DEMOCRATIC DATA: A RELATIONAL THEORY FOR DATA GOVERNANCE

Forthcoming, Yale Law Journal, Vol 131

Salomé Viljoen[*].

Data governance law—the legal regime that regulates how data about people is collected, processed, and used—is a subject of lively theorizing and several proposed legislative reforms. Different theories advance different legal interests in information. Some seek to reassert individual control for data subjects over the terms of their datafication, while others aim to maximize data subject financial gain. But these proposals share a common conceptual flaw. Put simply, they miss the point of data production in a digital economy: to put people into population-based relations with one another. This relational aspect of data production drives much of the social value as well as the social harm of data production and use in a digital economy.

In response, this Article advances a theoretical account of data as social relations, constituted by both legal and technical systems. It shows how data relations result in supra-individual legal interests, and properly representing and adjudicating among these interests necessitates far more public and collective (i.e., democratic) forms of governing data production. This theoretical account offers two notable insights for data governance law. First, this account better reflects the realities of how and why data production produces economic value as well as social harm in a digital economy. The data collection practices of the most powerful technology companies are primarily aimed at deriving population-level insights from data subjects for population-level applicability, not individual-level insights specific to a data subject. The value derived from this activity drives data collection in the digital economy and results in some of the most pressing forms of social informational harm. Individualist data subject rights cannot represent, let alone address, these population-level effects. Second, this account offers an alternative (and it argues, more precise) normative argument for what makes datafication—the transformation of information about people into a commodity—wrongful. What makes datafication wrong is not (only) that it erodes the capacity for subject self-formation, but also that it materializes unjust social relations: data relations that enact or amplify social inequality. This egalitarian normative account indexes many of the most pressing forms of social informational harm that animate criticism of data extraction yet fall outside typical accounts of informational harm. This account also offers a positive theory for socially beneficial data production. To address the inegalitarian harms of datafication—and develop socially beneficial alternatives—will require democratizing data social relations: moving from individual data subject rights, to more democratic institutions of data governance.

---

[*] Joint Fellow at NYU Law, Information Law Institute and Cornell Tech, Digital Life Initiative. Many thanks to the members of the 2020 Privacy Law Scholars Workshop, the Information Law Institute Fellows Workshop, and the Digital Life Initiative Fellows Group for their careful and generous comments. Additional thanks to Yochai Benkler, Elettra Bietti, Angelina Fisher, Jake Goldenfein, Ben Green, Lily Hu, Aziz Huq, Issa Kohler-Hausmann, Thomas Streinz, Duncan Kennedy, Lee McGuigan, Christopher Morten, Angie Raymond, Neil Richards, Katherine Strandburg, Mark Verstraete, Ari Waldman and Richard Wagner. An early version of this work was presented in 2018 at Indiana University's Ellen Ostrom Workshop.

Table of Contents

INTRODUCTION

In recent years the technology industry has been the focus of increased public distrust, civil and worker activism, and regulatory scrutiny.[1] Concerns over datafication—the transformation of information or knowledge about people into a commodity—play a central role in this widespread front of curdled goodwill, popularly referred to as the "techlash."[2]

As these firms mediate more of life and grow more economically dominant, the centrality they place on data collection in turn raises the stakes of data governance law—the legal regime that governs how data about people is collected, processed, and used. There is broad consensus that current data governance law has failed to discipline against the harms of data extraction, in part because it cannot account for the large and growing gap between data's *de jure* status as the subject of consumer rights and its *de facto* status as quasi-capital.[3] Data governance reform is the subject of much debate and lively theorizing, with many proposals being surfaced to address the status quo's inadequacy.[4]

This Article evaluates the legal conceptualizations behind these proposals—in other words, how proposed reforms conceive of what makes datafication worth regulating and whose interests in information ought to gain legal recognition. How datafication is conceptualized shapes and constrains how the law responds to datafication's effects. If data governance law is inattentive to

---

[1] Facebook's Cambridge Analytica scandal marked a turning point in the press coverage and popular sentiment towards technology companies. For more on Cambridge Analytica, see e.g. New York Times Editorial Board, *Mark Zuckerberg Testimony: Senators Question Facebook's Commitment to Privacy,* N. Y. TIMES April 10, 2018; Zeynep Tufecki, *Facebook's Surveillance Machine*, N. Y. TIMES, May 19, 2018. From 2015 to 2019, the number of Americans who held a positive view of technology fell by 21 percentage points. See Carol Doherty and Jocelyn Kiley, *Americans have become much less positive about tech companies' impact on the U.S,* Pew Research, July 29, 2019. Worker activism at tech companies has increased sharply since 2016, particularly in response to contracts between technology companies and the Department of Defense and Immigration Customs Enforcement (ICE). See #NoTechforICE, at https://notechforice.com/; the Tech Workers Coalition, at https://techworkerscoalition.org/; Jimmy Wu, *Optimize What?* COMMUNE, March 15, 2019; Drew Harwell, *Google to drop Pentagon AI contract after employee objections to 'business of war',* WASH. POST, June 1, 2018.

[2] The origin of the term "techlash" is commonly attributed to its use in the *Economist* in 2013. In 2018, both Oxford Dictionaries and the Financial Times deemed "techlash" to be a word of the year. See Oxford Languages. 2018. Word of the Year 2018: Shortlist. *Oxford Languages*. https://languages.oup.com/word-of-the-year/2018-shortlist/. Foroohar, Rana. 2018. Year in a Word: Techlash. *Financial Times*, available at https://www.ft.com/content/76578fba-fca1-11e8-ac00-57a2a826423e.

[3] JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM (2019).

[4] See infra. Parts I and III for extended discussion.

how data production creates social benefits and risk, it will be poorly equipped to foster the benefits and mitigate the risks of data's creation and use. Missing or misdiagnosing the effects of datafication can lead to reforms that, once achieved, fail to address the harms that motivated reform.

The Article's core argument is that the data collection practices of the most powerful technology companies are aimed primarily at deriving (and producing) population-level insights, not individual insights specific to the data subject. These insights can then be applied to all individuals (not just the data subject) that share these population features. This population-level economic motivation matters conceptually for the legal regimes that regulate the activity of data collection and use – it requires revisiting long-held notions of why individuals have a legal interest in information about them, and where such interests obtain.

The status quo of data governance law, as well as prominent proposals for its reform, approach these population-level effects as a byproduct or an externality, to the extent they conceive of these effects at all. As a result, both the status quo and reform proposals suffer from a common conceptual flaw: they attempt to reduce legal interests in information to individualist claims subject to individualist remedies, which are structurally incapable of representing the interests and effects of data production's population-level aims. This in turn allows significant forms of social informational harm to go unrepresented, and as a result, unaddressed in how the law governs data collection, processing and use.

Properly representing the population-level interests that result from data production in the digital economy will require far more collective, democratic modes of ordering this productive activity.[5] The relevant task of data governance is not to reassert individual control over the terms of one's own datafication (even if this were possible) or to maximize personal gain, as leading legal approaches to data governance seek to do. Instead, the task is to develop the institutional responses necessary to represent (and adjudicate among) the relevant population-level interests at stake in data production.

---

[5] The Article will refer variously to the "data political economy" the "data economy" and the "digital economy." While there are distinctions between these concepts in their own right, here these all refer to set of actors, products, business practices, and imperatives that depend on the ability to produce economic value (and political effects) through processes of data capture, transfer, and analysis. See MARK ANDREJEVIC, INFOGLUT: HOW TOO MUCH INFORMATION IS CHANGING THE WAY WE THINK AND KNOW (2013); Matthew Crain, *Financial Markets and Online Advertising: Reevaluating the dotcom investment bubble*, Information, Communication and Society 17(3), 371-384 (2014); OSCAR H. GANDY, THE PANOPTIC SORT: A POLITICAL ECONOMY OF PERSONAL INFORMATION(1993), Lee McGuigan and Vincent Manzerolle, *All the world's a shopping cart: Theorizing the political economy of ubiquitous media and markets,* New Media and Society, 17(11): 1830-48 (2015); Joseph Turow and Nick Couldry, *Media as Data Extraction: Towards a New Map of a Transformed Communications Field*, Journal of Communication, Vol. 68, Issue 2: 415-23 (April 2018).

Recognizing the significance of these population-level interests in information shifts the proper aim of legal reform. Current efforts aim to provide data subjects expanded opportunities for exit, payment, or recourse. Instead, data governance reform should aim to secure recognition and standing to shape the purposes and conditions of data production for those with material interests at stake in such choices. In other words, responding effectively to the economic imperatives and social effects of data production will require moving past proposals for individualist data subject *rights* and towards theorizing the collective *institutional forms* required for responsible data governance.

A. Informational Capitalism and the Expanded Task of Data Governance

Data plays a central role in both descriptive and critical accounts that characterize the contemporary digital political economy as informational capitalism.[6] Among competing technology companies, greater access to high-quality data is a key competitive advantage that allows them to build better algorithmic products, gain better insights into their customers (or the audiences their customers want to reach), and more advantageously price goods, services, or bids.[7]

---

[6] Informational capitalism, also called surveillance capitalism and data capitalism, refers to a mode of production *centrally oriented* around extracting and processing information in order to extract and amass wealth. This transforms information—particularly information in the machine-legible form of data—into a key productive resource. Manuel Castells defines informational capitalism as the alignment of capitalism as a mode of production with informationalism as a mode of development. See JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM (2019) at 6 ("In a regime of informational capitalism, market actors use knowledge, culture, and networked information technologies as means of extracting and appropriating surplus value, including consumer surplus"). See also MANUEL CASTELLS, THE INFORMATION AGE, VOL. 1: THE RISE OF THE NETWORK SOCIETY (1996) 14-16; DAN SCHILLER, HOW TO THINK ABOUT INFORMATION (2007). For an early discussion of these concepts, see VINCENT MOSCO AND JANET WASKO, ED., THE POLITICAL ECONOMY OF INFORMATION (1988), particularly "Cybernetic Capitalism," which provides a prescient analysis of what today is called surveillance capitalism.

[7] ERIC POSNER AND GLEN WEYL, RADICAL MARKETS: UPROOTING CAPITALISM AND DEMOCRACY FOR A JUST SOCIETY (2018). Business-facing publications emphasize the importance of data for maintaining and achieving competitive advantage. See e.g., Andrai Hagiu and Julian Wright, *When Data Creates Competitive Advantage,* Harvard Business Review, Jan-Feb 2020; Nitin Seth, *Analytics Are a Source of Competitive Advantage, If Used Properly*, Forbes, July 18, 2018; Antitrust scholars are paying increasing attention to the competitive effects of mass data collection and the locked-in advantages greater data offers incumbent computing technologies. See e.g., ALLEN P. GRUNES AND MAURICE E. STUCKE, BIG DATA AND COMPETITION POLICY, (2016); Dina Srinivasan, *Why Google Dominates Advertising Markets,* 24 Stan.L. Rev (2020). The near-monopolistic control of data flows by certain entities, and the competitive advantage this creates, has attracted growing regulatory attention in the EU. See Europen Commission, "Antitrust: Commission launches sector inquiry into the consumer Internet of Things (IoT)" European Commission, July 1, 2020 available at https://ec.europa.eu/commission/presscorner/detail/en/IP_20_1326 .

Critics similarly note data production's significance for the digital economy. Jathan Sadowski identifies data as a distinct form of capital, linking the imperative to collect data to the perpetual cycle of capital accumulation.[8] Julie Cohen traces how platform companies like Amazon and Facebook secure "quasi-ownership" through their enclosure of data and identifies the processing of information in "data refineries" as a "centrally important means of economic production."[9] In Polanyian tradition, Cohen argues that data about people represents a "fourth factor of production" that sets apart informational forms of capitalism.[10] Shoshanna Zuboff compares data production to conquest-based forms of wealth accumulation, likening people's inner lives to a pre-Colonial continent, invaded and strip-mined for profit by technology companies.[11] These accounts locate in datafication a particular economic process of value creation that demarcates informational capitalism from its predecessors.[12]

Datafication raises new legal challenges. Privacy and data governance law have traditionally governed forms of private interpersonal exchange in order to secure the benefits of data subject *dignity* or *autonomy*. Yet as data collection and use become key productive activities, new kinds of information-based harm arise. Data production structures processes that may marginalize social groups, amplify differences in wealth and power, and create social, political, and economic winners and losers. There is growing evidence of the role digital technology plays in facilitating social and economic inequality.[13] Digital

---

[8] Jathan Sadowski, *When data is capital: Datafication, accumulation, and extraction*, Big Data & Society (2019).

[9] JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM (2019), at 67.

[10] Cohen 2019 at 67. Cohen develops her account of data's role as a factor of production in informational capitalism from the three inputs Karl Polanyi identified as basic factors of production in a capitalist political economy: land, labor, and money. The movement to industrial capitalism transformed these three inputs into commodities. Cohen argues that the movement to informational capitalism reconstitutes them again, into new datafied inputs for profit extraction. At the same time, data flows about people become a vital, fourth factor of production.

[11] SHOSHANA ZUBOFF, AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER (2019). See also, Shoshanna Zuboff, *Big other: surveillance capitalism and the prospects of an information civilization*, Journal of Information Technology 30.1 (2015): 75-89. Others more explicitly engage the comparison between data extraction and colonialism. See e.g. NICK COULDRY AND ULISES MEJIAS, THE COSTS OF CONNECTION: HOW DATA IS COLONIZING HUMAN LIFE AND APPROPRIATING IT FOR CAPITALISM (2019);

[12] This Article will repeatedly refer to the terms "datafication" and "data extraction." Consistent with the definition above, it defines "datafication" as the transformation of information or knowledge into a commodity. It defines "data extraction" as the seamless and near-continual flow of such datafied knowledge from data subjects to data collectors (often platforms).

[13] See e.g., VIRGINA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR (2018); BEN GREEN, THE SMART ENOUGH CITY: PUTTING TECHNOLOGY IN ITS PLACE TO RECLAIM OUR URBAN FUTURE (2019); Ben Green and Salomé Viljoen, *Algorithmic Realism: Expanding the Boundaries of Algorithmic Thought,*

surveillance technologies used to enhance user experience for the rich simultaneously provide methods of discipline and punishment for the poor. Algorithmic systems may reproduce or amplify sex and race discrimination.[14] Seemingly innocuous data collection may be used in service of domination and oppression.[15] The pursuit for user attention and uninterrupted access to data flows amplify forms of identitarian polarization, aggression, and even violence.[16]

Such evidence suggests that social processes of datafication not only produce violations of personal dignity or autonomy, but also enact or amplify *social inequality*. Data production may facilitate social inequality in various ways. First, it may result in maldistribution of the material value from information production and use, such that data production facilitates or exacerbates economic inequality. Second, it may distribute the risks of information production unevenly by materializing forms of group-based injustice. Third, it may result in the underproduction of certain forms of socially valuable yet unprofitable data production that would equalize other spheres of life.

Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAT*)(2020); Frank Pasquale, *Two Narratives of Platform Capitalism,* 35 Yale Law & Pol. Rev. 309 (2016); Miriam Pawel, *You call it the gig economy. California calls it 'feudalsim,* N.Y. TIMES Sept 12, 2019; Neil Irwin, *To understand rising inequality, consider the janitors at two top companies, then and now,* N. Y. TIMES, Sept 3, 2017. Other arguments highlight how the negative effects of surveillance are apportioned along lines of privilege. Frank Pasquale, *Paradoxes of Privacy in an Era of Asymmetrical Social Control*, in Big Data, Crime and Social Control 31 (Aleš Završnik ed., 2018). Paul Blest, *ICE is Using Location Data form Games and Apps to Track and Arrest Immigrants, Report Says*, VICE NEWS, Feb 7, 2020; Solon Barocas and Andrew Selbst, *Big Data's Disparate Impact,* Cal. L. Rev., Vol. 671 (2016).

[14] See e.g., Joy Buolamwini and Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification.* In: Sorelle AF and Christo W (eds) *Proceedings of the 1st Conference on Fairness, Accountability and Transparency* Proceedings of Machine Learning Research: PMLR, (2018) 77—91; Safiya Noble, *Google search: Hyper-visibility as a means of rendering black women and girls invisible*, InVisible Culture (2013*)*; Ben Green, *The False Promise of Risk Assessments: Epistemic Reform and the Limits of Fairness,* Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAT*) 2020.

[15] See Paul Blest, *ICE is Using Location Data form Games and Apps to Track and Arrest Immigrants, Report Says*, VICE NEWS, Feb 7, 2020; Joseph Cox, "How the U.S. Military Buys Location Data from Ordinary Apps" Vice, Nov 16, 2020 (detailing how the U.S military buys location data from many sources, including a Muslim prayer app with over 98 million downloads).

[16] See e.g., THE MEDIA MANIPULATION CASEBOOK, ed. Joan Donovan, available at https://mediamanipulation.org/about-us; *Weaponizing the Digital Influence Machine,* Data & Society, October 2018; Ronan Farrow, *A Pennsylvania Mother's Path to Insurrection*, NEW YORKER, Feb 2, 2021; Chinmayi Arun, *On WhatsApp, Rumours, Lynchings, and the Indian Government*, Economic & Political Weekly vol. lIV no. 6, (2019).

Contending with the economic realities of data production requires reconceptualizing the task of data governance as one of managing population-level effects. This expands the task of data governance and the set of legal challenges facing data governance: from disciplining against forms of interpersonal violation to also structuring the rules of economic production (and social reproduction) in the information economy. How can (or should) data governance law respond to this shifting role of data collection and use and the heightened stakes of data production?

B. There's no I in "Our Data"

Properly answering this question requires attentiveness to the new role data plays in the digital economy, and how this role creates new forms of social value and social harm. Data production in the digital economy is fundamentally relational: a basic purpose of data production is to relate people to one another on the basis of relevant shared population features. This relational purpose of data production produces both considerable social value as well as many of the pressing forms of social risk that plague the digital economy. As the Article will explore further below, data's relationality results in widespread population-level interests in data collection and use that are irreducible to individual legal interests within a given data exchange.

However, the central economic and social role of data's relationality is largely absent in both current and proposed data governance law. Proposals for reform locate data at different points on the continuum from "person-like" to "object-like."[17] On one end of this spectrum, propertarian reforms conceive of data as "object-like". These reforms call for formalizing an alienable right to data as labor or property, to be bought and sold in a market for goods or labor. On the other end, dignitarian reforms conceive of data as "person-like;" an extension of data subject selfhood. These reforms call for stronger protections for data under human or civil rights law and encode a form of civic data relations. Yet despite their differences, both propertarian and dignitarian reforms—like the

[17] Luke Stark & Anna Lauren Hoffman, *Data is the New What? Popular Metaphors and Professional Ethics,* Journal of Cultural Analytics, May 2019. See also, Rob Kitchin, Big Data, *New Epistemologies and Paradigm Shifts,* Big Data & Society 1, no. 1 (April 2014); Rowan Wilken, *An Exploratory Comparative Analysis of the Use of Metaphors in Writing on the Internet and Mobile Phones,* Social Semiotics 23, no. 5, 632-47 (Nov. 2013); Dawn Nafus, *Stuck Data, Dead Data, and Disloyal Data: The stops and Starts in Making Numbers Into Social Practices,* Distinktion: Journal of Social Theory 15, no. 2, 208-22 (June 2014); Cornelius Puschmann and Jean Burgess, *Big Data, Big Questions: Metaphors of Big Data,* International Journal of Communication 8 (2014); Deborah Lupton, *Swimming or Drowning the Data Ocean? Thoughts on the Metaphors of Big Data,* The Sociological Life*,* October 29, 2013; Sara Watson, *Metaphors of Big Data,* DIS Magazine, May 28, 2016; Kailash Awati and Simon Buckinham Sham, *Big Data Metaphors We Live by*, Towards Data Science, May 14, 2015; Cory Doctorow, *Why Personal Data is Like Nuclear Waste,* THE GURADIAN, January 15, 2008; Lilly Irani, *Justice for 'Data Janitors',* Public Books, January 15, 2015.

status quo regimes their proponents seek to replace—persist in individualizing legal interests and remedies over data collection and use.

This poses two problems. The first problem is conceptual: a central economic imperative that drives data production goes unrepresented in both existing and proposed laws governing datafication. These laws consider data's individual-level, but not its population-level, effects. As a practical matter, this leaves the law out of step with many of the ways that information creates social value, allowing material forms of social informational harm to persist unaddressed. This presents U.S. data governance law with the *sociality problem*: how can data governance law account for data production's social effects?

The second problem is a matter of institutional design. Individualist theories of informational interests result in legal proposals that advance a range of new rights and duties with respect to information, but practically fall back on individuals to adjudicate between legitimate and illegitimate information production. This not only leaves certain social informational harms unrepresented (let alone addressed) but also risks foreclosing socially beneficial information production. This presents U.S. data governance law with the *legitimacy problem*: how can the legal regimes governing data production distinguish legitimate from illegitimate data use without relying on individual notice and choice?

The sociality problem demonstrates the need in data governance law for an expanded account of the interests at stake in information production, while the legitimacy problem points to the need for data governance to expand the remit of data governance. Addressing the second problem follows from addressing the first: once one identifies what interests the law ought to recognize and how such interests arise, one may develop an appropriate legal approach to address them.

1. Democratic data governance

This Article's two primary contributions offer a response to these problems. Conceptually, it offers an account of the sociality problem that recognizes the ubiquity and the relevance of the population-level interests that result from data production. From such recognition follows the Article's account of the legitimacy problem, which argues for governing many types of data as a collective resource that necessitates far more democratic, as opposed to personal, forms of institutional governance.

This in turn leads to a different line of inquiry regarding the legal challenges facing data governance law. Current debates center on how to secure greater data subject control, more robust protections for data subject dignity, or better legal expressions of data subject autonomy. An account of data social relations focuses future inquiry on how to balance the overlapping and at times competing interests that comprise the population-level effects of data production. This line of inquiry raises core questions of *democratic governance*: how to grant people

a say in the social processes of their mutual formation; how to balance fair recognition with special concern for certain minority interests; what level of civic life achieves the appropriate level of pooled interest; how to recognize that data production produces winners and losers and develop fair institutional responses to these effects.

In short, reckoning with data's relationality in data governance law contends with data production as an economic activity and a social process. It focuses not only on relevant legal interests when a person is reduced to a data flow (a moment of potential personal violation), but also on the legal relevance of data production as a process of socio-economic production: data production creates social and economic winners and losers, defines who wins and who loses, and determines the stakes of winning and losing.

2. Prior accounts of the social effects of privacy

This Article builds on prior digital privacy and data governance scholarship that points out the importance of social causes and social effects of privacy erosion.[18] This Article takes up these insights to offer an account of *why* the social effects of privacy erosion should be considered of greater relevance—indeed, central relevance—for data governance law. It argues for recognizing a greater set of legal interests in information in order to represent and address these social effects in our data governance law. In developing its theory of data relations, the Article also departs from prior accounts.

Prior accounts rightly identify the deep entanglement between the challenges of protecting autonomy in the digital economy and the realities of how data production operates as a social process: without securing better social conditions for data production for everyone, the personal benefits of robust privacy protection cannot be realized.[19] On this view, the supra-individual nature of digital privacy erosion matters because it raises additional complications for securing the benefits of robust digital privacy protection for individuals.

This Article departs from such accounts in that it places the inegalitarian effects of data extraction on equal theoretical footing with its autonomy-eroding effects. Privacy erosion's social effects *do* implicate the personal (and social)

[18] See e.g., PRISCILLA M. REGAN, LEGISLATING PRIVACY: TECHNOLOGY, SOCIAL VALUES, AND PUBLIC POLICY (1995); HELEN NISSENBAUM, PRIVACY IN CONTEXT (2010); Julie E. Cohen, *What privacy is for,* Harv. L. Rev., Vol 126 (2013). A more complete discussion of prior accounts will be explored in depth in Part I, infra.

[19] For more on the extended discussion of the democratic values at issue in data production, see Amy Kapczynski, *The Law of Informational Capitalism,* Yale L. J., Vol 129, No. 5, March 2020; Evgeny Morozov, *Digital Socialism?* NEW LEFT REVIEW 116/117, March/June 2019; Ben Tarnoff and Moira Weigel, *Why Silicon Valley can't fix itself,* THE GUARDIAN, May 3, 2018. PRISCILLA M. REGAN, LEGISLATING PRIVACY: TECHNOLOGY, SOCIAL VALUES, AND PUBLIC POLICY (1995); HELEN NISSENBAUM, PRIVACY IN CONTEXT (2010); Julie E. Cohen, *What privacy is For,* 126 Harv. L. Rev. 1904 (2013). Neil M. Richards, *Intellectual Privacy*, 87 Texas L. Rev. 387 (2008).

value of individual autonomy. But the inequality that results from data production should be considered relevant to the task of data governance for its own sake, and not only for the effects inequality has on data subjects' individual capacities for self-formation and self-enactment. Thus, the Article argues that, alongside traditional concerns over individual autonomy, the social inequalities that result from data production are *also* forms of informational harm.

C. Definitional Note

Three definitional and stylistic notes regarding this Article's use of key terms.

Data. For the sake of brevity "data" refers to data about people unless otherwise noted. Data about people is the data collected as people "invest, work, operate businesses, socialize" and otherwise go about their lives.[20] This data is of greatest interest to competing digital technology companies, and of greatest interest to observers of the business models built from data collection. It is also deliberately more expansive than U.S. definitions of "personal data" which commonly refers only to data that identifies or relates to a particular individual. Furthermore, the Article will refer to "data" as a singular, not a plural noun. This stylistic choice is in line with the common—rather than the strictly correct—usage.

Data subject and data collector. The Article will use the term "data subject" to refer to the individual from whom data is being collected—often also referred to in technology communities as the "user." "Data processor" is used synonymously with "data collector" to refer to the entity or set of entities that collect, analyze, process, and use data. The definitions of "data subject" and "data processor" are loosely derived from the European Union's General Data Protection Regulation (GDPR).[21] While the GDPR's definition of personal data offers some capacity for non-individualistic interpretation, any reference to "data subject' in this Article will refer to the individual from whom or about whom data is being collected.

Informational Harm. Individual informational harm refers to harm that a data subject may incur from how information about them is collected, processed, or used. In contrast, *social* informational harm refers to harms that third-party

[20] JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM (2019), at 38.

[21] Article 4 offers the following definition: "'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person." General Data Protection Regulation, Art 4.

individuals may incur from how information about a data subject is collected, processed, or used.

D. Roadmap

Part One describes the stakes and the status quo of data governance. It documents the significance of data processing for the digital economy. It then evaluates how the predominant legal regimes that govern data collection and use—contract and privacy law—code data as an individual medium. This conceptualization is referred to throughout the Article as "data as individual medium" (DIM). DIM regimes apprehend data's capacity to cause individual harm as the legally relevant feature of datafication; from this theory of harm follows the tendency of DIM regimes to subject data to private individual ordering.

Part Two presents the core argument of the Article, regarding the incentives and implications of data social relations within the data political economy. Data's capacity to transmit social and relational meaning renders data production especially capable of benefitting and harming others beyond the data subject from whom data is collected. It also results in population-level interests in data production that are not reducible to the individual interests that generally feature in data governance. Thus, data's relationality presents a conceptual challenge for data governance reform.

Part Three evaluates two prominent legal reform proposals that have emerged in response to concerns over datafication. Propertarian proposals respond to growing wealth inequality in the data economy by formalizing individual propertarian rights over data as a personal asset. Dignitarian reforms respond to how excessive data extraction can erode individual autonomy by granting fundamental rights protections to data as an extension of personal selfhood. While propertarian and dignitarian proposals differ on the theories of injustice underlying datafication and accordingly provide different solutions, both resolve to individualist claims and remedies that do not represent, let alone address, the relational nature of data collection and use.

Part Four proposes an alternative approach: data as a democratic medium (DDM). This alternative conceptual approach apprehends data's capacity to cause social harm as a fundamentally relevant feature of datafication. This leads to a commitment to collective institutional forms of ordering. Conceiving of data as a collective resource subject to democratic ordering accounts for the importance of population-based relationality in the digital economy. This recognizes a greater number of relevant interests in data production. DDM responds not only to salient forms of injustice identified by other data governance reforms, but also to significant forms of injustice missed by individualist accounts. In doing so, DDM also provides a theory of data governance from which to defend forms of socially beneficial data production that individualist accounts may foreclose. Part Four concludes by outlining some

examples of what regimes that conceive of data as democratic could look like in practice.

I. DATA GOVERNANCE: THE STAKES AND THE STATUS QUO

A. Data's Value in the Information Economy

Data about people produces revenue in three ways: companies can sell it directly, use it to improve services, or use it to predict, change or modify behavior.[22] Of these three, behavioral use represents by far the biggest source of revenue for technology companies.[23] Based on available evidence, the vast majority of this revenue comes from the ad tech industry—the business of buying and selling user attention.[24] In 2019, Google reported 134.81 billion U.S. dollars in advertising revenue out of 160.74 U.S. dollars in total revenue.[25] In the first quarter of 2020, Facebook's total advertising revenue amounted to 17.44

---

[22] The author wishes to thank and credit David Stein for this excellent and succinct description of how data is transformed into money in the digital economy.

[23] Some evidence pegs the global data brokerage industry at about $200 billion annually. However, a significant amount of the data being bought and sold via data brokers is not data about people. See David Lazarus, *Column: Shadowy Data Brokers make the most of their invisiblity cloak*, L. A. TIMES, November 5, 2019. Increasingly, data value comes not from direct sales of that data but its use to gain insights over consumers. High-quality objective and publicly available estimates of the value of global data flows are difficult to obtain and standardizing such measures is a subject of ongoing effort. See OECD, *A Roadmap toward a Common Framework for Measuring the Digital Economy*, Report for the G20 Digital Economy Task Force, 2020; Economic Statistics Centre of Excellence, "Measurement Issues in the Digital Economy," available at: https://www.escoe.ac.uk/projects/measurement-issues-in-the-digital-economy/; David Nguyen and Marta Paczos, *Measuring the economic value of data and cross-border data flows: A business perspective,* OECD Digital Economy Papers, No. 297, OECD Publishing, 2020; Diane Coyl and David Nguyen, *Cloud Computing, Cross-Border Data Flows and New Challenges for Measurement in Economics*, 249 National Institute Economic Review, Issue 1, 2019.

[24] The use of behavioral data to improve pricing and bidding strategies in online stores or advertising auction exchanges, with the aim of capturing a greater proportion of surplus value, is a lively topic of research among the data science and algorithmic mechanism design research communities and in the industry of programmatic advertising. See e.g., Hal R Varian, "Computer mediated transactions" *American Economic Review* 100(2): 1-10. 2010; Liran Einav and Jonathan D. Levin, "Economics in the age of big data. "*Science* 246(6210) (2014). Joseph Y Halpern and Rafael Pass, "Algorithmic rationality: Game theory with costly computation." *Journal of Economic Theory* 156: 246-26; (2015). Eric Sodomka ,"On how machine learning and auction theory power Facebook advertising." Simons Institute for the Theory of Computing, Berkeley, CA, November 17, 2015. Video, 54:21. https://simons.berkeley.edu/talks/eric-sodomka-2015-11-17; Tuomas Sandholm, "Automated mechanism design: A new application area for search algorithms" *International Conference on Principles and Practice of Constraint Programming* (September 2003): 19-36. For a legal treatment, see Dina Srinivasan, 'Why Google Dominates Advertising Markets," 24 STAN. L. REV (2020).

[25] Alphabet, Advertising revenue of Google from 2001 to 2019 (in billion U.S. dollars) Statista, https://www-statista-com.proxy.library.nyu.edu/statistics/266249/advertising-revenue-of-google/ (last visited July 16, 2020)

billion U.S. dollars, compared to 297 million U.S. dollars in revenues from other streams.[26]

Yet advertising techniques developed to predict or to influence behavior are increasingly gaining purchase for other industries. The same capabilities that help digital companies know (or claim to know)[27] what attributes make someone likely to buy an advertised product, or that are leveraged to increase a desired behavior, can be used for other tasks. For instance, to identify potential voters likely to engage on an issue or with a candidate, to identify what behaviors are associated with risky or risk-averse financial or health behavior, or to predict how much different people are willing to pay for a product. These point towards new avenues of growth for the data economy: in political consulting services, health insurance, financial services, and hiring.[28] Overall, the digital economy represents anywhere from 1.35 trillion U.S. dollars to 2.1 trillion U.S. dollars, making it between the seventh and fourth largest industry in the U.S.[29]

Data's value drives consequential business decisions in the digital economy. Consider just two recent examples. First, the streaming service HBO Max (owned by WarnerMedia) launched in May 2020 but for months was not available on two of the largest streaming platforms, Roku and Amazon Fire TV, which together comprise 63 percent of viewing time in the US. WarnerMedia wanted greater access and control over user data and resulting advertising than either Roku or Amazon are willing to provide. In order to maintain their position regarding this data, all parties are willing to forego considerable mutual gains.[30]

---

[26] Facebook, Facebook's global revenue as of 1st quarter 2020, by segment (in million U.S. dollars) Statista, https://www-statista-com.proxy.library.nyu.edu/statistics/277963/facebooks-quarterly-global-revenue-by-segment/ (last visited July 16, 2020)

[27] LEE MCGUIGAN, SELLING THE AMERICAN PEOPLE: DREAMS AND DESIGNS TO OPTIMIZE ADVERTISING, (forthcoming, copy of manuscript on file with author).

[28] Many instances of core digital services providers branching into other sectors exist. See e.g. John Hancock selecting Amazon's health wearable, Halo as the "featured, complimentary wearable" in their Vitality Program. John Hancock Insurance, "Amazon Halo Now Available for John Hancock Vitality Members," December 14, 2020, available at https://www.johnhancock.com/about-us/news/john-hancock-insurance/2020/12/amazon-halo-now-available-for-john-hancock-vitality-members.html; Verily Life Sciences, an Alphabet-owned company focused on health, launched a new health-insurance subsidiary Coefficient. Jay Peters, "Verily, Google's health-focused sister company, is getting into insurance" The Verge, August 25, 2020.

[29] In 2017, the Bureau of Economic Analysis estimated the digital economy's contribution to overall GDP at $1.3 trillion. In 2019, the Internet Association, an industry group that represents Facebook, Google, Twitter and many other technology firms, provided the $2.1 trillion estimate.

[30] Julia Alexander, *Why Peacock and HBO Max aren't on the biggest streaming platforms,* The Verge, July 15, 2020. ("The roadblock, like so many debates in the tech and media space, comes down to money and data. Essentially, both NBCUniversal (owned by Comcast) and WarnerMedia (owned by AT&T) want more control over user data and advertising generated by their apps").

Second, the decision of the Trump re-election campaign to partner with a small advertising software agency called Phunware to develop its 2020 re-election app was based on the company's ability to deliver valuable electoral data.[31]

> "The Trump campaign is not paying Phunware four million dollars for an app […] They are paying for data. They are paying for targeted advertising services. Imagine if every time I open my phone I see a campaign message that Joe Biden's America means we're going to have war in the streets. That's the service the Trump campaign […] have bought from Phunware. An app is just part of the package."[32]

B. Privacy Law's Individualism

The primary regime governing the collection of such data in the U.S. is digital privacy law, used here to encompass the suite of laws that together regulate how data about people is collected, processed, shared, and used.[33]

U.S. privacy law comprises federal and state contract law, consumer protection, privacy torts, and sector-specific consumer rights laws. Most data collected about people is governed by contractual terms of service, subject to laws of contract with FTC Section 5 and state consumer protection oversight.[34] A series of sector-specific privacy laws grant additional rights to consumers over particular kinds of data, such as consumer credit data, health and financial information, and educational information.[35] These include the Health Insurance Portability and Accountability Act (HIPAA), the Children's Online Privacy and Protection Act, the Family Educational Rights and Privacy Act (FERPA), the Gramm-Leach-Bliley Act (GLBA), and the Fair Credit Reporting Act (FCRA), alongside a few prominent state laws like Illinois' Biometric Information Privacy Act (BIPA), and California's Consumer Privacy Act (CCPA).[36]

---

[31] Sue Halpern, *How the Trump Campaign's Mobile App is Collecting Huge Amounts of Voter Data* NEW YORKER, September 13, 2020. Available at: https://www.newyorker.com/news/campaign-chronicles/the-trump-campaigns-mobile-app-is-collecting-massive-amounts-of-voter-data

[32] Ibid.

[33] Woodrow Hartzog and Neil Richards, *A Duty of Loyalty in Privacy Law* at 18, draft on file with author. Intellectual property and trade secrecy also play a significant role in structuring current data processing.

[34] See 15 U.S.C. § 45(a)(1)(2018)(prohibiting "unfair or deceptive acts or practices in or affecting commerce").

[35] Health Insurance Portability and Accountability Act (HIPAA), P.L. no. 104-104, 110 Stat. 1938 (1996); Children's Online Privacy and Protection Act (COPPA), 15 U.S.C. § 6501 et seq (2000); Family Educational Rights and Privacy Act (FERPA), 20 U.S.C. § 1232g (1974); Gramm-Leach-Bliley Act (GLBA), 12 U.S.C. § 78, § 377; 15 U.S.C. §80 (1999); Fair Credit Reporting Act (FCRA), 15 U.S.C. § 1681 (1970); Biometric Information Privacy Act (BIPA), 740 ILCS/14, Pub. Act 095-994 (2008); California Consumer Privacy Act (CCPA), Cal. Civ. Code, Div. 3, Pt 4, Title 1.81.5 § 1798.100-199 (2018).

[36] Health Insurance Portability and Accountability Act (HIPAA), P.L. no. 104-104, 110 Stat.

Many of these sector-specific laws are based on the Fair Information Practice Principles (FIPPS). FIPPS is an influential set of guidelines and recommendations from the FTC regarding the standards that typify fair data processing, and serve as continued model from which new privacy protections are developed.[37] FIPPS equates fair data processing with practices that grant individuals meaningful control over their data. This includes the ability to give informed consent to data being processed and being given notice regarding how data is being used. FIPPS is not enforceable by law but serves as an influential set of guidelines for how the FTC evaluates the self-regulation of privacy by industry and how it provides guidance for developing industry best practices. In lieu of enforcing the FIPPS, the FTC uses its authority under the FTC Act to enforce the promises companies make to data subjects. In practice, this results in the much-maligned privacy regime known as "notice and consent."[38] Under this regulatory regime, the terms and conditions of digital services like search engines, social networks, mobile phone apps, and other digitally mediated services are presumptively valid as long as consumers are offered notice of the data being collected about them and consent to this collection.[39]

As will be discussed below, the resulting privacy law regime conceives of data as an individual medium—it focuses legal inquiry and accords legal relevance to data's potential to cause personal harm and as therefore appropriately subject to private, individual ordering. This conceptualization of 'data as an individual medium' (DIM) privileges data processing's capacity to

---

1938 (1996); Children's Online Privacy and Protection Act (COPPA), 15 U.S.C. § 6501 et seq (2000); Family Educational Rights and Privacy Act (FERPA), 20 U.S.C. § 1232g (1974); Gramm-Leach-Bliley Act (GLBA), 12 U.S.C. § 78, § 377; 15 U.S.C. §80 (1999); Fair Credit Reporting Act (FCRA), 15 U.S.C. § 1681 (1970); Biometric Information Privacy Act (BIPA), 740 ILCS/14, Pub. Act 095-994 (2008); California Consumer Privacy Act (CCPA), Cal. Civ. Code, Div. 3, Pt 4, Title 1.81.5 § 1798.100-199 (2018).

[37] The FTC's Fair Information Practice Principles were originally named in an influential report commissioned to explore the ways in which entities using computational automated methods to collect and use personal information. See the US Secretary's Advisory Committee on Automated Personal Data Systems, Records, Computers and the Rights of Citizens (1973).

[38] See, e.g., NANCY KIM, CONSENTABILITY: CONSENT AND ITS LIMITS (2019); Elizabeth Edenberg & Meg Leta Jones, *Analyzing the Legal Roots and Moral Core of Digital Consent*, 21 New Media & Soc'y 1804 (2019); Woodrow Hartzog & Neil Richards, *Privacy's Trust Gap*, 126 Yale L.J. 1180 (2017); NANCY KIM, WRAP CONTRACTS (2013); MARGARET RADIN, BOILERPLATE: THE FINE PRINT, VANISHING RIGHTS, AND THE RULE OF LAW (2012); Neil Richards & Woodrow Hartzog, *Taking Trust Seriously in Privacy Law*, 19 Stan. Tech. L. Rev 431 (2016); Daniel J. Solove, *Privacy Self-Management and the Consent Dilemma*, 126 Harv. L. Rev. 1880 (2013); Andrea M. Matwyshyn, *Technoconsen(t)sus,* 85 Wash. U. L. Rev. 529 (2007); Elettra Bietti, *Consent as a Free Pass: Platform Power and the Limits of the Informational Turn* (2020). 40 Pace L. Rev. 307 (2020).

[39] See Fed. Trade Comm'n, Protecting Consumer Privacy in an Era of Rapid Change: A Proposed Framework for Businesses and Policymakers iii (2010); Woodrow Hartzog, *The New Price to Play: Are Passive Online Media Users Bound by Terms of Use?,* 15 Comm. L. & Pol'y 405 (2010).

transmit knowledge about the data subject over its capacity to transmit knowledge about others. Under DIM, this individualist knowledge transmission is the legally and normatively relevant feature of datafication.

1. Private individual ordering

Notice and consent structures the basic legal relationship between the individual consumer (the data subject) and the digital service provider (the data processor). Sectoral privacy laws affirmatively grant data subjects some additional rights and impose additional duties on data processors within this relationship, but most follow a notice and consent template. For instance, rights to greater detail regarding use of data and duties for companies to affirmatively get opt-in consent (as opposed to the more passive opt-out consent) are common features of such laws.[40] Other laws grant consumers rights that strengthen certain forms of individual choice and individual control. For example, the CCPA grants data subjects rights to request information about what data is being collected about them and whether any of their personal data is being sold or disclosed to third parties; it additionally grants data subjects the right to opt-out of the sale of their personal information.[41] FCRA grants data subjects the right to dispute the accuracy of information in their credit reports and to have inaccurate information be updated or deleted.[42] HIPAA gives patients the right to access their health information, to receive notice regarding how their information may be used and shared, and to consent to certain uses of their health information.[43] These consumer rights provide additional scope for the terms private ordering, but (absent a few notable but narrow exceptions) the onus remains on data subjects to exercise these rights.[44]

---

[40] See e.g., Health Insurance Portability and Accountability Act (HIPAA), P.L. no. 104-104, 110 Stat., Fair Credit Reporting Act (FCRA), 15 U.S.C. § 1681c. Children's Online Privacy and Protection Act (COPPA), 15 U.S.C. § 6501 et seq (2000); Family Educational Rights and Privacy Act (FERPA), 20 U.S.C. § 1232g (1974); Gramm-Leach-Bliley Act (GLBA), 12 U.S.C. § 78, § 377; 15 U.S.C. §80 (1999)

[41] California Consumer Privacy Act (CCPA), Cal. Civ. Code, Div. 3, Pt 4, Title 1.81.5 § 1798.100(a); 110(3); 135. See also, Salome Viljoen "The Promise and Pitfalls of California's Consumer Privacy Act," Critical Reflections, 2020.

[42] Fair Credit Reporting Act (FCRA), 15 U.S.C. § 1681c.

[43] Health Insurance Portability and Accountability Act (HIPAA), P.L. no. 104-104, 110 Stat. Like FCRA discussed below, HIPAA does also include a few affirmative data processing obligations and specify certain data uses that are not subject to individual consent. However, the majority of health data sharing do not rely on these exceptions, and instead use a combination of the law's anonymity rules and patient consent to share health data. See Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization,* 57 UCLA L. Rev. 1701 (2010).

[44] Certain elements of FCRA are a notable exception. For example, alongside consumer rights, FCRA places affirmative limits on who may use consumer reports for which purposes. Fair Credit Reporting Act (FCRA), 15 U.S.C. § 1681 (1970). Other exceptions include uses of personally-identifiable information forbidden under HIPAA. Health Insurance Portability and Accountability Act (HIPAA), P.L. no. 104-104, 110 Stat.

2. Individual harm

Existing privacy laws generally contemplate *individual* informational harm of the following forms.

Consentless collection. Collecting data about someone absent their consent is the most basic and the most fundamental form of informational harm contemplated by privacy laws. Obtaining personal information without consent is considered a violation of that person's right to control how information about them is used. This violation harms data subject autonomy and dignity by denying the data subject's right to informational self-determination.[45]

Sludgy consent. Closely connected to collection absent consent are ways in which a corrupted architecture or design process may result in an appearance of consent that in fact violates or undermines true consent. These may include engineering consent through design features that make opting out difficult or almost impossible or using behavioral insights to heavily influence data subjects towards granting consent.[46] Like consentless collection, sludgy consent undermines the true will of data subjects in ways that thwart their capacity for informational self-determination.

Harms of access. Harms of access may occur when people are denied access to information about themselves, violating notions of informational self-determination or when they are unable to limit or control access to information about themselves by others.[47] Harms of excessive access may include chilling effects on self-expression and harassment.[48]

---

[45] This concept of undermining informational self-determination is closely linked to articulations of privacy as control. Alan Westin, PRIVACY AND FREEDOM: LOCATING THE VALUE IN PRIVACY, New York: Atheneum (1967) at 7 (Defining privacy as "The right of the individual to decide what information about himself should be communicated to others and under what circumstances"). See Part I infra for further discussion of privacy as control. See also Julie E. Cohen, *What privacy is For,* 126 Harv. L. Rev. 1904 (2013), 1905 ("[Privacy] protects the situated practices of boundary management through which the capacity for self-determination develops").

[46] Daniel Susser, Beate Roessler, and Helen Nissenbaum, *Online Manipulation: Hidden Influences in a Digital World,* 4 Geo. L. Tech. Rev 1 (2019). These design features are frequently termed "dark patterns," see e.g. Harry Brignull, *Dark Patterns: Inside the interfaces designed to trick you,* The Verge, August 29, 2013; such designs frequently take advantage of behavioral insights from psychology and behavioral economics that are widely used to "nudge" individuals towards socially desirable outcomes, but deploy them for more socially dubious ends. See e.g., Richard H. Thaler, *Nudge, not sludge*, Science Magazine, Aug 2018.

[47] In the copyright realm, Shyamkrishna Balganesh makes a similar and related claim regarding a "disseminative harm," when creators' rights to determine whether and when their works are shared have been violated, which he identifies as "compelled authorship." *Private Copyright,* Vanderbilt L. Rev. 1 (2020).

[48] Privacy scholars arguing for better protection against online harassment and gender-based violence as privacy enhancing argue that harassment may have a chilling effect on the expressive freedoms of vulnerable groups. See e.g. Danielle Keats Citron, *Law's Expressive Value in*

Reidentification. Individuals may be harmed when their identifiable personal data is released, whether intentionally or as a result of a data breach or hack. In some cases, disclosure causes immediate harm, for example reputational harm. Harm may also result from various inappropriate uses, including identity theft or stalking. Many privacy statutes guard against reidentification harm by allowing freer processing and use of information that has been (at least nominally) stripped of identifiers that can be used to reidentify individuals.[49] Statutes also directly address data breaches and restrict or ban certain uses.

Inaccuracy and discrimination. Privacy laws also include a few thicker conceptions of individual informational harm that capture how certain forms of knowledge may cause people to unfairly lose out on important opportunities. For instance, FCRA includes a right to accurate information and the right to delete inaccurate information in credit reports. Ban the box initiatives similarly ban employers from asking about criminal convictions on employment applications, on the theory that this information may unjustly foreclose employment opportunities to deserving applicants.[50]

These forms of informational harm are individual; they identify how information flows may be produced or used in a way that may harm the data subject.

C. Critiques of privacy law and their motivating accounts

While there is general scholarly agreement that data governance is in need of repair, critiques of the digital economy offer up different diagnoses of *why* the status quo is insufficient, *what* the stakes of failure are, and *on what grounds* data governance fails. These diagnoses rest on different underlying claims about

---

Combating Cyber Gender Harassment, Mich, Law Review, Vol. 108, (2009); Danielle Citron and Jon Penney, *When Law Frees Us to Speak*, Fordham L. Rev (forthcoming).; See also, SCOTT SKINNER-THOMPSON, PRIVACY AT THE MARGINS (2020), on how privacy offers a form of expressive resistance to surveillance regimes.

[49] One prominent such example is the Health Insurance Portability and Accountability Act (HIPAA), P.L. no. 104-104, 110 Stat. Premising free processing of information on anonymization is widespread, though increasingly vexed. See Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization,* 57 UCLA L. Rev. 1701 (2010).

[50] Fair Credit Reporting Act (FCRA), 15 U.S.C. § 1681c. Currently thirteen states and the District of Columbia have ban-the-box laws that apply to private employers. See *Ban the Box: U.S. Cities, Counties and States Adopt Fair Hiring Policies,* National Employment Law Project, available at https://www.nelp.org/publication/ban-the-box-fair-chance-hiring-state-and-local-guide/ (last accessed Jan 31, 2021); however, research suggests that in jurisdictions that have passed ban the box laws, employers are more likely to discriminate against young Black applicants. See Osborne Jackson and Bo Zhao, *The Effect of Changing Employers' Access to Criminal Histories on Ex-Offenders' Labor Market Outcomes: Evidence from 2010-2012 Massachusetts CORI Reform,* Federal Reserve of Boston: Research Department Working Papers (2016); Amanda Agan and Sonja Starr, *Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment,* The Quarterly Journal of Economics, Vol. 133 Issue 1, pp 191-235 (Feb 2018).

how information may cause people harm, how information may benefit people, and what this implies for how legal reformers should approach the project of data governance.

1. Failure of notice and consent

Much ink has been spilled on how privacy law fails to secure data subject autonomy and thus prevents the individual and societal goods of privacy from being realized. Notice and consent's inability to protect privacy, to secure against informational harm and to limit flows of data extraction is well established. Like many "shrinkwrap" contracts of adhesion, privacy terms-of-service operate from a legal fiction; individuals do not read the privacy policies to which they consent and have no real way to bargain over the terms they contain.[51] Personal data is also non-rivalrous, non-extinguishable, and reusable, meaning that how it flows and how it is used can change as technologies and business models evolve. This makes data ill-suited to a regulatory approach premised on a one-time exercise of informed, individual choice.[52] Notice and consent's emphasis on personal control at the point of collection aligns poorly with contextually specific, fine-grained concerns over appropriate information flow.[53] Consent is easily circumvented, particularly in digital settings designing for optimal data extraction.[54] Notice and consent does not address the regulatory gap between how privacy may be protected in law and how privacy may be facilitated or eroded in technical design.[55]

---

[51] Katherine J. Strandburg, *Free Fall: The Online Market's Consumer Preference Disconnect,* University of Chicago Legal Forum: Vol. 2013, Article 5; J. A. Obar, & A. Oeldorf-Hirsch, *The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services*, Information, Communication & Society, *23*(1), 128-147. Privacy policies are often long and full of legalese. They are also pervasive. One study from 2008 found that it would take an average user 76 days to read all the privacy policies they encountered in one year alone, with a nationwide annual estimated opportunity cost of $781 billion. See Aleecia McDonald and Lorrie Cranor, *The Cost of Reading Privacy Policies,* I/S: A Journal of Law and Policy for the Information Society, vol. 4, no. 3 (2008), 543-568.

[52] This shortcoming has been given an exhaustive treatment by many privacy scholars. Neil M. Richards and Woodrow Hartzog provide a useful typology categorizing the different ways consent fails to secure privacy in the digital context. See *The Pathologies of Digital Consent*, 96 Washington Univ. Law Rev. 1461 (2019). Elettra Bietti provides a helpful exploration of the normative stakes of this failure. See *Consent as a Free Pass: Platform Power and the Limits of the Informational Turn,* 40 Pace L. Rev. 307 (2020).

[53] HELEN NISSENBAUM, PRIVACY IN CONTEXT (2010); see also Privacy as Commons: Case Evaluation Through the Governing Knowledge Commons Framework; https://www.jstor.org/stable/10.5325/jinfopoli.8.2018.0116#metadata_info_tab_contents

[54] Solon Barocas and Helen Nissenbaum, *Big Data's End Run around Anonymity and Consent*, In J. Lane, V. Stodden, S. Bender, & H. Nissenbaum (Eds.), Privacy, Big Data, and the Public Good: Frameworks for Engagement, 44-75 (2014).

[55] Ari Ezra Waldman, *Privacy Law's False Promise*, 97 Wash.U. Law Review 3 (2019); WOODROW HARTZOG, PRIVACY'S BLUEPRINT (2018).

2. Traditional accounts of the value of privacy

While critiques of privacy law have long focused on the failure of notice and choice to secure the benefits of privacy, there are competing accounts about what, exactly, those benefits are. [56] In general, legal and philosophical accounts consider privacy a predicate condition or instrumental right—part of what a just society offers by way of robust protection for individual autonomy or individual dignity.[57] On this view, privacy erosion threatens the vital conditions that foster the individual's ability to think for herself, enjoy a privileged relationship to her inner desires, know her own mind and express it as she chooses, and be in charge of her own formation as a social, political and economic being.

The focus on individual selfhood guides how laws aim to secure privacy, expressed in the canonical purpose of data governance: informational self-determination.[58] This purpose is consistent with classic legal views of privacy as control and privacy as access, both of which offer ways to secure and enact self-determination.

Privacy as control. Many early and influential legal theories of privacy adopted the view of privacy as control. Alan Westin's *Privacy and Freedom: Locating the Value in Privacy* defines privacy as "the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others."[59] Charles Fried defines privacy as "not simply an absence of information about us in the minds of others, rather […] the control we have over information about ourselves."[60] More recent scholarship has similarly adopted this view. Michael Froomkin defines privacy as "the ability to control the acquisition or release of information about oneself."[61] Jerry Kang defines it as "an individual's control over the processing— i.e., the acquisition, disclosure, and use— of personal information."[62]

Privacy as access. A distinct yet related account views privacy as access. Ruth Gavison is a proponent of this view and traces this account in privacy laws that share a concern with intrusions of knowledge and information: under what conditions knowledge of one may be gained, what may be known by whom, how

[56] DAN SOLOVE, UNDERSTANDING PRIVACY (2008) at 1.

[57] HELEN NISSENBAUM, PRIVACY IN CONTEXT (2010).

[58] Neil Richard and Woody Hartzog, "Duty of Loyalty in Privacy law"; Woodrow Hartzog and Neil Richards, Privacy's Constitutional Moment and the Limits of Data Protection, 61 B.C. L. Rev. (forthcoming 2020);

[59] ALAN WESTIN, PRIVACY AND FREEDOM: LOCATING THE VALUE IN PRIVACY (1967) at 7.

[60] Charles Fried, "Privacy: A Moral Analysis" 77 Yale L. J. 1, 1968, at 482.

[61] Michael Froomkin, *The Death of Privacy*, 52 Stan. L. R. 1461, 1464 (2000).

[62] Jerry Kang, *Information Privacy in Cyberspace Transactions,* 50 Stan. L. R. 1193, 1203 (1998).

such knowledge may be used, and what effects such uses of knowledge may produce.[63]

3. Alternative accounts: The social value of privacy

Others link the failure of notice and consent to the collective nature of privacy harm, noting the contextual nature of information flow,[64] the collective action problems and market failures it produces,[65] the externality effects from individual transactions,[66] and the epistemic constraints of individualistic frames of reference.[67] These diagnoses track alternative accounts of what privacy is for, emphasizing the social value of privacy and rejecting the atomistic conceptions behind privacy protection as informational self-determination. These accounts advance a thicker conception of autonomy that includes privacy's importance in fostering conditions of public citizenship and public governmentality.

For instance, Priscilla Regan argues for the social importance of privacy's ability to facilitate democratic political flourishing through its protection of free association and free speech. She also emphasizes the common stakes of privacy given market forces that make it difficult for any one individual to have privacy unless a minimum is guaranteed to everyone.[68] Helen Nissenbaum develops an account of privacy as appropriate information flow, where context-appropriate information sharing is determined by reference to socially developed norms.[69] Julie Cohen offers a variant of the social privacy account that aims to depart from the liberal conception of the autonomous subject, arguing for privacy as vital for the socially constructed subject instead. For this subject, privacy shelters, "dynamic, emergent subjectivity" from data-driven attempts to render these subjects "fixed, transparent, and predictable." This capacity is vital for self-definition, critical self-reflection, and informed citizenship—the necessary conditions for liberal democracy.[70]

---

[63] This list accords with Ruth Gavison's highly influential view of privacy as a measure of the access others have to you through information, attention, and physical proximity. See Ruth Gavison, "Privacy and the Limits of Law," 89 Yale L. J. 3, (1980).

[64] HELEN NISSENBAUM, PRIVACY IN CONTEXT (2010).

[65] Katherine J. Strandburg, *Free Fall: The Online Market's Consumer Preference Disconnect*, U. Chicago Legal Forum: Vol. 2013, Article 5. (2013)

[66] Daniel, J. Solove, *Privacy self-management and the consent dilemma*, 126 Harv. L. R. 1880-1883 (2013); Joel R. Reidenberg et al., *Privacy harms and the effectiveness of the notice and choice framework*, ISJLP 11 (2015): 485.

[67] JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM (2019), at 67.

[68] PRISCILLA M. REGAN, LEGISLATING PRIVACY: TECHNOLOGY, SOCIAL VALUES, AND PUBLIC POLICY (1995). Neil Richards advances a similar claim regarding the necessity of intellectual privacy for robust free expression. *Intellectual Privacy*, 87 Texas L. Rev. 387 (2008).

[69] HELEN NISSENBAUM, PRIVACY IN CONTEXT (2010).

[70] Julie E. Cohen, *What privacy is For,* 126 Harv. L. Rev. 1904 (2013).

These thicker accounts rightly identify the social effects that drive privacy erosion and the social consequences of privacy erosion that extend beyond its individual consequences; yet their normative account remains anchored around claims that privacy erosion is primarily wrong because it threatens the capacity for individual self-formation. While these accounts do take social effects into account, such effects serve to heighten the stakes or increase the challenges of the primary normative task—securing privacy protections in order to secure conditions of individual self-formation and self-enactment.

Both standard and thick accounts of autonomy inform how critics view the stakes of privacy law's failure. On these accounts, data production practices are wrong when they lead to manipulation, erode self-determination in the data market (and beyond), chill self-expression, or when they involve forms of data extraction and algorithmic governmentality that infringe on an individual's capacity to act as a moral agent.[71]

These accounts also guide views on how data governance should be reformed. As will be discussed in Part III, concerns over autonomy and dignity guide dignitarian efforts to reduce datafication's commodification of inner life and to increase the regulatory oversight of data flows that act back on users in ways that wrongfully undermine their self-will. Concern over loss of control and lack of clear legal rights to data's value motivate propertarian efforts to formalize market rights to data for data subjects. These reforms and others focus on increasing data subjects' capacity to determine how (and under what conditions) their data is collected, processed, and used.

II. DATA RELATIONS AND THEIR SOCIAL EFFECTS

One way to evaluate different theories of data governance is to examine how such theories conceive of (and propose to act upon) the social relations structured by data flows.[72] To understand the significance of data's relationality,

---

[71] Many privacy and digital rights activists focus on these effects, especially in the context of private systems that further personal violation for profit. On manipulation, see Daniel Susser, Beate Roessler, and Helen Nissenbaum, *Online Manipulation: Hidden Influences in a Digital World,* 4 Geo. L. Tech. Rev 1 (2019); Ryan Calo, *Digital Market Manipulation*, George Wash. L. Rev. Vol. 82, 995 (2014). On eroded self-determination, see Woodrow Hartzog, *Privacy's Blueprint*, Harvard University Press, 2018. See also, Woodrow Hartzog and Neil Richards, *The Pathologies of Digital Consent*, Wash. U. L. Rev. (2018). On chilling effects of self-expression, see Citron, Danielle Keats, *Law's Expressive Value in Combating Cyber Gender Harassment* (2009). Michigan Law Review, Vol. 108, (2009); Danielle Citron and Jon Penney, *When Law Frees Us to Speak*, Fordham L. Rev (forthcoming). On data extraction and algorithmic governmentality, see Elettra Bietti and Jennifer Cobbe, Rethinking Digital Platforms for the post-Covid 19 era, Centre for International Governance Innovation (May 12, 2019).

[72] This Article adapts its concept of "data relations" from prior work. Nick Couldry and Ulises Mejias use the term 'data relations' to describe the process of capturing and processing social data, which they argue results in a new social order based on continuous tracking. *Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject,* Television & New

let us consider with greater specificity how data relates people to one another, how such relations may produce social effects, and which of these relations are (and are not) accorded legal relevance by current and proposed forms of data governance law

A. Data Governance's Sociality Problem

1. Scenario: TattooView AI, Adam, and Ben.

In July of 2018, privacy activists reported that the Federal Bureau of Investigation (FBI) along with the National Institute for Standards and Technology (NIST) were evaluating the effectiveness of tattoo recognition technology.[73] To conduct this evaluation, the FBI provided access to their database TAG-IMAGE, which includes images of thousands of prisoner tattoos collected from prison inmates and arrestees, to 19 company and academic groups with the goal of developing image recognition technologies capable of identifying individuals by their tattoos, as well as identifying which tattoos are markers of gang affiliation.[74]

Consider a scenario where the FBI partners with a company, TattooView AI, to provide tattoo recognition products and automated matching—identifying not only a particular individual via their tattoo, but also whether this tattoo is a sign of gang membership more generally.[75] This tool is then used by law enforcement to identify potential gang members for heightened police observation.

Media, Vol 20 Issue 4: 336-349 (2019). Data social relations are constituted by both legal and technical systems that influence how data is created, collected, transmitted, and used.

[73] Dave Maass, *FBI Wish List: An App that Can Recognize the Meaning of Your Tattoos*, Electronic Frontier Foundation, July 16, 2018.

[74] TAG-IMAGE is one of several forms of biometric markers included in the Next Generation Identification (NGI) used to automate processes of biometric identification capabilities and extend the tracking of biometric markers beyond those included in the FBI's Automated Fingerprint Identification System (IAFIS). The Biometric Center of Excellence (BCOE) is the agency's primary group working to develop biometrics and identity management. The BCOE notes that tattoos and other biometric markers have possible uses for law enforcement well beyond purposes of identify verification. Most notably, automated recognition services allow investigators to use a probe or query image to find similar images. ("While the value of image-to-image matching technology is obvious from an identification perspective, the benefits of knowing the symbolism and background behind tattoos and graffiti can be equally valuable. From an intelligence standpoint, certain symbols or graffiti may be used to help establish whether an individual is associated with a particular gang, terrorist organization, or extremist group. This may help determine the extent to which the individual or gang poses a threat to law enforcement or the community, and possibly to recognize and link crimes across the country."), "Image-Based Matching Technology Offers Identification and Intelligence Prospects," CJIS Link, Vol 14, No 3 (December 28 2012), available at: https://www.fbi.gov/services/cjis/cjis-link/image-based-matching-technology-offers-identification-and-intelligence-prospects

[75] This scenario gains plausibility from three facts in addition to the 2018 NIST trial. First, law enforcement already tracks and identifies gang membership on the basis of certain shared tattoos [CITE- LAPD gang database, Georgia gang database, UK gang database]. Second, law

This biometric data flows across several parties—from the initial arrestee, to managers of TAG-IMAGE, to third parties such as TattooView, to the point of its use by a law enforcement officer to detain a suspected gang member. It also flows across several legal regimes: criminal law, government procurement and trade secrecy law, contract law and privacy law. Yet this flow begins and ends with two human events: first, a person has his tattoo photographed and added to TAG-IMAGE (let's call him Adam) and second, a person with the same tattoo is detained using that image data (let's call him Ben).[76]

Standard privacy critiques of data flows like this one emphasize not only the significant stakes of this data flow for both Adam and Ben, but also that this data was collected from Adam under highly coercive conditions (i.e., while being detained in prison), for a purpose (to identify other gang members) with which he may not agree and had no say over. Adam's lack of agency at the point of this data's collection is accorded significant moral and legal relevance in critiques of biometric surveillance.[77]

---

enforcement already partners with private companies to access automated biometric identity verification and investigation tools. See Kashmir Hill, *The Secretive Company that Might End Privacy as we Know It*, N. Y. TIMES, Jan. 18, 2020. Third, it was recently reported that Palantir already sells capabilities very similar to the hypothetical ones described below to law enforcement. Caroline Haskins, *Scars, Tattoos, and License Plates: This is What Palantir and the LAPD Know About You*, BUZZFEED NEWS, September 29, 2020, available at: https://www.buzzfeednews.com/article/carolinehaskins1/training-documents-palantir-lapd

[76] This Article will explore several scenarios where the data collected about one person may be used against another person. Through these scenarios we can examine the social effects of many common and widespread data practices. For example, data about one person may affect another person via the linkage of two datasets, by revealing data about social networks or genetic information that are by definition shared information, or by applying a prediction algorithm trained off of the data of one person and used against another. This relational effect can have a considerable outside impact. For example, Cambridge Analytica directly collected data from the 270,000 people who downloaded the "This Is Your Digital Life" application. Because Cambridge Analytica was able to receive those people's social network data (the profiles of their friends and family), they obtained the profile information of about 87 million users (70.6 million in the U.S.). This information was used to train an ad-targeting program that delivered micro-targeted political advertisements to some portion of Facebook's 190 million U.S. users (as well as users in the UK and elsewhere), based on their likelihood to respond to a given advertisement. See Nicholas Confessore, *Cambridge Analytica and Facebook: The Scandal and the Fallout So Far,* N. Y. TIMES, APRIL 4, 2018. Sam Meredith, "Facebook-Cambridge Analytica: A timeline of the data hijacking scandal," CNBC, April 10, 2018. Hannes Grassgegger & Mikael Krogerus, *The Data the Turned the World Upside Down*, January 28, 2017, Vice; Issie Lapowsky, *Facebook Exposed 87 Million Users to Cambridge Analytica*, April 4, 2018, Wired.

[77] See e.g., Complaint at 3, ACLU et al v. Clearview AI, Inc, No. 9337839 (Cook County Cir. Ct.) ("The ability to control their biometric identifiers and to move about in public, free from the threat of surreptitious unmasking or surveillance, is essential to Plaintiffs' members, clients, and program participants in Illinois…Clearview has captured more than three billion faceprints from images available online, all without the knowledge—much less the consent—of those pictured"), available at: https://www.aclu.org/legal-document/aclu-v-doj-face-recognition-surveillance-complaint.

Adam and Ben's interests are sufficiently aligned such that enhancing Adam's legal rights vis-à-vis TattooView will likely also protect Ben's interests. If Adam is granted a robust right to refuse inclusion of his tattoos in TAG-IMAGE then he is likely to exercise that right, and then his tattoo image data cannot be used to detain Ben. But consider an alternative scenario: TattooView AI develops its tattoo recognition algorithm not from the FBI's TAG-IMAGE dataset, but from a dataset it obtained when TattooView purchased TattooID. TattooID is a social platform where tattoo enthusiasts can share photos of their tattoos, tag their tattoo artist, and search for designs. Suppose that Adam, a former gang member who regrets his gang involvement, voluntarily shares his tattoo images on TattooID, and tags them as tattoos related to gang affiliation in the hopes that they can help identify other gang members.

In this alternative scenario, individualist conceptions of how this data may harm Adam do not capture the way this data flow affects Ben. Adam was not coerced into sharing this data, but instead did so willingly. Moreover, the purposes to which this data is being applied align with Adam's intent in sharing it and would in his view represent a valid outcome. And yet the fact remains that Ben faces significant consequences from this data flow. To the extent Ben's interests are of legal as well as normative relevance, this presents a problem for data governance law.

B. Mapping data social relations along vertical and horizontal axes

The relationships that arise among data subjects, data producers, and the third parties impacted by data use can be mapped along two axes.

1. Vertical data relations

Along the *vertical axis* lies the data relation between an individual data subject and an individual data collector (also known as a data processor). The vertical data relation describes the relationship between Adam and TattooView AI, when Adam agrees to the terms of data collection laid out by TattooView AI and shares his data with them. This vertical data relation structures the process whereby data subjects exchange data about themselves for the digital services the data collector provides.

This vertical social relation is expressed technically via the flow of data from a data subject to a data collector, and legally via the contractual terms that structure terms of exchange between data subject and data collector, as well as background consumer and privacy law regimes that allocate privileges, claims, and duties among the two parties. This vertical data relation is in some sense well-understood in data governance law. As will be discussed in greater detail in Part Three, proposals for data governance reform are attentive to how the law governs this vertical relation, how it may structure unequal relations among data subjects and data producers, and how duties and rights between these two parties may be reallocated to address this imbalance.

2. Horizontal data relations

Data production also relates people via the capacity of data flows to reveal information about third parties. The *horizontal axis* describes how data production relates data subjects not to data collectors, but to one another and to others that share relevant population features with the data subject. The relationship between Adam and Ben describes a horizontal data relation.

This horizontal relation is expressed technically through informational infrastructures that make sense of data subjects via group classification, and that operationalize classifications to act back on subjects. These technical expressions apprehend (and in apprehending, help to define) the social fact of group identity via shared preferences, social patterns, and behaviors that make people similar to one another. For example, the horizontal data relation between Adam and Ben apprehends a particular social meaning based on their shared tattoo. This horizontal data relation structures a social process whereby a relevant shared feature (i.e. tattoo) is operationalized to make a prediction and define a social meaning (i.e. gang member) and act back on a group member (Ben) according to this grouping.

Horizontal relations are not actually one-to-one relations between data subjects like Adam and those impacted by data flows like Ben. They are population-based relations. For instance, sharing his tattoo image puts Adam in horizontal relation not only with Ben, but with everyone who also has his tattoo and may be acted upon on the basis of this shared feature. We can call this population P_{use}, such that P_{use} = [Ben, B_1, B_2, B_3…], where B_n = other individuals in P_{use}. And the same holds the other way around. Ben is not only in horizontal relation with Adam, but with everyone who has a relevant population feature in common with him (in this case, his tattoo) and has shared data about ths feature with a data collector. We can call this population $P_{collect}$, such that $P_{collect}$ = [Adam, A_1, A_2, A_3…], where A_n= other individuals in $P_{collect}$.

These population-level relations give rise to *population-level* interests along the horizontal relation. For example, we can understand Ben's interest in Adam's data collection as one instance of the more general interest of P_{use} in $P_{collect}$'s data collection. This interest, unlike those along the vertical relation, *does not reduce to the individual provenance of the data*. Ben's interest in $P_{collect}$'s data sharing is based on the effect that use of this data with will have on him. This use may occur regardless of whether this data was collected from him, from Adam, or from someone else. In this sense, it does not matter who the data "came" from, but what such data says about Ben, and how such meaning is used to act upon Ben. This is the population-level interest Ben (and others like Ben) have in data that apprehends a relevant shared population feature about them. As the example shows, this interest may arise from data Ben shares, data Adam shares, or data someone else shares. Each individual instance of this interest may

be weak but they occur at scale throughout the data production economy, and link individuals to many other individuals via webs of horizontal relation.

3. The significance and the puzzle of data relations

One way to understand data governance's unsatisfying response to downstream social effects from data collection (the sociality problem) is data governance law's conceptual commitment to individualism (DIM). This commitment focuses data governance reform on how data production may harm data subjects and develops legal responses to such harm. While this may result in improvements to the vertical data relation between data subjects like Adam and data collectors like TattooView, it does not address the role horizontal data relations play in producing social value and social risk. This has several significant consequences discussed in greater detail below.

C. The importance of horizontal data relations in the digital economy

While horizontal data relations are minimally relevant to data governance law, they are central to how data production produces both social value and social risk. Data production for the digital economy is deeply, even fundamentally, relational.

Data flows are quite literally structured, collected, and produced so as to relate people to one another.[78] Data flows are useful, and do the work they are supposed to do, when they relate people to one another. Data flows are designed to represent the ways that people are like one another and reveal meaningful things about one another; how we are alike biologically, interpersonally, politically, and economically.[79]

---

[78] BEN GREEN, THE SMART ENOUGH CITY (2019). For more on the utility of data because of its ability to reveal information on others, see e.g. Sebastian Benthall and Jake Goldenfein, "Data Science and the Decline of Liberal Law and Ethics", Ethics of Data Science Conference (2020); Sebastian Benthall, Seda Gürses, and Helen Nissenbaum, *Contextual integrity through the lens of computer science*, Now Publishers, 2017.

[79] Almost all data harvested from an individual person or personal device has the capacity to be relational. Social media data reveals information (such as preferences and observations) not just about an individual, but also about her social networks. This information can have political as well as social consequences. For example, network data can be used to probabilistically identify support or opposition for a political candidate or position to target political advertising or get-out-the-vote efforts. Robert Bond et al, *A 61-million-person experiment in social influence and political mobilization,* Nature 489, 295–298 (2012), available at https://doi.org/10.1038/nature11421. Network data can be used predict a credit score and help proactively track and target a disease, suicides, and gun violence. Ben Green, Thibault Horel and Andrew Papachristos, *Modeling Contagion Through Social Networks to Explain and Predict Gunshot Violence in Chicago, 2006 to 2014*, JAMA Intern. Med. 177(3) 326-333 (2017). Genetic data from a consumer genome test reveals information about one's relatives that can help them detect disease early, or leave them vulnerable to ethnic or racial discrimination. ALONDRA NELSON, THE SOCIAL LIFE OF DNA: RACE, REPARATIONS, AND RECONCILIATION AFTER THE GENOME (2016). Location data reveals information about one's household. In fact almost all data harvested from one person that can be used to make a prediction about them or

Data flows classify and sort people along particular categories of group membership—this sorting and classifying is how an individual becomes rendered as a data subject and is how economic value from production is realized. In other words, data about individuals is useful in the digital economy because it helps to define relevant group categories. These categorizations are operationalized to make sense of people on the basis of their classifications and to act back on such insights.

Data about populations is used to develop models to predict and change behavior, to gain intimate consumer or competitor knowledge for market advantage, or to retain greater surplus value.[80] Such activities demonstrate an orientation towards data subjects that recognizes them not as individuals, but as members of groups that are constituted via their shared features and common patterns of behavior.[81] This process recasts people as assemblages of their social relations and group behaviors, and apprehends data subjects as patterns of behavior derived from group-based insights. This basic approach is what makes behavioral targeting, prediction tasks, at-scale risk assessment, and modulated feedback systems both possible and profitable.[82]

Data's relationality is central to how data collection produces economic value. This distinguishes the value of Adam's data for the machine learning (ML) or artificial intelligence (AI) applications of the contemporary digital economy from the value personal data has for older forms of consumer surveillance (and that inform the law's current approach to data privacy). Prior to the widespread availability of large-dataset computing technology, data about a data subject like Adam may have been valuable because it helped businesses,

---

attempt to change their behavior can be used, in the form of behavioral model, to make a prediction or attempt to change the behavior of others. For more on the utility of data because of its ability to reveal information on others, see e.g. Sebastian Benthall, Sebastian, Seda Gürses, and Helen Nissenbaum, *Contextual integrity through the lens of computer science*, Now Publishers, 2017. Institutional economics literature extols the competitive value of data via tailoring, prediction, personalization, nudging and marketplace design. See e.g. "The Rise of Data Capital" MIT Technology Review Custom Report, 2016. ("Data is now a form of capital, on the same level of financial capital in terms of generating new digital products and services.") This literature aligns with conceptions of human behavior as predictive and probabilistic developed in cybernetics. See J. BENIGER, THE CONTROL REVOLUTION: TECHNOLOGICAL AND ECONOMIC ORIGINS OF THE INFORMATION SOCIETY (1986). On the concept of data enclosure, see JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM (2019) at 67.

[80] Salome Viljoen, Jake Goldenfein and Lee McGuigan, *Economic Method, Digital Platform: When Mechanism Design Moves Online*, manuscript on file with author.

[81] See Jake Goldenfein, Monitoring Laws: Profiling and Identity in the World State (2019) Sebastian Benthall and Jake Goldenfein, "Data Science and the Decline of Liberal Law and Ethics", Ethics of Data Science Conference (2020); FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION, (2015).

[82] Ibid.

employers, the government, and insurers know things about Adam. And to some extent, this is still what makes data about Adam valuable.

But what makes data about Adam particularly and distinctly valuable for the contemporary digital economy is its capacity to help companies make predictions or change the behaviors of *others* based on relevant population features they share with Adam—in other words, on the basis of at-scale population-level horizontal data relations. It is this relational value of data that drives much of the imperatives to data access, processing, and use. The distinctive feature of ML- and AI-based systems is that they can be used to know things about Adam that Adam does not know, by inferring back to Adam from A*n*. And, of greater legal significance (or concern), data from A*n* can be used to train models that "know" things about B*n*, a population that may not be in any vertical relation with the system's owner. This is the key shift of at-scale data analysis, as compared to prior digital data collection and use approaches that did not have access to the scope and degree of data aggregation, computation capabilities, and inference models that typify digital economic activity in the past decade. It also highlights the importance of horizontal data relations not only for expressing an expanded set of interests in data flows, but in structuring the incentives of data collectors along vertical data relations with data subjects.

1. Two implications of data relationality's economic significance

Two implications follow from recognizing the significance of data's relationality in the digital economy. First, conceiving of data's horizontal relationality as incidental to the task of managing data production is wrong. Data's horizontal relationality does result in observable externality effects (from the perspective of the data subject and from that of status quo data governance); however, conceiving of these effects as "external" to the purposes and uses of data that drive entities to transact for it is incorrect.[83] Enacting horizontal relations is not like producing pollution; if polluters could magic away pollution they likely would (if only to save themselves some reputational harm). But the same cannot be said for data producers: data's relationality is central to the business of data production and constitutes much of what makes data production economically valuable to begin with.

Second, data's aggregate effects amplify the consequences of this disconnect. In a typical data flow, any one individual's data is essentially meaningless, and the marginal cost of any one individual defecting from collection is very low.[84] Yet in aggregate, data is highly valuable and grows in

[83] Daniel, J. Solove. *Privacy self-management and the consent dilemma.,* Harv. L. Rev. 126 (2013): 1880-1883. Reidenberg, Joel R., et al. *Privacy harms and the effectiveness of the notice and choice framework*, ISJLP 11 (2015): 485.

[84] Michael Mandel, *The Economic Impact of Data: Why Data is Not Like Oil*, Progressive Policy Institute Report, at 6 ("[U]nused data, by itself, has uncertain economic value. Its value depends on how it is combined and used with other data"); see also Alessandro Acquisti, J

value the more data can be combined with other kinds of data.[85] Across many different fields of algorithmic development and machine learning—from computer vision to natural language processing to adversarial machine learning—the rule of thumb is that quality and quantity of data in a model's training set is the biggest determinant of overall performance.[86] More data means better models, which result in digital products that make better predictions about both data subjects as well as others who share relevant features with data subjects. Large scale data collection and aggregation therefore becomes a key competitive advantage in the digital economy.

Treating data's relationality as an accidental byproduct of data creation in our legal conceptions of data misdiagnoses a feature as a bug. The combination of relational and aggregate effects from data production drives companies to collect as much data as possible from data subjects. Data subjects are in turn poorly equipped to exert meaningful coercive force back on to data collectors in the face of such strong incentives. But the issue is not simply a mismatch in the relative coercive power between parties, but the wide range of interests that are not represented in these transactions at all, even while the economic benefits of exploiting these interests motivate the data collection practices of digital firms.

The prevalence of horizontal interests in data thus creates a structural mismatch in vertical relations between data subjects and data collectors. Data subjects possess only a fraction of the interests in a given data flow (and as described above many of their interests in information do not reduce to their vertical transaction with a data collector either); meanwhile data collectors are highly motivated to collect as much data from as many data subjects as possible

---

Grossklags, *Privacy and rationality in individual decision making* IEEE Security & Privacy, vol 2, p. 24 – 30 (2005); Alessandro Acquisti, Curtis Taylor, Liad Wagman, *The Economics of Privacy*, Journal of Economic Literature, *54* (*2*), *442--492*, (2016); Alessandro Acquisti, Leslie John, George Loewenstein *What is Privacy Worth?,* The Journal of Legal Studies, *42*(*2*), *249--274*, (2013). See also Arrieta Ibarra, Imanol and Goff, Leonard and Jiménez Hernández, Diego and Lanier, Jaron and Weyl, Eric Glen, Should We Treat Data as Labor? Moving Beyond 'Free' (December 27, 2017). American Economic Association Papers & Proceedings, Vol. 1, No. 1, Forthcoming. Available at SSRN: https://ssrn.com/abstract=3093683; ERIC POSNER AND GLEN WEYL, RADICAL MARKETS: UPROOTING CAPITALISM AND DEMOCRACY FOR A JUST SOCIETY (2018).

[85] Ibid.

[86] Pedro Domingos, *A Few Useful Things to Know about Machine Learning,* Communications of the ACM, Vol. 55, No. 10 (2012) at 84 ("More data beats a cleverer algorithm"); see also Amandalynne Paullada et al, *Data and its _(dis)contents: A survey of dataset development and use in machine learning research,* NeurIPS 2020 Workshop: ML Retrospectives, Surveys & Meta-analyses (2020), at 1 ("The importance of datasets for machine learning research cannot be overstated."); Alon Halevery, Peter Norvig, and Fernando Periera, *The unreasonable effectiveness of data,* IEEE Intelligent Systems, 24(2): 8-12 (2009); Chen Sun, Abhinav Shrivastava, Saurabh Singh, and Abhinav Gupta, *Revisiting unreasonable effectiveness of data in deep learning era,* In Proceedings of the IEEE International Conference on Computer Vision, pp 843-852 (2017).

in order to realize the considerable benefits that accrue from exploiting the insights of horizontal data relations. Without accounting for horizontal relations in data governance law, the interests they represent and the behaviors they motivate from data collectors cannot be fully accounted. Misdiagnosing these effects as incidental to the task of preventing further privacy erosion risks developing reforms that are not up to the task of disciplining excessive or overly risky data production.

D. The absence of horizontal data relations in data governance law

While horizontal data relations are of primary importance in explaining why data collectors develop infrastructures to collect and monetize data flows, they do not feature much, if at all, in how current data governance law allocates claims, privileges, and duties among actors in the digital economy. Many of the relevant interests in data production that accrue along these population-level relations unrepresented in data governance law.

This has both practical and normative implications. First, as a practical matter, the absence of legal interests for horizontal data relations leaves the law out of step with the importance of these relations for the digital economy. As discussed above, this may preclude effective regulation of vertical relations as well. The imperatives to relate individuals along the horizontal axis motivate data collectors and influence the conditions of exchange between them and data subjects; horizontal relations are therefore relevant to the task of regulating data subject-data collector vertical relations.

Second, the absence of horizontal data relations in law may cause data governance law to miss—or misconceive—how data production results in particular kinds of injustice. Because these population-level interests are not represented, data governance law is not indexing forms of injustice that operate via horizonal relations. This misconception may also lead to regimes of data governance that inadvertently foreclose socially beneficial forms of data production. The second implication is discussed in greater detail below..

1. Horizontal relations and social informational harm

The legal marginality of horizonal data relations leaves many consequences of data production unaccounted for in data governance law. This includes externalities (such as Ben's lack of representation) in how the law accounts the sum of risks and benefits in the data flow from Adam to TattooView AI. But it also leaves unaddressed distributive effects: how data flows spread the benefits and risks of data production unevenly among actors in the digital economy, often along the lines of group identities that serve to inscribe forms of oppression and domination. For instance, if Ben is Black, the incapacity of data governance law

to represent Ben's interests in Adam's data flow presents problems that are of a different (arguably more significant) normative quality, given the way this data flow materializes a racialized social process (i.e. "detaining a Black man on the basis of suspected gang membership"). This distinction between non-representation of horizontal relation interests and the uneven *stakes* of non-representation are explored in greater detail below.

a. Unjust data collection may result in social informational harm

First, certain forms of data production may equally subject individuals to coercive forms of data collection but lead to unequally harsh consequences from the resulting data flows. While coercive collection practices may generally constitute unjust vertical relations, the resulting horizonal relations may enact normatively distinct group-based forms of oppression.

For example, consider the recent purchase of Immigration and Customs Enforcement (ICE) and Customs and Border Protection (CBP) of mobile location data from the company Venntel to identify and arrest suspected undocumented immigrants on the basis of mobile phone activity in remote borderlands.[87] Location data in Venntel's database tracks location information from millions of mobile phones, and is drawn from mobile applications like games and weather apps that request access to users' location data. ICE has also purchased licenses from Clearview AI, a facial recognition company that recently drew public scrutiny for its widespread use among law enforcement agencies and dubious—possibly even illegal—data collection practices.[88] In both instances, millions of data subjects are subject to data collection practices by Clearview and Venntel that may fail to meet the standard of meaningful consent. Many data subjects may find these data practices unfair or unjust, and express interest in reforming data collection and use practices to address them.[89]

However, the *risks* from how this data is used fall unevenly among the population of those they impact. This in turn presents a different class of harm

[87] Byron Tao and Michelle Hackman, *Federal Agencies Use Cellphone Location Data for Immigration Enforcement,* WALL STREET JOURNAL, Feb 7, 2020, available at: https://www.wsj.com/articles/federal-agencies-use-cellphone-location-data-for-immigration-enforcement-11581078600?mod=hp_lead_pos5. See also Paul Blest, *ICE is Using Location Data form Games and Apps to Track and Arrest Immigrants, Report Says*, VICE NEWS, Feb 7, 2020.

[88] Kim Lyons, *ICE just signed a contract with facial recognition company Clearview AI,* The Verge, August 14, 2020. Clearview AI built its facial recognition database by scraping publicly-available face images from the web, in violation of Illinois' Biometric Information Privacy Act, which requires companies to obtain notice from consumers before collecting and using their biometric information. Biometric Information Privacy Act (BIPA), 740 ILCS/14, Pub. Act 095-994 (2008).

[89] Andrew Perrin, "Half of Americans have decided not to use a product or service because of privacy concerns," Pew Research, April 14, 2020; Bruce Schneier, *We're Banning Facial Recognition. We're Missing the Point,* N. Y. TIMES, Jan 20, 2020.

than simply unjust conditions of data collection. The Venntel data flow enacts horizontal data relations whereby a relevant shared feature (i.e. movement patterns) is operationalized to make a prediction (i.e. undocumented immigrant) and act back on a group member according to this categorization (i.e. detain them). This amplifies the stakes of this data flow (and the shared feature it acts upon) on the basis of membership in a socially oppressed group. Individuals who—due to their race, ethnicity, religion, or language—are subject to heightened scrutiny from immigration officials face disproportionate risks to themselves and others like them from having their movement patterns (or those of people like them) apprehended via these data flows.

This is not due to some inherent oppressive feature of movement patterns. Instead "movement patterns" as a relevant shared population feature become constitutive of how members of this population are socially defined and acted upon in oppressive ways. Movement patterns become a usfeul identifying feature for undocumented immigrants and is then acted upon to detain group members on the basis of their immigration status. In other words, this horizontal relation materializes a social process of oppression. If one is *not* a member of the relevant group (undocumented immigrant), one faces negligible risk of this kind of social informational harm, even if one's location data is being collected.

These unevenly distributed risks suggest that even where data subjects are subject to equal conditions of collection, the benefits and risks from use may be spread unevenly, amplifying the harmful social consequences of minority group memberships. This harm is normatively distinct from potentially unjust data collection; it locates injustice in the social process this data flow enacts, not the conditions under which it was collected. Thus, reducing concerns over this data flow to the (unjust) conditions of collection alone under-represents both the overall stakes of collection, and the normative significance of how and why such stakes are distributed unevenly.

b. Voluntary data collection may amplify social inequality

Second, more socially advantaged groups may engage in voluntary data collection that benefits them yet results in greater risks of harm for socially disadvantaged groups. The horizontal relations between voluntary data subjects and involuntary third parties may materialize social processes that amplify the (oppressive) differences between groups. For example, consider a scenario where a homeowner (let's call her Alice) voluntarily installs the Amazon Ring, a popular internet- and video-enabled doorbell that allows residents to remotely record their front porch and speak to individuals. Like many Ring users, Alice also joins Ring's Neighbors app, which allows her to receive and post real-time

crime and safety alerts.[90] Alice knows and approves of the partnership between Neighbors and her local law enforcement agency.

Alice has two neighbors, Beatrice (who is white) and Cara (who is Black). Because data collected from Alice's Ring may be used to report and act on Beatrice and Cara on the basis of a shared feature (i.e. they all live in the same small radius in which Ring-based alerts may lead to intervention) both are in horizontal data relationships with Alice. Both Beatrice and Cara are third parties to Alice's transaction with Ring. Both bear externalities from Alice's relationship with Ring due to their unrepresented interests in this data flow. Both may benefit from having their porches under Alice's surveillance, but both also incur some risk: shared population data about them flows from Alice's Ring to the Neighbor App, Amazon, and local law enforcement. Yet Cara incurs greater risk of possible violence from this horizontal data relation than does Beatrice. Her data relation with Alice is one way the pre-existing unjust social processes of racial hierarchy are materialized. This materialized social process is what makes it more likely that this surveillance leads to violence against her from law enforcement or other neighbors. Thus, the two horizontal relations between Alice and Beatrice and Alice and Cara carry normatively distinct meanings: one may result in the productive or distributive inefficiencies that arise due to externalities, while the other may serve to reproduce or amplify racism.

These disproportionate risks suggest that even when data subjects voluntarily consent to data collection, relevant horizontal relations remain unrepresented in law in ways that can amplify the harmful and subordinating consequences of marginal group membership.

III. DIM REFORMS AND THEIR CONCEPTUAL LIMITS

Part Three evaluates two prominent legal reform proposals that have emerged in response to concerns over datafication. *Propertarian* proposals respond to growing wealth inequality in the data economy by formalizing individual propertarian rights over data as a personal asset. *Dignitarian* reforms respond to excessive data extraction's threat to individual autonomy by granting fundamental rights protections to data as an extension of personal selfhood. While both reforms have some merit they suffer from a common conceptual flaw: both attempt to reduce legal interests in information to individualist claims subject to individualist remedies that are structurally incapable of representing the horizontal, population-level interests of data production. This in turn allows

[90] Ring.com/neighbors, available at: https://store.ring.com/neighbors. Neighbors has entered into video-sharing partnerships with over 1300 local law enforcement agencies. See Atlas of Surveillance, EFF, available at: https://atlasofsurveillance.org/; see also Khaleda Rahman, P*olice Are Monitoring Black Lives Matter Protests With Ring Doorbell Data and Drones*, NEWSWEEK, August 9, 2020.

significant forms of social informational harm to go unaddressed and may foreclose socially valuable forms of data production.

A. Propertarian Data Governance Reform

1. Data governance reform as a response to inequality in the digital economy

In response to the harms of data extraction, scholars, activists, technologists and even presidential candidates have all advanced proposals for data governance reform. Many of these reforms are motivated by the connection between data extraction and wealth accumulation—and claim to redistribute wealth more broadly among data subjects and data processors.

Sir Tim Berners-Lee (inventor of the World Wide Web) began Solid from concern over how data extraction fuels the growing power imbalance online.[91] He notes that "for all the good we've achieved, the web has evolved into an engine of inequity and division; swayed by powerful forces who use it for their own agendas."[92] In response, Solid "aims to radically change the way Web applications work today, resulting in true data ownership as well as improved privacy."[93] Solid is a popular project within the blockchain community's #ownyourdata movement. Another is Radical Markets, a suite of proposals from Glen Weyl and Eric Posner that includes developing a labor market for data.[94] Weyl and others advocate for data as labor as a response to inequality: they aim to disrupt the digital economy's "technofeudalism," where the uncompensated fruit of data laborers is "distributed to a small number of wealthy savants rather than to the masses."[95]

---

[91] Solid aims to respond to the de facto enclosure of data via a system that ensures personal data control via local storage, mediated by a series of contractual agreements for access to the user's data. See Solid, available at: https://inrupt.com/solid.

[92] Solid, available at: https://inrupt.com/solid.

[93] Solid, available at: https://inrupt.com/solid.

[94] Jaron Lanier is one of the earliest to propose conceiving of data as labor. He similarly "worries about the distributional and social consequences of the failure to pay for data and online creative production." See JARON LANIER, WHO OWNS THE FUTURE (2013) (quoted in Posner and Weyl 2018, supra at 222). His proposal is taken up by Glen Weyl, Eric Posner and others as preferable over data as property or capital, because (unlike data as capital) it captures the role individuals have in generating value in the data economy. On this view, it is necessary to conceive of data as labor, not capital, to restore a functioning market for user contributions. ERIC POSNER AND GLEN WEYL, RADICAL MARKETS: UPROOTING CAPITALISM AND DEMOCRACY FOR A JUST SOCIETY (2018). Imanol Arrieta Ibarra, Leonard Goff, Diego Jiménez Hernández, Jaron Lanier and Glen Weyl, "Should we treat data as labor? Let's open up the discussion" Brookings (2018).

[95] Posner and Weyl, supra, at 209. See also Arrieta Ibarra et al, *Should We Treat Data as Labor? Moving Beyond 'Free' (*December 27, 2017). American Economic Association Papers & Proceedings, Vol. 1, No. 1, Forthcoming. Available at SSRN: https://ssrn.com/abstract=3093683 (Data as labor helps resolve the problems of "distributing the gains from the data economy unequally").

Progressive politicians, concerned over inequality in the information economy, have advanced similar proposals. Former Presidential candidate Andrew Yang included a right to data property in his campaign platform, and he recently launched the Data Dividend Project to push companies like Facebook and Google to pay users a "data dividend" for the wealth their data capital generates.[96] Representative Alexandria Ocasio-Cortez has also posited data ownership as a solution to inequality, tweeting "the reason many tech platforms have created billionaires is [because] they track you without your knowledge, amass your personal data& sell it without your express consent. You don't own your data, & you should."[97]

2. Propertarian reforms

The proposals above all advance a version of data governance reform that grants a propertarian entitlement to data. Propertarian reforms formalize the right to data as an individual's entitlement to their data-assets. Most reforms propose a property right over data about the subject, in which the data subject may then sell usage or full ownership rights. Alternatively, data production may be conceived of as a form of the subject's labor that entitles the data subject to command a wage in a data-labor market.

Propertarian data reform posits a particular legal solution to the problems of data extraction that transforms data *about* the subject into an asset that generates wealth *for* the subject.[98] On this view, data is already being "coded" as quasi-capital in law (through a combination of contractual agreements and trade secrecy law) in a manner that serves to create wealth for its holders but excludes the individuals from whom data originated.[99] The problem isn't the conceptualization of data as capital *per se*, but who has legal rights to benefit from this capital. As a legal matter, enacting propertarian reforms would code data with features considered more amendable to wealth creation for data subjects.[100] Data governance therefore becomes the governance (via contract

[96] The Data Dividend Project, available at datadividendproject.com. For Andrew Yang's proposal to grant a property right to data, see https://www.yang2020.com/policies/data-property-right/.

[97] Alexandria Ocasio-Cortez,https://twitter.com/AOC/status/1230352135335940096?s=20

[98] Glen Weyl and Eric Posner. Radical Markets: Uprooting Capitalism and Democracy for a Just Society. Princeton University Press, 2018, at 207.

[99] See JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM (2019) at 63; For a detailed treatment of how assets are coded in law to become capital, see KATHARINA PISTOR, THE CODE OF CAPITAL (2019) at 2-3 ("Fundamentally, capital is made form two ingredients: an asset, and the legal code…With the right legal coding, any of these assets can be turned into capital and thereby increase its propensity to create wealth for its holder(s).").

[100] As Pistor aptly describes, once data is conceived of as an asset of any kind in law, any conceptual distinction between capital (K) and labor (L) is reduced. In law, both render data as the subject of an exchange relation between data subject and data processor for data's alienable value. As a legal conceptual matter, L is easily turned into K with a bit of legal engineering.

law, property law, employment law, and labor law) of property relations or wage relations. This translates into a legal reform agenda to change the legal code being *applied* to data-assets, not to reject the concept of data *as* an asset.

a. The case for propertarian reforms

Moving from *de facto* to *de jure* property rights over data is meant to secure several benefits classically associated with propertarian reforms. First, they clarify rights of self-determination and control over data by allocating legal entitlements over data to data subjects. Call this the *data control claim*. This has the corollary effect of establishing at least some alienable claims to data. Second, propertarian reforms allow for bargaining between data subjects and collectors in a marketplace for personal data with the aim of achieving a Pareto efficient allocation of the benefits (and hence, Pareto efficient levels of production and consumption) of data extraction. Call this the *market efficiency claim*. Third, by compensating individuals for the value they help create, such entitlements are meant to spread the benefits of the digital economy more widely.[101] Call this the *redistribution claim*.

Together, these claims make propertarian reform an intuitively appealing response to the quasi-enclosure and de facto ownership of data resources by technology companies. By formalizing the informal propertarian status of data, such reforms directly counteract the quasi-propertarian claims to personal data flows of large data collectors like Google, Facebook, and Amazon, and directly invalidate the current practice of capturing data-value from subjects without compensation.[102]

Take for example partners in an LLP. They contribute their labor to the corporate entity as in-kind services and take out dividends as a shareholder in lieu of a salary, thus benefitting from the better legal protections and a lower tax rate afforded K for the same exact work that would be preformed where it coded as L instead. Beyond law, the concept of "human capital" also serves to collapse this distinction. KATHARINA PISTOR, THE CODE OF CAPITAL (2019).

[101] Propertarian reforms have long been motivated by claims that they can extend material benefits to those who are currently excluded from enjoying them. Development economist Hernando de Soto was a prominent proponent of granting the poor in developing countries property rights as a way to achieve economic security. Such rights, he argues, can turn "dead land" into "life capital," granting owners the opportunity to mortgage land or other assets to invest in new ventures and begin to accrue wealth. This general theory of widespread and shared wealth creation via property rights experienced a "surge" in the 1980s, when the idea of "clear property rights and credible contract enforcement," to create "conditions by which everyone would prosper" was widely adopted by development economists and politicians throughout the world. HERNANDO DE SOTO, THE MYSTERY OF CAPITAL: WHY CAPITALISM TRIUMPHS IN THE WEST AND FAILS EVERYWHERE ELSE (2003), at 46. For discussion of de Soto and the popularity of this reform in development economics, see Katharina Pistor, 2019, supra at 14 and 2.

[102] Large platform companies assert quasi-propertarian claims to data flows vi their "de facto appropriation and enclosure" of personal data flows. JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM (2019) at 25.

Propertarian reforms also dovetail nicely with certain diagnoses of—and responses to—problems of competition and fairness in the data political economy.[103] Such views identify either too much corporate control over data assets, or too much *concentration* in the corporate control of data assets, as key barriers to competition.[104] One popular response to such problems is "data portability," a set of technical interoperability requirements and legal rights that allow users to transfer their data.[105] Under this view, empowering users to "shop" for new digital services will encourage market discipline (due to enhanced user exit options) and give new market entrants the opportunity to attract users and their valuable user data.[106] Data portability combines elements of the data control claim and the market efficiency claim to enhance competitive opportunity via individuals' market actions.

Finally, propertarian reforms respond to an important claim of injustice levied against the digital economy: that individuals play a role in generating a materially valuable resource from which they see no value (and which at times further places them at risk). As detailed above, calls for entitlement reform are often made in response to frustration over the wealth amassed by companies that harvest data for which they pay nothing. At a time when technology companies are widely accused of wielding too much economic and political power over the lives of others, the redistributive claim may contribute to their enduring and widespread appeal.[107] Even for those who may view the redistributive claim as purely an instrumental effect of achieving data control and market efficiency, it serves a justificatory role in advocating for propertarian entitlements.

b. The critiques levied against propertarian data governance reform

There are several reasons to be skeptical of propertarian data reforms. One is impracticability. Operationalizing the kind of complex and comprehensive micro-payments system at the scale required may simply not be feasible or cost-effective.[108] Moreover, it is empirically unclear whether propertarian solutions

---

[103] Lina Khan, *Amazon's Antitrust Paradox,* 126 Yale L.J. (2017). The FTC is holding hearings and workshops on the concept. See "FTC announces September 22 Workshop on Data Portability" FTC.gov. Senators Mark Warner and Josh Hawley introduced the *Augmenting Compatibility and Competition by Enabling Service Switching (ACCESS) Act*, a bill to encourage market competition among social media platforms that includes data portability. See https://www.hawley.senate.gov/sites/default/files/2019-10/ACCESS-Act-Bill-Text.pdf

[104] MAURICE STUCKE, AND ALLEN GRUNES, BIG DATA AND COMPETITION POLICY (2016).

[105] Gabriel Nicholas, *Taking It With You: Platform Barriers to Entry and the Limits of Data Portability*, Mich. Telecom. and Tech. L. R., (forthcoming).

[106] Ibid. Portability can be seen as an "exit" enhancing market response. ALBERT O. HIRSCHMAN, EXIT, VOICE, AND LOYALTY: RESPONSES TO DECLINE IN FIRMS, ORGANIZATIONS, AND STATES (1970).

[107] See Part II supra.

[108] Critics should be wary of overreliance on convenient arguments of implementation. While such a payments system may appear impracticable from a consumer perspective, as a

will materially address data extraction given current conditions of datafication.[109]

Second, propertarian data reforms may be unlikely to address the privacy erosion that motivates many concerns over data extraction.[110] Payment provides additional incentive for people to share data about themselves and thus may further degrade privacy, not only for themselves but also for others. A data subject may decide that the risk of their privacy loss is worth the payment provided and thus sell their data mutually beneficial exchange. Putting aside the effects this sale has on others, privacy risk is notoriously easy to under-value at the point of exchange.[111] Privacy risk associated with data isn't static, nor is

---

technical matter such a system may not be all that different from the highly complex algorithmic auction platform systems and exchanges through which advertisers purchase views, impressions and clicks from consumers, and which pose similar challenges of managing billions of instantaneous pricing and transaction actions at scale. See e.g. Salome Viljoen, Jake Goldenfein, Lee McGuigan, "Economic Method, Digital Platform: When Mechanism Design Moves Online," manuscript on file with author.

[109] See Zoë Hitzig et al, *The Technological Politics of Mechanism Design,* U. Chicago L. R. Online, 2019, for a more detailed discussion. In short, the mere granting of a labor or property right does not guarantee that the conditions underlying the sale of that labor/property will be non-extractive and un-coerced. And the conditions of the current data market do not inspire confidence. Large data collectors are highly concentrated and are able to leverage their existing superior knowledge to design exchanges and prices to their advantage. In contrast, data subjects are widely dispersed and isolated from one another, and have little insight into how data value is created from which to bargain. Personal data from any one data subject is essentially value-less, reducing the capacity for individual data subjects to meaningfully exert bargaining power. Moreover, data subjects do not (yet) identify as a common social group from which to build political bargaining power. Finally, datafication does not result in the kinds of visceral oppression that may motivate moral outrage and build counter-power—in contrast with oppressive workplace domination or highly impoverished conditions of production, data extraction is designed to occur as seamlessly and painlessly as possible, transmitting flows of data in parallel with data subjects living their online and offline lives. On the challenges of the US labor market in general, see Matthew Desmond, *Americans Want to Believe Jobs Are the Solution to Poverty: They're Not,* NY TIMES, Sept 11, 2018, archived at http://perma.cc/V64B-R36B. On the essential valuelessness of any one individual's data see Kenneth Bamberger et al, *Can you pay for privacy? Consumer Expectations and the Behavior of Free and Paid Apps,* 35 Berkeley Tech. L. J. 327 (2020).

[110] See e.g., Pamela Samuelson, *Privacy as Intellectual Property*, 52 Stan. L. Rev. 1125 (2000); *see also,* Jessica Litman, *Information Privacy/Information Property*, 52 Stan. L. Rev. 1283 (2000) (offering a critique of property approaches to privacy); Julie E. Cohen, *Examined Lives: Information Privacy and the Subject as Object,* 52 Stan. L. Rev. 1373 (2000); Jane Bambauer, *The Perils of Privacy as Property: The Likely Impact of the GDPR and the CCPA on Innovation and Consumer Welfare: GDPR & CCPA: Opt-ins, Consumer Control, and the Impact on Competition and Innovation*, 116th Cong. ___ (2019) (written testimony submitted to the Senate Committee on the Judiciary, Tuesday, Mar. 12, 2019).

[111] Daniel Solove, *The Myth of the Privacy Paradox*, 89 G.W. .L. Rev 1 (2021); Alessandro Acquisti, J Grossklags, *Privacy and rationality in individual decision making* IEEE Security & Privacy, vol 2, p. 24 – 30 (2005); Alessandro Acquisti et al, *The Economics of Privacy*, Journal of Economic Literature, *54* (*2*), *442--492*, (2016); Alessandro Acquisti, et al, *What is Privacy Worth?,* The Journal of Legal Studies, *42*(*2*), *249--274*, (2013).

privacy loss linear—it accumulates and grows over time based on the composition effects from multiple sources of data, varied downstream uses, and new applications.[112] People tagging online photos of themselves and their friends in 2009, for example, could not have known that companies contracting with law enforcement in 2019 would use such information for facial recognition products.[113]

Finally propertarian reforms place greater marginal pressure to sell data on those least able to forego the income it offers—transforming privacy into an even greater privilege than it is today.[114]

Whether in the name of privacy or no, propertarian reforms concede existing processes of data commodification in the digital economy: this ship having sailed, what data subjects can and should secure is their fair share of the value such processes produce.

B. Dignitarian Alternatives

1. Data governance reform as a response to commodification and legibility

Refusal to concede data commodification lies at the heart of dignitarian critiques of both the status quo and propertarian alternatives. Where propertarian reforms conceive of data as the subject of individual ownership (data as object-like), dignitarian data governance conceives of data as an expression (or extension) of individual selfhood (data as person-like).[115]

Some of the most vivid normative critiques of informational capitalism and privacy erosion invoke dignitarian arguments against datafication. For instance, a highly criticized aspect of information capitalism is that it rewards economic

---

[112] Fluitt, A. et al, *Big Data's Composition Problem*, European Data Protection Law Review , 2019

[113] Kim Lyons, "ICE just signed a contract with facial recognition company Clearview AI," the Verge, August 14, 2020

[114] Michelle Gilman & Rebecca Green, *The Surveillance Gap: The Harms of Extreme Privacy and Data Marginalization*, 42 NYU Rev. L. & Soc. Change (2018). Proponents of propertarian reforms (to the extent they advocate on the basis of privacy at all) adopt the contested position of privacy as control. Under this view, clarifying data subjects' legal rights over data grants them more control over such data, and is by extension more privacy protective. See e.g. Inrupt.com ("Users control which entities and apps can access their data"). For accounts of privacy that contest theories of privacy as control, see HELEN NISSENBAUM, PRIVACY IN CONTEXT (2010); Mark Verstraete, I*nseparable Uses,* North Carolina L. Rev., Vol. 99, 2021, provides an interesting account of privacy as control via a theory of separability that severs the claim of control from a basis in alienability.

[115] The European Union's data governance regime derives its theory of privacy and data protection from Kantian dignitary conceptions of data as an expression of the self, and thus subject to deontological requirements of human dignity. This normative and conceptual account anchors the robust European regime, including its suite inalienable rights over personal data. See Article 88 of the General Data Protection Regulation; Luciano Floridi, *On Human Dignity as a Foundation for the Right to Privacy,* Philos. Technol. 29 307–312 (2016).

imperatives to apprehend (and act on) individuals in machine-readable form, often in ways that occur without meaningful consent and for purposes that may violate the wishes of data subjects. Datafication, and the seamless and continual data extraction it relies on, reconstitute individuals into "data doubles," representing them in algorithmically legible forms.[116] In doing so, datafication renders individuals as patterns of behavior, identified as amalgams of categories or classifications (e.g. "Woman," "Millennial" "Lawyer"). This violates basic notions of individuals as autonomous beings.

A closely related subject of critique is the affordances of datafication for algorithmic governmentality: the cycle of rendering individuals as patterns of behavior based on certain categories and features, and then algorithmically and iteratively acting back on individuals on the basis of these classifications in a state of constant feedback and fine-tuning. This cycle re-inscribes algorithmic ways of understanding the subject back onto the subject herself, undermining her capacity for self-formation and the enactment of self-will.[117]

In response to such concerns, dignitarians like Zuboff argue that datafication and data extraction represent the end of the relationship we enjoy with our innermost selves.[118] The "dark continent" of inner life is invaded and transformed into a "collectivist vision that claims the totality of society."[119] Zuboff diagnoses the injustice of informational capitalism as its endeavor to commodify, colonize, and rule this inner self for profit, via monetization schemes that rely on behavioral prediction and control. This new capitalist imperative violates human dignity and destroys personal agency.

Zuboff's repeated invocation of apprehension as violation and behavioral modification as colonization suggests her concern is with datafication itself, not

---

[116] JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM (2019) at 67. This Article's use of the terms "legible" and "legibility" is informed particularly by James Scott's SEEING LIKE A STATE (Yale 1998) and Michel Foucault's DISCIPLINE AND PUNISH: THE BIRTH OF THE PRISON (Random House 1995).

[117] For an excellent treatment of the subject of how data science theorizes its subject as patterns of behavior, and the disconnect this produces from the subject theorized by law, see Sebastian Benthall and Jake Goldenfein, "Data Science and the Decline of Liberal Law and Ethics", Ethics of Data Science Conference (2020. See also Marion Fourcade and Kieran Healy, "Seeing Like a Market," Socio-Economic Review, Vol 15:1 (2017) at 10. See also Dan Burk, Algorithmic Legal Metrics, NOTRE DAME L. REV., forthcoming; Julie E. Cohen, *What privacy is For,* 126 Harv. L. Rev. 1904 (2013), 1905 ("[Privacy] protects the situated practices of boundary management through which the capacity for self-determination develops").

[118] SHOSHANA ZUBOFF, AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER (2019). See also, Shoshanna Zuboff, *Big other: surveillance capitalism and the prospects of an information civilization,* Journal of Information Technology 30.1 (2015): 75-89. For an excellent review of Zuboff's enlightenment ideals and their limitations, see Quinn Slobodian, *False Promises of Enlightenment,* BOSTON REVIEW, May 29, 2019.

[119] SHOSHANA ZUBOFF, AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER (2019).

merely the ends to which it is put or the relations under which it occurs. This posits datafication as *legibility harm*—inflicting on individuals a depth of representation that violates the dignity of their personhood.

This diagnosis is consistent with dignitarian accounts that identify datafication—the commodification and alienation of the inner self—as a central injustice of informational capitalism. On this view, rendering a person legible via datafication represents a form of personal violation. Datafication, data extraction, and algorithmic governmentality are wrong because these processes manipulate people, invade and violate the sanctity of their inner being, and undermine their capacity to express and enact their free will.[120] To dignitarians, these injustices present ontological and existential threats to personhood, and are therefore wrong on their own basis, regardless of how the resulting data may be used.

2. Dignitarian reforms

In response, dignitarians aim to invigorate legal protections of individual autonomy (or, as is common in the European context, individual dignity) in the digital economy. The strongest such accounts advance legal rights over personal data as akin to natural rights, and thus advocate for fundamental rights to data as an extension of the data subject's moral right to dignity and self-determination. Such rights are contained within the European Union's data governance regime, which (alongside other affirmative data processing obligations) affords universal and inalienable rights over personal information and enshrines data protection and privacy as fundamental rights.[121] Advocates of this approach in the EU and beyond argue for extending the human rights framework to data governance as a way to strengthen fundamental data protection in law.[122] Fundamental rights provide individuals with inalienable

[120] Several other critiques of the digital economy similarly focus on how existing processes of data production undermine individual autonomy. See e.g., BRETT FRISCHMANN AND EVAN SELINGER, RE-ENGINEERING HUMANITY (2018); WOODROW HARTZOG, PRIVACY'S BLUEPRINT (2018); Daniel Susser, Beate Roessler, and Helen Nissenbaum, *Online Manipulation: Hidden Influences in a Digital World,* 4 Geo. L. Tech. Rev 1 (2019).

[121] European Union Charter, Articles 7&8, available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:12012P/TXT&from=EN. The EU's General Data Protection Regulation (GDPR) includes more than just dignitarian data reform. Alongside its suite of individual rights, the GDPR includes a number of affirmative data processing obligations that apply to data processors regardless of individual consumer choices, and affirmatively requires a lawful basis for any data processing to occur (individual consent is one of six). While there is considerable debate regarding how broad the scope of the GDPR's lawful bases are, and how they interact with individual consent and the individual right to restrict processing and the right to erasure, at a minimum they provide a legal framework for data protection beyond individual ordering. For a helpful explainer, see the UK's Information Commission Office, "Guide to the General Data Protection Regulation" available at https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/

[122] See e.g. Dazza Greenwood and Elizabeth M. Renieris, *Do we really want to "sell" ourselves? The risks of a property law paradigm for personal data ownership*, Sept 23, 2018,

rights of control over their information, including more stringent consent requirements and ongoing rights of access to data for the data subject.[123] Granting human rights standing to data subjects would ensure a "minimum standard that cannot be waived by consent, even if all potential uses of data could be foreseen."[124]

Dignitarian reforms posit a robust legal solution to the problems of data extraction: they enhance the protection of data *about* the subject by making these protections more akin to those afforded the subject *herself*. Dignitarian reforms therefore aim to encode data with features more like those afforded a natural person. On the basis of this quasi-personhood, these reforms would extend inalienable rights and impose on others certain duties that ensure personal data is granted a legal baseline of civil and political status. This would formally abolish the quasi-ownership claims to data and instead recognize data's quasi-personhood status, subject to the range of civil libertarian protections afforded individuals in public life.

Many dignitarian reformers claim that data extraction involves not only individual stakes, but also societal ones. Zuboff says the world's digital information is a public good; the EU Data Protection Supervisor notes that privacy is not "only an individual right but a social value." [125]Yet in practice, the legal solutions advanced under dignitarian conceptions of data governance still subject data to individual ordering and protect data subjects from individualist

---

Medium, available at https://medium.com/@hackylawyER/do-we-really-want-to-sell-ourselves-the-risks-of-a-property-law-paradigm-for-data-ownership-b217e42edffa

[123] The GDPR grants individual the following rights over data: the right to be informed, the right of access, the right to rectification, the right to erasure, the right to restrict processing, the right to data portability, the right to object, and rights in relation to automated decision making and profiling. All of these rights are subject to some overriding exceptions and are undergoing active interpretation in EU law. General Data Protection Regulation, Regulation (EU) 2016/679.

[124] Human Rights Watch, "The EU General Data Protection Regulation," June 6, 2018, available at: https://www.hrw.org/news/2018/06/06/eu-general-data-protection-regulation#. Other forms of data governance adopt a less fundamental approach to enacting a minimum standard, seeking instead to heighten the duties owed to data subjects by data collectors in virtue of data's capacity to enduringly and significantly affect the data subject. For instance, several promising reforms aim to invigorate theories of fiduciary obligation, or extend helpful theories of separability from property theory, to individual data governance. See e.g. Jack M. Balkin, *Information Fiduciaries and the First Amendment*, 49 UC Davis L. Rev 1185 (2016); Woodrow Hartzog & Neil Richards, *Privacy's Trust Gap*, 126 Yale L.J. 1180 (2017); Neil Richards & Woodrow Hartzog, Taking *Trust Seriously in Privacy Law,* 19 Stan. Tech. L. Rev 431 (2016); Mark Verstraete*, Inseparable Uses*, North Carolina L. Rev, Vol. 99, 2021. Such reforms, while they share certain relevant features with dignitarian approaches, are not as directly link to the dignitarian normative basis for reform and thus warrant separate analysis beyond the scope of this Article.

[125] Alvin Powell, *An awakening over data privacy,* THE HARVARD GAZETTE, Feb 27, 2020, available at: https://news.harvard.edu/gazette/story/2020/02/surveillance-capitalism-author-sees-data-privacy-awakening/; European Data Protection Supervisor, "Data Protection," available at: https://edps.europa.eu/data-protection/data-protection_en#DP_Law

informational harm.[126] Dignitarian reforms secure negative rights for data subjects *against* certain downstream uses (e.g., use without consent, use that goes beyond the purposes originally given, or use once consent has been withdrawn), and that obtain with respect to data collected about *them*.[127] These negative freedoms secure personal data's quasi-personhood status in law governed by civil, not contractual, rights.

C. Conceptual Limitations of DIM reforms

While dignitarian reforms offer a more robust individualist regime for data protection than propertarian reforms, like propertarian reforms they still conceive of data as an individual medium (DIM). As a result, both propertarian and dignitarian reforms attempt to reduce legal interests in information to individualist claims subject to individualist remedies that are structurally incapable of representing the population-level interests that arise due to data horizontal relations. This fails to fully account for significant forms of social informational harm, and risks foreclosing socially beneficial forms of data production

1. Absence of horizontal relations

Failing to account for these horizonal relations presents a problem for DIM reforms even on their own terms. Using shared population features derived from data about Adam to act upon Ben is what makes such data collection so desirable. This relationality is part of why data collectors face such strong incentives to extract continual data streams from data subjects like Adam to begin with. Horizontal relations, whether explicitly accounted for or not, motivate data collectors to engage in such continual and fine-grained data extraction. Ignoring the interests that result from horizontal relations therefore not only sidelines Ben's interests in such data, but also fails to account for structural conditions that influence the terms of exchange between Adam and TattooView, and that in turn index many of the interests that Adam also has in the information collected from him.

---

[126] Mark Scott et al, *How Silicon Valley Gamed Europe's privacy rules,* POLITICO, May 22, 2019, available at: https://www.politico.eu/article/europe-data-protection-gdpr-general-data-protection-regulation-facebook-google/ (noting that despite being previously banned, Facebook's facial recognition technology is once again permitted in Europe because users are given the choice to opt in to the service under the consent rules of the GDPR).

[127] Under Europe's General Data Protection Regulation, these dignitarian rights are accompanied by a series of affirmative obligations imposed on data processors regarding the storage, transmission, and processing of data. Whether these affirmative obligations are sufficient to accord European data subjects more than individualist protections is a subject of active scholarly debate. See e.g., Jef Ausloos, René Mahieu R. & Michael Veale, *Getting Data Subject Rights Right,* Journal of Intellectual Property, Information Technology and Electronic Commerce Law (2019) 10: JIPTEC 283. Margot Kaminski, *The Right to Explanation, Explained*, 34 Berkeley Tech. L. J. (2019).

Consider again the prior scenario involving Adam, TattooView, and Ben. Propertarian reforms would require that TattooView pay Adam for his data. Yet payment at the point of collection for Adam does nothing to address how his data is used to detain Ben. Ben incurs significant harm but receives none of the benefit from propertarian data reforms. In granting Adam right to payment, propertarian reforms seek to rebalance the terms of Adam's vertical relation with TattooView. Such reforms may ameliorate the worst excesses of data subject exploitation (and result in some degree of redistribution) but in failing to apprehend both Ben and Adam's legal interests that accrue along horizontal relations, they do not grant Adam (or Ben) the ability to address the conditions structuring the terms of this exchange. Given all the practical realities discussed above, such reforms are highly unlikely to produce more equal data relations along either axis. [128]

Dignitarian reforms would admirably extend protection to downstream uses that violate Adam's protected interests in data collected from him. Extending fundamental protections to Adam grant him standing to argue that use of his data to detain him violates his fundamental rights, or alternatively may grant him stronger up-front rights to refuse collection.[129] Yet similar to propertarian reforms, dignitarian rights leave third parties like Ben unaccounted for. Granting Adam rights against having data collected from him used against him does not affirmatively prevent against Adam's data—or the category of tattoo image data generally—being used against *others* like Ben for purposes of detention. And yet presumably the interest Ben and Adam have in this information is the same (i.e. an interest against having tattoo image data about their tattoo being used to 1) classify them as a suspected gang member and 2) detain them on the basis of this classification); to grant legal protection to one while excluding the other is arbitrary and nonsensical. In both instances, the relevant set of legal interests in this data flow do not reduce to the individual rights granted to Adam by DIM reforms.

Like propertarian forms, dignitarian reforms fail to apprehend the structural conditions driving the behavior they aim to address. In granting Adam inalienable rights over the terms of his data collection and use, dignitarian reforms seek to rebalance the terms of Adam' vertical relation with TattooView. Dignitarian reforms may ameliorate some forms of data subject violation. But in failing to index the many horizontal interests at stake, they fail to account for the role horizontal relations play in the economic imperatives of data extraction, as well as the forms of social informational harm such relations may materialize. The observation that data production may violate individual autonomy does nothing to further our understanding of *why* or *how* this violation has become an

[128] For a more detailed treatment of these conditions, see Salome Viljoen, *Data as Property?,* Phenomenal World, October 16, 2020.

[129] The merits of such a case are unclear, and beyond the scope of this analysis.

imperative of competitive market behavior in the data political economy.[130] Acting on this observation with attempts to strengthen rights of individual data subject control is thus unlikely to address the structural conditions driving this state of affairs.

2. Missing or misdiagnosed theories of harm

The absence of horizontal data relations in law may cause data governance law to miss—or misconceive—how data production results in particular kinds of injustice. As detailed above, datafication gives rise to two classes of critique or claims of injustice: the inequality diagnosis and the commodification diagnosis. The *inequality diagnosis* locates the injustice of data production in the unfair distribution of wealth that datafication creates. It conceives of the injustice of datafication as one of unjust enrichment. The *commodification diagnosis* locates the injustice of datafication in the excessive legibility of data subjects that results. This diagnosis conceives of the injustice of datafication as the wrongful control this excessive legibility grants data collectors over data subjects. This control in turn undermines data subject autonomy and violates their dignity by reducing their inner lives to transactions mined for value.

These two articulations or diagnoses of what makes datafication wrongful in turn motivate the two DIM agendas for reform. Propertarian reforms aim to respond to the inequality diagnosis by granting data subjects a right to reclaim some portion of the material benefits created from data production. Dignitarian reforms aim to respond to the commodification diagnosis by re-asserting greater control for data subjects over if, when, and how they may be rendered legible by data collectors.

Yet each reform fails to respond to the diagnosed injustice of the other. Propertarian reforms by design concede extensive data subject legibility as a necessary condition of securing some redistributive benefit. Dignitarian reforms by their own commitments cannot provide data subjects material redistributive value, as this would violate dignitarian prescriptions against commodifying knowledge of the inner self. Even if one assumes each reform can address its diagnosed form of injustice (and as the previous subsection notes, there are significant reasons not to make such an assumption), choosing one leaves the other diagnosis of injustice unaddressed. If one believes both capture compelling concerns regarding data production, then pursuing the either/or path of DIM reforms presents a dilemma.

3. Unjust data production as unequal data relations

Each diagnosis and related agenda for reform present both a normative issue (i.e. not addressing a valid aspect of what makes datafication wrongful) as well

[130] For example, the GDPR does not outlaw the advertising-driven business model that predominantly drives datafication; it requires companies to be more transparent about this use and gives users greater access to how their data is being used.

as an operational issue (i.e. missing relevant features in its attempt to address its own diagnosis of what makes datafication wrongful) that leave each unlikely to materially address the problems motivating reform.

Reconceptualizing these diagnoses of injustice to account for data relations may resolve these issues in helpful and clarifying ways. What makes datafication wrongful is not *either* that it represents unjust enrichment *or* that it is an instance of wrongful self-commodification. Datafication (or more precisely, data production) is wrongful if and when it materializes unjust social relations along either the vertical or horizontal axis. These unjust social relations may take the form of exploitative data relations that generate unfair wealth distributions, as well as data relations that materialize forms of group oppression like racism, xenophobia, and sexism. By centering data relations in our diagnosis of injustice, we can recast the reform agenda of data governance law as managing (ideally equalizing) these data relations.

This alternative normative diagnosis also helps pinpoint what DIM legal agendas miss. Data production's role in enacting or amplifying inequality is not simply a matter of data subject nonpayment, but concerns the *unjust social relations* being amplified or enacted on the basis of shared population features. Payment at the point of collection may redistribute some portion of the profit that results from exploitative data collection but does nothing to address how data production itself may amplify or enact social oppression as a means to generate that profit.[131] Even if payment were to distribute the gains from data production in a completely egalitarian manner, datafication as a process materializing unequal and oppressive social relations would remain.

The focus on social inequality (as opposed to unjust enrichment) also captures relevant aspects of dignitarian concerns regarding algorithmic governmentality. Governmentality via data-driven feedback systems is wrong not only because it undermines processes of self-formation (though it may well have this effect), but also because such systems enact unjust social relations that serve to dominate, marginalize, and demean.[132] Recasting the injustice of

[131] In fact, by legitimating the marketplace for data, payment may serve to legitimate downstream practices that result from lawful engagement in that marketplace. Because data is commoditized to begin with, ICE was able to purchase access to this database from its provider, Venntel, as opposed to gathering this data itself. This commercial exchange provides ICE strong legal protection for using this data. Under *Carpenter v. United States* 585 U.S. __ (2018) (No 16-402), ICE may have needed a warrant to obtain this data from carriers or app companies directly. Yet because ICE simply purchased access to the database from a data broker, as could any other entity, any potential Constitutional challenge is weakened.

[132]For example, consider the growing literature on how algorithmic forms of self-knowing enact cultural imperialism. See, e.g. NICK COULDRY AND ULISES MEJIAS, THE COSTS OF CONNECTION: HOW DATA IS COLONIZING HUMAN LIFE AND APPROPRIATING IT FOR CAPITALISM (2019); Dan M. Kotliar, "Data orientalism: on the algorithmic construction of the non-Western other," Theory and Society (2020). Cultural imperialism refers to the universalization of a dominant group's experience or culture and its establishment as the norm. This grants the

surveillance data flows as that of unequal social relations brings into view the structural forces driving personal instances of violation as well as the mutual stakes we have in the injustice of such conditions.

Recasting data governance reform as equalizing data relations also helpfully clarifies a distinction glossed over in dignitarian accounts between "commodification" and "legibility" regarding what makes legibility wrong: namely, the goals motivating apprehension, and the substantive and procedural conditions that determine those goals. This distinction vanishes in critiques against private companies like Facebook (which are currently the subject of the fiercest dignitarian critiques), but is relevant for distinguishing the data collection and use of private companies from those of publicly (or otherwise collectively) accountable data infrastructures.

The relevant inquiry is not whether and to what degree a data subject has been rendered legible (and whether they had the opportunity to exert control over this process), but to what *ends* and under what *conditions* legibility occurs—and most importantly whether these have been determined in ways that enact more equal data relations. Under this account, permissible legibility is not simply a matter of individual data subject consent or control, but one of the *institutional forms* that adjudicate between and determine the legitimate and illegitimate bases for data production.

Consider again the example of TattooView collecting user data to detain suspected gang members. What makes the tattoo data flow potentially unjust is not that the population at the point of data collection wasn't paid, but that information about one group (the data subjects) is being used to oppress and dominate others on the basis of their ascribed group membership (i.e. "gang member", a group membership informed by racial, ethnic, class and linguistic difference). This tattoo data flow is not (only) unjust because its collection or its use renders Adam legible in ways that may violate Adam's autonomy and his right to self-determination. It also materializes a social category (i.e. "gang member") that, when acted upon, results in the domination and oppression of others. Under propertarian and dignitarian reforms this *social effect* continues to have no bearing on how information law regulates what data may be collected, stored, exchanged, or used.

---

dominant group primary access to what Nancy Fraser calls the "means of interpretation and communication in a society." Nancy Fraser, "Social Movements vs. Disciplinary Bureaucracies: The Discourse of Social Needs." CHS Occassional Paper No. 8, Center for Humanistic Studies, University of Minnesota, 1987. Often without realizing it, dominant groups project their experiences as the experiences of humanity; the result is cultural products of communication and sense-making that reflect dominant experience, values, goals, and achievements. This creates for the culturally oppressed the experience W.E.B. Du Bois called "double consciousness," the sense of "always looking at one's self through the eyes of others, of measuring one's soul by the tape of a world that looks on in amused contempt and pity." W.E.B. DU BOIS, THE SOULS OF BLACK FOLK, (1969 ed. 1903), at 45.

4. DIM reforms and socially beneficial data production

Reducing interests in the digital economy to individual data subject interests may inadvertently foreclose socially beneficial forms of data production. Currently a predominant purpose that draws critiques of datafication is that of private wealth creation.[133] Wealth creation is one purpose for collecting data, but there are others: for example, running social welfare enterprises that require at scale distribution and management of precious resources like water, or those that require time-sensitive predictions for overriding public interests, such as public health strategies to limit the spread of COVID-19. These urgent public tasks require high quality population data to ensure public welfare obligations are met effectively and fairly.

Yet the diagnoses of harm under DIM reforms do not index these distinctions, and neither do the the legal agendas that result from them. Under these accounts, purposes of datafication for the public interest and those for private wealth creation pose the same risk of individual violation and are subject to the same forms of individualized governance. Under propertarian regimes, if a public agency cannot pay data subjects a fair price for this data, it may be well be argued such datafication constitutes a public taking or should be subject to individual decisions to donate such data or not. Under dignitarian regimes, an individual may disagree with the public purpose (e.g., they believe the government efforts to trace COVID-19 violate their medical liberty) and deny access to their data on the basis that this use violates their individual will and their fundamental rights. In both instances—taking data for free or collecting it absent consent and for a purpose the data subject disagrees with—violate individualist conceptions of how information's collection and use should be ordered, and what conditions of datafication are legitimate. Yet deferring to voluntary adoption into these systems may significantly undercut their capacity to realize the broader social benefits they are meant to achieve.[134]

---

[133] Another is law enforcement and government surveillance. In the context of the US however, most high-profile scandals regarding law enforcement use of technology involved private entities selling surveillance products to law enforcement. This again, is one particular business model under the organizing principle of datafication for the purpose of private wealth creation, which fuels the ubiquity of personal data-based surveillance products available for sale on the private market. Kim Lyons, *ICE just signed a contract with facial recognition company Clearview AI,* THE VERGE, August 14, 2020; Dana Goodyear, *Can the manufacturer of tasers provide the answer to police abuse?,* NEW YORKER, August 20, 2018.

[134] For instance, public health authorities deploying Covid-19 digital contact tracing apps targeted a 60% population threshold for the systems to work most effectively to counteract the pandemic. Although lower numbers of app users are still estimated to reduce the number of coronavirus cases, getting closer to the 60% threshold significantly increases the efficacy of digital tracing systems. "Digital contract tracing can slow or even stop coronavirus transmission and ease us out of lockdown," Oxford Coronavirus Research, April 16, 2020 (available at https://www.research.ox.ac.uk/Article/2020-04-16-digital-contact-tracing-can-slow-or-even-stop-coronavirus-transmission-and-ease-us-out-of-lockdown). Patrick Howell O'Neill, *No,*

DIM reforms thus suffer from being simultaneously overly narrow and overly broad. By focusing on datafication's violation of self or uncompensated value creation, they do not address the economic imperatives that drive such harm nor do they provide an effective agenda for addressing inequality in the data political economy. At the same time, the focus on datafication paints over meaningful distinctions regarding the purposes of data production and the conditions under which such purposes are determined.

IV. DATA AS A DEMOCRATIC MEDIUM

A. The Legitimacy Problem

Both current and proposed individual-level rights in data cannot address the population-level interests that arise from data production. As a result, these reforms are unable to resolve the legitimacy problem that, alongside the sociality problem, continues to vex U.S. data governance law. The legitimacy problem asks: how can data governance law distinguish legitimate from illegitimate data use without relying on individual adjudication?

1. Watercorp and Waterorg

Consider the following example. Suppose that an entity (Watercorp) is collecting data on household water consumption. Every instance when one drinks water from the tap, sets the kettle to boil, waters one's herb garden, or brushes one's teeth is collected, processed, analyzed, with the goal of changing future water usage behavior for the households in a given municipality. This data reveals intimate facts about people's lives—from this data emerges a detailed portrait of their daily habits. The resulting data may be used for any number of reasons: to help households set and meet water reduction goals, to calculate "surge" prices for water usage based on peak consumption, to use feedback data to shift people's water consumption patterns towards the bottled drinks of a client, to can sell insights about people's daily habits to advertisers, insurers, creditors, and employers, or to test what strategies make people most likely to pay their utility bills. [135]

Now suppose instead of Watercorp, Waterorg—the municipal public authority for the drought-prone area—engages in this data collection to understand the municipality's water usage, develop strategies to reduce water consumption, and (as droughts grow more severe) develop plans to ensure water will be distributed fairly and responsibly as it becomes scarcer. Suppose further that the risks of drought-based water shortages and shutoffs are highest in the

---

coronavirus apps don't need 60% adoption to be effective, MIT TECH. REV., June 5, 2020.

[135] A controversial recent randomized-controlled trial ran an experiment to see whether shutting off tenants' water makes landlords more likely to pay their water bills; see https://twitter.com/joshbudlender/status/1292170843389386761 (the paper has been temporarily withdrawn, but screenshots and discussion available on the twitter thread link provided).

driest and poorest districts of the municipality, where a higher proportion of minority residents live.[136] Finally, suppose that a handful of citizens of the municipality object to the coercive power of the state in collecting this data from them, or to using this data to inform water allocation strategies that affect them. They argue that this data collection violates their dignity and autonomy, extracting their intimate water consumption data against their own interests, since it is collected in furtherance of future water policies that will almost certainly reduce their capacity to freely access and use water as they choose, free from observation. Like Watercorp, Waterorg is exerting coercive power to render such citizens legible to Waterorg against their will, without payment, and for purposes that go against their interests.

If one is of the view that Watercorp's data production may be potentially concerning, but that it is permissible (even responsible) for Waterorg to engage in this data production for its stated purposes, analysis under DIM accounts of data governance presents a challenge. Watercorp's basic governance structure allows for broader (democratic) representation in the determination of societal goals, accompanied with constitutional constraints against certain forms of individualized use. For Watercorp, we have means for neither democratic decisions about the collective societal goals, nor the constitutional forms of oversight that serve as substantive backstops against impermissible collection and use. Focusing on the preferences or rights of each individual or household regarding whether to participate in this collection does not apprehend the normatively distinct purposes and conditions of data production from these two entities. Further, this approach fails to accord relevance to the mutual and overlapping interests these households have in one another's choices.

B. Horizontal Relations and Institutional Design

To address the relevant distinctions between Watercorp and Waterorg's data production and to adjudicate their legitimacy requires recourse to population-level, democratic evaluation of these proposed data production schemes.

---

[136] Though at a smaller scale, this distribution is not unrealistic. Evidence suggests that water-stress as a result of a changing climate will disproportionately impact poorer communities. The World Health Organization estimates by that 2025, half of the world's population will be living in water-stressed areas, and one-quarter of the world population face "extremely high" levels of water stress. North Africa and the Middle East represent 12 of the 17 most water-stressed countries; India ranks 13th internationally for water stress and has more than three times the population of the other 17 most-stressed countries combined. In the U.S., New Mexico faces extreme water stress, and California, Arizona, Colorado and Nebraska all face high water stress. See Rutger Willem Ofste, Paul Reig, Leah Schleifer, "17 Countries, Home to One-Quarter of the World's Population, Face Extremely High Water Stress," World Resources Institute, August 6, 2019, available at: https://www.wri.org/blog/2019/08/17-countries-home-one-quarter-world-population-face-extremely-high-water-stress; World Health Organization, "Drinking Water," 14 June 2019,available at: https://www.who.int/news-room/fact-sheets/detail/drinking-water;

1. Individualist conceptual account

Under DIM, evaluation of the legitimacy of such a scheme would lead to the following kinds of inquiries: did these households adequately consent to this tracking? Are the purposes to which this water data is being used purposes that uphold the rights of household members, or do not violate duties owed household members? Alternatively, are households being adequately compensated for this data collection? In response to the citizens who object to data being collected, robust DIM reforms would grant the ability to deny collection, the right not to have data about them used in ways that violate their interests, or payment for the data they provide.

Under this analysis, both Waterorg and Watercorp's behavior may be diagnosed as wrongful (and if addressed via legal reform, unlawful) if either entity collects and commodifies household water data against the wishes or interests of households from whom it is collected. If household members feel wrongfully commodified, under robust dignitarian DIM protections they would have the right to object to and opt out from this data production; under robust propertarian DIM protections, they would have a right to demand a greater share of the wealth their data creates for Watercorp, or fair repayment under takings law from Waterorg. On the other hand, Waterorg or Watercorp's behavior under this analysis is *not* wrongful (and not unlawful) if it collects this data under robust conditions of meaningful consent, does not use this data in ways that violate the protected legal interests of household individuals, provides real options for households to opt out of water collection, or alternatively, provides a fair wage or sale price for household water data. In sum, these protections, done right, secure for households the rights to payment, exit or recourse, regardless of which entity is collecting their data, or which purposes guide this collection. This may empower individual households against either entity, but still practically falls back on individual choice to determine the legitimacy of data collection.

2. Population-based relationality

Even robust DIM-based responses miss how population-level interests in data production work. Consider the citizens who object to Waterorg's data collection due to the possible adverse use of such data against them; let's call one such citizen Cate. Cate's concern over the adverse use of household water data neither reduces to a right to prevent such data from being collected from her home nor to a right to restrict how data from her home may be used. Instead, her concern presents a population-level interest in *all* household water data. Waterorg doesn't need her data to get population-level insights about water consumption habits for households like hers—they may easily derive such insights from households that share relevant features (e.g. same household size, same neighborhood). For Cate's concern to be effectively expressed, it would

need to be accounted for at the population-level: for municipal water data production as a whole.

Nor is Cate's interest the sole interest at stake in water usage data collected from her home. Without enough quality household water data, Waterorg may not be able to make sufficiently fair or accurate water allocation plans as droughts grow more severe. This stymies Waterorg's plans to develop drought-conscious water management not just for those who withheld their data, but for everyone in the municipality. Fair and effective water management is particularly important for those who live in the poorer, drought-prone areas. The risks of non-collection will fall disproportionately on them, amplifying the material hardship experienced by the community who lives there.[137] These interests also accrue at the population-level, for water data production as a whole.

The legitimacy of Waterorg's data production cannot be determined via the conditions of data subjects' interpersonal exchanges since neither do their interests reduce to such choices nor are theirs the sole interests implicated by such choices. Only by representing these interests as relevant to the task of governance, can we begin to address the forms of social informational harm that may arise as a result of them.

C. Democratic data governance

Reconceptualizing the project of data governance from securing individual rights to institutionalizing collective ordering shifts the relevant line of inquiry: from how to secure greater data subject control or better legal expressions of data subject autonomy, to how to balance the overlapping and at times competing interests that comprise the population-level effects of data production. This raises core questions of *democratic governance*: how to grant people a say in the social processes of their own formation, how to balance fair recognition with special concern for certain minority interests, how to identify the relevant "public" or institutional level of civic life at which to coalesce and govern such collective interests, how to not only *recognize* that data production produces winners and losers, but also develop fair institutional *responses* to these effects.

This in turn theorizes a different approach to data in law—from an individual medium expressing individual interests, to a democratic medium that materializes population-level, social interests. Like other mediums of social relation, the governance of data raises political questions regarding what

[137] Note: this example explores a positive purpose for data collection (water allocation) which stands to disproportionately benefit this poorer community. One can also imagine a negative example which may produce disproportionate risks to this poorer community and would give rise to an interest for this community in non-collection. But again, this interest in data production would obtain at the institutional level concerning *all* water collection data.

individuals are owed and owe one another on the basis of these material relations, and how to distribute relevant benefits and risks among one another. This conceptualization of data is referred to below as "data as democratic medium" (DDM).[138]

1. Democracy as normative (egalitarian) standard

Asserting that data relations are democratic goes beyond descriptive claims regarding data's relationality to capture distinctly political and normative criteria for how this relationality and its attendant social effects should be negotiated and managed.[139] Conceptualizing data as a democratic medium therefore asserts both a positive and a normative claim: describing the kinds of interests that *do* result from data production as well as how such interests *ought* to be governed.

Democracy as a normative standard offers criteria for evaluating how data relations are ordered, and should be ordered, by data governance law. It provides one theory of what features define unjust data relations and distinguish them from just relations. "Thorough social and political democracy," writes Iris Marion Young, "is the opposite of domination."[140] Democratic equality is achieved, argues Elizabeth Anderson, under conditions in which "people stand in relation of equality to others."[141] Developing democratic institutions whereby people relate as equals does not merely secure the social conditions of individual freedom; it also addresses the institutional arrangements by which people's opportunities are generated over time, and "reflects a deontic requirement grounded in our equal moral status as persons."[142] Institutional recognition of competing interests therefore operationalizes the normative force of the population-level effects one's personal choices over data have on others, and may express not only what individuals are owed, but also what their obligations are to one another.[143] This posits an egalitarian political standard for legitimacy in place of individual choice, that considers the *quality of relations* under which data production occurs and those it seeks to enact.

Democratic ordering can therefore also provide one substantive standard by which to evaluate and distinguish different goals of data production, on the basis

---

[138] The author wishes to credit a conference held by Christine Desan, "Money as a Democratic Medium," at Harvard Law School in December 2018 for inspiring this phrase's application in the data economy context.

[139] Samuel Scheffler, *The Practice of Equality*, In C. Fourie, F. Schuppert and I. Walliman-Helmer (eds.) *Social Equality* (2015), at 31; KARL LIPPERT-RASMUSSEN, RELATIONAL EGALITARIANISM: LIVING AS EQUALS (2018); IRIS MARION YOUNG, JUSTICE AND THE POLITICS OF DIFFERENCE (2011 ed. 1990).

[140] IRIS MARION YOUNG, JUSTICE AND THE POLITICS OF DIFFERENCE (2011 ed. 1990) at 38

[141] Anderson 1999, supra at 289.

[142] KARL LIPPERT-RASMUSSEN, RELATIONAL EGALITARIANISM: LIVING AS EQUALS (2018) at 19; Anderson 1999, supra.

[143] Lippert-Rasmussen supra; Samuel Scheffler, *The Practice of Equality*, In C. Fourie, F. Schuppert and I. Walliman-Helmer (eds.) *Social Equality* (2015) at 17.

of the goals it seeks to achieve and the social relations under which production occurs. In the context of data production, the general egalitarian case for democratic ordering is bolstered by the specific, empirical significance of population-level interests in data production. DDM expresses not only the general case in favor of more democratic ordering, but also something akin to empirical fact: personal choices over data sharing should reflect the effects this choice has on others, not only because of the political and moral benefits of considering others, but also because under current conditions of datafication, individuals *do* directly relay information relating to others, which *is* used to predict and influence the behavior of others.

2. Democratic evaluation of Waterorg vs. Watercorp

On this view, Watercorp's data production is not only unjust if it extracts household data without consent, underpays households, and/or renders household's legible against their will. These injustices stem from a more fundamental problem: that households under Watercorp's data production scheme have no ability to meaningfully determine the social processes via data production to which they are being subjected—to exercise equal power *back* onto Watercorp and over one another with respect to the population-level decisions that affect them all. In other words, Watercorp does not have to consider the normative force of its decisions or actions on others—and nor do individual households who may choose to opt in or out of this data production.[144] Under the Watercorp scheme, households have (at best) an incomplete say in the institutional arrangements that structure the scope of their choices and the social processes to which they are subjected. Securing negative rights of exit or payment are not the same as securing affirmative rights to representation in the conditions and purposes of data production.

Alternatively, Waterorg's data production may be legitimate even if it subjects data subjects to mandatory data collection if this fundamental condition of institutional recognition is satisfied in the terms of collective obligation that bind data subjects. What population-level interests make clear is that the relevant task of data governance is not to reassert individual control over the terms of one's own datafication (even if this were possible) or to maximize personal gain, but instead to develop the institutional responses necessary to represent the relevant population-level interests at stake in data production. This shifts the task of reform, from providing opportunities for exit, payment or recourse, to securing recognition and standing to shape the purposes and conditions of data production for those with interests at stake in such choices, and thus establish the terms of legitimate mutual obligation.[145]

---

[144] Anderson 2009, supra.

[145] Nancy Fraser, *From Redistribution to Recognition? Dilemmas of Justice in a 'Postsocialist' Age*, Justice Interruptus (1997), at 11-39; AXEL HONNETH, THE STRUGGLE FOR RECOGNITION, TRANS. JOEL ANDERSON (1995).

Population-level representation also clarifies the tradeoffs among competing interests in data production. In the Waterorg scheme, shifting from individual rights to institutional governance represents both the interests of the citizens who oppose data collection and the interests of citizens who stand to suffer the most in the absence of such collection. This clarifies who stands to lose and who stands to benefit from data production, as well as the potentially distinct normative stakes of these relative wins and losses.

D. Conceptual benefits of DDM

1. Social informational harm

Reconceptualizing what interests are relevant for data governance clarifies what makes data production, as a core economic activity in the digital economy, potentially wrongful. Data production may indeed be unjust if data subjects are manipulated at the point of collection, or subject to governmentality at the point of use. Such acts may wrongfully violate data subject autonomy. But data production may *also* be unjust when it enacts or amplifies social processes of oppression along horizonal data relations—which evidence suggests is a large and growing problem in the digital economy, and a significant source of the political and social critique levied against large data producers.[146]

As an unjust *social* process datafication denies individuals (both data subjects and those with whom they are in horizontal data relations) a say in the social processes of their mutual formation. Data relations can materialize unjust group-based relations like racism, sexism, and classism.[147]

Let us take again the example of ICE detaining undocumented immigrants on the basis of their movement patterns. "Movement patterns" as a shared feature becomes one defining feature of the category of "undocumented immigrant" (a category which in turn is defined via racial, class, and linguistic difference). By identifying this common feature and operationalizing it to detain people, this data flow materializes a particular oppressive social meaning onto the category of "undocumented immigrant." Such data flows thus become social fibers of domination; they help to create, organize express, and direct the meaning of this social category ("undocumented immigrant") as the experience

[146] This point is covered in some detail in the Introduction. For further reading, see Salomé Viljoen, "The Promise and Limits of Lawfulness: Inequality, Law, and the Techlash," (forthcoming, manuscript on file with author).

[147] This theory of injustice is far from new. Several political philosophers and legal theorists (cited throughout) similarly view social relations as the primary basis of (in)justice. This view also builds on social constructivist accounts of group membership; these accounts center the social meaning of group membership—the cultural practices, institutions, norms, and material conditions that make group membership coherent indicators of identity and experience, and for relevant forms of group membership (race, gender, caste, nationality, etc) also define forms of oppression that attend (and constitute) group membership.

of systematic violence and oppression for those who occupy this category.[148] This gives social meaning to the category of "undocumented immigrant," such that part of what group membership becomes *is* the fact of having the movement patterns of yourself and others weaponized against you.[149]

This form of injustice is a fellow traveler of personal violation—it denies individual undocumented immigrants the chance to determine their own social formation—but it also represents a distinct form of *social injustice*. It structures a hierarchical group relationship between undocumented immigrants and others. DDM's conceptual account thus helpfully identifies *why* patterns of datafied personal violation re-inscribe existing social arrangements of patterned disparity on the basis of race, sex, class, and national origin. Focusing legal inquiry on data production's population-level effects brings into view both *how* and *why* the risks of personal violation are not randomly distributed but determined via existing social patterns of power distribution that occurs along the lines of group membership. [150]

In short, by forming and then acting on population-level similarities in oppressive and dominating ways, datafication may materialize classificatory acts of oppressive category formation that are themselves unjust. This adds a social dimension to the personal violations of governmentality. Datafication is not only unjust because data extraction or resulting datafied governmentality may violate individual autonomy; datafication may also be unjust because it violates ideals of social equality. Social informational harm thus represents an additional and fundamental form of potential injustice of relevance for data governance law. Locating material forms of social injustice in datafication also helps to identify data production as important terrain for debating theories regarding why social processes that enact group oppression may be wrong, and how they may be addressed via law.[151]

---

[148] MacKinnon, supra at 516.

[149] See IRIS MARION YOUNG, JUSTICE AND THE POLITICS OF DIFFERENCE (2011 ed. 1990) at 61 on systematic violence. Young defines a particular form of systematic violence as a system of social oppression. Members of oppressed groups often live with knowledge that they must fear random, unprovoked attacks on the basis of group membership. The social practice of violence serves to reproduce social oppression through its assertion onto the meaning of group identity and make a feature of group membership the experience of fearing a particular form of violence. Catherine A. MacKinnon famously advances this argument regarding the social construction of sexuality via hierarchical relations of desire. See *Feminism, Marxism, Method and the State: An Agenda for Theory*, Signs, Spring 1982 Vol 7, No 3 pp. 515-44.

[150] CHARLES TILLY, DURABLE INEQUALITY (1999); IRIS MARION YOUNG, JUSTICE AND THE POLITICS OF DIFFERENCE (2011 ed. 1990) at 37-8.

[151] See e.g. Issa Kohler-Hausmann, *Eddie Murphy and the Dangers of Counterfacutal Causal Thinking About Detecting Racial Discrimination*, 113 Northwestern L. Rev. 1163 (2019); Lily Hu, *Direct Effects,* Phenomenal World, September 25, 2020; MARTHA MINOW, MAKING ALL THE DIFFERENCE (1990); IRIS MARION YOUNG, JUSTICE AND THE POLITICS OF DIFFERENCE (2011 ed. 1990); The rich and lively debate in political and social philosophy

2. Socially beneficial data production

DDM also offers an opportunity to conceptually distinguish purposes and priorities of data production for socially worthwhile ends. This offers a robust positive agenda for data governance law to expand on existing practices of data production for the public interest, undertaken with strong forms of public accountability, purpose limitations, and confidentiality standards.

a. Expanding on existing practices

Public data collection and use has long served a key role in the institutional management of state welfare and in other instances of public knowledge management for public benefit. Public health care information systems like the UK's national health data sets, or the Veteran's Affairs Administration's open source electronic health records system VistA, facilitate high-quality public health research.[152] Statistics on U.S. demographics and economic activity collected by the Bureau of Labor Statistics and other U.S. statistical agencies offer invaluable insight into the changing patterns of American life. The basic task of governance could not be achieved without the massive collection of tax information by the Internal Revenue Service, nor could financial regulation occur without the disclosure requirements overseen by the Securities Exchange Commission.

Governance with any commitment to public welfare will always require balancing the necessity of collecting important, at times highly personal and consequential, information from citizenry, and the risk of oppression and undue coercion that accompanies any such collection. Yet as the Article argues above, the absence of public oversight does not signify the absence of potentially coercive and harmful effects from data production. Indeed, existing best practices and several proto-democratic proposals for data governance offer promising examples of how to achieve robust legal protections against socially harmful data production while preserving the societal benefits data production may facilitate.

---

regarding whether properly attending to group membership requires a group-based methodology for identifying features of justice, or a group-based theory of justice is ongoing. This worthwhile debate is complex and beyond the scope of this piece, which will simply identify here the importance of group membership and the role of category construction in social processes of injustice for many theorists (a few of which are citeed above) in understanding how social injustice works, and thus what justice may require for groups qua group membership. See LISA SCHWARTZMAN, CHALLENGING LIBERALISM: FEMINISM AS POLITICAL CRITIQUE (2006); Elizabeth Anderson, *Towards a Non-Ideal, Relational Methodology for Political Philosophy: Comments on Schwartzman's Challenging Liberalism,* Hypatia vol. 24, no. 4 (Fall, 2009); SALLY HASLANGER, RESISTING REALITY: SOCIAL CONSTRUCTION AND SOCIAL CRITIQUE (2012).

[152] On the VA, see PHIL LONGMAN, BEST CARE ANYWHERE, (2010). See also Arthur Allen, *A 40-year conspiracy at the VA*, Politico, March 19, 2017. The author wishes to thank and credit Chris Morton for this excellent example. On the NHS, see "Data Sets" at digital.nhs.uk.

There are several proto-democratic data governance proposals and projects from which to draw inspiration regarding how responsible collective governance of information flows may be realized.

Several proposals aim to assert public management and control over existing proprietary data flows, often via mandated public access or by reverting such data to the public domain to be managed via public trust. One possibility is that data governance legislation could require private data companies to provide national statistical officers (appropriately safeguarded) access to private data sets under specifications set by law or agency determination.

A bolder alternative is to build on examples like the Human Genome Project to develop public data management for public benefit rather than for proprietary gain. Former German Social Democrat leader Andrea Nahles has argued for a national data trust, likening digital technology companies to pharmaceutical companies that enjoy a limited monopoly right to their data. After a set period of years, such data would revert to the public domain to be governed by a public trust or independent agency for use in service of the public good.[153]The UK and Canada have explored public data trusts as a way to collectively govern citizen data as a national resource from which to develop competitive technology industries.[154] Barcelona has implemented a civic data trust to manage its data commons, democratizing data governance while also using its data infrastructures to deepen democratic engagement.[155]

---

[153] Evgeny Morozov, *There is a leftwing way to challenge big tech for our data. Here it is,* THE GUARDIAN, 2018; Andrea Nahles, *Die Tech-Riesen des Silicon Valleys gefährden den fairen Wettwerb,* HANDELSBLATT, August 13, 2018; Hetan Shah, *Use our personal data for the common good*, NATURE, Vol 556, 5 April 2018 (in which then-executive director of the Royal Statistical Society argues in favor of public data governance for the common good).

[154] Dame Wendy Hall and Jerome Pesenti, Growing the Artificial Intelligence Industry in the UK, Independent Report Commissioned by UK Dept of Digital, Culture, Media and Sport and UK Dept for Business, Energy and Industrial Strategy (15 Oct 2017), available at https://www.gov.uk/government/publications/growing-the-artificial-intelligence-industry-in-the-uk, at 4 (recommending data trusts to improve secure and mutually beneficial data exchanges); Ontario has commissioned a series of discussion papers for the region's Data Strategy, which includes discussion of the merits of data trusts, and launched a public consultation session in August 2020 to seek public input. See *Ontario Launches Consultations to Strengthen Privacy Protections of Personal Data,* August 13, 2020, available at https://news.ontario.ca/en/release/57985/ontario-launches-consultations-to-strengthen-privacy-protections-of-personal-data; The Open Data Institute is prominent international non-profit group that works with governments and other entities to develop more open data ecosystems and has worked with the UK government (among others) to research and implement data trusts. See Open Data Institute, *Data Trusts: Lessons from three pilots*, April 15, 2019, available at https://theodi.org/article/odi-data-trusts-report/; for more on data trusts generally, see Bianca Wylie and Sean McDonald, *What Is a Data Trust,* CIGI Online, October 9, 2018, available at https://www.cigionline.org/articles/what-data-trust.

[155] Evgeny Morozov and Francesca Bria, *Rethinking the Smart City: Democratizing Urban Technology*, 2018 at 26 (detailing Barcelona's approach to building a "city data commons"); SmartCityHub (2018).

Such proposals can be distinguished from individualist propertarian approaches in that they do not extend individual rights to data subjects as a way to break open the walled gardens of corporate-held consumer data. Instead, they conceive of citizen data as a public resource (or infrastructure), to be managed via public governance and in furtherance of public goals. Such proposals also depart from dignitarian approaches; they advance legal responses to citizen data not only as a subject of potential violation, but also a potential resource for citizen empowerment. Dignitarian governance systems like the General Data Protection Regulation may establish standards of violation and pathways for exit, but these proto-democratic forms of public data governance offer a promising (and largely though not always complementary) addition to grow and develop public capacity to utilize data infrastructure for public ends.[156] In other words, rather than a governance approach that establishes what private entities may *not* do to German, Canadian or Barceloní citizens' data, these alternative approaches consider what data as a public resource *can* do for German, Canadian, or Barceloní citizens. Indeed, the highly-attuned feedback structures that data production allows offer new possibilities for public governance and social coordination.

Not all proposals advocating for new collective data institutions envision traditionally public forms of data management. Others seek to democratize governance of data production as part of ongoing efforts to democratize other spheres of life, most notably the workplace. Labor activists are developing worker data collectives to counter growing workplace surveillance by employers by monitoring forms of workplace oppression, documenting OSHA violations and wage theft, with the goal of collectively negotiating how algorithms govern life at work.[157] Other advocates are developing alternative worker-based data streams to better document the economic value and impact of essential workers, or to give workers greater ability to document and trace supply chains for their products.[158] These non-governmental collective alternatives may be particularly

---

[156] The dignitarian data subject rights granted under the GDPR may provide a complementary backstop to the kinds of affirmative data production envisioned by such proposals, but as discussed in the Waterorg example, strong individual data subject rights may also foreclose them. In fact, many commentators believe the proposed Data Governance Act in the EU, which provides the basis for some collective forms of data governance, would violate fundamental data subject rights in the EU, because it would allow data subjects to devolve inalienable rights over their data to the data institutions.

[157] Weclock.it/about ("[WeClock] offers a privacy-preserving way to empower workers and unions in their battle for decent work"); Lighthouse: a guide to good data stewardship for trade unions, available at https://lighthouse.prospect.org.uk; The National Domestic Workers' Alliance developed its alternative platform for domestic workers to help house cleaners get benefits by providing clients a platform to contribute to a cleaners' Alia count. In turn, cleaners can use the collective contributions from clients to purchase benefits that domestic workers may not otherwise be entitled to by law. See ndwalabs.org, "Alia" available at https://www.ndwalabs.org/alia.

[158] National Domestic Workers' Alliance, "La Allianza," available at:

attractive in places and with respect to data flows where individuals have little faith either in private companies or the government to safeguard collective interests.[159] Private data governance mechanisms may also face certain challenges in realizing the ideals of democratic data governance. Most notably, many proposals for private trusts work by pooling individual data subject rights. This only recognizes the interests of data subjects from whom data is collected, rather than also considering those on whom data products may be used—and who therefore also have a relevant interest in the terms that govern how data is collected and processed.[160]

Finally, existing forms of trusted public data collection and management, like those of the US Census and its statistical agencies, the Library of Congress, and state and local municipal libraries may be expanded into more general data governance bodies.[161] Public statistical agencies and libraries have established professional expertise around responsible information and knowledge management for the public good, and adhere to strict purpose limitations as well as high confidentiality standards.[162] Alternatively, public data management for the public good may be achieved via an expanded remit for scientific research agencies such as the National Institutes of Health and the National Science Foundation, or public agencies that already hold public data like the Food and Drug Administration. These agencies already have institutional expertise in stewarding data and managing scientific resources in service of the public good.[163] While none are perfect, each stem from long professional histories of managing collective knowledge in the public interest.

b. The possibility of democratic data

The data economy has resulted in massive collection of information regarding consumer purchasing preferences and social networks, for instance,

---

https://www.ndwalabs.org/alianza; see Katya Abazajian for Mozilla Insights, What Helps? Understanding the Needs and the Ecosystem for Support, March 2021 at 37 (Abalobi gives South African fishing communities access to data that helps them track where their fish is sold and connect with restaurants and other patrons who buy their stock; the platform is managed by fishing labor cooperatives that make collective decisions regarding the platform)

[159] In an international survey of several organizations developing alternative data governance regimes conducted by Mozilla, almost all respondents suggest that users would trust a collective of peers more than they would trust themselves or government to appropriately use their data. Katya Abazajian for Mozilla Insights, What Helps? Understanding the Needs and the Ecosystem for Support, March 2021

[160] See e.g. Sidealk Labs proposed data trust.

[161] Salome Viljoen, Jake Goldenfein, Ben Green, *Privacy vs. Health is a False Trade off*, JACOBIN, April 17, 2020; JULIA LANE, DEMOCRATIZING OUR DATA (2020).

[162] U.S. Census Act, 13 U.S.C. §§ 8(b), 8(c), 9. 68 Stat. 1012 (1954); Eun Seo Jo and Timnit Gebru, *Lessons from archives: Strategies for collecting sociocultural data in machine learning*, Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, pp 306-316.

[163] The author wishes to thank and credit Christopher Morton and Amy Kapczynski for drawing her attention to the example of the Food and Drug Administration.

but has contributed little to further knowledge about waste production, water usage, or how wealth from financial instruments flows globally.[164] Companies know a great deal about their consumers, but consumers still have little insight into the supply chains, ownership structures, and operating practices of companies. Workers are subject to increased surveillance at the workplace and in the screening process for employment, but know comparatively little about the hiring practices, quality of workplace life, and histories of discrimination and harassment of employers. Ensuring greater recognition can expand the set of interests considered relevant to setting the agendas of data production, and in turn how data infrastructures are funded and developed. In short, conceiving of data's democratic possibilities can provide greater standing for a wider range of priorities and goals to motivate how and why information is produced. This may result not just in less consumer-preference data production, but in the proliferation of other kinds of socially useful data production.

As the Waterorg example shows, DDM also affords stronger conceptual footing for data production conditions that may require mandatory data collection, as long as the purposes and the conditions of such collection are derived from legitimate forms of collective self-willing and further legitimate public ends. This has important implications for other public reform projects that will almost certainly rely upon data infrastructures and citizen data. Conceptually distinguishing and defending data production for core public functions is especially valuable for data governance reform projects that aim to act from a political position that citizens are owed more, not less, from the state by virtue of their status as citizens.[165] Public provisioning will require making productive and distributive decisions over social resources—decisions that should be (indeed, likely must be) informed by citizen data. The data infrastructures necessary to responsibly produce and allocate goods and services such as healthcare, education, housing, clean air, and fresh water, will require some degree of mandatory citizen data collection to manage this provision efficiently and fairly.

---

[164] Compare e.g., Liran Einav and Jonathan Levin, *Economics in the age of big data,* Science 346 (2014) ("Private companies that specialize in data aggregation, such as credit bureaus or marketing companies such as Acxiom, are assembling rich individual-level data on virtually every household"); Richard Henderson and Owen Walker, BlackRock's black box: the technology hub of modern finance, Financial Times, Feb 24, 2020 (discussing how BlackRock's tech platform Aladdin, a "central nervous system for many of the largest players in the investment management industry." BlackRock is not required to disclose how many of the world's assets sit on the system. They last did so in 2017, at which time they reported $20 trillion; since then BlackRock has added scores of new clients).

[165] Elizabeth Anderson, *How Should Egalitarians Cope with Market Risks?* Theoretical Inquiries in Law 9(1) (2008). See also Elizabeth Anderson, *Common Property: How Social Insurance became confused with socialism,* BOSTON REVIEW, July 25, 2016.

3. Democratic regimes and individual data subject rights

The discussion above highlights a few key insights regarding the relationship between legal agendas for democratic data governance and those that prioritize individualized data subject rights.

First, the theory of democratic regimes advocated in this Article is agnostic regarding the ontological commitments implied by individualist regimes (that view data either as "thing-like" or "person-like"). There is a long philosophical (and legal) tradition that makes sense of both property and persons as constitutive of and constituted by social relations. Where democratic governance proposals depart from individualist ones is in their conception of where interests in information adhere, and the legal agendas that flow from this conception.

For instance, democratic governance regimes do not repudiate the notion that individuals have dignitary interests in information; it repudiates the idea that legal protection of these interests is reducible to the vertical relation between data subject and data collector.[166] Consider for example data collected by a fertility tracking app suggesting a person (let's call her Amy) is in her first trimester of pregnancy. One may consider it a dignitary violation for an advertising company or employer to gain downstream access to this data. But Amy's dignitary interests in keeping her pregnancy private are implicated whether the company gains access to Amy's data via her fertility tracking app, or whether the company contracts with a service that analyzes and infers from several relevant features Amy shares with known pregnant people that there is a 95% chance that Amy is in her first trimester of pregnancy. Amy has a dignitarian interest against people seeking to learn her pregnancy status, but this interest resides—both for Amy and for others—at the category level of first trimester pregnancy data.

Democratic regimes also allow us to recognize (and adjudicate among) competing dignitarian interests with respect to the same data. For instance, responding to Amy's dignitarian interests by restricting the collection of first trimester pregnancy data may be in tension with the dignitarian interests of others to enact their informational self-determination—to share data about their first trimester pregnancy with a fertility app to enjoy its services.

As this Article has endeavored to show, people do not only have dignitarian interests in information; they also have egalitarian ones. These interests index concerns over social informational harm: that people have a collective interest against the unjust social processes data flows may materialize, against being

[166] It is important to note that even under an expansive democratic regime, certain dignitarian interests in information *do* rightly reside with individual data subjects. For instance, democratic data regimes should grant individuals rights against being singled out or re-identified by aggregate data processing meant to provide insight into population-level trends, and rights over unique biometric identifiers for purposes of identification and verification.

drafted into the project of one another's oppression as a condition of digital life, and against being put into data relations that constitute instances of domination and oppression for themselves or others on the basis of group membership. Casting all relevant concerns regarding information as individual claims to payment or self-determination masks collective egalitarian social interests in enacting data relations of equality rather than oppression.

By recognizing such interests, democratic data regimes in turn apprehend potential tensions (both conceptual and institutional) between achieving more egalitarian data relations and robust dignitarian informational protections. Consider for example Amy's first trimester pregnancy data. The dignitarian account may well express that companies or employers gaining access to this information would violate a privileged relationship Amy enjoys to this sensitive information. But a relational account of this data flow also captures *why* this information is so sensitive to begin with.

One may find this data flow (i.e. first trimester pregnancy data) particularly sensitive because of its significance in constituting a relevant group identity—of materializing a key aspect of what it means (legally and socially) to occupy the status of "woman" in this particular historical context.[167] Part of the social construction of womanhood involves the contested legal and social terrain of early pregnancy; and data flows that impart knowledge of an early pregnancy bring Amy onto this terrain. This in turn leaves her vulnerable to certain forms of social oppression on the basis of this category membership. It may implicate or constrain the choices she makes (including sensitive and contested ones like terminating her pregnancy) that are intimately bound up with how legal, cultural, and social institutions construct and condition womanhood. In sum, early pregnancy data flows are sensitive and require governance because these flows help to materialize social relations of sex, gender, and fertility—and depending on how these data flows are governed, they can exacerbate or reduce the inegalitarian condition of these relations.

Put another way, many of the intuitions currently cast as dignitarian interests in protecting data flows actually have a great deal to do with the (egalitarian) social significance of data flows. Pregnancy data is sensitive because womanhood (and even more so non-woman pregnant person) is a historically oppressed social category. The data flow "redheads who like potato chips" likely implicates far fewer (and far less significant) legal interests because "redheads who like potato chips" are not a social category historically constituted through domination.

[167] For the sake of this argument, it is assumed that Amy identifies as a woman. However, this materialized social relation becomes even more sensitive and more significant in the case where Amy does not identify as a woman.

But to distinguish between data flows that constitute socially innocuous categories and socially consequential ones, and to distinguish between (and adjudicate among) social egalitarian interests and individual dignitarian ones, requires comprehensive data governance mechanisms that can apprehend these various interests at the population-level.

Where democratic governance regimes depart from individualist alternatives is recognizing this plurality of (population-level) interests in information production, and providing a normative theory for adjudicating among them. The underlying claims of injustice that motivate individualist agendas for reform are important but incomplete. Reducing these interests to individual data subject rights in a data transaction do not do these interests justice, nor do they apprehend when other interests may conflict with and at times supersede such interests.

CONCLUSION: REORIENTING THE TASK OF DATA GOVERNANCE

If the aim of data governance is to account for population-level interests in the digital economy, then different legal conceptions of informational harm (and our legal responses to them) may be required. This is not to say injustice may not also occur along vertical relations—it may, and it does. But as shown in Part II, the imperatives to relate people to one another place pressure on the conditions of exchange that structure vertical relations; accounting for population-level horizontal interests are thus relevant to the task of addressing these forms of injustice too.

As argued in Part III, theories of data governance that stem from individualist conceptions of informational harm do not represent the social effects of data production as a result of the pervasive population-level horizonal relations that data production enacts. Such theories thus cannot address the ways these effects may cause harm nor how these effects could be structured to produce shared benefits. This presents a methodological limitation and an epistemic deficiency, since such notions of informational harm fail to provide adequate tools for identifying and addressing the harmful social effects that datafication produces.

The conceptual account offered by this Article foregrounds data's relationality, which results in a few helpful reorientations regarding the task of data governance. First, it clarifies that social inequality is not a byproduct of unjust data collection but is an injustice of concern in data production in its own right. This informs a different diagnosis of data governance failure. On this account, datafication may be wrong not only because it manipulates people; it may be also wrong (or even be *primarily* wrong) because the social effects it produces or materializes violate standards of equality. As an economic process, datafication may lead to unfair wealth inequality that violates distributive ideals

of justice. As a social process, datafication may reproduce and amplify forms of social hierarchy that violate relational standards of justice.

The prevalence of population-level interests in data production mean that one's actions in the data political economy necessarily impact others, in uneven ways over which one has no direct control, often re-creating or exacerbating the durable inequalities that operate along the lines of group identity.[168] This raises quintessentially democratic questions: it requires negotiating tradeoffs among groups of people with competing and at times normatively distinct interests. Hence, datafication gives rise not only to personal claims regarding risk of personal violation that justify personal ordering, but also to population-level claims about the risk of social effects that justify political ordering.

The unsettled status of data in law presents both a challenge and an opportunity: a challenge for addressing the injustices that arise from digital life, and an opportunity to experiment with the kinds of social ordering the law may enact in response. Far from offering terrain on which to re-impose forms of private market ordering, data governance may plausibly retrieve spheres of life from private governance and begin to develop new alternatives.

[168] CHARLES TILLY, DURABLE INEQUALITY (1999); IRIS MARION YOUNG, JUSTICE AND THE POLITICS OF DIFFERENCE (2011 ed. 1990); Nick Couldry and Ulises A. Mejias offer one interesting account theorizing the data social relation as that of colonizer and colonized, an account they refer to as 'data colonialism'. See THE COSTS OF CONNECTION (2019).