extracted and what purposes are served by large data aggregations; (3) not exploiting the promise of big data to its fullest will be costly to society; and (4) to address harms and risks typically associated with threats to privacy the regulation of data *use* is sufficient.

This chapter disputes BDE on conceptual, normative, and descriptive grounds. To begin with, ambiguity in the key terms *use* and *collection* challenges the coherence of the distinction. Take collection. As an inevitable byproduct of functioning digital technologies, it resembles the imprint of a foot in wet sand. If that were all collection entailed, BDE would not amount to much, but if collection entails more, how much more? At minimum, one would expect the digital imprint to mean something—that is, be conceptualized and classifed—and beyond this, for collection to allow a degree of permanence and recovery, hence storage in an indexed or searchable database, allowing for later access. An ability to organize, amass, aggregate, and curate seems also inevitably to follow. Data flow from a frst-party collector and a third party might appear to be a use instance, but third parties could argue that they are "collecting" data, albeit not from data subjects directly. And while something may seem wrong with placing data brokers, in their data gathering mode, outside the remit of regulation, it is consistent with the industry practice of acquiring data through mergers and acquisitions.

Although proponents rarely acknowledge these ambiguities, where one draws the line can change what BDE means and poses a dilemma: an attenuated defnition of collection that keeps more activities within the scope of regulation is more palatable to privacy advocates but reduces the scope of al-

lowable practices; more activities outside the scope means greater freedom for big data processors and is farther from traditional privacy needs. This chapter has assumed a more inclusive defnition of collection.

In support of deregulating collection, proponents cite grandiose forecasts—progress on the world's direst problems of economy, health, and security—and offer a handful of dramatic applications that have given cause for optimism. My chapter counters this logic. Data available for public and public interest research is a small fraction of data held in private, commercial hands—for that matter, concentrated in the hands of a few dominant actors. These actors effectively hoard the data, obligated neither to pursue benefcial and progressive applications nor to open their troves to third parties to do so, or even to allow access for inspection and scrutiny of their internal practices. Such powers of use and exclusion, sustained and enforced through a combination of property rights, commercial freedom, contracts, and technologies, will not be dislodged without a concentrated effort on several fronts, including regulation. Without it, the illusion cannot be sustained that these holders will ask questions about their data and address problems to serve the common good. Similarly, risks to data subjects will be addressed only to the extent that this aligns with the interests of data holders. (Such an alignment was ingeniously achieved in the case of credit card fraud liability.)

The BDE proposition comprises two interdependent halves: lifting restrictions on collection, counterbalanced by restrictions on use. I have concluded that collection without accountability is not currently justifed. Furthermore, past failures to harness and guide information use in ethically legitimate directions, or even to audit data holdings as a precursor to prevention, cast into serious doubt the grounds for faith in use restrictions.[113] The greatest challenge yet to the positive promise of use restrictions is (and will be) contestation over what uses should be allowed and what should be restricted.[114] Controversial use regulation will be no less subject to stakeholder manipulation in all the data practices that concern us—from holding data processers responsible for breaches and preventing insurance companies from incorporating preexisting medical conditions to using proxies to reveal union sympathizers and inferring pregnancy to determine marketing strategies.

Finally, I have challenged the notion that collection itself is innocuous. On the contrary, the mere holding of data by powerful parties bolsters their domination over the subjects of this data; a metaphorical sword dangles

overhead, but we know neither the nature of the weapon nor what will trigger its plunge. Such threats, for centuries understood in the relation of government to citizens, increasingly characterize the relation of individuals to powerful corporate actors.

A paragraph of recommendations is inadequate to address the issues presented in this chapter. In broad brushstrokes, one clear candidate is that we sustain and strengthen efforts to regulate both collection and use. Another is that we regulate collection and use along contextual lines, not lines of data ownership. This means resisting the common practice of companies accruing data through acquisition, which might have raised eyebrows if achieved through the sharing of the same data across company lines. It also means regulating the willy-nilly merging of data sets accumulated from different domains already within a single company. It means meaningfully holding data proprietors responsible for data breaches. It means insisting that whereas there can be flexibility in how data may be used, we can still insist that all data holders commit to broad purposes and that, depending on the nature of the purposes, we will regulate. Finally, with the accumulation of vast data holdings comes the responsibility to allow this data to serve the public interest, not merely at the discretion of the data holder (i.e., not as "data philanthropy") but at the determination of the people's will.

Notes

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.

2. Polly Sprenger, "Sun on Privacy: 'Get Over It,'" *WIRED*, January 26, 1999, 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, Mass.: Perseus Books, 1998).

4. See also Christian Heller, *Post-Privacy: Prima leben ohne Privatsphäre* (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, Conn.: Yale University Press, 2011).

6. Recommendation 1 in President's Council of Advisors on Science and Technology (PCAST), *Big Data and Privacy: A Technological Perspective*, (North Charleston, S.C.: CreateSpace, 2014), 49.

7. Michael Seemann, *Digital Tailspin: Ten Rules for the Internet After Snowden* (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.

10. Seemann, *Digital Tailspin*, 55.

11. Craig Mundie, "Privacy Pragmatism: Focus on Data Use, Not Data Collection," *Foreign Affairs*, March/April 2014, 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*, ed. Mireille Hildebrandt and Katja De Vries (New York: Routledge, 2013), 197.

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): 259–60.

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, 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. For a detailed account of contextual integrity, see, e.g., Helen Nissenbaum, *Privacy in Context: Technology, Policy, and the Integrity of Social Life* (Palo Alto, Calif.: Stanford University Press, 2010); Nissenbaum, "Respect for Context as a Benchmark for Privacy Online: What It Is and Isn't," in *Social Dimensions of Privacy: Interdisciplinary Perspectives*, ed. Beate Roessler and Dorota Mokrosinska (Cambridge: Cambridge University Press, 2015), 278–302.

17. Nissenbaum, *Privacy in Context*.

18. See Lawrence Lessig, "The Law of the Horse: What Cyberlaw Might Teach," *Harvard Law Review* 113, no. 2 (1999): 501–49.

19. I owe thanks to Jason Schultz for drawing these cultural meanings to my attention.

20. 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"; Executive Office of the President, *Big Data: Seizing Opportunities, Preserving Values* (North Charleston, S.C.: CreateSpace, 2014); and a host of newly minted trade books.

21. See, e.g., Katherine J. Strandburg, "Monitoring, Datafcation, 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, 2013).

22. See, e.g., Thomas S. Kuhn, *The Structure of Scientific Revolutions*, 3rd ed. (Chicago: University of Chicago Press, 1996), and Paul Feyerabend, *Against Method*, 4th ed. (New York: Verso, 2010).

23. Some have gone so far as declaring the end of theory. See Chris Anderson, "The End of Theory: The Data Deluge Makes the Scientifc Method Obsolete," *Wired*, June 23, 2008, https://www.wired.com/2008/06/pb-theory/.

24. See, e.g., Charles Duhigg, "How Companies Learn Your Secrets," *New York Times*, February 16, 2012, http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html; Mara Hvistendahl, "Can Predictive Policing Prevent Crime Before It Happens?" *Science*, September 28, 2016, http://www.sciencemag.org/news/2016/09/can-predictive-policing-prevent-crime-it-happens. Nate Silver, *The Signal and the Noise: Why So Many Predictions Fail—But Some Don't* (New York: Penguin Books, 2012).

25. I have not been persuaded by efforts to defne and distinguish the terms *data* and *information*, and since a rigorous distinction is not necessary for the overall argument of this chapter, I have used the terms interchangeably.

26. See Lisa Gitelman, ed., *"Raw Data" Is an Oxymoron* (Cambridge, Mass.: MIT Press, 2013); Geoffrey C. Bowker, *Memory Practices in the Sciences* (Cambridge, Mass.: MIT Press, 2005), 184 (suggesting that "Raw data is both an oxymoron and a bad idea; to the contrary, data should be cooked with care"); Geoffrey C. Bowker and Susan Leigh Star, eds., *Sorting Things Out: Classification and Its Consequences* (Cambridge, Mass.: MIT Press, 1999); Rob Kitchin, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences* (London: Sage, 2014).

27. Ahead of his time, in 2011 Ken Farrall was the frst person I heard making this argument, though unfortunately he did not publish it.

28. Katherine Strandburg has dubbed this as "data fcation," that is, "the recording, aggregation, and organization of information into a form that can be used for data mining" ("Monitoring, Data fcation," 5).

29. Joris van Hoboken, "From Collection to Use in Privacy Regulation? A Forward-Looking Comparison of European and US Frameworks for Personal Data Processing," in *Exploring the Boundaries of Big Data*, ed. Bart van der Sloot, Dennis Broeders, and Erik Schrijvers (Amsterdam: Amsterdam University Press, 2016), 233.

30. See, e.g., Richard P. Kusserow, "The Government Needs Computer Matching to Root Out Waste and Fraud," *Communications of the ACM* 27, no. 6 (June 1984): 542–45, and John Shattuck, "Computer Matching Is a Serious Threat to Individual Rights," *Communications of the ACM* 27, no. 6 (June 1984): 538–41.

31. I include impacts on people which are derived from algorithmic models generated from information that is not necessarily drawn from them, as discussed in Barocas and Nissenbaum, "Big Data's End Run."

32. See van Hoboken, "From Collection to Use," for a clear discussion of collection deregulation in relation to FIPs.

33. For instance, in 2013 Google faced lawsuits from six different European countries over the unifcation of its privacy policies across platforms. See Charles Arthur, "Google Facing Legal Threat from Six European Countries over Privacy," *Guardian*, April 2, 2013, https://www.theguardian.com/technology/2013/apr/02/google-privacy-policy-legal-threat-europe. For a related discussion of the disjuncture of privacy norms and political economy, see Helen Nissenbaum, "Respecting Context to Protect Privacy: Why Meaning Matters," *Science and Engineering Ethics*, July 12, 2015, http://link.springer.com/article/10.1007%2Fs11948-015-9674-9.

34. See, e.g., Amit Datta, Michael Carl Tschantz, and Anupam Datta, "Automated Experiments on Ad Privacy Settings," *Proceedings on Privacy Enhancing Technologies 2015* 1 (2015): 92–112.

35. See Kirsten Martin, "Privacy Notices as Tabula Rasa: An Empirical Investigation into How Complying with a Privacy Notice Is Related to Meeting Privacy Expectations Online," *Journal of Public Policy and Marketing* 34, no. 2 (Fall 2015): 210–27. See also Aleecia McDonald and Lorrie Faith Cranor, "The Cost of Reading Privacy Policies," *I/S: A Journal of Law and Policy for the Information Society* 4, no. 3 (2008): 540–65. See also Joel R. Reidenberg and Lorrie Faith Cranor, "Can User Agents Accurately Represent Privacy Policies?" Discussion Draft 1.0. August 30, 2002. http://papers.ssrn.com/sol3/papers.cfm?abstractid=328860.

36. 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): 1027–72; Joel R. Reidenberg et al., "Privacy Harms and the Effectiveness of the Notice and Choice Framework," Fordham Law Legal Studies Research Paper No. 2418247, March 29, 2014, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2418247.

37. See Helen Nissenbaum, "A Contextual Approach to Privacy Online," *Daedalus* 140, no. 4 (Fall 2011): 32–48; Barocas and Nissenbaum, "Big Data's End Run."

38. See, e.g., Brendan Meeder, Jennifer Tam, Patrick G. Kelley, and Lorrie F. Cranor, "RT @IWantPrivacy: Widespread Violation of Privacy Settings in the Twitter Social Network," *Proceedings of Web 2.0 Security and Privacy (W2SP 2010)*, Oakland, Calif., 2010.

39. See Vincent Toubiana et al., "Adnostic: Privacy Preserving Targeted Advertising," paper presented at the 17th Annual Network and Distributed System Security Symposium, San Diego, March 2010.

40. Great Firewall of China, http://www.greatfrewallofchina.org/.

41. Signal demonstrates that one can choose not to collect data about users—that data collection is a choice rather than a technological requirement. All websites make certain choices, including Wikipedia, which can provide editors with a certain amount of anonymity but will deny access to Wikipedia to an editor using Tor unless the user requests special permission and reveals personal information to the Wikimedia Foundation while making her case. See "Advice to Users Using Tor," Wikipedia, December 17, 2016, https://en.wikipedia.org/wiki/Wikipedia:Advice_to_users_using_Tor.

42. "Unintended Consequences: Fifteen Years under the DMCA," *Electronic Frontier Foundation*, https://www.eff.org/pages/unintended-consequences-ffteen-years-under-dmca.

43. For example, LiveRamp's purchase of several publisher databases to track consumers across devices. See Kate Kaye, "Acxiom's LiveRamp Buys Two Publisher Data Firms in Race to I.D. Consumers Across Devices," *Adage*, November 17, 2016, http://adage.com/article/datadriven-marketing/acxiom-s-liveramp-buys-publisher-data-frms/306831/. See also Microsoft's acquisition of LinkedIn; Nick Wingfeld, "With LinkedIn, Microsoft Looks to Avoid Past Acquisition Busts," December 8, 2016, https://www.nytimes.com/2016/12/08/technology/with-linkedin-microsoft-looks-to-avoid-past-acquisition-busts.html.

44. Thomas P. Hughes, *Networks of Power* (Baltimore: Johns Hopkins University Press, 1983), 6.

45. 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.

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

47. Proposed in 1965 by the Social Science Research Council (SSRC) 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. *Report of the Committee on the Preservation and Use of Economic Data, Social Science Research Council* (April 1965). The SSRC urged the creation of a federal data center to make basic statistical data from all federal agencies available to nongovernmental users and other federal agencies.

48. For an overview, see, e.g., Nissenbaum, *Privacy in Context*, especially part 2.

49. See, e.g., *Sorrell v. IMS Health*, 564 U.S. 552 (2011). See also Julie E. Cohen, "The Zombie First Amendment," *William and Mary Law Review* 56, no. 3 (2015): 1119–58.

50. 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): 404–8.

51. 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): 1203–5.

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

53. See Michael Lewis, *Moneyball: The Art of Winning an Unfair Game* (New York: W. W. Norton, 2004).

54. See Duhigg, "How Companies Learn Your Secrets."

55. This was recently discussed at the Harnessing "Big Data" to Stop HIV conference cohosted by the NIAID Division of AIDS, NIMH Division of AIDS Research, NIH Big Data to Knowledge, and the Bill and Melinda Gates Foundation. See "Harnessing

Big Data to Stop HIV," National Institutes of Health, accessed June 13, 2018, https://www.datascience.nih.gov/node/249.

56. See, e.g., "IBM Watson Health," IBM Think, http://www.ibm.com/smarterplanet/us/en/think/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* (Washington, D.C.: U.S. Department of Education, 2012), https://tech.ed.gov/wp-content/uploads/2014/03/edm-la-brief.pdf.

57. See, e.g., "IBM Watson Health"; Rachel Willcox, "Big-Data Analytics: The Power of Prediction," *Public Finance*, January 27, 2016, http://www.publicfnance.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, 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, http://www.scientifccomputing.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, 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, 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: W. W. Norton, 2014).

58. Executive Office of the President National Science and Technology Council Committee on Technology, "Preparing for the Future of Artifcial Intelligence," *Office of Science and Technology Policy*, October 2016, https://www.whitehouse.gov/sites/default/fles/whitehouse_fles/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf.

59. See, e.g., Tene and Polonetsky, "Big Data for All"; Executive Office of the President, *Big Data*.

60. 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 put protocols in place that required an articulation of benefts while allowing stakeholders (or their representatives) to identify potential harms.

61. 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): 662–79; Neil M. Richards and Jonathan H. King, "Big Data Ethics," *Wake Forest Law Review* 49, no. 2 (Summer 2014): 393–432.

62. 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, no. 3 (2016): 671–732.

63. See boyd and Crawford, "Critical Questions for Big Data."

64. See Tad Friend, "Sam Altman's Manifest Destiny," *New Yorker*, October 10, 2016, http://www.newyorker.com/magazine/2016/10/10/sam-altmans-manifest-destiny. "Y Combinator has even begun using an A.I. bot, Hal9000, to help it sift admission applications: the bot's neural net trains itself by assessing previous applications and those companies' outcomes. 'What's it looking for?' [Tad Friend] asked Altman. 'I have no idea,' he replied. 'That's the unsettling thing about neural networks—you have no idea what they're doing, and they can't tell you.'" My thanks to Ira Rubenstein for drawing this to my attention.

65. See Jenna Burrell, "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms," *Big Data and Society*, January 5, 2016.

66. 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.

67. See Pasquale, *Black Box Society*.

68. See Pasquale, *Black Box Society*. See also Cathy O'Neill, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York: Crown, 2016).

69. See, e.g., Karen Levy, "The Future of Work: What Isn't Counted Counts," *Pacific Standard*, August 3, 2015, 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, http://www.nytimes.com/interactive/2014/08/13/us/starbucks-workers-scheduling-hours.html.

70. An (albeit extreme) example of how this information may be used is the "social credit system" that China wishes to roll out nationwide by 2020. See Josh Chin and Gillian Wong, "China's New Tool for Social Control: A Credit Rating for Everything," *Wall Street Journal*, November 28, 2016, http://www.wsj.com/articles/chinas-new-tool-for-social-control-a-credit-rating-for-everything-1480351590.

71. See, e.g., William A. Gorton, "Manipulating Citizens: How Political Campaigns' Use of Behavioral Social Science Harms Democracy," *New Political Science* 38, no. 1 (2016); "Cross-Device Offers Political Advertisers Great Promise—and Signifcant Challenges," *AdExchanger*, January 12, 2016, https://adexchanger.com/politics/cross-device-offers-political-advertisers-great-promise-and-signifcant-challenges/.

72. See, e.g., Ira Rubinstein et al., "Data Mining and Internet Profling: Emerging Regulatory and Technological Approaches," *University of Chicago Law Review* 75, no. 1 (2008): 261–85.

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

74. See https://www.niaid.nih.gov/about/organization/daids/Pages/big-data.aspx.

75. See Bernard Marr, "The 7 Most Data-Rich Companies in the World," *Data Science Central*, April 18, 2015, http://www.datasciencecentral.com/profles/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, 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.

76. See, e.g., boyd and Crawford, "Critical Questions for Big Data"; Tene and Polonetsky, "Big Data for All." Mandatory data retention differs from country to country; see https://www.eff.org/issues/mandatory-data-retention/us.

77. See Craig Konnoth, "Health Information Equity," *Penn Law Review* 165, no. 6 (2017): 1317–76.

78. To understand how serious this motivation is, consider Facebook's confrontation with ad blockers; see Josh Constine, "Facebook Rolls Out Code to Nullify AdBlock Plus' Workaround Again," *TechCrunch*, August 11, 2016, https://techcrunch.com/2016/08/11/friendblock/.

79. Compare Google's monitoring of Gmail inboxes for child pornography. With the practice of scanning emails for advertising keywords, "Google's creepy data practices have helped police catch who they think is an even bigger creep." Kevin Roose, "Google Detected a User Sending Child Porn from His Gmail Account and Alerted the Police," *New York Magazine*, August 4, 2014, http://nymag.com/daily/intelligencer/2014/08/google-scans-users-email-fnds-child-porn.html. See also Samuel Gibbs, "Gmail Does Scan All Emails, New Google Terms Clarify," *Guardian*, April 15, 2014, https://www.theguardian.com/technology/2014/apr/15/gmail-scans-all-emails-new-google-terms-clarify.

80. Experts seem to be unanimous in worrying about the egregious wealth disparities in the United States and around the world. Perhaps data disparities are at the root of the issue. This seems to be an important issue worth studying.

81. "Increasingly, it has begun to seem as though there is one set of rules for the ordinary consumer and institutional investors serving that consumer and a very different set for the financial cognoscenti." Julie E. Cohen, "The Regulatory State in the Information Age," *Theoretical Inquiries in Law* 17, no. 2 (2016): 15.

82. See Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock, "Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks," *Proceedings of the National Academy of Sciences* 111, no. 24 (2014): 8788–90; Adam Tanner, "How Data Brokers Make Money Off Your Medical Records," *Scientific American*, February 1, 2016, https://www.scientifcamerican.com/article/how-data-brokers-make-money-off-your-medical-records/; Erin Digitale, "On the Records: Tapping

into Stanford's Mother Lode of Clinical Information," *Stanford Medicine* (Summer 2012), http://sm.stanford.edu/archive/stanmed/2012summer/article5.html.

83. See, e.g., Bev Harris et al., *Black Box Voting: Ballot Tampering in the 21st Century* (High Point, N.C.: Plan Nine, 2003).

84. See Executive Office of the President, *Big Data*, 40–43.

85. See Joseph Turow et al., "Americans Reject Tailored Advertising and Three Activities That Enable It," University of Pennsylvania Departmental Papers of the Annenberg School of Communication, September 2009, http://repository.upenn.edu/cgi/viewcontent.cgi?article=1138&context=asc_papers.

86. See Privacy Rights Clearing House, "Data Breaches," accessed October 7, 2016, https://www.privacyrights.org/data-breaches.

87. Nicole Perlroth, "Yahoo Says Hackers Stole Data on 500 Million Users in 2014," *New York Times*, September 22, 2016, http://www.nytimes.com/2016/09/23/technology/yahoo-hackers.html?_r=0. BBC, "Yahoo 'State' Hackers Stole Data from 500 Million Users," September 23, 2016, http://www.bbc.com/news/world-us-canada-37447016. Harriet Taylor, "Yahoo CEO Mayer Knew About Data Breach in July: Report," CNBC, September 23, 2016, http://www.cnbc.com/2016/09/23/yahoo-ceo-mayer-knew-about-data-breach-in-july-report.html. Jeff John Roberts, "Yahoo Has Been Hacked: What You Need to Know," Fortune, September 22, 2016, http://fortune.com/2016/09/22/yahoo-hack-qa/.

88. Courts have difficulty dealing with litigation involving data breaches, as a central component of standing (a required element for judicial review) is an articulation of harm or injury. For an in-depth explanation of this difficulty in the courts, see Daniel J. Solove and Danielle Keats Citron, "Risk and Anxiety: A Theory of Data Breach Harms," December 14, 2016, https://ssrn.com/abstract=2885638.

89. See Pasquale, *Black Box Society*, 149.

90. See, e.g., Federal Trade Commission, "Big Data: A Tool for Inclusion or Exclusion: Understanding the Issues," FTC Report, January 2016, https://www.ftc.gov/system/fles/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf.

91. See David Robinson et al., "Civil Rights, Big Data and Our Algorithmic Future," *Upturn*, September 2014, https://bigdata.fairness.io/.

92. See, e.g., the 1974 Fair Credit Billing Act. I am indebted to Finn Brunton for pointing out these connections.

93. For an overview of state legislation regarding security breach notifcations, see "Security Breach Notifcation Laws," *National Conference of State Legislatures*, March 29, 2018, http://www.ncsl.org/research/telecommunications-and-information-technology/security-breach-notifcation-laws.aspx. For two recent high-profle hacking cases, see David E. Sanger et al., "Attack Gave Chinese Hackers Privileged Access to U.S. Systems," *New York Times*, June 20, 2015, 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, http://dealbook.nytimes.com/2014/10/02/jpmorgan-discovers-further-cyber-security-issues/.

94. See 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 (Washington, D.C.: Author, 1973), http://www.justice.gov/opcl/docs/rec-com-rights.pdf.

95. This, of course, is not unique to the United States. Equivalent principles exist in other liberal democracies; see, e.g., Article 8 of the Canadian Charter of Rights and Freedoms (stipulating that "Everyone has the right to be secure against unreasonable search or seizure").

96. For a discussion of predictive policing, see Sarah Brayne et al., "Predictive Policing," *Data and Civil Rights*, October 27, 2015, 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, http://www.wired.com/insights/2013/08/predictive-policing-using-machine-learning-to-detect-patterns-of-crime/.

97. The 1974 Privacy Act was one effort to maintain separation among the databases accrued by disparate government agencies.

98. For a comprehensive overview of the revelations, see *Guardian*, "The NSA Files," accessed October 7, 2016, https://www.theguardian.com/us-news/the-nsa-fles.

99. Compare the concept of *gezeihra* as a "fence around the Torah" in Jewish law; see "Halakhah: Jewish Law," *Judaism 101*, accessed February 9, 2016, http://www.jewfaq.org/halakhah.htm.

100. 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, Mass.: MIT Press, 2015), 79–80.

101. Finn Brunton and I have argued that information yields power to the holder, which is particularly dangerous when that holder is already powerful. See Brunton and Nissenbaum, *Obfuscation*.

102. E.g., the 1968 Omnibus Crime Control and Safe Streets Act, 1974 Privacy Act, 1988 Computer Matching and Privacy Protection Act, 1996 Health Insurance Portability and Accountability Act, 1999 Financial Services Modernization (Gramm-Leach-Bliley) Act, etc.

103. See *Birchfield v. North Dakota*, 136 S. Ct. 2160, 2178 (2016). Thanks to Kiel Brennan-Marquez for drawing my attention to this case. For further discussion on anxiety as an articulation of injury sufficient to satisfy standing requirements, see Solove and Citron, "Risk and Anxiety."

104. "The ability to harvest the wealth of information contained in biomedical Big Data will advance our understanding of human health and disease." "Big Data to Knowledge," National Institutes of Health, accessed January 19, 2017, https://datascience.nih.gov/bd2k.

105. See Federal Trade Commission, "Data Brokers: A Call for Transparency and Accountability," May 2014, https://www.ftc.gov/system/fles/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf. See also Katherine J. Strandburg, "Home, Home on

the Web and Other Fourth Amendment Implications of Techno-Social Change," *Maryland Law Review* 70, no. 3 (2011): 614–80.

106. *United States v. Jones*, 565 U.S. 400, 415 (2012) (Sotomayor, J., concurring).

107. See David C. Gray and Danielle Keats Citron, "The Right to Quantitative Privacy," *Minnesota Law Review* 98 (2013): 62–144; see also Helen Nissenbaum, "Toward an Approach to Privacy in Public: Challenges of Information Technology," *Ethics and Behavior* 7, no. 3 (1997): 207–19; Daniel Solove, *The Digital Person: Technology and Privacy in the Information Age* (New York: NYU Press, 2004), chap. 8.

108. See Mayer-Schönberger and Cukier, *Big Data*.

109. See discussion of this and related points in Barocas and Nissenbaum, "Big Data's End Run," and also Katherine J. Strandburg, "Freedom of Association in a Networked World: First Amendment Regulation of Relational Surveillance," *Boston College Law Review* 49 (2008): 741–821.

110. For a discussion of how this played out in the Volkswagen scandal, see generally Cohen, "Regulatory State."

111. See Philip N. Howard, *Pax Technica: How the Internet of Things May Set Us Free or Lock Us Up* (New Haven, Conn.: Yale University Press, 2015). See also Michael D. Birnhack and Niva Elkin-Koren, "The Invisible Handshake: The Reemergence of the State in the Digital Environment," *Virginia Journal of Law and Technology* 8, no. 2 (Summer 2003): 1–57; Kiel Brennan-Marquez, "Private Searches in an Age of Big Data," July 12, 2016, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2808829.

112. See Tech 2, "EU Says Firms Like Google and Facebook Must Meet Privacy Laws," Firstpost, June 7, 2014, http://tech.frstpost.com/news-analysis/eu-says-frms-like-google-facebook-must-meet-privacy-rules-225348.html.

113. 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, http://www.nytimes.com/2015/08/22/world/middleeast/prospect-of-self-inspections-by-iran-feeds-opposition-to-nuclear-deal.html.

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

References

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

Anderson, Chris. "The End of Theory: The Data Deluge Makes the Scientifc Method Obsolete." *Wired*, June 23, 2008. https://www.wired.com/2008/06/pb-theory/.

Arthur, Charles. "Google Facing Legal Threat from Six European Countries over Privacy." *Guardian*, April 2, 2013. https://www.theguardian.com/technology/2013/apr/02/google-privacy-policy-legal-threat-europe.

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, no. 3 (2016): 671–732.

BBC. "Yahoo 'State' Hackers Stole Data from 500 Million Users." September 23, 2016. http://www.bbc.com/news/world-us-canada-37447016.

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

Birnhack, Michael D., and Niva Elkin-Koren. "The Invisible Handshake: The Reemergence of the State in the Digital Environment." *Virginia Journal of Law and Technology* 8, no. 2 (Summer 2003): 1–57.

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): 662–79.

Bowker, Geoffrey C. *Memory Practices in the Sciences*. Cambridge, Mass.: MIT Press, 2005.

Bowker, Geoffrey C., and Susan Leigh Star, eds. *Sorting Things Out: Classification and Its Consequences*. Cambridge, Mass.: MIT Press, 1999.

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

Brennan-Marquez, Kiel. "Fourth Amendment Fiduciaries." *Fordham Law Review* 84, no. 2 (2015): 611–59.

———. "Private Searches in an Age of Big Data." July 12, 2016. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2808829.

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

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

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

Burrell, Jenna. "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms." *Big Data and Society*, January 5, 2016.

Calo, Ryan "Against Notice Skepticism in Privacy (and Elsewhere)." *Notre Dame Law Review* 87, no. 3 (2013): 1027–72.

Cate, Fred H. "The Failure of Fair Information Practice Principles." In *Consumer Protection in the Age of the Information Economy*, edited by Jane K. Winn, 341–78. Burlington, Vt.: Ashgate, 2006.

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

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.

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

Cohen, Julie E. "The Zombie First Amendment." *William and Mary Law Review* 56, no. 3 (2015): 1119–58.

Constine, Josh. "Facebook Rolls Out Code to Nullify AdBlock Plus' Workaround Again." *Tech Crunch*, August 11, 2016. https://techcrunch.com/2016/08/11/friendblock/.

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.

Datta, Amit, Michael Carl Tschantz, and Anupam Datta. "Automated Experiments on Ad Privacy Settings." *Proceedings on Privacy Enhancing Technologies 2015* 1 (2015): 92–112.

Digitale, Erin. "On the Records: Tapping into Stanford's Mother Lode of Clinical Information." *Stanford Medicine* (Summer 2012). http://sm.stanford.edu/archive/stanmed/2012summer/article5.html.

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

Executive Office of the President. *Big Data: Seizing Opportunities, Preserving Values*. North Charleston, S.C.: CreateSpace, 2014.

Federal Trade Commission. "Big Data: A Tool for Inclusion or Exclusion: Understanding the Issues." FTC Report. January 2016. https://www.ftc.gov/system/fles/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf.

———. "Data Brokers: A Call for Transparency and Accountability.," FTC Report. May 2014. https://www.ftc.gov/system/fles/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf.

Feyerabend, Paul. *Against Method*. 4th ed.. New York: Verso, 2010.

Gitelman, Lisa, ed. *"Raw Data" Is an Oxymoron*. Cambridge, Mass.: MIT Press, 2013.

Gray, David C., and Danielle Keats Citron. "The Right to Quantitative Privacy." *Minnesota Law Review* 98 (2013): 62–144.

Guardian. "The NSA Files." Accessed October 7, 2016. https://www.theguardian.com/us-news/the-nsa-fles.

"Halakhah: Jewish Law." *Judaism 101*. Accessed February 9, 2016. http://www.jewfaq.org/halakhah.htm.

Heller, Christian. *Post-Privacy: Prima leben ohne Privatsphäre*. 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, Conn.: Yale University Press, 2015.

Hvistendahl, Mara. "Can Predictive Policing Prevent Crime Before It Happens?" *Science*, September 28, 2016. http://www.sciencemag.org/news/2016/09/can-predictive-policing-prevent-crime-it-happens.

"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. http://www.scientifccomputing.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. http://www.nytimes.com/interactive/2014/08/13/us/starbucks-workers-scheduling-hours.html.

Kitchin, Rob. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: Sage, 2014.

Konnoth, Craig. "Health Information Equity." *University of Pennsylvania Law Review* 165, no. 6 (2017): 1317–76.

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*, edited by Mireille Hildebrandt and Katja De Vries, 196–220. New York: Routledge, 2013.

Kramer, Adam D. I., Jamie E. Guillory, and Jeffrey T. Hancock. "Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks." *Proceedings of the National Academy of Sciences* 111, no. 24 (2014): 8788–90.

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

Kusserow, Richard P. "The Government Needs Computer Matching to Root Out Waste and Fraud." *Communications of the ACM* 27, no. 6 (June 1984): 542–45.

Lazar, David, Ryan Kennedy, Gary King, and Alessandro Vespignani. "The Parable of Google Flu: Traps in Big Data Analysis." *Science* 343, no. 6176 (2014): 1203–5.

Lessig, Lawrence. "The Law of the Horse: What Cyberlaw Might Teach." *Harvard Law Review* 113, no. 2 (1999): 501–49.

Levy, Karen. "The Future of Work: What Isn't Counted Counts." *Pacific Standard*, August 3, 2015. 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: W. W. Norton, 2004.

Manovich, Lev. "Trending: The Promises and the Challenges of Big Social Data." Manovich [blog], April 28, 2011. 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. http://www.datasciencecentral.com/profles/blogs/the-7-most-data-rich-companies-in-the-world.

Martin, Kirsten. "Privacy Notices as Tabula Rasa: An Empirical Investigation into How Complying with a Privacy Notice Is Related to Meeting Privacy Expectations Online." *Journal of Public Policy and Marketing* 34, no. 2 (Fall 2015): 210–27.

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

McDonald, Aleecia, and Lorrie Faith Cranor. "The Cost of Reading Privacy Policies." *I/S: A Journal of Law and Policy for the Information Society* 4, no. 3 (2008): 540–65.

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

Meeder, Brendan, Jennifer Tam, Patrick G. Kelley, and Lorrie F. Cranor. "RT @IWantPrivacy: Widespread Violation of Privacy Settings in the Twitter Social Network." *Proceedings of Web 2.0 Security and Privacy (W2SP 2010)*. Oakland, Calif., 2010.

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

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

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

———. "Respect for Context as a Benchmark for Privacy Online: What It Is and Isn't." In *Social Dimensions of Privacy: Interdisciplinary Perspectives*, edited by Beate Roessler and Dorota Mokrosinska, 278–302. Cambridge: Cambridge University Press, 2015.

———. "Respecting Context to Protect Privacy: Why Meaning Matters." *Science and Engineering Ethics*, July 12, 2015. http://link.springer.com/article/10.1007%2Fs11948-015-9674-9.

———. "Toward an Approach to Privacy in Public: Challenges of Information Technology." *Ethics and Behavior* 7, no. 3 (1997): 207–19.

O'Neill, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown, 2016.

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

Perlroth, Nicole. "Yahoo Says Hackers Stole Data on 500 Million Users in 2014." *New York Times*, September 22, 2016. http://www.nytimes.com/2016/09/23/technology/yahoo-hackers.html?_r=0.

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

President's Council of Advisors on Science and Technology (PCAST). *Big Data and Privacy: A Technological Perspective*. North Charleston, S.C.: CreateSpace, 2014.

Privacy Rights Clearing House. "Data Breaches." Accessed October 7, 2016. https://www.privacyrights.org/data-breaches.

Reidenberg, Joel R., and Lorrie Faith Cranor, "Can User Agents Accurately Represent Privacy Policies?" Discussion Draft 1.0. August 30, 2002. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=328860.

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

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

Roberts, Jeff John. "Yahoo Has Been Hacked: What You Need to Know." *Fortune*, September 22, 2016. http://fortune.com/2016/09/22/yahoo-hack-qa/.

Robinson, David, Harlan Yu, and Aaron Rieke. "Civil Rights, Big Data and Our Algorithmic Future." *Upturn*, September 2014. https://bigdata.fairness.io/.

Rubens, Paul. "Is Big Data Dating the Key to Long-Lasting Romance?" BBC, March 25, 2014. 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, Ronald D. Lee, and Paul M. Schwarz. "Data Mining and Internet Profiling: Emerging Regulatory and Technological Approaches." *University of Chicago Law Review* 75, no. 1 (2008): 261–85.

Rudin, Cynthia. "Predictive Policing: Using Machine Learning to Detect Patterns of Crime." *WIRED*, August 2013. 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. http://www.nytimes.com/2015/08/22/world/middleeast/prospect-of-self-inspections-by-iran-feeds-opposition-to-nuclear-deal.html.

Sanger, David E., Nicole Perlroth, and Michael D. Shear. "Attack Gave Chinese Hackers Privileged Access to U.S. Systems." *New York Times*, June 20, 2015. 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. 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, Calif.: O'Reilly Media, 2014.

"Security Breach Notifcation Laws." National Conference of State Legislatures, March 29, 2018. http://www.ncsl.org/research/telecommunications-and-information-technology/security-breach-notifcation-laws.aspx.

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

Shattuck, John. "Computer Matching Is a Serious Threat to Individual Rights." *Communications of the ACM* 27, no. 6 (June 1984): 538–41.

Silver, Nate. *The Signal and the Noise: Why So Many Predictions Fail—But Some Don't*. New York: Penguin Books, 2012.

Silver-Greenberg, Jessica, Matthew Goldstein, and Nicole Perlroth. "JPMorgan Chase Hacking Affects 76 Million Households." *New York Times*, October 2, 2014. 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. http://archive.wired.com/politics/law/news/1999/01/17538.

Strandburg, Katherine J. "Freedom of Association in a Networked World: First Amendment Regulation of Relational Surveillance." *Boston College Law Review* 49 (2008): 741–821.

———. "Home, Home on the Web and Other Fourth Amendment Implications of Techno-Social Change." *Maryland Law Review* 70, no. 3 (2011): 614–80.

———. "Monitoring, Datafcation, and Consent: Legal Approaches to Privacy in the Big Data Context." In *Privacy, Big Data and the Public Good: Frameworks for Engagement*, edited by Julia Lane, Victoria Stodden, Stefan Bender, and Helen Nissenbaum, 5–43. Cambridge: Cambridge University Press, 2015.

Tanner, Adam. "How Data Brokers Make Money Off Your Medical Records." *Scientifc American*, February 1, 2016. https://www.scientifcamerican.com/article/how-data-brokers-make-money-off-your-medical-records/.

Taylor, Harriet. "Yahoo CEO Mayer Knew About Data Breach in July: Report." CNBC, September 23, 2016. http://www.cnbc.com/2016/09/23/yahoo-ceo-mayer-knew-about-data-breach-in-july-report.html.

Tech 2. "EU Says Firms Like Google and Facebook Must Meet Privacy Laws." Firstpost, June 7, 2014. http://tech.frstpost.com/news-analysis/eu-says-frms-like-google-facebook-must-meet-privacy-rules-225348.html.

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): 239–73.

Toubiana, Vincent, Arvind Narayanan, Dan Boneh, Helen Nissenbaum, and Solon Barocas. "Adnostic: Privacy Preserving Targeted Advertising." Paper presented at the 17th Annual Network and Distributed System Security Symposium, San Diego, March 2010.

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

Turow, Joseph, Jennifer King, Chris Jay Hoofnagle, Amy Bleakley, and Michael Hennessey. "Americans Reject Tailored Advertising and Three Activities That Enable It." University of Pennsylvania Departmental Papers of the Annenberg School of Communication, September 2009. http://repository.upenn.edu/cgi/viewcontent.cgi?article=1138&context=asc_papers.

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*. Washington, D.C.: Author, 1973. http://www.justice.gov/opcl/docs/rec-com-rights.pdf.

Van Hoboken, Joris. "From Collection to Use in Privacy Regulation? A Forward-Looking Comparison of European and US Frameworks for Personal Data Processing." In *Exploring the Boundaries of Big Data*, edited by Bart van der Sloot, Dennis Broeders, and Erik Schrijvers, 231–59. Amsterdam: Amsterdam University Press, 2016.

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): 1049–124.

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

White, Ryen W., Nicholas P. Tatonetti, Nigam H. Shah, Russ B. Altman, and Eric Horvitz. "Web-Scale Pharmacovigilance: Listening to Signals from the Crowd." *Journal of the American Medical Informatics Association* 20, no. 3 (2013): 404–8.

Willcox, Rachel. "Big-Data Analytics: The Power of Prediction." *Public Finance*, January 27, 2016. http://www.publicfnance.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. 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.