Coercion and Distribution in Big Data: A Law and Political Economy Approach in Privacy Law

Salome Viljoen, Fellow, Berkman Klein Center for Internet and Society, Harvard University

Abstract

The big dollars of Big Data result in widespread data collection that fuel new technological innovations but have rendered traditional notions of privacy largely obsolete. And what privacy laws do exist fail to adequately capture the full extent of harms that result from users’ relationship to the collection and dissemination of their data. This paper uses a distributional analysis to examine how privacy law has facilitated the collection of user data, how the benefits and risks of such collection are unevenly distributed among individuals in society, and what role it may play in efforts to shift market power away from data collectors and toward data creators.

Note to Ostrom workshop participants

This paper is a working draft that is under active revision. Part I reviews relevant trends in the data marketplace and Part II introduces a distributional analysis of the current dominant legal regime governing data collection by private actors. Part III outlines a set of three proposed further research projects to address the issues highlighted by Parts I and II. Although this draft is still very much a work in progress, I welcome your comments and suggestions, and look forward to our discussion.
Introduction

In 1923, Robert Hale wrote about the “so-called individualist” for whom “[t]he practical function of economic theory is merely to prove to statesmen the wisdom of leaving [economic] matters alone, not to aid them in the process of interfering. And . . . in domestic affairs, [statesmen] should make no effort to control the natural working of economic events.” Much the same could be said of technologists today. Like the industrialists of the early 20th century, the technologists of the early 21st century generally find legal restrictions inhibiting the collection of data useless in the face of the ubiquity and immense economic value of the practice.

They may be right. The rise of big data analytics has resulted in widespread data collection that fuels new technological innovations but has rendered traditional notions of privacy largely obsolete. And what privacy laws do exist fail to adequately capture the full extent of harms that result from users’ relationship to the collection and dissemination of their data. Nevertheless, neither advocates nor legal scholars should succumb to the temptation to declare that ‘privacy is dead’; privacy provides an ethical and legal basis for understanding and enforcing the limits to what others may know about us and how such information may be obtained and used. As such, reformed privacy law is of central importance in articulating and limiting the harms of ubiquitous information extraction.

This paper uses a distributional analysis to examine how privacy law has facilitated the collection of user data, how the benefits and risks of such collection are unevenly distributed among individuals in society, and what role it may play in efforts to shift market power away from data collectors and toward data creators. Part I will review trends in the market of data collection. The last decade has been characterized by the growth and consolidation of tech companies whose business models depend on the commodification of user data. These companies engage in such a ubiquitous practice of data collection that many in the tech industry claim privacy is dead. Part II examines how privacy law distributes power between data collectors and data creators. This distributional approach comes from the work of Legal Realist scholars, particularly that of Robert Hale, and is has recently been revived in legal

theoretical work pioneering a “law and political economy” approach to legal scholarship. It will analyze how current privacy law distributes power among collectors of data and creators of data. Ubiquitous data collection results from the relative market power between data creators and data collectors, and the laws that condition such relationships. Privacy law’s intervention in private sector data collection to date is significantly limited. Notice-and-consent regimes belie an intuition that non-state actors pose only a particularized, narrow risk of inflicting privacy harm. Privacy law does not account for the immense economic imperatives that drive data collection and how that incentivizes exploitation, nor does it consider the disparate risks different users face from data collection. As a result, legally actionable privacy harms represent just a fraction of the full breadth of harms users may face as a result of the datafication of their lives. Part III will review proposed privacy law interventions and analyze their potential distributive effects.

Part I: Trends in Big Data

Powerful new data analysis techniques are combining advancements in information-processing hardware with new machine learning techniques to identify patterns, develop predictions, and continually improve in response to new data. The result, colloquially known as Big Data “[is] a technique for converting voluminous, heterogeneous flows of physical, transactional, and behavioral information about people (or about anything else) into a particular, highly data-intensive type of knowledge.”\(^2\) In 2008 Randal E. Bryant, Randy H. Katz, and Edward D. Lazowika wrote that big data computing “can and will transform the activities of companies, scientific researchers, medical practitioners, and our nation’s defense and intelligence operations…. Big-data computing is perhaps the biggest innovation in computing in the last decade. We have only begun to see its potential to collect, organize, and process data in all walks of life.”\(^3\)

Big Data may fuel new technological innovations, but the promise such new technologies hold cannot be disentangled from the current structural reality of the information market—concentrated informational power that underpins and amplifies concentrated economic


\(^3\) Randal E. Bryant, Randy H. Katz, and Edward D. Lazowska, “Big-Data Computing: Creating Revolutionary Breakthroughs in Commerce, Science and Society”
power. Big Data holds immense analytical promise, but much of the most sophisticated analytical innovations are also driving the consolidation of immense power in the hands of platform companies who collect data and sell the insights they hold to advertisers, or use it to achieve market dominance.

*The Market of Data Collection*

Never in the course of human history have so few known so much about so many. A handful of companies – Google, Facebook, Amazon, Microsoft and Apple – collect the data of billions of users, and use that data to make fine-grained predictions about us. Amazon knows how long you hover your mouse on a particular item and how many days an item sits in your shopping basket before you purchase it. Google, Facebook and Amazon are able to see the blogs you visit before looking for those same items through a search engine, or connect the advertisements you saw on their platform to your search for those items later on. Google knows your location since you began using Google on your phone. It knows everything you’ve searched for and which searches you’ve deleted. It holds on to your entire YouTube history. Facebook know every page, photo, and post you liked and every message you’ve sent. And with this data, companies train machine-learning algorithms to learn other things about us, too. From Facebook likes alone, these algorithms can infer things like ethnicity, religious and political views, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender.\(^4\)

---

\(^4\) Lina Khan, Amazon’s Antitrust Paradox, *Yale L. Journal* 126, 710 (2017)

\(^5\) See Michal Kosinski, David Stillwell and Thore Graepel, Private traits and attributes are predictable from digital records of human behavior, PNAS April 9, 2013; Zeynep Tufekci, “We’re building a dystopia just to make people click on ads,” Tiny Ted Talk, available at https://en.tiny.ted.com/talks/zeynep_tufekci_we_re_building_a_dystopia_just_to_make_people_click_on_ads
There’s big money in Big Data. In 2017 Facebook generated revenue of over $40 billion; Amazon recorded $178 billion and Alphabet (Google) $111 billion. Together Alphabet, Amazon, Apple, Facebook and Microsoft—the largest platform companies—have a combined market capitalization of about $3.5 trillion.6 Other companies like Netflix, Airbnb and Uber have leveraged their data-driven business models to create outsize impacts in their respective industries. Netflix has a market capitalization of $100 billion and accounts for 75% of online streaming.7 It now has more subscribers than cable television.8 AirBnB is valued at $31 billion, but in terms of awareness (measured by Google search), it is reaching similar levels to mainstays of the $550 billion hotel industry, and its market growth has grown by several percentage points each year. In less than a decade ridesharing apps like Uber and Lyft have all but displaced the taxi industry.

The Data Imperative

Platforms, digital advertisers and data brokers face strong economic incentives to undertake the mass collection and analysis of data; in various ways they depend on user surveillance and the targeted insights it generates for their revenue.9 Companies that are not in the business of monitoring and selling attention are also adopting ubiquitous data collection and analysis—companies like Amazon, Netflix and Uber commodify and monetize data differently from companies like Facebook and Google, but they face a similarly powerful

---

6 Based on calculations in April 2018 combined market capitalization was $3.494.23.
8 Fortune, Netflix has more subscribers than Cable TV, June 25, 2017, available at http://fortune.com/2017/06/15/netflix-more-subscribers-than-cable/
9 Google and Facebook together now command approximately 20% of global advertising revenue, 65% of digital advertising revenue and 85% of every new dollar spent on advertising.
economic incentive to do so. These companies leverage their superior data-analysis capacities (based on the mass surveillance of users) to scale, price target and lock in users to gain market control.10

As technology companies have used data-driven techniques to disrupt non-traditional tech markets, they push non-technology companies to adopt similar practices or risk their market positions. Pharmacies, car companies and appliance makers are all embracing the collection of user data and hiring data scientists in the arms-race to maintain and expand their market positions.11 Much like the rise and dominance of financial instruments in the US economy led to the “financialization” of non-finance industries, the rise and dominance of data as an asset is beginning to lead to the “datafication” of non-technology industries, and the economy as a whole.12

10 It is suspected that Amazon is also able to tailor prices to individual consumers, known as first-degree price discrimination. There is no public evidence that Amazon is currently engaging in personalized pricing, but online retailers generally are devoting significant resources to analyzing how to implement it. See, Lina Khan, Amazon’s Antitrust Paradox, at 763; at 763, fn 276“As a group of authors stated in a recent letter to the Justice Department: [T]he corporation’s detailed knowledge of the buying habits of millions of readers—which it amasses through a minute-by-minute tracking of their actions online—puts it in a powerful position to use such ‘personalized’ pricing and marketing to influence the decisions of readers and thereby extract the most amount of cash possible from each individual. Letter from Authors United to William J. Baer, Assistant Att’y Gen., Antitrust Div., Dep’t of Justice (July 14, 2015), http://www.authorsunited.net/july/longdocument.html [http://perma.cc/L9RN-YESR]; see also David Streitfeld, Accusing Amazon of Antitrust Violations, Authors and Booksellers Demand Inquiry, N.Y. TIMES (July 13, 2015), http://www.nytimes.com/2015/07/14/technology/accusing-amazon-of-antitrust-violations-authors-and-booksellers-demand-us-inquiry.html [http://perma.cc/G8QF-5LYY] (reporting on the Authors United letter to the Assistant Attorney General and its claim that Amazon seems to be “engag[ing] in content control” in its decisions to sell certain books).


This creates a profound change in economic organization around the harvesting of data. Personal information is transformed into a “fourth factor of production.” Transforming data into a store of value imposes design imperatives “that shape the emergence of smart mobile devices, wearable computing, and the Internet of Things,” all of which feature “seamless tracking” and “fine-trained measurement of patterns of behavior and attention” that work to transform our “[c]ommunications networks” into “sensing networks.” As more devices perform computing tasks in exchange for continual measurement, the nature of personal information becomes more “datafied” – translated into configurations and flows of data “best suited to analysis and commercial exploitation.”

This transformation signals a shift in the nature of personal information itself. Data is “now a kind of capital, on par with financial and human capital in creating new digital products and services.” Some scholars believe that for “most companies, their data is their single biggest asset” and that the success of companies like Uber and Amazon comes from their superior ability to leverage data as an asset. Treating data as a form of capital pushes companies to hoard, commodify, and monetize as much data as they can, in a competitive race to datafy valuable activities before rivals. German tech firm Siemens notes: “We need to understand that data is everywhere, and it is generated every second of the day. We need to understand data as an asset – and turn it into a value.”


14 Id at 142.
15 Id. Also, see generally Anne Helmond, The Platformization of the Web: Making Web Data Platform Ready, SOC. MEDIA & SOC’Y, July-Dec. 2015 at 1.
17 From Id, quoting Andrew W. Lo, Charles E. and Susan T. Harris Professor of Finance at the MIT Sloan School of Management.
18 Id.
20 Siemens, Siemens Smart Data, YouTube, Sept. 4, 2014, available at https://www.youtube.com/watch?v=ZxoO-DvHQRw
The mass collection and analysis of data, driven by business models that harvest data as a key asset, structures an entire marketplace around what can be called the “data imperative.” This imperative demands the extraction of as much data as possible, from as many sources as possible, and the development of new techniques to extract such data. It demands the increasing enclosure of information, from a free-flowing exchange to a one-way accumulation of a valuable, proprietary asset. It exports the platform logic of winner-take-all networks to more and more sectors of the economy, where the “winners” are those best able to use data capital to achieve market dominance. The concentration of informational power has become both a means and an end of the informational economy.

Part II: How Privacy Law Distributes Power

A century ago, Louis Brandeis coined the term the “curse of bigness,” to express the widespread concerns of Progressive Era reformers regarding the effect of “new mega-corporations trusts, monopolies” on “economic well-being.” Progressive Era legal scholars like Louis Brandeis and Robert Hale sought to interrogate and analyze how the law contributed to, and could work to undo, the structural dominance of concentrated private power. Increasing corporate concentration in the information platform economy has caused a number of scholars, journalists and policymakers to “revist[] this Progressive Era concern.”

---


22 See Rise of Big Data Capital, supra, for a discussion of the “Three Principles of Data Capital” (“Digitize and datafy key value-creating activities before rivals do.”) Cf. Paul Ohm, Broken Promises of Privacy, 57 UCLA L. Rev. 1701, 1736 (2010) arguing for the importance of privacy protections that honor the value of free information exchange (“The free flow of information fuels the modern economy, nourishes our hunger for knowledge, shines a light on the inner workings of powerful institutions and organizations, and represents an exercise of liberty.”)

23 See generally, Lina Khan, Amazon’s Antitrust Paradox, YALE L. JOURNAL 126 (2017); Maurice E. Stucke, Should We Be Concerned About Data-opolies?, GEORGETOWN LAW TECHNOLOGY REVIEW (forthcoming).


25 Id.
Much like the laissez-faire approach markets of the Progressive Era led to the structural dominance of industrial capital, the current set of laws governing the mass extraction of data has led to the structural dominance of informational capital. As the political philosopher Iris Young argues, structural domination arises “when social processes put large groups of persons under systemic threat of domination [. . .], at the same time that these processes enable others to [. . .] have a wide range of opportunities for developing and exercising capacities available to them.” The structure in which the dominance occurs “can seem involuntary and natural, and yet be the product of many independent policies and decisions which end up driving inequality.”

The effect of the data imperative has been to subject individuals to data extraction on the one hand, while the immense promise of analyzing and building from such data is controlled by a handful of powerful entities on the other. This continual one-way flow of information is the result of processes, largely occurring within the bounds of law, that drive the concentration of informational power. This market momentum is not particular but diffuse – the combined result of independent, competing companies operating inside legal norms. As a result, the legal regime governing data collection does not merely serve to police breaches and bad actors, but also underpins the general ecosystem of data practices of which breaches are merely symptomatic.

This should make us consider the role privacy law has played in the emergence of the informational economy. Is the notion of privacy that laws are meant to protect consistent with the scope of collection and degree of commodification that typify current Big Data practices? And if so, how have current privacy laws – laws meant to create limits to what can be known about us – facilitated this kind of structural concentration of informational power?

*Brief History of Privacy Law*

---

Privacy has long been invoked as a “substantive limit on state action.” While the term privacy is rightfully criticized for being “vague and evanescent” and “suffering from an embarrassment of meanings,” the instincts behind privacy harms include the desire to avoid observation, keep personal matters secret, prevent discriminatory action on the basis of personal information, and is also often extended to the individual’s choice to selectively share information without opening it up to general public use.

New technologies have long provoked anxiety about privacy. Louis Brandeis and Samuel Warren wrote “The Right to Privacy” in response to the “numerous mechanical devices” that had invaded the “sacred precincts of private and domestic life.” It is almost universally regarded as the origin of the American privacy torts, and the foundation of a “right to privacy” in American law more generally. The common law basis for a right to privacy, the authors asserted, was an already-existing right for each person to determine if, when and to what extent their “thoughts, sentiments and emotions shall be communicated

---

29 Kim Lane Scheppele, LEGAL SECRETS 184-85 (1988).
30 President’s Council of Advisors on Science and Technology Report on Big Data and Privacy (hereafter PCAST) at 2. Available at https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy__may_2014.pdf
32 See, e.g., Pavesich v. New England Life Ins. Co., 50 S.E. 68, 69 (Ga. 1905) (“[A] claim to a right of privacy, independent of a property or contractual right, or some right of a similar nature, had, up to [1890], never been recognized in terms in any decision.”); Roberson v. Rochester Folding Box Co., 64 N.E. 442, 443 (N.Y. 1902) (informing that the right to privacy has not “been asserted prior to about the year 1890, when it was presented with attractiveness, and no inconsiderable ability, in the Harvard Law Review”); Elbridge L. Adams, The Law of Privacy, 175 N. Am. Rev. 361, 364 (1902); Ken Gormley, One Hundred Years of Privacy, 1992 Wis. L. Rev. 1335, 1342-47; Prosser, supra note 4, at 383. For more recent pronouncements on the foundational and seminal nature of the article, see, for example, Anderson v. Romero, 72 F.3d 518, 521 (7th Cir. 1995)(including Judge Posner’s comment that “the legal concept of privacy... originated in a famous article by Warren and Brandeis”), and West v. Media Gen. Convergence, Inc., 53 S.W.3d 640, 642 (Tenn. 2001) (remarking in a false light invasion of privacy case that “Warren and... Brandeis... established the concept of the right to privacy in the common law”).
33 This paper uses “them/they/their” as a non-binary alternative to “she/he”, and “her/his”. 
Motivating this notion of harm was the idea that the use of information about one’s self is powerful and can be harmful, a violation not simply of a proprietary claim but of one’s “inviolate personality.”

But the landmark case establishing a constitutional right to privacy arose due to changing legal conceptions of the family, not new technological advances. Seventy-five years after Warren and Brandeis wrote their article, the Supreme Court in *Griswold v. Connecticut* struck down a Connecticut law banning contraception as a violation of “marital privacy.” Justice Douglas, writing for the majority, found a “right to privacy” in the “penumbras” and “emanations” of other constitutional protections. Concurring opinions by Justices Goldberg, Harlan, and White found protections of privacy articulated in the Ninth Amendment and the due process clause of the Fourteenth Amendment.

**Privacy as Autonomy**

In both its public and private legal protections, contemporary digital privacy law relies centrally on the notion of individual autonomy. This translates basic arguments securing individual freedom against state interference from contracts, torts and property law to the digital age. Private law stemmed from the liberal notion that the "state ought to and largely did in fact define the rules of law so as to guarantee the free exercise of individual will, subject to the constraint that willing actors respect the like rights of other willing actors."
The will theory underpinning legal principles was later updated in private law via autonomy,

---

34 Warren & Brandeis at 198.
35 Warren & Brandeis at 205
36 381 U.S. 479 (1965),
37 381 U.S. 479 (1965).
38 Justice Goldberg wrote a concurrence where he used the Ninth Amendment to support the ruling. He and Justice Harlan wrote a concurring opinion finding privacy protected by the due process clause, and Justice White's concurring opinion was also based on the due process clause.
which reconceived individual will “as a value rather than as a legally or constitutionally binding principle.”

Ruth Gavison notes that privacy is "essential" to democracies because it "fosters and encourages the moral autonomy of the citizen." Whether in reference to intimate personal decisions, characteristics, secrets and seclusion, anonymous expression, or controlling access to personal information, the legal protection of privacy enforces the sanctity of an individual’s sphere of autonomy against interference. This creates a private sphere of free, uncoerced action where the state acts as the guarantor against its own intrusion.

The importance of autonomy is reflected in what types of privacy violation the law protects against at the point of collection. Positioning the individual as the best judge of her privacy needs, the law becomes a guarantor of the ability to make free, uncoerced decisions about when our personal information is shared with another. As a result, the primary way privacy law places substantive limits on information collection is through individual consent. In

---

40 Id at 161.
42 The barring of some kinds of negative consequences from the use of an individual’s personal information is forbidden by a number of regulations, particularly in the realm of medical data. For example, job discrimination on the basis of personal DNA is forbidden by the Genetic Information Nondiscrimination Act of 2008, PL 110–233, May 21, 2008, 122 Stat 881. California’s Confidentiality of Medical Information Act limits disclosure of patient information, Cal. Civ. Code § 56–56.37. More directly, the Affordable Care Act forbids insurers from refusing coverage or charging more for patients with a pre-existing condition. See Health and Human Services, “About the Law”, available at http://www.hhs.gov/healthcare/about-the-law/pre-existing-conditions/index.html.
43 The “right to be left alone” of Brandeis’ dissenting opinion in Olmstead v. United States, 277 U.S. 438 (1928).
44 Especially (but not only) in political speech, as in McIntyre v. Ohio Elections Commission, 514 U.S. 334, 340–41 (1995). The decision reads in part, “Protections for anonymous speech are vital to democratic discourse. Allowing dissenters to shield their identities frees them to express critical minority views . . . Anonymity is a shield from the tyranny of the majority . . . It thus exemplifies the purpose behind the Bill of Rights and of the First Amendment in particular: to protect unpopular individuals from retaliation . . . at the hand of an intolerant society.”
45 For example, protection of the ability to control access by others to personal information after it leaves one’s exclusive possession by the FTC’s Fair Information Practice Principles, found in Federal Trade Commission, “Privacy Online: Fair Information Practices in the Electronic Marketplace,” May 2000.
theory, this places the data creator in the position of ultimate authority regarding the terms of permitted information collection.

Two notable exceptions rely on classical liberal notions of the main threats to individual autonomy—government action, which can exert the government monopoly of force to coerce the individual against their wishes, and private action that occurs outside the realm of mutual agreement. As a result, substantive privacy controls on collection of information focus on 1) government action and 2) malicious or “bad actor” private action.

The main control on government collection of information is the Fourth Amendment protection against unreasonable search and seizure. Early Fourth Amendment decisions limited its protection to physical intrusion of property or persons, but *Katz v. United States* (1967) extended Fourth Amendment protections to any area where a person as a reasonable expectation of privacy.\(^{46}\) Though exceptions exist, the Fourth Amendment requires the government to obtain a warrant, before collecting any private information from an individual, meeting the standards of probable cause, and stating with specificity the scope of the government intrusion.

The law also creates several rights against the impermissible collection of information by private actors. Private rights of action in tort law allow individuals to sue for invasions of privacy. Modern privacy torts, as categorized by William Prosser, fall into four categories: intrusion of solitude, public disclosure of private facts, false light, and appropriation.\(^{47}\) All four privacy torts require intentional, malicious, or otherwise unpermitted intrusion into the private sphere of another, often but not always for personal gain through broad dissemination of the impermissibly collected information.

The Computer Fraud and Abuse Act (CFAA) creates substantive limits on impermissible collection of information by unauthorized actors. Although the CFAA covers many different actions, the general framework of the law prohibits accessing a computer without

---


authorization, or in excess of authorization.\textsuperscript{48} The CFAA creates an explicit statutory extension of fraud law in the digital realm, but other classic and statutory fraud regimes also govern the impermissible collection of information, most notably under state and federal consumer protection statutory regimes. Since 2002, the Federal Trade Commission (FTC) has pursued enforcement action against companies for violating the privacy promises made in their terms of service under Section 5 of the FTC Act, which bars “unfair and deceptive acts and practices.”\textsuperscript{49}

However, the vast majority of private collection of data is governed under by the contractual regime of notice and consent contained in a company’s terms of service. These terms provide notice – laying out for data creators what data will be collected, how it will be used, and under what conditions it will be shared – and become the binding terms of the relationship once the data creator has given her consent.

Under the notice and consent contractual regime, privacy controls, to the extent they exist, occur later in the pipeline, controlling not what data is collected, but how data is shared, and what forms such data must take. These statutory privacy obligations fall under a patchwork of regulations based on the notion that certain kinds of sector-specific data, like medical and financial data, are particularly sensitive, and therefore require special protections in the storing or sharing of such data to ensure that data creator confidentiality is maintained. Much of the statutory requirements detailing these protections stem from the influential privacy framework called the “Fair Information Principles” (FIPS).\textsuperscript{50} FIPS have informed the dominant federal statutory framework for the collection and management of data by government, focused on balancing the social benefit of sharing data with the private interests of data creators. To achieve this balance, the FIPS framework details preventative measures

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{49} Federal Trade Commission, Enforcing Privacy Promises, available at https://www.ftc.gov/news-events/media-resources/protecting-consumer-privacy/enforcing-privacy-promises
\item \textsuperscript{50} U.S. DEPT OF HEALTH, EDUC., & WELFARE, RECORDS, COMPUTERS, AND THE RIGHTS OF CITIZENS (1973).
\end{itemize}
\end{footnotesize}
to reduce the overall risk of privacy harm based on the “inherent riskiness” of certain types of data.\textsuperscript{51}

The FIPs framework and subsequent patchwork of privacy laws do not conceive of limits to how data may be collected, or what kinds of data can be collected, once a data creator has given her consent to the collection. Privacy laws typically require substantive processing practices to limit access to or anonymize data categorized as sensitive, not at the point of collection, but in the risk of release. In effect, despite a fairly robust framework for protecting confidentiality after collection, the FIPS creates almost no substantive limit to what may be known about an individual, or to what ends such information may be used, other than each individual’s own consent.

Nevertheless, it is at the point of collection that the initial risk to autonomy is introduced – when another party gains access to information about us, and may subsequently go on to use that information to violate our sense of self either by making still further inferences and predictions about our behavior or by using that information in ways that go against our wishes.

\textit{Coercion and consent}

Privacy regimes fail to account for relationships of private power between data creators and data collectors in much the same manner the private property regimes described by Robert Hale in 1923 failed to account for the extractive relationships between property owners and laborers. Hale examines the fallacy undergirding most laissez-faire arguments extolling the necessities of freedom from government regulation. These arguments are premised on the view that government is the main, indeed almost sole, coercive force against which the legal regime must be vigilant. Government "should not coerce" individuals, but it may intervene to "prevent any private person or group from exercising any compulsion"\textsuperscript{52} This traditional account of regulation suggests the main form coercion we must guard against is government action and extreme compulsion from individuals – so called “bad actors”. Hale argues that

\textsuperscript{51} Paul Ohm, Broken Promises of Privacy, 57 UCLA L. REV. 1701, 1734 (2010).
\textsuperscript{52} Robert Hale, “Coercion and Distribution in a Supposedly Noncoercive State” 38 \textit{Political Science Quarterly} 470 (1923) in The Canon of American Legal Thought, 88
this approach misses the extensive and essential nature of coercion as the daily outcome of decisions operating within the bounds of law. People are exposed to coercion at the "hands of both" private and state actors all of the time.\textsuperscript{53}

Hale argues that upon examination, any "influences exerted to induce another to act against his will" are essentially indistinguishable from coercion.\textsuperscript{54} These influences define the basic relationship between property owners and non-property owners, and from private coercion, Hale illuminates all sorts of constraints on action and behavior. The distribution of power between any set of individuals, industry, and government is a result of the net effect of each actor's coercive force. Which actor has the dominant coercive force is partly a function the legal regime that prohibits some behavior from some actors but permits it from others. In this way, the choices faced in regulating spheres of autonomy are not between more or less "freedom" but between different distributions of coercive power among private actors enforced by background legal regimes.

It is worth considering in detail the coercive forces at play at the point of data collection between a data creator and a data collector. Like other consumers, data creators do not have the luxury of making bespoke contractual agreements with large technology companies. Their choice-signaling mechanisms are limited to exit and voice.\textsuperscript{55} Of these two, the only one that is explicitly recognized by law as governing the collection of data is exit – in other words, revoking consent, or refusing to give it to begin with.

The price of revoking or refusing consent is high: industry data collectors may not have a monopoly of force like the state, but they exert the power of network effects, which more or less create monopolies of access. And as these companies accrue more power, the scope and pull of their network effects only grow. Amazon controls 46 percent of the online retail

\textsuperscript{53} Id at 94.

\textsuperscript{54} Id at 98.

\textsuperscript{55} Albert O. Hirschman first detailed these two primary reactions to signaling dissatisfaction with organizations in Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States. Harvard University Press, 1970.
market, and its share is only growing. Google dominates search, maps and online email. Facebook is not merely dominant in social media; news outlets, volunteer groups, political campaigns and community organizers are increasingly accessible or organized only via Facebook. With the advent of Facebook's global Free Basics program and the proliferation of WhatsApp internationally, “in many countries Facebook and its products simply are the internet.” Despite the fact that an individual constructively has no other choice regarding network companies, the information they collect and the ways in which it can be used, the notice and consent regime does not recognize this lack of choice as relevant to a privacy loss inquiry. Although consent is premised on an idea of individual autonomy, in practice it grants individuals the choice to accept the terms of data collection given by companies, or opt out of full participation in digital life.

Even if consent is granted, the meaning of this action is unclear. As an expressed preference, it signals very little about an individual’s true understanding of all the implications and outcomes of their decision, particularly with respect to privacy. Assume for the sake of argument that an individual had read all the language in a company’s terms of service and agreed to them. Should we assume this person keeps up with the latest academic research on computational inference? Do they know that algorithms do a fairly good job of inferring personality traits, sexual orientation, political views, mental health status, substance abuse history and more based on the collection and analysis of their Facebook “likes”? How closely do they follow the data broker marketplace? Do they know that health insurance companies purchase information on race, education level, television habits, marital status, social media posts, online purchasing habits and payment histories about hundreds of millions of Americans to make predictions about health care costs? As prominent privacy theorist Helen Nissenbaum notes “[i]t’s not that we don’t know what consent means; it’s

56 Olivia LaVecchia & Stacy Mitchell, *Amazon’s Stranglehold*, INST. FOR LOC. SELF-RELIANCE 10 (Nov. 2016)
that getting to a point where we understand the true sense of what consent means is impossible.”

Instead of seeking to minimize the coercive effects from monopolies of access, conditioned by a limited notion of consent, the law enforces these effects. The law prohibits an individual from bypassing networked monopolies through the steady expansion of copyrights, patents, and similar exclusive rights that serve to enclose the digital commons and distribute control of access away from individuals to industries. Privacy law further facilitates a one-way knowledge acquisition and manipulation that distributes power away from individuals to industries. Although it is generally accepted that data and algorithms are not subject to property rights, “[a]pparances can be deceptive.” The coercive option users face – access to “essential social, commercial and cultural connectivity” for the data online platforms need to “create and sustain competitive advantage” – has resulted in the “quiet revolution in the legal status of data as (de facto if not de jure) proprietary informational property.” Platforms “jealously guard access” to the data they have collected, and citing the privacy concerns of users is one way to credibly prevent or restrict access to valuable data. In this way, privacy breaches and the ensuing public outcry serve to shore up the private power of online platforms. It is in the interest of companies like Facebook to “lock down its stores of data,” in response to scandals like Cambridge Analytica, since the extensive data it collects allows it to develop its unique “access to [users] and [their] attention.”

Enclosure of data allows companies to retain the value of data, while the ability for companies to unilaterally set the terms of access shifts the risks associated with information

---

61 Benkler, Wealth of Networks at 380-81. See also, Benkler Degrees of Freedom, Dimensions of Power
63 Id.
64 Id.
65 Id.
66 Id at 155.
68 Id.
collection and storage disproportionately onto data creators. As the possibilities of what can be known in digital space about digital selves is expanded, the scope and magnitude of harm that this can do to physical selves in physical space is expanded.69

The transformation of data into an asset has resulted in a system-wide increase in data collection, and as a result, an increased risk of privacy harm. The wholesale increase in everyday lives being transformed into continual structured flows of personal data has increased both the risk and the stakes of privacy harms for the users who are the source of information. “Data is constructed by people, from people.”70 Or, as digital studies scholar Karen Gregory puts it: “Big data, like Soylent Green, is made of people.”71

Consider Daniel Solove’s taxonomy of the privacy harms under Big Data market conditions. He identifies four major types of problem: information collection, information processing, information dissemination and invasion, under which 16 sub-types of privacy harm are organized.72 The mass collection, analysis and uses of user data that occur under Big Data result in activity that typifies almost every single kind of privacy problem he identifies – from the massive increase in surveillance and interrogation (the “pressuring of people to divulge information”73), to the countless ways that Big Data information processing relies on linking, aggregating and repeated uses of data. Fine grained predictive inferences about our behavior, and extensive use of behavioral psychology to keep users engaged and permanently connected may constitute forms of intrusion and decisional interference.

69 The concept of digital space affecting the physical space is borrowed from Daniel J. Solove, The Digital Person.
73 Id at 500.
But mass collection of data does not just threaten individual harm. The harmful effects from the data imperative “extends beyond individual user privacy to encompass a wider range of possible harms from the potentially skewed, biased, discriminatory, misleading, or manipulable dynamics of the information results themselves.”74 Facebook’s micro-targeting ability, for example, allows rental advertisers to discriminate against African Americans, the disabled, religious minorities, and non-English speakers.75 The poor are disproportionately harmed by the emerging use of social media data in the context of employment, education and policing.76 And as more data is collected and ever-finer predictions of behavior become feasible, the concentration of informational power will only grow, as ever more data sets are combined, their insights used to build out more architectures of datafication and more seamless tracking, via the inevitable spread of smart home devices and the Internet of Things.

The consent regime is based on a conception of individual autonomy that does not align with reality: the collection of information is limited by terms that reflect the stark power imbalance between data creators and data collectors. Moreover, the broad powers granted to companies under those terms is used to extract extensive amounts of information about data creators that result in even more unequal distributions of power between individuals and companies. In this way, a regime governing data collection premised on individual autonomy in theory, leaves most data creators vulnerable to coercion from private actors in practice. By overlooking these relationships, privacy law enforces existing distributions of power.

**Part III: Proposed privacy law interventions and their distributive effects**

*The reform mindset*

---

74 Frank Pasquale Internet Nondiscrimination Principles at 264
The incredible amount of personal information that has already been collected, along with the amazing analytic potential of Big Data, means that preventing data collection and analysis is neither feasible nor desirable. But regimes of data governance should strive to empower a more comprehensive set of actors. Legal solutions to the data imperative should correct for power imbalances between individuals and firms, and refocus Big Data’s analytic potential based on collective priorities.

If the data imperative is considered a form of structural domination, proposals for legal reform should address the harms that result from concentrated private power via data extraction. As Iris Young notes, “individuals may experience economic and social structures as objective and exogenous, but they are in fact the result of hidden and accumulated decisions, policies and actions. The accumulation of these human choices congeals into a larger structure”, which concentrates power in the hands of some at the expense of others.77

The first step, then, in doing something about the data imperative is recognizing that it is a man-made phenomenon. Early online communities had neither the desire nor the capability to make surveillance synonymous with online life.78 It was only when social networking and gaming communities that “relied on cookies to manage login information”79 converged with the “new and powerful data analysis capabilities” of Big Data that platforms looking to monetize their popularity in response to “demands of capital markets” developed their present-day business models and the data imperative emerged. Cohen notes, “Advertisers who might provide revenue wanted results and users were learning to value personalization. Personalized tracking and predictive modeling seemed the logical way to satisfy both imperatives.”80

---

79 Id.
80 Id.
Because structural forms of dominance are diffuse and impersonal, courts and policymakers typically have trouble providing remedies to address them. Courts, in particular, are designed to address harms from power disparities and domination from a single actor acting to assert its will against others. Yet diffuse systems can, in the aggregate, create immense disparities even in absence of a malicious or even merely intentional assertion of will. The accumulation of individual business decisions, technical innovations, and governing policies, all operating within the bounds of the law, can develop into a set of everyday marketplace practices that create significant, ongoing harm.

**Proposed Agenda for Further Work:**

1. Lay out an explicitly distributive analysis of proposed reforms to data governance and privacy law:
   a. central to the analysis of any legal regime is an understanding of its impact; this mode of analysis was central to Progressive Era reformers, and as policymakers look to address inequality and concentrated power in the digital economy, are important to revive today
   b. This normative evaluation of law is essentially empirical and material in nature, and stems from the position that the relevant inquiry for judging a legal regime is the outcomes it produces. That is, it cannot simply be judged by the rights and privileges it extends to individuals, but by what outcome is achieved via those rights and privileges.

2. Proposed alternative framework
   a. Reasonable consumer of data proposal:
      i. Motivation: To align law governing data collection with how people actually make choices about data collection. Implicit here is also a normative claim: to the extent the law enforced so as to achieve socially beneficial outcomes, it should be built from some understanding or basic prediction of how people actually make choices
      ii. Definition: when entering into data collection agreement, the reasonable consumer believes and expects the transaction they are entering into to be:
         1. Reasonably safe - for themselves and for others
         2. Reasonably for the thing/service they are engaging with - for themselves and for others
         3. And, as a corollary to 2; reasonably “not for” something else
      iii. If this definition is not met, then market solution (aka contract solution premised on consent) will fail to provide the mutually expected outcome.

---

iv. “Reasonably safe” – extending existing ideas in law to digital setting
   1. Tort law on product liability
   2. Regulation to limit harmful effect of externalities
v. “Reasonably for the thing/service they are engaging with”
   1. Extends intuition behind fraud/deception law, but these 19th and early 20th century legal tools do not capture the kind of deception that arises out of legal and technical complexity
vi. Hypothesis 1: This is empirically true – when people engage in transactions of digital life,

3. Taxonomy of solutions – analyzed via reasonable consumer of data framework
   a. Individual-focused reforms:
      i. Goals: Some individual reforms focus on giving individuals more power through access to more knowledge about what the data imperative means with respect to their data. Other individualist reforms aim to distribute more power to individuals by offering various degrees of a property or labor right over data.
      ii. Knowledge Access reforms: Increasing transparency, improving literacy and embracing meaningful user consent
         1. Aim to address information asymmetry between data creator and data collector; allow data creator to make a more informed, and therefore better, choice about how their data may be gathered and used.
         2. Information asymmetries do currently exist; which may in theory lead to moral hazard.
         3. However, unlikely informational reforms alone can address structural harms of data imperative. Doesn’t change essential nature of relationship or lower the cost of exit.
         4. Susceptible to design manipulation and false choice
      iii. Property/labor reforms: mandating access to data, interoperability, individual ownership of data, or wage payment for contributing data.
         1. Property/labor right reforms seek to address the extractive nature of data collection by re-aligning rights of determination and exploitation over data with data creators. In this way, users can “take back” from platforms some degree of control over their data, or some degree of the value their data generates.
         2. True that currently companies extract significant value from data users turn over for free, and seems right to redistribute some of the gains of that value generation back to data creators.
         3. However, property/labor reforms provide limited answer to the structural mass extraction of data.
            a. Access, interoperability, ownership or a wage doesn’t change the market for data, just reshuffles the first order property right, and in return for a property or
wage payment forecloses any moral claim about how fruits of data used downstream.

b. Introduce significant administrability and enforceability concerns
c. Seems to rely on misconception of data and data value - There is almost no data about us that is not also about others. Only when data about us is combined and analyzed alongside that of others can the rich analytical insights of Big Data be achieved.

4. Individual property rights also serve to further enclose access to data

b. Structural reforms:
   i. Goals: At a broad level, structural reforms aim to 1) protect against the potential of Big Data to result in either extractive or invasive harms and 2) empower substantive uses of Big Data based on criteria of social value.
      1. Resulting focus on collection and use, rather than confidentiality-preserving techniques
   ii. Ecosystem approaches: environmental approach, public health approach
      1. Limit toxic resource collection or spread, encourage efficient use and use that improves overall 'ecosystem'
      2. Produce market-wide shift in behavior to reduce risk and align individual interests with socially beneficial outcomes
   iii. Antitrust/Market Power approaches: antitrust, public utility approaches
      1. Either break up concentrated market power, or allow natural monopolies as long as semi-publicly governed
   iv. Fiduciary approaches: Data trusts, information fiduciaries
      1. Create duty of care either in entity itself, or in data trust governed by set of terms agreed to by data creators

---

82 The Open Data Institute, No One Owns Data: We Need to Strengthen our Rights, March 24, 2018, available at https://theodi.org/article/no-one-owns-data-we-need-to-strengthen-our-rights/