University of Chicago Law School

Chicago Unbound

Coase-Sandor Working Paper Series in Law and Economics

Coase-Sandor Institute for Law and Economics

2018

Data Pollution

Omri Ben-Shahar
dangelolawlib+omribenshahar@gmail.com

Follow this and additional works at: https://chicagounbound.uchicago.edu/law_and_economics

Part of the Law Commons

Recommended Citation

Ben-Shahar, Omri, "Data Pollution" (2018). *Coase-Sandor Working Paper Series in Law and Economics*. 905.
https://chicagounbound.uchicago.edu/law_and_economics/905

This Working Paper is brought to you for free and open access by the Coase-Sandor Institute for Law and Economics at Chicago Unbound. It has been accepted for inclusion in Coase-Sandor Working Paper Series in Law and Economics by an authorized administrator of Chicago Unbound. For more information, please contact unbound@law.uchicago.edu.

DATA POLLUTION

*Omri Ben-Shahar**

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

ABSTRACT

Digital information is the fuel of the new economy. But like the old economy's carbon fuel, it also pollutes. Harmful "data emissions" are leaked into the digital ecosystem, disrupting social institutions and public interests. This article develops a novel framework—*data pollution*—to rethink the harms the data economy creates and the way they have to be regulated. It argues that social intervention should focus on the external harms from collection and misuse of personal data. The article challenges the hegemony of the prevailing view—that the injuries from digital data enterprise are exclusively private. That view has led lawmakers to focus solely on privacy protection as the regulatory objective. The article claims, instead, that a central problem in the digital economy has been largely ignored: how the information given by people affects others, and how it undermines and degrades public goods and interests. The data pollution concept offers a novel perspective why existing regulatory tools—torts, contracts, and disclosure law—are ineffective, mirroring their historical futility in curbing the harms from industrial pollution. The data pollution framework also opens up a rich roadmap for new regulatory devices—"an environmental law for data protection"—which focuses on controlling these external effects. The article examines how the tools used to control industrial pollution—production restrictions, carbon tax, and emissions liability—could be adapted to govern data pollution.

"Data are to this century what oil was to the last one"
—*The Economist*, May 2017

1. INTRODUCTION

Digital information is the fuel of the new economy. It is the resource that creates new products and companies, new markets and currencies, and endless new

* University of Chicago Law School, University of Chicago, Chicago, IL 60637, USA. Tel: +1 773-702-2087, E-mail: omri@uchicago.edu. I am grateful to Ronen Avraham, Oren Bar-Gill, Karen Bradshaw, Daniel Hemel, Jaime Hine, William Hubbard, Florencial Marrota-Wurgler, Jennifer Nou, Lisa Larimore Ouillette, Ariel Porat, Eric Posner, Ricky Revesz, Lior Strahilevitz, Mark Templeton, and workshop participants at the University of Chicago, the Federal Trade Commission, Harvard, Stanford, and Tel-Aviv University for helpful discussions, and to Brenna Darling and Jason Grover for research assistance.

© The Author(s) 2019. Published by Oxford University Press on behalf of The John M. Olin Center for Law, Economics and Business at Harvard Law School.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
doi:10.1093/jla/laz005

Electronic copy available at: https://ssrn.com/abstract=3191231

opportunities to create great social value.[1] But like the old economy's carbon fuel, it also pollutes. Harmful "data emissions" are spilled into the digital ecosystem, disrupting social institutions and public interests. This article develops a novel framework—*data pollution*—to understand the harms of the data economy and the regulatory responses to address these harms.

Digital information assembles every possible compilation of facts, but perhaps the most treasured content is personal data. Digital platforms are learning who and where people are at any given time, what they did in the past and how they plan their future, what and who they like, and how their decisions could be influenced. The widespread aggregation of such personal data creates new personalized, social environments with enormous private and social benefits. But they also produce potential harms. The potential injury to privacy interests—a type of private harm—has been widely remarked upon. The external harms, on the other hand, are less concrete and far less noticed. Understanding the scope of these external harms and reducing their magnitude are among the biggest policy challenges of our era.

Two phenomena have added urgency to this challenge. The first is the *intentional release* of personal data, which the events surrounding the 2016 U.S. presidential election dramatically illustrated. Facebook's database of personal information was used to spread false political ads.[2] Political lies are not new, but their effect is magnified and harder to detect when propelled and pinpointed by data-rich processes. The second phenomenon is the *nonintentional release* of personal data—the failure of companies to secure their databases. The Equifax security breach, in which entire financial dossiers of 143 million consumers were stolen, is a prominent exemplar of this data emission problem.

Societies are searching for paradigms to understand, and techniques to address, the actual harms and potential misuses of personal data. This search is largely conducted in one place. The dominant, perhaps the sole, criterion currently used to evaluate the harm from the personal data enterprise is *privacy*. Under the data privacy paradigm, the collection and use of personal data create various harms to those whose data are collected. The privacy paradigm says that when personal and private matters are known or inferred about these individuals, their well-being, rights, autonomy, dignity—in short, their personal spheres—are impaired (Schwartz & Peifer 2017, p. 126). The privacy paradigm is founded on the premise that the injury from the personal data enterprise is private in nature—injury to the "core self"—but by sheer aggregation (or by

1 See *Economist* (2017). See also DeVries (2003) (the digital technological change on a scale matching or exceeding the industrial revolution) and Isenberg (1995, p. 15).

2 See Granville (2018)

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

more nuanced channels), these deeply private injuries have a derivative, super-additive, social impact (Westin 1967; see also Reiman 1982, p. 300).[3]

The privacy paradigm is disturbingly incomplete because the harms from data misuse are often far greater than the sum of private injuries to the individuals whose information is taken. If indeed "data are to this century what oil was to the last one," then—I argue—data pollution is to our century what industrial pollution was to the last one. Pollution, whether industrial or digital, creates public harms and destroys public goods, in addition to the impact felt by private people who use products that pollute. The methods to control pollution and to protect public interests are distinctly different than the legal redress for private harms.

The concept of data pollution invites us to expand the focus and examine the ways that the collection of personal data affects institutions and groups of people—beyond those whose data are taken, and apart from the harm to their privacy. Facebook's data practices lucidly illustrated the impact of data sharing on an ecosystem as a whole. When Facebook granted advertisers access to personal data, allowing them to distort voting decisions, the negative effect was not fully captured by private injuries to the specific individuals who received data-driven ads and whose voting was influenced (many of whom, indeed, regard themselves as unharmed). The critical negative effect was far broader, captured by the damage to an entire electoral and political environment, including nonprivacy-related harms and harms to other members of society. Even when not abused, platforms that provide people "personalized news" are increasingly viewed as fragmentizing and polarizing, harmful to democratic deliberation, and degrading the "social glue" (Sunstein 2017, 2018).

Data sharing also pollutes in other, more concrete, ways. When people allow websites to collect information about their emails, social networks, and even DNA, they provide information about other individuals who are not party to these transactions. In personalized environments, the experience of each individual depends in part on the data shared about others.

The concept of data pollution helps organize three ambitious contributions that this article advances. The first contribution is to characterize the nature of data's social harm problem. A vast literature has combed through every aspect of the imaginable private harms from data collection—the potential privacy

3 See, e.g., Schwartz (1999, p. 1653) (database privacy is necessary for democratic deliberation), Cohen (2000) (privacy is necessary for a thriving civil society, free expression, and collective comfort), Nehf (2003, pp. 69–71) (privacy is necessary to the proper functioning of a democratic political system), Ashenmacher (2006) (characterizing the dignity harm caused by data breach and speculating that it could make people "hesitant to share data, which would frustrate stated policy goals"). See, generally, Solove (2002) (discussing the private and public domains of privacy protection).

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

injuries to the people whose data are collected. The externality problem, however, has often been neglected: how allowing one's data to be harvested affects other indviduals and the public as a whole. Section 2 exposes the various facets of this external, societal effect. It distinguishes data's less recognized social harm from its widely remarked upon private harm, thus beginning to build a new and complementary justification for data regulation. The discussion in Section 2 also helps solve a profound puzzle—how to reconcile the ubiquitous unease people harbor toward personal data collection with the universal indifference they exhibit by continuing to "pay with their data." This misalignment is largely regarded as a "privacy paradox" (Wittes & Liu 2015; Hermstrüwer 2017, p. 17; Athey et al. 2018).[4] Data pollution solves the paradox: people care about data's social harm, how it affects society as a whole—not so much about the potential private harm. Privately, they find data's private benefits irresistible.

The second intended contribution of this article is to explain the failure of existing legal tools in addressing the problems of data pollution. Section 3 argues that private law and private enforcement are unable to control data pollution for precisely the same reasons that they failed to control industrial pollution. The failures of private causes of action in this case are primarily due to the public nature of the harm. Pollution is an externality; it affects the entire environment, not merely the individuals with whom the polluter transacted, or whose data it emitted. Data pollution, like its industrial ancestor, creates harms to the public, and by the time specific individuals are injured, it becomes hard to identify the cause or the full magnitude of the social harm. In the industrial pollution context, plaintiffs historically have had difficulties attributing pollution-caused diseases to specific emissions, just as current data-misuse victims are finding it difficult to prove which data emitters caused their harms (Dewees 1992, p. 429). Even when causation is established, the magnitude of the harm suffered by specific victims of pollution—both in the industrial and digital areas—is often too speculative for the assessment of private law remedies.

I further argue in Section 3 that the shortfall of private law in regulating data pollution is due not only to the limits of tort law; it is also a failure of contracting. For the very same reasons that voluntary transactions over polluting products have failed to adequately reduce industrial emissions, markets for digital products fail to give meaningful attention to reduction of data pollution. Consumers buy products with excessive carbon footprints, and likewise leave excessive data footprints in digital domains. In both the industrial and digital arenas, people do not contract optimally over pollution for a variety of reasons,

4 For attempts to explain the privacy paradox as a problem of asymmetric information, see Froomkin (2015, pp. 1732–1735), Acquisti et al. (2015) (discussing the uncertainty and complexity of privacy decisions as the cause for behavior unprotective of privacy).

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

but primarily because pollution reduction is a public good. The optimism that contracts and behaviorally informed choice architecture would help people make wise data sharing decisions and reduce data pollution is fundamentally misplaced, because private bilateral contracts are the wrong mechanisms to address the harm to third parties. It is not people who need to be protected from a mesh of data-predatory contracts; but rather, *it is the ecosystem that needs to be protected from the data sharing contracts that people endlessly enter into.*

If Section 2 offers a new diagnosis of data's harm, and Section 3 explains the failure of existing approaches to address this harm, Section 4 presents the third contribution of this article, and a shamelessly ambitious one: developing an alternative regulatory roadmap for the control of data pollution. The pollution metaphor introduces a richness of regulatory devices and an organized set of prescriptions that until now have been either unnoticed or justified on other grounds.[5]

The primary regulatory method for pollution control is a predetermined set of production restrictions, most often in the form of quantity caps and quotas. The scope of the pollution-causing activity could be limited by requiring permits for the emitting activity or by subjecting the inputs and outputs to quantity regulations. Data pollution could, by analogy, be controlled by restricting production of data services along various dimensions: which data can be collected and by whom, how much and for what reasons, how may it be used or

5 A few writers have previously and thoughtfully invoked the environmental context as a framework to examine the regulation of data. However, in sharp contrast to this article, they focused on privacy harms and privacy law—how data collection causes private injuries to the people whose data are collected. Closest to the analysis of this article are Hirsch (2006, 2014), Hirsch & King (2016), Froomkin (2015), and Nehf (2003).
Like the analysis here, these authors examine a so-called "externality" caused by data collection, but define it as the diminishment of privacy brought upon by data-gatherers' surveillance practices. See, e.g., Froomkin (2015, p. 1732) ("If the parties being surveilled care about their privacy, then the surveilling party is imposing an un-bargained for cost on his target in order to achieve an end of his own. Whether or not that perfectly fits the classic model of an externality, it can certainly be modeled as one."), Hirsch (2006, p. 28) ("Companies benefit from the information they collect, but do not face the costs they impose . . . (i.e., the violation of consumers' privacy . . . In economic terms, the companies collecting personal information impose a negative externality on consumers"), Hirsch (2014, p. 375) ("Big Data . . . makes the social environment less conducive to the growth of the human personality"). Both Hirsch (2006, 2014) and Froomkin (2015) look to command-and-control regulatory devices used in environmental law as models for data-harm regulation, but because they view those harms as solely privacy-related (what they call the "inner environment," "privacy pollution," or "human personality"), their analyses of the regulatory methods lead them to different conclusions than the ones discussed in this article. Nehf (2003) by contrast, examines the societal value of privacy. Although he primarily focuses on the social derivatives of private/privacy injury ("alienation" and loss of power vis-à-vis "large institutions"), (*id.*, pp. 69–71), he also recognizes the "external costs beyond the direct injury to the individuals involved," like the pass-through societal costs of data breach (*id.*, pp. 79–80).

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

transferred, when it must be deleted, how should it be secured, and more. Such ex ante commands are increasingly favored by European privacy regulators[6] and in some narrow bands of U.S. privacy law—for example, in dealings with children.[7] Quantity regulations are the archetypical command-and-control regulations, and they are effective in obtaining a pollution-reducing result, but often at a substantial cost. They reduce not only the negative externalities, but also the positive ones, and they stifle innovation. In the data sphere, it is particularly challenging to apply a cost–benefit analysis to the design of data restrictions because the costs and benefits are hard to evaluate.

Recognizing the predicament of quantity regulations, Section 4 then turns to examine another central technique to control pollution: pricing the social cost. It is widely thought that "Pigouvian taxes" on an activity, on the inputs that feed it, or on the outputs it generates, could correct the distortion produced by a negative externality. In industrial production, this approach prescribes a carbon tax, and in the digital economy it could suggest a data tax. The social cost of private data collection could be internalized through a tax on the data activity. Section 4 explores some basic design problems with data tax: who would pay it, how would it be set, and what might be some of its intended and unintended effects. It also recognizes that data sometimes produce off-setting positive externalities, which might affect the size of data tax. Importantly, the data tax approach varies from (and conflicts with) recent proposals to require firms to pay people for their personal data (Kaiser 2018; Posner & Weyl 2018, pp. 246–249). Pay-for-data is a zero-sum transfer between two parties who are jointly producing data pollution, and it therefore does not reduce the underlying activity nor does it encourage pollution-reduction investment.

A third approach for pollution control is to design a liability regime for data harms, in particular to address problems of inadvertent data emissions. Like toxic waste released by industrial production, data spills are rapidly becoming a major externality that private law is struggling to address. Environmental law uses various tools to shift the harm of toxic waste to the emitters, and data pollution law could similarly focus on liability and prevention. While cleanup of spilled data is largely impossible, the harm from the release can be mitigated by post-spill actions and adequate preparedness. And the expected harm could be reduced by a proper system of deterrence. Liability equal to the social cost of

6 European Directive on Data Protection, 95/46/EC of the European Parliament and of the Council, Art. 25, OJ L 281, 23 November 1995, 56–57 (1995); General Data Protection Regulation (GDPR), 2016/679 of the European Parliament and of the Council, OJ 119, 4 May 2016, 60–62 (2016).

7 Children's Online Privacy Protection Act, 15 U.S.C. § 6501 (1998).

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

spills (bolstered by compulsory liability insurance) would lead to better precautions and self-regulation. One of the main contributions of Section 4 is the proposal to shift to a proportional liability regime.

Some of these methods of regulation have been previously examined, but only through the lens of privacy protection. Privacy is an alluring framework because the polluting databases are constructed from personal, sometimes private, information. So alluring is privacy—so plainly does it seem to be the sole issue at stake—that lawmakers and advocates have neglected to address the broader societal impact, which extends well beyond any effect on the private parties whose personal data are harvested. Section 4 considers how to correct his oversight, exploring a regulatory design for data pollution law designed to minimize the external social cost.

Environmental law was born in the industrial era because private law dealing with private harms failed to protect public goods and the environment (Abraham 2008, p. 149). We now need a twenty-first century version of environmental law—data pollution regulation—to expand the focus of data protection and begin to address the *public* harms from the personal data enterprise. This article offers a roadmap for such a transformation.

2. DATA'S HARM: PRIVATE OR SOCIAL?

For decades, a dominant concern in a world fueled by data technology has been privacy. Under the privacy paradigm, the collection of personal data by commercial entities may cause harm to the people whose information is being collected and used. Companies collecting personal information learn and infer things about people, and use this knowledge in ways that sometimes benefit individuals—but may also subject them to personal risks and harm.

An immense literature has labored to define the contours of this privacy harm. At times, the emotional injury to victims is clear and present, as when a website used by people to find partners for extramarital affairs is hacked.[8] Other times, people may feel harmed when a data-fueled algorithm presents them to the world in a manner that affects their reputation or financial opportunities (Keats Citron 2019). Episodes of clear and present emotional and reputational harm help sustain a ubiquitous premise—that the injury arising from the assembly of databases containing personal information is personal and private in nature. While some privacy theorists have articulated avenues by which, they think, this intimate and dignitary harm is also social—for example, by demoralizing people and thus degrading "democratic deliberation" or

8 See Thomsen (2015).

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

undermining a "thriving civil society"—the public harms they identify are still derivatives of the personal injury: the demoralization to the individuals whose private information is taken (e.g., Schwartz 1999, p. 1653; Cohen 2000; Solove 2002; Nehf 2003, pp. 69–71; Ashenmacher 2016).

The diagnosis that data's problem is privacy and its solution is privacy protection is alluring because the databases that firms deploy consist primarily of private information that people do not openly share. Much of the data are collected through procedures characterized by critics as "surveillance," whereby companies place "a permanent foothold in a person's home from which he can be monitored."[9] If the problem is that smart devices and apps are "spying on you (even in your own home)"[10] then the first harm that comes to mind is personal in nature, and the obvious redress to this potential harm is the protection of private domains.

But this reigning notion—that data technology inflicts privacy harms—faces a nagging difficulty, sometimes referred to as the "privacy paradox." Despite the vast attention lawmakers and advocates lavish on privacy risks and privacy protection, and despite widespread popular sentiment documented through survey evidence that data privacy matters, people largely behave as if it does not (Morey 2015).[11] They say that they greatly value their personal data, but they turn around and give it up for meager quid pro quo (Acquisti, John & Loewenstein 2013; Strahilevitz & Kugler 2016; Athey et al. 2018). The revealed preference for data privacy is distinctly lower than the declared valuation. There is not enough evidence, in short, that the privacy concerns commonly articulated—emotional health, personal dignity, autonomy, reputation—are impaired by the digital data enterprise in a discernable, measurable manner.

And yet, concern over the harm caused by data technology is dramatically increasing. The American political system has been shaken to its core by the recognition that Facebook's immense database was likely misused and may have influenced and even distorted election results. Additionally, the American consumer financial system has been jolted by the massive leakage of consumers' personal and financial data and its potential fraudulent misuses. Major jurisdictions around the world are enacting widely popular, sweeping reforms

9 See Silverman (2016).

10 See Steinberg (2014).

11 See Pollack & Sullivan (2018), Dell Technologies (2014), IBM News Room (2018) ("85 Percent of Consumers Say Businesses Should Be Doing More to Actively Protect Their Data"); see generally, Morey et al. (2015).

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

intended to make data collection more difficult.[12] The longstanding concerns over privacy violations are reverberating louder than ever.

How to reconcile these two conflicting empirical observations—the universal anxiety among people over the power of data with the universal indifference to sharing their own private information? This is the fundamental question haunting the field of data privacy law, for which a variety of explanations have been proposed (e.g., Acquisti et al. 2015; Froomkin 2015, pp. 1732–1735; Matthews & Tucker 2017). An explanation largely overlooked, and a centerpiece of this article, focuses on the nature of the harm. If an important component of data's harm is public, then the two sentiments are perfectly consistent. People worry about the power of data to cause social harm. They are not as worried about private harm, and thus they continue to share their data.

Agnostic to the question whether the private injuries caused by data are significant, I develop in this article the view that data's external harms should form a primary justification for data law. The effects of a database consisting of personal information could be felt by an entire ecosystem, not merely by those whose data are misused or emitted. Accordingly, the remainder of this section examines the patterns by which collection of personal data affects the public—how data pollute.

2.1 Effects on Public Interests

Industrial pollution degrades a public good. It is the quintessential negative externality, afflicting many who are not party to the polluting activity. It impacts an ecosystem as a whole, as well as the health of many third parties.

Emissions of data are like emissions of other pollutants; the costs are often external, degrading social interests. A digital database is not like the library card catalog of generations past, but merely the simple indexed sum of individual items of information. A digital data base is super-additive; new information can be learned that were not known when the information was atomized, including aggregate information that affects public interests. Other pieces of new information created by the aggregation include hints about individuals whose own information is *not* in the database, which are then used in ways that could harm these individuals or society as a whole. Let me use several examples to illustrate these negative externalities.

First, Facebook. When the social media giant allows app developers and other parties to access its users' database, the impact is only partially experienced by the individual users whose data are exposed. If, as may have happened in the Cambridge Analytica case, the data were used to spread political lies and fake

12 For example, California Consumer Privacy Act of 2018, AB 375.

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

news more effectively, the infected public interest was the integrity of the voting process. This effect reaches far beyond the private interests of the exposed parties. (In fact, it is quite possible that those whose data were used and whose behavior was influenced ended up satisfied with their conduct, not experiencing any personal harm.) The new, digitally propelled ability of hostile foreign governments to deploy digital platforms and pervert democratic outcomes is a primary threat to the public good.

A second example illustrating how a database can reveal information affecting public interests other than the users' privacy is the Strava fitness app—the self-proclaimed "social network for athletes." Strava enables millions of users to post map depictions of their physical workouts online, to be then viewed en masse in a "heat map" that may be accessed by anyone. Because the map exposes large concentrated clusters of users in areas of dense activity, it allows detection of secret geographic locations of U.S. military operations around the world.[13] What else could a cluster of physical workouts in the Sahara Desert, or in an outskirt of an Afghan city, stand for? It is through the aggregation of the personal data that a meta-picture emerges, and it threatens a public good—national security—not the individual privacy of any specific data sharer.

Figure 1. Strava heatmap of Kabul, Afghanistan, displaying a patch of activity south of the city.

13 See Perez-Pena and Rosenberg (2018) and Brown (2018).

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

The concern about the public harm caused by data aggregation is reflected in governments' efforts to limit cross-border transfers of commercial databases by establishing data exit controls and requiring "data localization" (Cohen et al. 2017, p. 107). The concerns driving such policies range from national security and law enforcement to trade protection and domestic industry prop up. The Chinese government, for one, declared that "data has become a national basic strategic resource" and mandated that personal information databases about Chinese citizens be stored within China.[14] It regards the huge amount of user information stored by the likes of Alibaba "a serious threat to national security" if leaked or exposed in unwanted manners (Yanqing 2017).

The potential for databases to be used in ways that harm public goods and publicly shared values is also illustrated by data's ability to personalize treatment and enable new forms of harmful discrimination. In general, the correlations in the database provide information about people that their individual data alone may not reveal, and these inferences could be translated into formulae for tailored services. Such granular treatments are the defining feature of personalized marketing and many other data-driven services. They deliver enormous social benefits—for example, when hospitals use digital medical records to provide better treatment faster and with less waste.[15]

But correlations inferred from digital databases could also allow for personalized treatments that discriminate against groups of people in disturbing ways. For example, when online ads promoting STEM careers are shown less often to women than to men (Lambrecht & Tucker 2018; Datta et al. 2015), or when online ads suggestive of arrest records appear more often along search results of black-sounding names (Sweeney 2013), the discriminatory impact could be toxic to society. The businesses advertising STEM careers on Facebook are rationally profiting by restricting the ads to men, and advertisers that help users find people's arrest records are "optimizing" their business by boosting their placement alongside pages that include black-sounding names. Such campaigns are made possible by personalized data analytics, and they merely respond to people's demand for information (Datta et al. 2018). But maximizing private valuation of ad placement in a society with preexisting discrimination and inequality does not guarantee socially optimal transmission of information. Instead, it can help optimize discrimination. It allows for patterns of discrimination that, in a world of small data, could not be achieved.

14 Cybersecurity Law of the People's Republic of China, Art. 37

15 See, e.g., Miller & Tucker (2017, pp. 51–54) (finding digital medical records reduce neonatal mortality).

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

It is not always easy to distinguish harmful discrimination from desirable personalization, as both are data-driven, tailored treatments of individuals. Personalized medicine, education, and nutrition help cure, teach, and feed people more effectively. Even personalized advertising helps people get more relevant information. It is quite possible that the benefits from data-driven personalization far exceed the negative impact from data-driven discrimination—that we should be talking about "data greens" rather than data pollution. But the benefits of data are often appropriated and internalized: firms creating such benefits have the incentives and technical tools to commercialize and monetize them (before competition dissipates such gains). The negative externalities, in contrast, remain orphan. They affect groups too broad and dispersed and cause injuries that are too abstract for private remedies to be effective.

Discrimination is a primary, but not the only, public degradation caused by data-driven environments. The rapid dissemination of hateful, manipulative, and polarizing news and beliefs through social media, the segregation of information communities, and the vanishing basis of the common experience—what Sunstein (2017) characterized as "echo chambers" and "information cocoons"—are phenomena that digital data have exacerbated (see also Benkler, Faris, & Roberts 2018).

2.2 Effects on Other Individuals

Pollution could have negative external effects not just on an ecosystem or a public good, but also on identified private victims. An asbestos contamination, for example, affects the health of the specific individuals who were exposed to it. Similarly, the externalization of costs associated with the digital data enterprise could occur through mechanisms that affect, not just the system in general, but identified individuals—individuals apart from the data givers.

The most common way such external effects occur is when users provide specific information about others to data gatherers. Consider, for example, Google's collection and use of personal data by scanning the texts of Gmail messages sent and received by its users. Any effect this has on individual Gmail users is internalized, accounted for by the transacting parties. Users are choosing to pay for email service with data rather than money (they have other options). But what about non-Gmail users who correspond with a Gmail account holder? The contents of their messages are viewed and collected by Google as well, pursuant to authorization by Gmail users. Any discomfort felt by these users—perhaps the same discomfort that drove them away from signing up for a free Gmail account in the first place—is an externality of the Gmail transaction. If these externally affected users were able to contract with

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

Gmail users (and with other email users who use email services hosted by Google) and "price in" the discomfort they experience, the external effect would be internalized. But such "Coasian" contracting is defeated by a host of transactions costs.

Another example of data sharing affecting third parties is the DNA information people give to genetic testing services like 23andMe or ancestry.com. The information stored in these databases reveals important facts about other people in the users' circles of biological relationships, who never agreed to give information to participate in such personal-origins discovery. Possibly, that information could be life-saving to third-party relatives. It could also be socially desirable when "genetic informants" help solve crimes or when the data help reunite families.[16] But the information could also affect others negatively, especially in circumstances where genetic anonymity is sought.[17]

Finally, consider a social network that gains authorized access to its users' data, which includes valuable information about these users' "friends" (including those who try to limit their exposure).[18] Short of exiting the networks altogether, there is little that the affected friends could do. Their efforts to remain anonymous are undermined once the data are harvested through the portals of others. Precautions, put differently, are jointly produced; failure of some members of the network to match the precaution level undermines the efforts by others.

Whether data pollution creates negative effects on the entire ecosystem or merely on an identified set of third-party individuals is relevant to the design of the regulatory response. Public law remedies, like quantity restrictions or taxes, may work for both categories of externalities, as discussed in Section 4 below. Private law solutions, in contrast, are ill-suited to redress harm to public goods. Some private remedies could potentially work when the externality targets specific and identified third parties. But when the operators of databases are shielded from liability, private remedies are ineffective.

2.3 Precaution and Insurance Externalities

The social impact of exposure to harm includes the cost of precautions. Some of data's external effects are preventable, but at a cost. These expenditures, too, have

16 See, e.g., May (2018) (cases of solved crimes), Lamott (2017) (reunited family).

17 See Crossland (2018).

18 Often, third-party apps allowed users to log in using their Facebook account. Prior to 2015, when an individual elected to use Facebook Login, they, often unwittingly, granted the app's developer a range of information from their Facebook profile—things such as their name, location, email or friends list. Facebook enabled this practice before suspending it in 2015, but third parties were not required to delete the previously collected data. See Constine (2015) and Lewis (2018).

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

a public good aspect. It is not surprising that pollution reduction is a public good. Victims of pollution are often part of a large pool of individuals who share common exposure, which in turn depends on the contributions of each member of the pool. In the environmental context, people who could deploy private emission-prevention measures fail to engage in the optimal levels, discounting the positive impact their effort has on other people. Indeed, one of the challenges of climate policy is to figure out how to induce free riders to comply with emissions targets (Ben-Shahar & Bradford 2012, p. 376).

A public good problem could arise even when prevention measures create only private and no external benefit—through an insurance externality. When consumers are protected from harm by insurance, they may take less care (the typical moral hazard problem).[19] This incentive problem becomes an externality when the cost of coverage for the inflated harm is spread across all members of the insurance pool. In environmental contexts, the cumulative exposure to pollution-caused illnesses, resulting from private emissions decisions, is spread across the entire pool of health insurance buyers.

Both the prevention and insurance externalities are present in the data pollution context. The insurance externality is particularly acute. When a security breach occurs and loads of sensitive personal data are released, people could suffer significant private harm in the form of identity theft, financial fraud, and having to make post-breach remediation efforts. But they are largely insured against these private fraud-related losses, through various statutory insurance programs[20] and covered for the residual loss through typical homeowners insurance policies.[21] The economic costs of data spills are significant,[22] but only a small fraction of it is borne by the consumers whose data is stolen.

Separately, companies subject to data breaches are often held hostage, demanded to "ransomeware" to their attackers. They are tempted to take the

19 Insurance contracts can mitigate and even overcome the moral hazard problem by creating incentives for care. See generally, Ben-Shahar & Logue (2012). I discuss below how some forms of insurance can substitute for public regulation of data pollution.

20 For explanation as to how fraud-related losses suffered by consumers are limited, see Pierce (2016, p. 982) (citing 15 U.S.C. §§ 1643(a), 1693(g), which limit the maximum amount of fraudulent charges that banks can pass along to cardholders, and concluding that "harm to consumers largely consists of inconvenience"). See generally, Weiss & Miller (2014) and Federal Trade Commission (2013) (outlining the protections of the Fair Credit Billing Act (FCBA) and the Electronic Fund Transfer Act (EFTA) in the event credit, ATM, or debit cards or data are lost or stolen).

21 Standard homeowners insurance policies cover unauthorized use of credit card or fund transfer, including forgery. See Insurance Information Institute (1999). Identity theft insurance is an optional endorsement available under a homeowner's policy, See, e.g., Liberty Mutual Insurance (2019).

22 See McAfee (2017). See also Center for Strategic and International Studies (2014)—McAfee Report ("a conservative estimate would be $375 billion in losses, while the maximum could be as much as $575 billion"); Norton Security (2012).

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

private action of paying ransom, not accounting for the external effects of making hackers more likely to more targets more often. Indeed, the FBI recommends against paying digital ransoms, yet companies continue to privately buy cyber insurance that covers such payments.[23]

Consumers pay indirectly for the protection they receive. They are insured, for example, against credit card fraud, but they pay higher fees and prices for credit card services and products, which operate as implicit insurance premiums for the data spill coverage. Critically, a consumer's cost is invariant to its own private precautions. A prudent consumer may decide to enroll in a service that provides better protection against data spills, but this added and sometimes costly precaution would not reduce the implicit insurance premium the consumer pays. The incentive to enroll in anti-spill precaution programs is crippled.

The public good aspect of data pollution prevention is evident not only in the context of data spills. In general, people entering data-intense environments online have some degree of reported anxiety over the impact from the potential exposure of their private information (Morey et al. 2015). But whatever caution this sentiment might arouse, it is defeated by the (correct) anticipation that, one way or another, their information would be exposed anyway, by the actions of others. If the same personal data profiles can be assembled from other sources—friends, service providers, predictive analytics—individuals will underinvest in data protection.

In sum, in this section, I began to assemble the argument that data's harm is, in an important part, public. It is public not merely in the derivative, secondary, sense that the privacy literature suggest—that deeply personal privacy injuries demoralize and degenerate the civic functioning of individuals and as a result impoverish public spheres and institutions.[24] Instead, the harm directly affects public ecosystems, and it is often unrelated to, nor channeled through, any impact on the specific individuals whose data are used. The digital economy creates digital smog; the question is what to do about it.

3. THE FAILURE OF PRIVATE LAW

Section 2 identified a problem—data pollution—that Sections 3 and 4 will now try to solve. I begin Section 3 by explaining what NOT to do—what regulatory approaches are not suited to deal with data pollution. It is a necessary first step, because it appears that much of the current regulatory response falls into this

23 See Kramer (2019).

24 See *supra* note 1.

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

category of ineffective law. Specifically, I argue that optimal control of data pollution cannot be achieved by private law or by enactments intended to prompt people to be more cautious about their data sharing practices. This discussion will subsequently be followed in Section 4 by a set of ideas on more effective solutions.

The failure of private law in the data pollution area is remarkable, because personal data rights are robustly defined by a manifold of statutes and are subject to intense and detailed private contracting. No less than an entire area of the law—data privacy law—is dedicated to the creation and enforcement of private rights in data. And the most common type of consumer contracts is the "privacy policy" of websites and apps, governing the transaction over the private rights in data (Bar-Gill, Ben-Shahar & Marotta-Wurgler 2017, pp. 25–30). When the baseline rights are so crystal clear and contracting over them is so explicit and rampant, why is private law failing? The reasons are explained below, and they are an exact replay of private law's failure to regulate industrial pollution.

3.1 Failure of Contracting

People care about their data ecosystem.[25] Usually, when consumers care about an attribute of a product, firms compete to provide it. We face a puzzle, then: why is data pollution not subject to preference-satisfying contracts? Why is the prevention of data emissions not bargained for?

The puzzle is heighted by the fact that much of the focus of data privacy law is to encourage parties to contract. The law is packed with statutes that allow firms to collect and use people's personal data only if they receive contractual permission.[26] A centerpiece of Europe's General Data Protection Regulation (GDPR) is the requirement that data gatherers give consumers more control over their personal data and enable them to restrict and to personalize its collection.

And indeed, many firms offer a menu of data control options to their customers. Some offer, for example, "premium" services in which customers may

25 Pollack & Sullivan (2018).

26 See, e.g., Electronic Communications Privacy Act, 18 U.S.C. §§ 2510–2522 (2018); Health Coverage Availability and Affordability Act of 1996, Pub. L. No. 104-191, 110 Stat. 1936; 45 C.F.R. § 164.502 (2018); Personal Information Protection Act, 815 Ill. Comp. Stat. 530/1 (2006); California Financial Information Privacy Act, Cal. Fin. Code. §§ 4050–4060 (prohibiting financial institutions from sharing or selling personally identifiable nonpublic information without obtaining a consumer's consent); California Online Privacy Protection Act, BPC § 2275.

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

pay with money instead of data.[27] Others offer "privacy consoles" that explain to consumers what data are collected and for what purposes, allowing consumers to turn off some of the data spigots.[28] Every website, app, or store has a "data policy" that explains to consumers what data are collected and how they are used. The market environment is sizzling with intensive contracting over data and with endless opportunity to protect personal data. Why, then, is there so much data pollution?

For the same reason that people do not contract enough over industrial pollution. Three primary market failures explain the shortcoming of markets in producing contracts with socially optimal levels of pollution: externalities, misinformation, and imperfect rationality. Because each of these factors has been richly discussed in the past to explain the failure of contracting over industrial pollution (and over public goods more generally), my focus below is to show how these factors apply to the data pollution context.

3.1.1 Externalities

Pollution harms people not party to the transaction. The production of meat, for example, emits toxic waste, and as long as these negative externalities are not felt by the producers or the meat eaters and not reflected in price, they are undercounted in purchasing decisions (Nesheim et al. 2015, Ch. 4; Kohn & Kruger 2016). Even during occasional and exceptional spikes in concern—when the environmental toxicity associated with the production of a particular product becomes so salient and disturbing that consumers shun the product—the reaction is rarely calibrated to reflect the magnitude of the harm.

I argued in Section 2 that emissions of data are like emissions of pollutants—the costs are often external. These externalities are the fundamental market failure that explains why private contracts are not the solution to data pollution. True, people are contracting all the time over data, but with complete indifference to the data pollution problem. They are given options to share less data—pay with money rather than with personal information—but rarely choose them, and rarely display any affirmative interest to bargain over data pollution. Legal default rules that prohibit companies from harvesting personal data have been thoroughly and methodically reversed—because consumers do not seem to care enough about the potential privacy harm and have no incentive to do much about the public harm.

27 See Bode (2016).

28 For example, Google's (2019) *Privacy Checkup* allows users to manage their data settings and limit the data activity Google tracks.

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

Indeed, exceptions help prove the rule; in the special cases when the harm from data emissions is *not* external, and when the privacy concerns are salient and acute, consumers are more inclined to contract into heightened reduction of data pollution. In the same way that consumers are careful not to buy kerosene heaters that emit pollutants inside their own homes (because the cost is primarily private), they are careful with their most personally sensitive data and demand greater security. If the personal data collected by a website is particularly embarrassing—for example, one's browsing preferences in adult websites—it is governed contractually by tighter data protection standards (Marotta-Wurgler 2016, p. 13). Likewise, cloud storage services that invite people to deposit their entire records for remote safekeeping implement tighter data security (*id.*, pp. 30–35). Here, when the harm is purely private, contracts are written to provide more effective data protection (and pollution reduction).

3.1.2 Information

People may fail to contract for optimal pollution emission even when the harm is internal, because of misinformation.[29] This problem, of course, extends well beyond emission of pollutants. Products harm people or perform poorly in a variety of ways that may become known only after their consumption. The trans-fat epidemic and the breast implant mass exposure are two well-known examples. Because many harmful effects (or warranted benefits) from products gestate slowly and manifest latently, uncertainty over the true causes leads to contracting failure.

Data security is a credence good. Consumers cannot know how lavish the data sharing and how loose the security practices are—until there is a sharing or security crisis. They are rarely aware what data are collected and by whom (Froomkin 2000, pp. 1501–1502). When their personal data are emitted, consumers often do not know if it is harmful, and indeed many emissions turn out to be harmless. When rating their experiences with the service provided by a firm, consumers rarely, if ever, consider the data handling practices of the firm, making it all the more difficult for others to base their contracting decisions on these factors. Even when consumers learn through experience about a harm caused by data, it is typically a private harm. People remain largely unaware of any public harmful effect.

In many markets, consumers overcome their lack of information by relying on informed intermediaries. People who care about environmental pollution could seek certifications and rating by the likes of ISO, and people who care

29 Froomkin (2015, pp. 1732–1737) describes people's "myopia" about the long term private harms they would suffer from sharing personal information, and their failure to recognize the true "average value" of their data to those who collect it.

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

about data emissions may similarly consult TRUSTe.[30] But such services provide only some information. They may tell consumers what is being collected and protected, but they are not able to identify the potential external harms. Also, they comingle many factors to generate the ratings, and consumers rarely know what weights each rating index gives to the various underlying factors. For example, some data privacy certifiers focus on rating the *promises* made by the data collectors, not their actual *practices*.[31] (It is easier to read and rate data policy statements posted by businesses than to monitor the actual protections implemented and the actual data sharing practices each follows.[32]) Without knowing how the ratings are generated, people could be lulled into a false sense of security. Rating services, in other words, are also credence goods.

Finally, when consumers are ill-informed about the overall riskiness of a product, it is less likely that competition among firms would lead to contracts that address the risk. It is more profitable to invest in salient features that create marketing advantages than to expensively bolster the hidden traits. Besides, firms do not tend to compete over risks that consumers undercount, even when if it is efficient to reduce these risks. A firm offering high-end data-pollution protection, for example, could be reluctant to brandish its advantage because highlighting such aspects could alert uninformed consumers to risks these consumers otherwise tend to ignore. Such alarm could chill consumers' demand for the entire class of products. The advantage to the low-emission firm in terms of increased market share could be more than offset by the disadvantage due to decreased market size.

3.1.3 Imperfect Rationality

Contracting over pollution fails for another reason: poor judgment. Environmental harms are the classic case of uncertain outcomes, where cognitive biases abound. Pondering pollution-related harm, people may be overly optimistic or overly pessimistic; they respond excessively to salient events and then gradually forget them; they discount future payoffs, but not along a systematic scale; they fall prey to framing manipulations; they are irrationally loyal

30 TRUSTe provides privacy risk assessments and certifications. See https://www.trustarc.com/products/enterprise-privacy-certification/.

31 For example, one of the most objective grading services is PrivacyGrade.org (2014), designed by Carnegie Mellon University. It measures "the gap between people's expectations of an app's behavior and the app's actual behavior" (*id.*). Yet some of the biggest data polluters get shining grades. Facebook's and Strava's apps score "A" grades—is that simply because people have such low privacy expectation from Facebook and Strave?

32 Many rating services do not perform audits of websites to ensure that the promises they make, or the rating standards, are being satisfied (see Nehf 2003, p. 65).

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

to status quos; they are averse to making any inquiry or decision, and more (Johnson & Levin 2009, p. 1597). Making a good choice that truly advances one's personal environmental goals is hard enough, often requiring subtle tradeoffs along multiple dimensions. It becomes insurmountable when the other side of the transaction is a sophisticated firm that recognizes the cognitive biases and amplifies them to profit from the individual's misperception.

Decisions over personal data face similar degrees of uncertainty and are similarly prone to imperfect rationality. Even more than chemical toxicity, digital risks are hard to ascertain (Nehf 2003, p. 62). There are no digital illnesses or deaths; the dimensions of the data emissions risks are many and complex, ripe for endless behavioral biases (Acquisti, Brandimarte & Loewenstein 2015, p. 509; Adjerid 2016). The manifestations of data-caused harms are sometimes subtle and easy to disregard, other times splashy and easy to exaggerate. Even if firms were to write data policies in clear and legible language (which they rarely do—and when they do, the texts are usually written at college level (Jenson & Potts 2004, p. 471), the underlying issues remain abstruse, confusing, and constantly shifting. Ironically, researchers have found that the very presence of a "privacy notice" document on a website soothes consumers' privacy worries and causes them to trust a website more (Pan & Zinkhan 2006, pp. 331–338). This, despite the fact that privacy policies of websites rarely carry any good news for consumers, they almost always reduce the protection relative to the default rules that would govern the transaction otherwise.

The difficulty in making rational, informed decisions regarding data pollution often drives consumers to ignore the data dimension altogether. Is this massive indifference phenomenon irrational? Or, against the background of insurmountable complexity, is inattention rational? Even if people wanted to contract smartly over data, to accord inquisitive attention to the management of personal information, they would be defeated by what elsewhere Carl Schneider and I called the "quantity problem": each website visit, app use, and even physical transaction presents consumers with its own overloaded set of data issues.[33] The problem of overload within each transaction and accumulation across multiple transactions are problems too implacable to solve in a world of private contracting. And, they are made exponentially more difficult by the fact that similar attention is required to address other daily contracting

33 See Ben-Shahar & Schneider (2014). According to one estimate, the average person encounters so many privacy disclosures every year that it would take 76 days to read them, and the lost time would cost the economy $781 billion. See also McDonald & Cranor (2008).

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

risks, some of which are much more urgent. In environments cluttered with layers upon layers of technical information, who is to say that ignorance and inattention are irrational?

Contracting fails, I conclude, and the solutions for this failure cannot come from within contract law. It might be thought that contracting failure could be corrected by "choice architecture"—namely, that behavioral economics could be the solution, not the problem. But these gentle solutions meet a formidable foe—the companies that benefit from people's data sharing gullibility. Ultimately, data sharing is done on platforms designed by parties that benefit from the data, whose interest is to counteract any legally required anti-sharing nudge. Anti-business default rules have proven futile in many contexts because they are just one click away from deletion—a click that businesses are eager to solicit from their customers (Radin 2013; Willis 2013; Ben-Shahar & Strahilevitz 2016).

Data-protective rules could be effective if mandatory, but this means (paradoxically) that the only way for contract law to overcome the contract failure would be to remove the matter from the bounds of permissible contracting all together. How to design such mandatory rules is the focus of Section 4 of this article. For private law to continue to have relevance in the environment of mandatory data pollution rules, victims must be granted enforcement powers. I therefore turn to examine why private enforcement of nondisclaimable anti-pollution rights in data fails.

3.2 Failure of Tort Law

Section 3.1 explained why contracts and markets fail to provide optimal levels of data pollution. But private law could overcome the market failure with other tools. It can render some data emissions actionable and rely on private enforcement to implement the commands. Data pollution could be, and often is, illegal—for example, when firms harvest personal information from people without their consent, use it in impermissible ways, negligently fail to secure it, or engage in deceptive data practices. All of these violate people's rights over their personal information, and could be redressed by tort law.

But tort law fails to address these wrongs, for the same reason that it historically has failed to redress many environmental wrongs. In theory, nuisance law has been available to deal with industrial pollution. But, as widely recognized, it has not succeeded (Abraham 2008, p. 149). Tort law has failed to deter and compensate industrial pollution harms for three primary reasons: causation, valuation, and societal externalities. I argue that these reasons are equally central in tort law's failure to control data pollution.

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

3.2.1 Causation

Tort law is effective when harm is immediate and visible. Pollution harm is neither. It is not immediate: environmental liability suffers from an acute problem of "long tail"—latent harms that are difficult to causally match with precise wrongs (Rosenberg 1984, p. 919; Dewees et al. 1996, pp. 293–294; Esty 2004, p. 131). And it is not visible: neighbors to pollution can show that they are exposed to a new *risk*, but have difficulty showing that they suffered actual *harm* (Dewees 1992, p. 429).

Episodes of data emissions are often afflicted with a similar problem of uncertainty over causation. Consider security breaches in which financial data of millions of consumers are taken due to negligent safekeeping by a website.[34] No doubt, private harm would accrue to specific individuals once this information ends up in the hands of identity thieves. But who within the data pool will be the actual victims? Courts—and even the victims themselves—may never have the necessary information. The immediate post-emission lawsuits are usually filed before any actual victims are identified (and indeed these suits often claim—largely unsuccessfully—damages primarily for the increased *risk*).[35] It could take years for the misuse of the data to occur, and by then it would be hard to attribute the harm to any specific data spill. No single source of data emission would be "more likely than not" to have caused the harm; many of the emission episodes will have been forgotten by the time they gestate.

The slow manifestation and the uncertainty over causal links defeat any attempt to apply a negligence-based regime. But they also blunt some of the more ambitious proposals to expand the reach of tort law into the data pollution area. It is sometimes thought, for example, that a shift from negligence to a strict liability regime would make firms more accountable for data emissions. Not having to prove emitters' negligence, victims would more easily collect tort compensation, which in turn "would force database operators to internalize the full costs of their activities" (Keats Citron 2007, pp. 241, 266). Unfortunately, strict tort liability still requires proof of causation. If it is difficult to identify the causal chain connecting particular data emission to specific victims, the desired deterrent and activity-regulating effects of strict liability

34 Data emissions differ from environmental emissions due to the existence of intentional hacking as the primary cause. The responsibility of companies for the release is thus secondary. Nevertheless, the collection and storage of sensitive data without adequate anti-hacking protection could be regarded as negligent in a manner analogous to the inadvertent preventable releases of environmental pollutants.

35 See, e.g., *Indep. Cmt. Bankers of Am. V. Equifax, Inc.* 1:17-cv-04756-MHC (N.D. Ga. Feb 20, 2019). See also Koo (2017).

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

would not occur. It is the inadequate proof of harm, not of negligence, that precludes tort liability.

Tort law could, in principle, compensate victims for *exposure*, rather than harm—assuming information is available about the general toxicity of an emission. People exposed to data pollution would be compensated for the risk, not the actual injury. But the information necessary for such a scheme is often unavailable in litigation because the harms caused are often latent (Viscusi 2000, p. xi; Schroeder 2002, p. 601; Lin 2005, p. 1452). If the data were handy—and perhaps in the context of data security breach they are, since those whose data are stolen face a known risk of identity theft—statistical evidence could be called upon to assess the aggregate injury to the class of affected consumers, and award fractional damages to each member of the class. Such a scheme, if based on good actuarial information, could provide optimal deterrence (Shavell 1987, pp. 115–118). For example, the Justice Department estimates that an average loss to victims of identity theft is $1,500.[36] To adjudicate a tort lawsuit for security breach by a website, a court would need survey evidence to estimate the increased likelihood of identity theft to the average member of the affected pool. With that, a remedy to the entire class could be crafted.

But for the same reasons that such exposure-damages claims failed in pollution lawsuits,[37] they are unlikely to succeed in data pollution lawsuits. Plaintiffs have been making such claims in data spill lawsuits—but with little success.[38] A tort remedy based on *expected harm* is exceedingly uncommon in courts,[39] and found more often in the remedial arsenal of public law (e.g., fines for speeding). And forward-looking injunctive remedies have little value for private litigants, and are more often pursued by agencies like the Federal Trade Commission (FTC) or state Attorneys General. Indeed, developing a scheme of exposure-based remedies injunctive relief for data emissions is a prominent

36 See Harrell & Langton (2017).

37 For further explanation, see Brennan (1988, pp. 491–493) ("Courts are troubled by the probabilistic evidence of causation with regard to hazardous substance injury . . . Courts rely on mechanistic notions of causation and are confused by probabilistic ones."). But see *Norfolk & Western Railway Company v. Ayres*, 538 U.S. 135 (2003).

38 Despite it being central to the issue of standing to sue in federal court, courts have reached inconsistent conclusions on the issue of harm and injury-in-fact in data breach cases. Plaintiffs argue that data breach "creates a risk of future injury, such as identity theft, fraud, or damaged reputations," and that they experience anxiety about this risk. See Solove & Keats Citron (2018). For analysis of courts' treatment of claims alleging increased risk of future harm, see Silverman (2017, p. 226).

39 Some courts refuse to accept any statistical evidence in tort suits. See Rosenberg (1984, p. 857). See e.g., *Smith v. Rapid Transit, Inc.*, 58 N.E.2d 754 (Mass. 1945) (holding that statistical evidence alone cannot prove bus company's causal role). See also Gelpe & Dan Tarlock (1974, p. 374).

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

motivation of Section 4 of this article—focusing on public law solutions to data pollution.

3.2.2 Valuation

A second problem with tort liability for pollution is valuation. Even when the impact of the emissions is proven, it is often qualitative and difficult to measure in dollars. In the area of environmental harms, problems of valuation forced tort law to deploy arbitrary exclusions, based on proof of loss.[40] Strong *physical* manifestations of injury make it possible to compensate some losses because they are easier to value. Valuation problems could also be overcome if the remedy is tailored toward *restoration*, rather than compensation (Ben-Shahar & Porat 2018, p. 1901). And even when some of the losses from pollution are quantifiable (e.g., loss of fishermen income due to oil spill), other major losses arise from the deterioration of the surrounding ecosystem, where the loss is more speculative.

The problem how to measure the injury is even more perplexing in the data pollution context. People say that data safety is important to them, but often behave as though it is not—the privacy paradox.[41] Should tort law compensate them on the basis of what they say, or what they do? This problem of private valuation is due to the deep uncertainty people have about the private consequences of personal data emissions—who will use it and how, and what would be the consequences of unauthorized uses. Even when the injury is traceable—like identity theft resulting from a particular data emission—the perception of financial harm could be drastically different from reality.

Data emissions lawsuits regularly confront the difficulty of demonstrating ascertainable injury. In a typical data security breach case, plaintiffs allege emotional harm as well as risk of future private harm posed by the spill, but many courts hold that such injury is too speculative to be compensated, and deny standing to sue (Solove & Schwartz 2017, pp. 960–962).[42] Even costs incurred by victims of data breach to monitor their financial information were considered insufficient to establish standing, "because costs incurred to watch for a speculative chain of future events based on hypothetical future criminal acts are no more 'actual' injuries than the alleged 'increased risk of injury.'"[43]

40 Courts use standards of proof to exclude some harms. See American Law Institute (1991, pp. 319–321) (surveying the relatively low total damages awarded for environmental injuries).

41 See *supra* note 2.

42 See, e.g., *Beck v. McDonald*, 949 F.3d 262 (4th Cir. 2017); *Amburgy v. Express Scripts, Inc.*, 671 F. Supp. 2d 1046 (E.D. Mo 2009).

43 *Reilly v. Ceridian Corporation*, 664 F. 3d 38 (3d Cir. 2011).

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

The difficulty of evaluating individual harms and distributing monetary compensation to victims could be set aside if the goal of tort law is to deter, rather than compensate. The polluter can be made to pay, even if the victims cannot collect. Such decoupling of liability and compensation could be achieved, for example, by *cy pres* settlements, whereby the court directs nondistributable portions of class-action settlements to third-party beneficiaries that work to advance the interests of the class.[44] But such methods are the exception, possibly a short-lived exception. They are thought to impermissibly push the boundaries of courts' constitutional authority in adjudication private law claims.[45] Indeed, it is precisely in the context of a claim over data pollution that the Supreme Court has been examining the legality of such private enforcement model.[46]

At the core, the problem of valuation is due to the external, societal impact of data pollution. The harms to various public goods, discussed in Section 2, are difficult to translate into the monetary redress coinage of private law. It is not clear which individuals should bring the complaints, what the concrete injury is, and it ultimately is hard to assess the total harm.

3.2.3 Societal Harm

A third major obstacle for regulation of data pollution by tort law is the broad societal reach of the ensuing harms. The existence of societal externalities is a key factor in my explanation as to why contracting over data pollution fails to provide optimal arrangements. In general, externalities do not doom tort law to fail—on the contrary, tort law is a primary social device to internalize negative externalities. But pollution creates a type of externality too widespread for tort law to control.[47]

In the environmental context, harms to air, public lands, or public water do not give rise to a robust response in the form of private compensation. True, tort law actions are not completely shackled: private and public nuisance doctrines, the public trust doctrine, and *cy pres* settlements permit tort recovery for

44 *Nachshin v. AOL, LLC*, 663 F.3d 1034, 1038 (9th Cir. 2011); American Law Institute (2010); Barnett (1987) (giving case examples).

45 The questionable constitutional foundations of *cy pres* distributions was acknowledged by Chief Justice John Roberts, noting "fundamental concerns surrounding the use of such remedies in class action litigation." *Marek v. Lane*, 134 S. Ct. 8, 9 (2013), *cert. denied* (No. 13–136).

46 See, e.g., *Frank v. Gaos*, 139 S. Ct. 1041, No. 15-15858. In that case, Google is being sued for sharing user search terms with third parties.

47 In the industrial context, externalities like pollution were the primary reason that enforcement of environmental law has shifted from to the torts system to public law. See Abraham (2002, p. 379, n. 2) and Butler & Macey (1996, p. 29).

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

societal harms.[48] Additionally, scholars have suggested innovative ways to expand tort law's private harm model to societal injuries.[49] These anecdotal expansions notwithstanding, tort law still limits private claimants to sue solely for private harms (Swanson & Hughes 1990; Dewees 1992, p. 428). To collect recovery for public nuisance under the public trust doctrine, for example, public enforcement is still necessary (Lin 2012, p. 1093). In environmental contexts, a tort-like remedy for natural resource damages yields large sums of compensation to restore injured natural resources, but it is only available to public agencies under the public trust doctrine (Bradshaw 2016). Generally, it is widely recognized that "the law of nuisance was not up to the task of protecting the environment" (Dewees 1992, pp. 428–429; Abraham 2008, p. 149).[50]

Similar to environmental pollution, data emissions create societal harms. These are the negative externalities discussed in Section 2—harms arising from databases, or from the public good aspects of the digital data enterprise. The harm to the integrity of the American elections from Facebook's data practices was a purely public harm—affecting a political ecosystem rather than any single user. What tort remedy could capture it? The harm to victims of financial data leaks is largely the expanded sense of insecurity—again, a type of injury that tort law does not readily remedy, and that courts have repeatedly rejected. And the harm from stereotyping and discrimination that people with black-sounding names experience when prison-related information attaches to search results of their names is so profoundly societal that it would be hard to imagine how it could be vindicated through tort actions. Like natural resource damages, a compensatory framework for data pollution would have to rely on public enforcement actions.

3.3 Failure of Mandated Disclosures

Tucked amidst these two pillars of private law—contracts and torts—are numerous public law enactments intended to help people self-protect against data misuse. Many federal and state statutes require companies that collect and process personal data to disclose their practices to consumers. Such disclosure mandates rest on the ubiquitous but unrealistic hope that people would then be able to give "informed consent" to the practices. For example, the Video Privacy

48 For a discussion of the history and application of the public trust and public nuisance doctrines as they relate to environmental protection, see Lin (2012).

49 For example, a new measure of damages—"societal damages"—could be awarded as part of a private tort suit to non-plaintiffs, to compensate victims of the same wrongdoing who are not before the court, or to the advancement of societal interests impaired by the wrongdoing. See Sharkey (2003).

50 Public nuisance doctrine was similarly ineffective in protecting the environment of Great Britain in the nineteenth century.

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

Protection Act prohibits service providers from sharing customers' personal data without written consent (imposing a penalty of $2,500 for each violation), with the result that prominent disclosures are meticulously attached to all membership contracts.[51]

Similarly, mandated disclosures are the primary response to data security breaches. Once such leakages occur, affected users are informed, with the hope that they would then be able to take precautions and mitigate the harm. California, for example, requires that the disclosure shall be made "in the most expedient time," shall be titled "Notice of Data Breach," and include clearly labeled sections like "What Happened," "What Information Was Involved," "What We Are Doing," and "What You Can Do," in a format "designed to call attention to the nature and significance of the information it contains."[52]

Mandated disclosure is without doubt the dominant regulatory approach in American data privacy law.[53] While it is a public form of regulation on its own, mandated disclosure is also widely regarded as a precondition for private contracting and private controls, and the violation of disclosure requirements is often a tort.

There is no evidence that mandated disclosures of data practices affect people's conduct in the data sphere, or that it renders their consent to the practices more informed. In fact, there is ample evidence that it does not achieve any of those goals (Ben-Shahar & Schneider 2014, p. 69; Ben-Shahar & Chilton 2016). The notice requirements fail because they primarily seek to harness the two mechanisms that, I argued above, are prone to fail in protecting against emissions. The requirement of informed consent harnesses protection-via-contract, aspiring to help people secure better consensual arrangements. And the requirement of post-breach notification harnesses tort law, letting people know when exposure occurred, prompting them to engage in precautions and to seek compensation. But consumers are not entering better contracts. And no matter how expedient post-breach notices, there are few if any precautions they can take in response to the notification of a breach, and they are largely unsuccessful in post-breach tort suits.

The failure of data pollution disclosures is not surprising. The same technique when applied to environmental emissions has not inspired great hope. California Proposition 65, for example, which requires advance warning about carcinogens, had been widely criticized for its many costs and few benefits

51 18 U.S.C. § 2710.

52 Cal. Civ. Code § 1798.29, 1798.82; see also, Cal. SB-46, Ch. 396.

53 Principles of the Law, Data Privacy, §3-4, American Law Institute (2019).

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

(Barnhill 1989; Barsa 1997, p. 1248). Public disclosures of toxic releases may support the clean-up response by public authorities, but the thought that these notices actually reduce spills or that they are necessary for post-spill mitigation is, as others have put it, "overstated" (Bui 2005; Bae et al. 2010).

A widely held but naïve belief supposes that if only the disclosure were simplified and targeted better, they could succeed in helping people make better choices. If disclosures are too long, shorten them. If too technical, make them more user-friendly. If poorly presented, improve the formatting. Much of regulatory effort in the area of data protection is thus focused on encouraging "best practices" in the presentation of data practices.[54] But the results are disappointing. When decisions are complex, simplification of formats cannot have a meaningful effect on people's understanding of the underlying tradeoffs. Moreover, if the harmful effects result from the collective behavior of all participants, and they impact an entire ecosystem, why bother to read even the simplest of disclosures?

4. PUBLIC REGULATION OF DATA POLLUTION

The pollution model as applied to the harms caused by data emissions is a powerful framework that I relied on in Section 3 to explain why private law is not suitable to solve the problem that Section 2 identified—data's external harm. Can the pollution model be equally instructive in pointing to public law solutions? Can it borrow the environmental regulatory framework to begin constructing data pollution law? This section explores this challenge.

Upon first reflection, it might seem that the central techniques used in environmental law to regulate pollution are not readily applicable in the data ecology. There are crucial differences between physical and digital pollution. First, physical pollution can often be cleaned up; digital pollution probably cannot. A "superfund" for data spills might therefore make little sense.[55] Second, the effects of environmental pollution are always negative (even if the underlying activity causing it is beneficial), whereas data emissions could actually be good—data create enormous positive externalities as well. Environmental law thus bans substances too toxic for people to use, a technique inapplicable to data. Third, environmental impacts can be measured scientifically as a basis for cost–benefit analysis, whereas data externalities are often

54 See, for example, Federal Trade Commission (2012), National Telecommunications and Information Administration (2013), and White House (2012).

55 In reality, clean-up is often inadequate even in environmental contexts. Pollution that gets to the groundwater, travels far away, or degrades air quality, cannot be cleaned up.

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

qualitative and conjectural. What is the cost of a distorted Presidential election or of discriminatory racial profiling?

These differences might suggest that a replication of the regulatory response to environmental pollution in the data sphere—for example, by simply establishing a new EPA-like agency for data protection (DPA) and granting it analogous powers to combat data pollution—would not work. But despite the superficial differences, this section is dedicated to finding instructive clues in environmental law on how to design social policies to deal with data's social harm. In fact, the tools of environmental law are merely concrete applications of more general regulatory techniques that deal with any kind of externalities. Combining these abstract techniques with the specific regulatory framework used to control industrial pollution provides an organizing paradigm for public regulation of data pollution.[56]

In some ways, what this section does is not novel. There are flickers of public enforcement actions in the data pollution area, by agencies authorized to regulate some consequences of data emissions. The FTC, for example, has long been active in enforcing data rights and has recently taken action against Facebook's fake-ads data pollution. But the FTC's actions have largely focused on deception and unfairness, which are largely muted if polluting companies like Facebook do not breach their posted practices. Likewise, agencies and public prosecutors sometimes investigate the more egregious data spills, but their mandates are largely limited to a narrow class of wrongs, like delays in sending notices of a data breach.[57]

A public enforcement model of data pollution is not novel for another reason—it is an important complement to private enforcement in the protection of privacy. In fact, the European Union has an elaborate public enforcement branch of privacy law.[58] The European approach, discussed below, implements a set of principles known as Fair Information Practices (Schwartz 2013, pp. 1974–1975). These principles contain various prohibitions on data collection, use, and transfer, through requirements like necessity and purpose limitation and through bans on aggregation of databases.

56 As mentioned in the Introduction, prior work proposed an adaptation of environmental regulatory tools to address data privacy concerns. For example, Hirsch (2006) examines regulatory tools that would encourage regulated parties to reduce the harmful effect of their data practices. This article goes beyond Hirsch's illuminating analysis by considering harms not related to privacy, and by focusing on different regulatory tools.

57 See, e.g., Robinson (2018).

58 The EU enacted two data protection mandates. The first is the European Directive on Data Protection, and the second is the GDPR.

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

Public enforcement templates exist in data law. But they have been preoccupied with the concern for individual privacy, helping people gain more control over their own personal data. These templates have not addressed data's external harm. Recognizing that data pollution is also a public problem degrading an entire ecosystem and not merely the individual spheres of the data givers, offers a new and rich perspective on the existing solutions—and introduces new ones.

This section organizes the arsenal of public law's countermeasures to combat external harms into three distinct families of regulatory devices, mirroring the three primary techniques utilized by environmental law. The first is command-and-control regulation—imposing strict limits on the polluting activity. The second is a data tax—a Pigouvian solution to the externality problem. And the third approach is the design of liability for data spills that could provide optimal deterrence and compensation.

4.1 Command-and-Control Regulation

The primary regulatory technique in controlling environmental pollution is to prohibit harmful emitting activity beyond legally set limits—primarily by prescribing quantity restrictions, requiring permits, or mandating better technology. These ex ante forms of regulation are the archetypical command-and-control methods, and they are usually effective in obtaining the restrictive results, but often at substantial, and sometimes unintended, costs. They can be tailored to combat data pollution by restricting which data firms may collect, for what purposes they might be used, or how they may be stored, shared, or transferred. Such regulatory controls would aim to identify the risks and reduce the harmful effects of databases.

At the outset, a conceptual problem must be addressed. Environmental law does not usually regulate production inputs as much as it focuses on the outputs emitted. Factories are generally free to use any input, as long as they comply with emissions restrictions. For example, under the National Ambient Air Quality Standards of the Clean Air Act, the EPA set caps on how many parts per million of a pollutant can be emitted.[59] Data, it might be thought, cannot be sorted along the input/emission divide—the information inputted is the same that potentially may be emitted. Quantity and activity restrictions, therefore, would have to apply to the input side of digital production and limit the data the companies may collect.

Although data are not toxic in the same way that industrial pollutants are, the environmental analogy continues to hold. Even notorious industrial pollutants have ancillary benefits (Graham & Baert Weiner 1997; Revesz & Livermore

59 42 U.S.C. § 7401; 40 C.F.R. 50.

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

2011). Asbestos, for example, can improve building insulation and reduce fire hazards, and carbon dioxide emissions facilitate agricultural production in cold areas like Siberia. Environmental law takes positive externalities attributed to pollutants can be taken into account in the cost–benefit analysis. Data's externalities are similarly bi-directional. Even when emitted (and used for purposes beyond those for which they are primarily collected), data create benefits. For example, Google Trends—a service that uses Google's search data for purposes different than those for which it was collected and stored—provides valuable clues about social phenomena such as the spread of medical and social ills (Jun, Yoo, & Choi 2018). Similarly, databases assembled by genetic testing services both help and harm nonmembers. Thus, the key challenge for the command-and-control approach to data pollution is to determine in advance which uses of the data are net socially harmful and ought to be restricted. Can the law rise to this towering challenge?

The European Union thinks it can. In particular, various quantity restrictions are a key part of the GDPR. Prominent among those are the principles of "data minimization" and "purpose limitation." The Regulation requires that data be "processed fairly" and only for "specified, explicit, and legitimate purposes" and even then the collected data must be "adequate, relevant, and not excessive in relation to the purpose or purposes for which they are processed."[60] Retail stores, for example, may collect personal information about the products their customers buy so as to personalize the offering and improve the shopping experience, and they may also collect information about payment methods so as to speed up the checkout process. But under the data minimization restriction, they would not be permitted to collect drivers' license data or information about their shoppers' social contacts, and they must delete information about people who have deactivated their accounts.[61] Unless used for personalized service, personal data should be anonymized or aggregated.

Quantity restrictions not only regulate the collection and storage of data, but also their processing and various uses. Currently, one primary use of databases is selling or renting third-party access to them for a variety of purposes. Regulatory restrictions might be employed to disallow or at least limit these transfers. Facebook's transfer of data to Cambridge Analytica is the type of use that could be restricted. The regulation would have to establish categories of circumstances under which data transfer may not occur. It could also

60 GDPR, Art. 5.

61 For an example of a data minimization restriction law limiting data retention, see New York's Cybersecurity Requirement for Financial Services Companies, 23 NYCRR 500.13 (2018).

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

implement "data localization" standards—the limits on shipping databases for storage or use in other countries.[62]

The challenge for principles like data minimization and purpose limitation is to determine what constitutes "fair" and "legitimate" purposes, and what counts as "adequate, relevant and not excessive." Applying these principles to privacy harms—as the GDPR does—is hard enough; the difficulty is compounded when applied to external harms. In the privacy context, these restrictions aim to restore people's control over their personal information. In the data pollution context, the commands must be justified by anticipating the net external effects of a database. This is a staggering challenge: Big Data allows investigations that reveal connections not previously known or anticipated. Who could have thought that a database of Internet searches would provide clues for the detection of major epidemics? (Ginsberg et al. 2009). Finding correlations within data has enormous upsides, and limiting the use of a database only to known and anticipated purposes runs the risk of critically stifling innovation.

Perhaps the solution is to apply quantity restrictions like the GDPR's only to "sensitive" data or uses. Environmental and natural resource laws focus their restrictions on the most toxic pollutants and the most vulnerable habitats. Similarly, data pollution law could set its sights on restricting collection and processing of information that, if used irresponsibly, would have the most damaging, socially toxic, impact. Arguably, some of the greatest social harms could come from data models that undermine basic constitutional protections and defeat the goals of antidiscrimination laws. Heightened restrictions may target the collection and processing of sensitive personal data like race and ethnic origin, religious or political beliefs, and various types of information about health and sexual preferences.[63]

Limitations on how sensitive data are collected and used are, however, a double edge sword: they shield protected groups from potential harms, but they also deny them potential benefits. The value of learning from Big Data clues about the spread of epidemics among underprivileged groups or about discriminatory patterns in crime and law enforcement could be large It is not until such discoveries are excavated from the data that their value becomes known. If data restrictions slow down the creation of new knowledge, their heightened application to protected groups could have the unintended effect of disproportionately harming these groups. Because data produce both positive

62 GDPR, Art. 5.

63 GDPR, Art. 9.

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

and negative externalities, command-and-control restrictions that target the latter would inevitably sweep the former.

Another possible way to mitigate the stifling effect of across-the-board quantity regulations is through a system of case-by-case permits, as done in environmental law. Under the Clean Water Act,[64] any discharge of specific pollutants to waters must comply with explicit permits. Permits could be required for particular data activities which pose greater risks. If a website wants to run an algorithm that collects and uses, for example, information about race, it would be required to secure a permit, issued only if the website can justify its use of the data and demonstrate that it is harmless to the protected group.

A permit regime has the advantage of better information: it fine-tunes the restrictions based on each data collector's goals and circumstances, as well as the particular potential harms that the applicant's database may cause. It is well-suited to solve problems after they manifest, like political uses of Facebook's database. And a permit regime could be designed to elicit information that can be used by the regulators. In the same way that would-be environmental polluters are required to submit environmental impact statements to identify the potential harms and costs,[65] would-be data polluters seeking data permits would have to provide more information about their purposes and practices, exposing the harms they might impose.[66]

Regulation by permits is one of the most intense and costly forms of command-and-control, and it has many proven drawbacks. First, the administrative burden of reviewing each data service through an IRB-like system is daunting, and it would have a chilling effect on the underlying regulated activity. Second, a licensing agency that is asked to balance the risks and benefits has the tendency to engage in over-protection (harms from over-denial are less salient). Third, instead of being overprotective, a licensing agency could focus primarily on formalistic task, like conditioning the permits on firms obtaining users' "meaningful" consent. This is largely what IRBs do, and it has never been proven an effective regulatory safeguard (Schneider 1998, Ch. 4). It is particularly futile in protecting against externalities.

If not through permits, command-and-control regulation could operate by focusing on the technology that firms use in their data practices. Environmental law controls emissions by forcing cleaner technology. Operators of plants that

64 33 U.S.C. § 1342.

65 National Environmental Policy Act of 1969, 42 U.S.C. §§ 4321–4347.

66 Froomkin (2015, pp. 1745–1747) proposes a requirement of "Privacy Impact Notice" modeled on existing NEPA requirements, arguing that it would "create the conditions for a more informed debate". Unlike the analysis in this Article, Froomkin views the problem as measured by the impact on individual privacy, not as a pollution-like harm to the ecosystem. See generally, Calo (2013).

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

emit air pollution are required to install the "best available control technology" to achieve the "lowest achievable emission rate."[67] In the data realm, regulations could require companies to adopt data processing and security technologies with desired attributes.[68] This could fit well with two central data pollution concerns—transparency and security. Algorithms used for personalized services could be asked to meet transparency standards that would enable watchdogs and lawmakers to scrutinize the process. Similarly, concerns with data security could be addressed through "best available technology" rules.

Environmental law recognizes the inefficiency and innovation-chilling effects of quantity regulations, and sometimes addresses these problems through a system of trading. By limiting the quantity or requiring permits, emissions are capped; and by allowing trade of allowances and permits, the highest value activities take precedence. Efficient production is further achieved by cap-and-trade because it enhances the incentive for polluters to improve their pollution control method (Congressional Budget Office 2001; Stavins 2003). Could data emissions restrictions be subject to trade?

Probably not. Cap-and-Trade succeeded in controlling air pollution because a specific group of emitters—utility power plants—were identified, and each received a complex but well-specified initial allowance of a unique pollutant (sulfur dioxide) (Burtraw & Szambelan 2009, pp. 9–40). Who are the utility power plants of the data economy, and what are the sulfur dioxides emitted by their production? Unlike the production of electricity, which is done by a few major plants emitting well-known pollutants, production of digital services can be done by virtually any company. If entry into digital production is almost costless, how could the quantity be controlled? Moreover, the principles of data minimization and purpose limitation that underlie the data collection caps are not easy to particularize and quantify to generate bright line allowances ready for trade.

The problem for data cap-and-trade is more fundamental than merely defining a data endowment. Quantity restrictions seek to limit the accumulation of databases that reveal too much, that give too much power, that allow too much manipulation, and that increase the risk of misuse. It is the compilation of various layers of data that creates social impact (both good and bad), suggesting that trade could be potentially harmful. If the principle of data minimization forbids, for example, a retailer from collecting drivers' license data or amassing

67 42 U.S.C.S. §§ 4321–4347.

68 Hirsch (2006, p. 37) proposes a policy to require data-gatherers to come up with their own cost-effective approaches to achieving emission goals and "allow these self-directed actions to count towards regulatory compliance."

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

personal data of anyone other than account holders, it would be self-defeating to allow the retailer to purchase this data from someone else. A quantity-restrictive law may permit service A to collect only data X and service B to collect only data Y because adequate partitioning prevents some perceived social harm. But if A and B can trade the allowances (or merge), a single firm might end up with both data X and Y, defeating the intended protections. When massive data compilation is a principal source of data pollution, a trading system does not address it.

If databases are more likely to pollute the larger they are, data pollution provides a rationale for limiting the growth of data giants. Currently, the leading concerns surrounding mega-data companies like Facebook and Google are market power and potential anticompetitive behaviors. But if large platforms are more likely to cause disproportionately greater external harm, size limits are justified even without demonstrating abuse of market power. Targeting big companies may well be a good first step within a command-and-control scheme. It would avoid a problem already observed under the GDPR, that subjecting both small and large firms to the same set of data restrictions disproportionately burdens the small entities, for whom the fixed costs of compliance are overwhelming.[69]

While there are ways to pinpoint quantity restrictions narrowly, the discussion in this section suggests that it would be hard to control data in a satisfying manner through commands that prohibit specific substantive uses of the data, without also suffocating productive data collection (Hirsch 2006, pp. 33–37). Under the privacy paradigm that currently underlies data regulation, the harms from restrictive regulation are largely avoided because data platform may continue much of their practices as long as they give users more "control." Data pollution law's restrictions could not be supplanted by users' consent or control, because the harms impact other people, rather than the users themselves. Mandatory limits on data practices could crush important, still-undiscovered benefits that data creates.

4.2 Data Tax

Could pollution be restricted without the administrative burden and the innovation-chilling effects of command-and-control regulations? In theory, yes: by using price instead of quantity as the regulatory target. A prominent method to control pollution is to price it. The external harm is internalized by a

69 *How Facebook and Google Could Benefit From the G.D.P.R., Europe's New Privacy Law, New York Times* (April 23, 2018), at https://www.nytimes.com/2018/04/23/technology/privacy-regulation-facebook-google.html; https://www.nytimes.com/2018/04/23/technology/privacy-regulation-facebook-google.html.

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

"Pigouvian tax" either directly on the activity or on the specific product that fuels the activity and is responsible for the pollution.

In industrial production, carbon is responsible for much pollution and a carbon tax is thus the best candidate for a Pigouvian tax, widely regarded as an efficient method to regulate pollution (Metcalf & Weisbach 2009, p. 500). In the digital economy, data are the fuel that generate the activity and all its beneficial value, but also the potential harm. The external harm can thus be internalized through a data tax.[70]

The most natural occasion to levy the tax is at the time of data collection. Consider a purchase transaction between a consumer and a retailer. When a consumer buys a product in cash at a physical Walmart store no personal data is collected. But when the same consumer purchases the same product for the same price at Walmart.com, a rich data component is bundled into the transaction. The website collects and stores information about the consumer, including browsing interests, payment information, and potentially loads of snoopier data harvested from the consumer's device.[71] Indeed, in the occasion of this data "exchange" (and in large part due to it), the Walmart.com transaction is subject to standard contract terms uniquely tailored for the online environment, which would not be adopted as part of the physical store transaction over the same goods. If a long contract can be affixed to the digital transaction, so could a small tax.

How might the tax rate be set? Carbon tax seeks to approximate the social cost of carbon; similarly, a data tax would seek to approximate the social cost of data. But the analogy stops there, because the conceptual and practical differences are striking. The social cost of carbon, while at times highly uncertain and disputed, is at least based on rigorous measurements.[72] Society can deploy ballpark estimates concerning emissions levels, statistical links, and magnitudes of harms. The social cost of data is harder to measure. Until harms occur, it might be impossible to predict which data practices would be harmful, let alone their severely.

70 A data tax should be distinguished from an emissions fee for email spam, which is levied not on the collection and build of data inventories, but on a particular use of it. See Hirsch (2006, pp. 42–48), Mossoff (2004), Zhang (2005, p. 304).

71 Walmart.com (2017), for example, collects IP address, location data, the type of hardware and software the consumer uses to complete the transaction, prior browsing history. By planting cookies and beacons Walmart.com can collect further information about future browsing, regardless of whether an email was opened or if ad was effective.

72 The models used to develop social cost estimates do not currently include all of the important physical, ecological, and economic impacts because of a lack of precise information. See Environmental Protection Agency (2017). See also Johnston (2016).

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

Moreover, unlike carbon, which creates mostly negative externalities, data have many positive social effects. A data tax designed to align the private and social costs of a data service would have to be scaled down to reflect the positive externalities. Note, however, that while data have large unexpected positive benefits, many of them are not externalities and thus do not have to be deducted from the data tax. Database owners have the incentive to capture and monetize the positive externalities by selling personalized access to these benefits. Not so with the negative effects, emitters have no incentive to "capture" those. Government intervention is needed as a counterweight to this asymmetry. Still, it is possible that even with such one-sided incentives, many benefits are too diffuse to capture and a net positive externality remains. Information is a public good, and if the owners of databases do not capture the benefits, they might underinvest in their creation. In general, the net social value of data emissions is not zero, suggesting that if administrative costs are not too large, some form of data tax or subsidy is justified.

This is not the place to develop a comprehensive framework for the design of data tax. It may be that practical constraints could render it impossible to even roughly identify the social costs of data required to calculate the appropriate financial surcharge over digital data collection and production. Not to mention that political interests would compound the already daunting conceptual problems. Still, it might be wise to institute, as a baseline, at least a "small" data tax on large databases. Even a nominal tax would force firms to carefully rethink the necessity of collecting certain data.

In general, firms can assess the potential benefits of data more accurately than the government, whereas the government is (perhaps) more sensitive and attentive to the potential harms. In the current zero-tax regime, there is no reason for firms to scale their data activity to the perceived benefits, and no reason to stop short of "data maximization"—of collecting all possible information. In contrast, in a command-and-control regime, the opposite problem arises: government would be called upon to assess not only the harms but also the potential benefits from data, despite not having the necessary information. A "small tax" regime harnesses firms' private information about benefits. And if the government has some crude assessment of concrete risks associated with some data collection, it could adjust the tax accordingly.

A data tax could reflect both the quantity and the quality of the information collected. Obviously, the more information a firm collects about more people, the greater the tax. The marginal tax curve need not be linear: it ought to reflect the marginal risk of additional data. It is possible, for example, that the tax rate would be higher the more data is collected, to reflect the heightened social concern (including competitive concerns) with mega databases. It stands to

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

reason that the tax on Amazon need not be equal to the tax paid by a local bookstore on the same bit of data.

A data tax could also reflect the varying sensitivity of information. A tax on collecting a user's race or health history could be higher than a tax on location data. And within a category of data, the tax amount could depend on relevance. For collecting health data, a hospital would pay less than a gym, which in turn would pay less than social network. Biometric data should be free to collect when used to give employees access to a building, but more expensive if also commercialized. DNA data is highly sensitive, and firms creating DNA banks might create significant externalities, both positive and negative. If a tax is imposed on data collection, it is critical to allow DNA firms to charge for some of the positive external value their databases create.

Who will pay the tax? It is natural to think that firms collecting data are those who would have to pay the data tax. But upon further reflection, a data tax could be levied directly on the people who provide it. A tax is levied on a *transaction* and in real economic terms it does not matter who among the two parties—the data taker or the data giver—pays for it, since it would be incorporated either way into the overall price. If a carbon tax on gasoline is paid by the gas station, the station would charge a higher price and roll at least part of the tax onto consumers.

With that said, there are compelling reasons to frame the data tax as a charge levied on the data givers (consumers) rather than on data takers. Data givers are providing information not only about themselves, but also about their social contacts. Gmail users expose not only their own personal emails but also their pen pals'.[73] Ancestry.com clients expose genetic information about their relatives. And Facebook users create portals for data about their friends: a user with 1,000 friends is exposing more external data than one who has only 100—and should be taxed more.

Giving data is like grazing a common pasture. A database reveals general attributes about people—not only about the data giver and its immediate social circle. The price of participating in the activity should therefore reflect the social impact. In a typical commons scenario involving the protection of a natural resource (like a fishery), we worry about people (over)using it. Taxing people who give data that implicate others mirrors the typical response to the problem of protecting a commons.

Data, it is often said, are the new money. People receive valuable digital services, paying with their personal data instead of money. Not long ago,

73 In *Daniel Matera v. Google Inc.*, No. 5:15-cv-04062 2016 WL 454130 (N.D. Cal. Sept. 4, 2015), non-Gmail users filed a class action lawsuit against Gmail and its user for exposing their emails without consent.

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

they paid upwards of $200 for cars' navigation devices. Then came the free Maps apps, "paid" for by geo location data that advertisers greatly value. Cash is a currency that is personally costly (money paid can no longer be used elsewhere) but has no externalities. Data, in contrast, are a currency that is personally cheap (private information given can still be given elsewhere), but could have a social cost. Even users who are queasy about using their personal data as quid pro quo and care about their own privacy would use such currency with disregard for the social impact. A user data tax could correct the distortion in the choice among monies.

There is a symbolic aspect to a data tax paid by users who give their data rather than firms who take them. It represents a normative shift—that the problem of data pollution is *not* about protecting people's privacy but rather protecting the public ecosystem. Under the data pollution paradigm, data givers are not those in need of protection but those from whom the ecosystem has to be protected. They give too much data too easily and too often, and have to be restrained. The problem is not that they care so much and receive so little protection for their privacy, but rather that they care too little about sharing polluting data, and thus emit too much. True, data givers often do not realize that they are paying with data. A child who downloads the Angry Birds game-app for 99 cents does not know that personal data will be extracted as further quid pro quo. A data tax would surely correct this oversight and alert users to the implicit choice they are otherwise making.

In a dramatic way, a data tax could offset the current "data discounts" that digital users and data givers are offered. Internet companies sometimes present their users with a menu of payment options: pay-with-data versus pay-with-cash. "Basic" options cost less money but more personal data, whereas "premium" accounts cost more in money but involve less or no personal data collection.[74] For example, AT&T and Comcast offer broadband plans that cost more (roughly double) but liberate the bounty-paying users from data collection and from data-driven ads.[75] The great majority of consumers shun these plans—they prefer to pay with data and enjoy the price discount. Such choice ignores the negative social impact of data pollution and should thus trigger the data tax. The appeal of pay-with-data plans would justly decline.

As said, data tax is an idea, not a proposal ready for implementation. The practical problems are profound, but perhaps the biggest reason for caution is the possibility that data's overall external benefits far exceed the harms. If so, the

74 Data discounts are a specific instance of a more general "pay-for-data" model, in which companies would have to pay the users for their personal data. See, e.g., Posner and Weyl (2018).

75 See Bode (2016).

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

specific harmful uses—not data as a unified input—would have to be the targets of regulation.

4.3 Management of Data Spills

Command-and-control rules and data tax are the two primary regulatory techniques that target the market failure that occurs at the time a database is created. They resemble the two central techniques of environmental law—quantity and price regulation. But environmental law has another major device in its arsenal—waste management regulation. Beyond the methods used to control emissions ex ante, prior to release, environmental law has an elaborate framework for how to address the harm ex post, especially in the aftermath of an unanticipated release.

If release of toxic waste was the major problem of the industrial era, data spills are rapidly becoming the major social problem of the digital era.[76] According to one cybercrime report, half billion people around the world are subjected to cybercrime per year, costing them $110 billion.[77] Data spills are often caused by intentional criminal hacking,[78] and they could be prevented, at least in part, by tighter security. Indeed, recent statutory enactments mandate that companies adhere to higher prevention standards (Agelidis 2016, p. 1057; Chen, Liu, & Yao 2017, p. 1). In addition, even when breach occurs, the magnitude of harm caused can be mitigated by organized preparedness: by collecting less and deleting more data, and by activating post-breach mitigation response.

Environmental law has an ambitious post-release objective—the clean-up of hazardous waste sites[79]—which does not have a digital match. In general, data emissions cannot be scrubbed. Digital matter does not exist in a well-confined, excludable, removable space. It is infinitely replicable by a costless touch of a button or a simple line of algorithmic code. Once released, it is not containable. The regulatory response has to focus, instead, on other mitigation techniques to reduce the harm, and on ex post liability to deter it.

76 Estimates of the cost of data breaches vary greatly. A report by New York Attorney General (2014) quotes an estimate that "in 2012, direct and indirect identity theft losses totaled $24.7 billion in the United States, a figure that exceeded the losses in all other categories of property crime combined." See also Department of Justice (2015, p. 7), Ponemon Institute (2017), and Juniper Research (2017).

77 See *supra* note 22.

78 According to the Identity Theft Research Center (2017), criminal hacking dwarfs all other methods of data compromise and constitutes almost 60% of data breaches.

79 Comprehensive Environmental Response, Compensation, and Liability Act of 1980, 42 U.S.C. §§ 9601–9616.

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

4.3.1 Mitigation

The recent onslaught of security breaches has led to a corresponding surge of legislation imposing response duties on the owners of hacked databases. One typical duty imposed on a spilling company is disclosure, notifying the government and the affected parties "as expediently as possible" that the data was spilled, in the hope that such "transparency" would help set in motion private mitigation actions by the victims.[80] Bolstering these post-release notification schemes is a prominent theme in various proposals to deal with data emissions.[81]

Post-breach notifications are not completely useless (Romanovsky et al. 2011). There are things people can do to reduce their exposure to private harms arising from the theft of their data. People may sign up for credit monitoring (which provides alerts when a fraudulent application for new credit in their name is made), place a credit freeze or fraud alert (which blocks new accounts in their name), diligently cancel and replace stolen credit cards or social security numbers, regularly check their credit reports, file tax returns early, and more. Yet consumers' response to written security breach notifications is, at best, sluggish.[82] This is not laziness or some cognitive misjudgment. Their indifference is rational because they are largely shielded, through private or social insurance, from the monetary harms arising from security breach (Kiernan 2015; Pierce 2016, p. 982).[83] Their indifference is also inevitable in environs in which these notifications come in lengthy standard form and look like just another pre-printed disclosure, the likes of which consumers typically ignore (Ben-Shahar & Schneider 2014).

Mitigation of post-emission harm can be done without consumers' active participation, but ordinarily still requires consent. After a security breach, the spilling company can enroll people in shield programs. For example, in the aftermath of Equifax' massive data breach, the company offered free credit

80 See, e.g., Cal. Civ.Code § 1798.29(a), 1798.82(a); Consumer Privacy Protection Act of 2017, H.R. 4081 115th Cong. § 211

81 Hirsch (2006, p. 58) proposes a new federal "Data Release Inventory (DRI)" program that would require companies to report annually how much data they released, both intentionally and unintentionally. See also Schwartz & Janger (2006), Froomkin (2009) (proposing data breach notice rules that would apply to governmental bodies).

82 The Ponemon Institute (2014) found that "the most frequent response to a notification is to ignore it and do nothing."

83 As a result of these insurance arrangements, the consequences for victims are greatly moderated. It is estimated that approximately 25 percent of victims of data breaches subsequently suffer identity theft, and 14 percent of victims of identity theft experience personal out-of-pocket financial losses of $1 or more, with half suffering less than $100 loss. See LexisNexis. (2013) and Harrell & Langton (2013).

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

monitoring through a program called TrustedID, which required only a minimal effort to enroll. Under the proposed Consumer Privacy Protection Act, spilling companies would have to provide "five years of appropriate identity theft prevention and mitigation services" at no cost to any individual who asks for it (but auto-enrollment is still prohibited).[84]

Mitigation can reduce potential private harm to consumers whose data were spilled, but the social cost might still be substantial. For one, identity theft and other violations still occur. Furthermore, mitigation itself is costly—people spend time and money before, and especially after, their data is misused. While only a fraction of the exposed consumers end up suffering actual harm, the entire pool is harmed if people experience a heightened overhanging sense of financial risk or if they are required to take costly precautions. Indeed, the average victim spends approximately seven hours to clear up problems arising from identity theft, and some spend much more. Overall, 15 percent of people experience identity theft at some point in their life, and the risk is associated with nontrivial emotional anxiety (Harrell & Langton 2013, pp. 10–13).

The prevalence of social programs aimed at reducing and insuring the private harm from data spills is one of the mechanisms that makes data pollution a social, not private, cost. For example, misused credit card data is a cost that the issuing bank, not the issued-to consumer, bears. It is a cost, however, that the bank recoups by increasing the charges it levies on either consumers or merchants. Either way, all consumers pay: the data-fraud insurance premiums bundled into credit card fees are higher, the more severe the data spills are. This is the insurance externality identified in Section 2. The ex post regulatory tools achieve valuable loss-shifting and loss-spreading, but other tools are needed to achieve loss reduction. Liability may be one such tool, and private regulation may be another.

4.3.2 Liability for Exposure

Environmental law imposes stiff ex post liability on toxin-spilling companies. Exxon's liability for the Valdez oil spill totaled over $1 billion (in addition to heavily litigated punitive damages of $507 million)[85] and BP's Deepwater Horizon oil spill in 2010 cost the company well over $40 billion. Can data spills be subject to similarly stiff liability?

Section 3 explained why tort law has failed to hold companies liable for data breach. Problems of uncertainty over causation, societal harm, and valuation

84 Consumer Privacy Protection Act of 2017, H.R. 4081 115th Cong. § 211.

85 *Exxon v. Baker*, 554 U.S. 471 (2008).

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

make it difficult for potential victims to establish standing in court and receive compensation reflecting the harm caused by the spill. A public enforcement scheme, however, is not constrained by the same remedial and evidentiary standards. Liability could reflect the *expected* social harm without requiring actual victims to prove and measure their injury in fact, and without divvying up the damage payment across victims.

To induce optimal precautions, liability must reflect the total expected cost arising from the emission. Whether a criminal fine, a civil emissions fee, or even a statutory measure of damages awarded in a class action, it should equal the best estimate of the *risk* that the data release inflicts on society. The long-tail problem of data emission harms would no longer bar liability, if aggregate measures of exposure are available.

One way to compile aggregate estimates of social harm is through survey information. The Justice Department estimates, for example, that the average loss to victims of identity theft is approximately $1500 (Harrell & Langton 2017). Estimates of the likelihood of identity theft to consumers whose social security numbers were exposed vary, perhaps in the range of 14–30 percent (New York State Attorney General 2014). With such estimates, a fixed charge could be established for each consumer exposed. The fine would then need equal the per-capita expected harm, multiplied by the number of individuals exposed. Indeed, in the same way that many risk-producing activities in society are subject to fines reflecting their average gravity, data emission fines could be preset, with different dollar amounts per stolen credit card information, social security number, or other sensitive data.

Ex post liability could be designed with sufficient contours to create proper incentives. The magnitude of the fine could reflect the financial sensitivity of information, the quantum of affected records, the degree of carelessness in securing them, the steps taken to mitigate the harm, and more. Various standards of data protection are now defined in statutory enactments, and the degree of liability could be reduced—even eliminated—if the affected company was not at fault. Part of the inquiry would focus on technical protections employed to secure the database. But another part of the inquiry could focus on the justification for obtaining the information in the first place. Spilling information that was collected unnecessarily should result in higher fines.

In the end, the total liability on all spilling companies has to equal the total harm from data breach suffered by the scattered victims. There are reliable estimates of this harm—for example, one survey estimates a $16.8 billion harm from identity fraud in the U.S. in 2018 (see Pascual et al. 2018)—and the only remaining problem is how to divide the liability among polluters. Various criteria could measure the contribution of each polluter, focusing on

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

the amount and quality of data released and the security shortfalls. While tort law solves similar problems of apportionment in cases with joint tortfeasors or concurrent exposure, data pollution law cannot leave this solution to private tort suits. The uncertainty over causation that would defeat tort liability could be overcome through a statutory liability scheme.

4.3.3 Mandatory Insurance

Ex post liability can accomplish its deterrence goals, but only if firms can afford to pay the liability and have the information necessary to choose cost-justified precautions. In the area of cybersecurity, both problems—solvency and information—could undermine the effect of liability, as they have threatened to do in the area of environmental liability (Abraham 1986). Liability insurance may therefore be required, and it can help solve both problems.

It is widely recognized that the obligation to buy insurance against harms arising from their activity forces actors who are potentially judgment-proof to account for the external costs that they would otherwise ignore, such as the cost of liability that they could otherwise not afford to pay. Insurance has the effect of a Pigouvian tax: the differentiated premiums that firms pay reflect the different external costs they impose (Ben-Shahar & Logue 2012, p. 207). Through mandatory insurance, the equivalent of the data tax discussed above is imposed, not directly by a government ex ante but instead indirectly by insurers pricing the risk of liability.

Less widely recognized is the effect of liability insurance on prevention efforts. A common concern with insurance is moral hazard—the idea that a party who is insured against risk has a suboptimal incentive to reduce it. But moral hazard occurs only if insurers cannot monitor the prevention efforts made by their policyholders and price the policies accordingly (Shavell 1979; 2000, pp. 168–169). Accurate actuarial pricing that accounts for the expected harm given actual prevention efforts made by the policyholder gives firms incentive to reduce risk. Additionally, insurers can use their technical expertise to advise their clients as to which prevention measures are most effective and cost-justified—information that many commercial parties lack.

Cybersecurity liability insurers engage in "cyber health checks" to help firms "harden their data security" (Talesh 2017). Using rigorous tools developed in the insurance industry, firms' security practices receive rating scores, which affect both the premiums and the advice they get how to fix problems. Insurers sometimes test their clients' protection by trying to penetrate the firewalls remotely. They require the insured parties to comply with audits and data protection "best practices" developed by third-party experts. And they

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

intervene early enough in the aftermath of a security breach to reduce both the magnitude of harm and the legal liability exposure (Talesh 2017).

Cybersecurity insurance is a new form of commercial liability insurance. Like its much more mature sibling, environmental liability insurance, it is a specialized policy that covers harms to third parties caused by commercial activities, which are otherwise excluded from coverage in the standard commercial liability insurance. Environmental law has an elaborate scheme of risk management, requiring sites to implement spill prevention, control, and countermeasures.[86] Even with such robust regulatory background, environmental liability insurance policies often require firms to adhere to stricter private environmental codes than those imposed by the EPA (Kunzman 1985, p. 477; Richardson 2002; Ben-Shahar & Logue 2012, pp. 225–226). Given the embryonic stage of cybersecurity law, the private development of risk reduction standards could be a major benefit emerging from a regime of stiff cyber-emissions liability coupled with mandatory liability insurance.

5. CONCLUSION

Digital data law should not be only about privacy. Too often, the exchange of data between giver and taker affects third parties; the data could contain information about others; or, more importantly, the database can be used or misused in ways that affect public interests beyond individual users' privacy. Data pollution is the name of this problem, and data pollution law is the set of legal tools to combat it.

Data pollution law might borrow regulatory solutions designed for privacy protection, but more often it would require different devices. For example, "user control" and "informed consent"—two longstanding pillars of data privacy law—are irrelevant to data pollution law. Control and consent tools are enacted under the heroic assumption that they help people protect themselves. Even if that were true, there is no reason to think that people would cease giving data in a way that harms others. Different interventions, including some regulations included in recent privacy law reforms, may help reduce data pollution. The concern with any type of intervention is that it would throw the baby out with the bathwater, suffocating data's positive externalities.

Perhaps the most novel technique to reduce data pollution is therefore data tax. This is where data pollution law clearly diverges from data privacy law. Whereas privacy violations have to be stopped, pollution only needs to be

86 Environmental Protection Agency; Oil Pollution Prevention Spill Prevention Countermeasure, 40 C.F.R. § 112 (2010).

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

priced. The design of a rational data tax is extremely challenging, and Section 4 of this article made some initial nods toward that mission. Two relatively simple strategies could set us in the right direction. First, let us stop harmful data subsidies. People are paid for their personal data all the time, primarily by the services data accumulators offer in return, and proposals for even bigger data subsidies—to require businesses to pay people for the data they harvest—are proliferating (Kaiser 2018; Posner & Weyl 2018). These subsidies are equivalent to paying people to pollute. Second, a small nominal data tax would go a long way toward stopping the mindless hoarding of unneeded data, without stifling meaningful innovation. Even a small tax would prompt data polluters to embrace some critical strategies for pollution reduction.

Data pollution law is urgently needed because data privacy has so far proven thoroughly ineffective. True, data privacy law's search for new, more impactful mandates might succeed in the future, where previous devices (primarily disclosure) failed. People might begin to care more and surrender less of their digital privacy. But securing privacy does not solve data's social harms. Data pollution harms could occur even when privacy is protected.

The contribution of this article is not in solving the problem of data pollution. Despite the article's combative tones against the dominance of privacy concerns in data law, it is not calling to diminish the concern with digital privacy. Rather, the article's contribution is in recognizing that a data pollution problem exists. If, as I have argued, data pollution is caused by the impact of databases on people other than those included within, lawmakers must begin to do the hard work of carefully distinguishing data's external effects from data's privacy harm, and look for the best ways to reduce these social costs.

REFERENCES

Abraham, Kenneth S. 1986. *Distributing Risk: Insurance, Legal Theory, and Public Policy*. New Haven: Yale University Press.

——— 2002. The Relation Between Civil Liability and Environmental Regulation: An Analytical Overview. **41** *Washburn L.J.* 379–398.

——— 2008. *The Liability Century: Insurance and Tort Law from the Progressive Era to 9/11*. Cambridge, MA: Harvard University Press.

Acquisti, Alessando, Laura Brandimarte & George Loewenstein. 2015. Privacy and Human Behavior in the Age of Information. **347** *Science* 509–514.

Acquisti, Alessandro, Leslie K. John & George Loewenstein. 2013. What Is Privacy Worth? **42** *J. Legal Stud.* 249–273.

Adjerid, Idris *et al.* 2016. A Query-Theory Perspective of Privacy Decision Making. **45** *J. Legal Stud.* S97–S121.

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

Agelidis, Yasmine. 2016. Protecting the Good, the Bad, and the Ugly: Exposure Data Breaches and Suggestions for Coping with Them. **31** *Berkeley Tech. L.J.* 1057–1078.

American Law Institute. 2010. *Principles of the Law of Aggregate Litigation*. Philadelphia, PA: American Law Institute.

———. 1991. *Enterprise Responsibility for Personal Injury*. Philadelphia: PA: American Law Institute, 319–321.

———. 2019. Principles of the Law, Data Privacy: §§ 3–4. American Law Institute.

Ashenmacher, George. 2016. Indignity: Redefining the Harm Caused by Data Breaches. **51** *Wake Forest L. Rev.* 1–56.

Athey, Susan *et al.* 2018. *The Digital Privacy Paradox: Small Money, Small Costs, Small Talk*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2916489.

Bae, Hyunhoe *et al.* 2010. Information Disclosure Policy: Do State Data Processing Efforts Help More than the Information Disclosure Itself? **29** *J. Pol'y Anal. Mgmt.* 163–182.

Barnett, Kerry. 1987. Equitable Trusts: An Effective Remedy in Consumer Class Actions. **96** *Yale L.J.* 1591–1614.

Barsa, Michael. 1997. California's Proposition 65 and the Limits of Information Economics. **49** *Stan. L. Rev.* 1223–1247.

Bar-Gill, Oren, Omri Ben-Shahar & Florencia Marotta-Wurgler. 2017. Searching for the Common Law: The Quantitative Approach of the Restatement of Consumer Contracts. **84** *U. Chi. L. Rev.* 7–35.

Barnhill, Allison Rosser. 1989. The Unraveling of California's Proposition **65**. 24 *Wake Forest L. Rev.* 367–408.

Benkler, Yochai, Robert Faris, & Hal Roberts. 2018. *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. Oxford: Oxford University Press.

Ben-Shahar, Omri & Anu Bradford. 2012. Efficient Enforcement in International Law. **12** *Chi. J. Int'l L.* 376–431.

Ben-Shahar, Omri & Ariel Porat. 2018. The Restoration Remedy in Private Law. **118** *Colum. L. Rev.* 1901–1952.

Ben-Shahar, Omri & Adam Chilton. 2016. Simplification of Privacy Disclosures: An Experimental Test. 45 *J. Legal Stud.* S42–S67.

Ben-Shahar, Omri & Carl Schneider. 2014. *More than You Wanted to Know: The Failure of Mandated Disclosure*. Princeton, NJ: Princeton University Press.

Ben-Shahar, Omri & Kyle Logue. 2012. Outsourcing Regulation: How Insurance Reduces Moral Hazard. 111 *Mich. L. Rev.* 197–248.

Ben-Shahar, Omri & Lior J. Strahilevitz. 2016. Contracting over Privacy. 45 *J. Legal Stud.* S1–S11.

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

Bode, Karl. 2016. AT&T Charges Steep Premium for Privacy, Calls it a 'Discount.' *DSL Reports* (March 17, 2016). https://www.dslreports.com/shownews/ATT-Charges-Steep-Premium-for-Privacy-Calls-it-a-Discount-136511.

Bradshaw, Karen. 2016. Settling for Natural Resource Damages. 40 *Harv. Env. L. Rev.* 211–253.

Brennan, Troyen A. 1988. Causal Chains and Statistical Links: The Role of Scientific Uncertainty in Hazardous-Substance Litigation. **73** *Cornell L. Rev.* 469–533.

Brown, Daniel. 2018. Here are Some of the Biggest Reveals from a Fitness-tracker Data Map That May Have Compromised Top-secret US Military Bases around the World. *Business Insider* (January 29, 2018). https://www.businessinsider.com.au/strava-heatmap-most-revealing-images-2018-1.

Bui, Linda T. 2005. Public Disclosure of Private Information as a Tool for Regulating Environmental Emissions: Firm-Level Responses by Petroleum Refineries to the Toxics Release Inventory. *Center for Economic Studies, U.S. Census Bureau, Working Papers 05-13.*

Burtraw, Dallas & Sarah Jo Szambelan. 2009. U.S. Emissions Trading Markets for SO2 and NOx. *Resources for the Future*, Discussion Paper 09-40.

Butler, Henry N. & Jonathan R. Macey. 1996. Externalities and the Matching Principle: The Case for Reallocating Environmental Regulatory Authority. **14** *Yale L. & Pol'y Rev.* 23–66.

Calo, Ryan. 2013. *Consumer Subject Review Boards: A Thought Experiment. 66 Stan. L. Rev.* Online 97–102.

Center for Strategic and International Studies. 2014. Net Losses: Estimating the Global Cost of Cybercrime: Economic Impact of Cybercrime II. *Intel Security* (June 5, 2014). https://csis-prod.s3.amazonaws.com/s3fs-public/legacy_files/files/attachments/140609_rp_economic_impact_cybercrime_report.pdf.

Cheng, Long, Fang Liu & Danfeng (Daphne) Yao. 2017. Enterprise data breach: causes, challenges, prevention, and future directions. **7** *WIREs: Data Mining and Knowledge Discovery* 1–14.

Cohen, Bret *et al.* 2017. Data Localization Laws and Their Impact on Privacy, Data Security and the Global Economy. **32** *Antitrust* 107–114.

Cohen, Julie E. 2000. Examined Lives: Informational Privacy and the Subject as Object. **52** *Stan. L. Rev.* 1373–1437.

Congressional Budget Office. 2001. Evaluation of Cap-and-Trade Programs for Reducing U.S. Carbon Emissions. *Congressional Budget Office* (June 2001). https://www.cbo.gov/publication/13107.

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

Constine, Josh. 2015. Facebook Is Shutting Down Its API for Giving Your Friends' Data to Apps. *TechCrunch* (April 28, 2015). https://techcrunch.com/2015/04/28/facebook-api-shut-down/.

Crossland, Kiley. 2018. The Hidden Risks of At-home DNA Testing. *World* (January 5, 2018). https://world.wng.org/content/the_hidden_risks_of_at_home_dna_testing.

Datta, Amit *et al.* 2015. Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination **2015** *PoPETs* 92–112.

——— *et al.* 2018. Discrimination in Online Advertising a Multidisciplinary Inquiry. **81** *Proc. Mach. Learn. Res.* 1–15.

Dell Technologies. 2014. EMC Privacy Index. *Dell.* https://www.emc.com/campaign/privacy-index/global.htm.

Department of Justice. 2015. *Victims of Identity Theft, 2014, NCJ 248991.* Department of Justice, Office of Justice Programs, Bureau of Justice Statistics (September 2015). https://www.bjs.gov/content/pub/pdf/vit14.pdf.

DeVries, Will Thomas. 2003. Protecting Privacy in the Digital Age. **18** *Berkeley Tech. L.J.* 283–311.

Dewees, Donald N. 1992. The Role of Tort Law in Controlling Environmental Pollution. **18** *Can. Public Pol'y.* 425–442.

——— *et al.* 1996. *Exploring the Domain of Accident Law: Taking the Facts Seriously.* Oxford, UK: Oxford University Press.

Economist. 2017. Data Is Giving Rise To A New Economy. *Economist* (May 6, 2017). https://www.economist.com/news/briefing/21721634-how-it-shaping-up-data-giving-rise-new-economy.

Environmental Protection Agency. 2017. The Social Cost of Carbon: Estimating the Benefits of Reducing Greenhouse Gas Emissions. *Environmental Protection Agency.* https://19january2017snapshot.epa.gov/climatechange/social-cost-carbon_.html.

Esty, Daniel C. 2004. Environmental Protection in the Information Age. 79 *NYU L. Rev.* 115–211.

Experian. 2017. Delivering Value in the Digital Age: Exploring UK Attitudes Towards Data. *Experian.* https://engage.experian.co.uk/delivering-value-in-the-digital-age/.

Federal Trade Commission. 2012. Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers. *Federal Trade Commission* (March 2012). https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf.

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

———. 2013. Lost or Stolen Credit, ATM, and Debit Cards. *Federal Trade Commission, Consumer Information*. https://www.consumer.ftc.gov/articles/0213-lost-or-stolen-credit-atm-and-debit-cards#Limit.

Froomkin, A. Michael. 2000. The Death of Privacy? **52** *Stan. L. Rev.* 1461–1543.

——— 2009. Government Data Breaches. **24** *Berkeley Tech. L.J.* 1019–1059.

——— 2015. Regulating Mass Surveillance as Privacy Pollution: Learning from Environmental Impact Statements. **2015** *U. Ill. L. Rev.* 1713–1790.

Gelpe, Marcia R. & A. Dan Tarlock. 1974. The Uses of Scientific Information in Environmental Decisionmaking.**48** *S. Cal. L. Rev.* 371–427.

Ginsberg, Jeremy *et al.* 2009. Detecting Influenza Epidemics Using Search Engine Query Data. **457** *Nature* 1012–1014.

Google. 2019. Google Privacy Checkup. *Google*. https://myaccount.google.com/privacycheckup.

John D. Graham, and Jonathan Baert Weiner eds. 1997. *Risk versus Risk: Tradeoffs in Protecting Health and the Environment*. Cambridge, MA: Harvard University Press.

Granville, Kevin. Facebook and Cambridge Analytica: What You Need to Know as Fallout Widens. *The New York Times* (March 19, 2018). https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html.

Harrell, Erika & Lynn Langton. 2013. *Victims of Identity Theft 2012*. U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics. https://www.bjs.gov/content/pub/pdf/vit12.pdf.

———. 2017. *Victims of Identity Theft 2014*. U.S Department of Justice, Office of Justice Programs, Bureau of Justice Statistics (Revised November 13, 2017) https://www.bjs.gov/content/pub/pdf/vit14.pdf.

Hermstrüwer, Yoan. 2017. Contracting Around Privacy: The (Behavioral) Law and Economics of Consent and Big Data. **8** *J. Intell. Prop. Info. Tech. & Elec. Com. L.* 9–26.

Hirsch, Dennis D. 2006. Protecting the Inner Environment: What Privacy Regulation Can Learn from Environmental Law. **41** *Ga. L. Rev.* 1–63.

——— 2014. The Glass House Effect: Big Data, the New Oil, and the Power of Analogy. **66** *Me. L. Rev.* 373–395.

Hirsch, Dennis D. & Jonathan H. King. 2016. Big Data Sustainability: An Environmental Management Systems Analogy. **72** *Wash. & Lee L. Rev.* Online 406–419.

IBM. 2018. New Survey Finds Deep Consumer Anxiety over Data Privacy and Security. *IBM News Room* (April 16, 2018). https://newsroom.ibm.com/2018-04-15-New-Survey-Finds-Deep-Consumer-Anxiety-over-Data-Privacy-and-Security.

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

Identity Theft Research Center. 2017. 2017 Annual Data Breach Year-End Review. *Identity Theft Research Center*. https://www.idtheftcenter.org/images/breach/2017Breaches/2017AnnualDataBreachYearEndReview.pdf.

Insurance Information Institute. 1999. *HO3, Section I.E.6*. Insurance Information Institute. https://www.iii.org/sites/default/files/docs/pdf/HO3_sample.pdf.

Isenberg, Howard. 1995. The Second Industrial Revolution: The Impact of the Information Explosion. **27** *Ind. Eng.* 14.

Jensen, Carlos & Colin Potts. 2004. Privacy Policies as Decision-Making Tools: An Evaluation of Online Privacy Notices. In *Proceedings of the 2004 Conference on Human Factors in Computing Systems*, 471–478. Vienna, Austria: ACM Press.

Johnson, Dominic & Simon Levin. 2009. The Tragedy of Cognition: Psychological Biases and Environmental Inaction. **97** *Curr. Sci.* 1593–1603.

Johnston, Jason Scott. 2016. The Social Cost of Carbon. **39** *Regulation* 36–44.

Jun, S-P, H.S Yoo & S. Choi. 2018. Ten Years of Research Change Using Google Trends: From the Perspective of Big Data Utilizations and Applications. **130** *Technol. Forecast. Soc. Change* 69–87.

Juniper Research 2017. Cybercrime Will Cost Businesses over $2 Trillion by 2019. *Juniper Research: Press Releases* (March 2017). https://www.juniperresearch.com/press/press-releases/cybercrime-cost-businesses-over-2trillion.

Kaiser, Brittany. 2018. Facebook Should Pay Its 2bn Users for Their Personal Data. *Financial Times* (April 9, 2018). https://www.ft.com/content/7a99cb46-3b0f-11e8-bcc8-cebcb81f1f90.

Keats Citron, Danielle. 2007. Reservoirs of Danger: The Evolution of Public and Private Law at the Dawn of the Information Age. **80** *S. Cal L. Rev.* 241–297.

——— 2019. Sexual Privacy. **128** *Yale L.J.* 1870–1961.

Kiernan, John. 2015. Fraud Liability Study: Which Cards Protect You Best? *Wallethub* (January 20, 2015). https://wallethub.com/edu/fraud-liability-study/25726/.

Kohn, Jeff & Kelsey Kruger. 2016. Understand Pollution, Environmental Impacts from Food in 6 Charts. *GreenBiz* (November 17, 2016). https://www.greenbiz.com/article/understand-pollution-environmental-impacts-food-6-charts.

Koo, Jimmy H. 2017. Equifax Negligent in Data Breach, Community Banks Allege. *Class Action Litigation Report, Bloomberg BNA* (December 12, 2017). https://news.bloomberglaw.com/class-action/equifax-negligent-in-data-breach-community-banks-allege.

Kramer, Ann. 2019. Ransomware, Data Breaches Expose Gaps in Cyber Insurance Market. *Bloomberg Law* (July 24, 2019).

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

Kunzman, Steven A. 1985. The Insurer as Surrogate Regulator of the Hazardous Waste Industry: Solution or Perversion? **20** *Forum* 469–488.

Lambrecht, Anja & Catherine Tucker. 2018. *Algorithmic Bias? An Empirical Study into Apparent Gender-Based Discrimination in the Display of STEM Career Ads*. https://www.ssrn.com/abstract=2852260.

Lamotte, Sandee. 2017. After 60 Years of Friendship, They Learned They're Biological Brothers. *CNN* (December 27, 2017). https://www-m.cnn.com/2017/12/27/health/friends-brothers-dna-discovery-hawaii-trnd/index.html?r=https%3A%2F%2Fwww.google.com%2F.

Lewis, Paul. 2018. 'Utterly Horrifying': ex-Facebook Insider Says Covert Data Harvesting Was Routine. *The Guardian* (March 20, 2018). https://www.theguardian.com/news/2018/mar/20/facebook-data-cambridge-analytica-sandy-parakilas.

LexisNexis. 2013. LexisNexis True Cost of Fraud Study: Merchants Struggle Against an Onslaught of High-Cost Identity Fraud and Online Fraud. *LexisNexis* (September 2013). https://www.lexisnexis.com/risk/downloads/assets/ true-cost-fraud-2013.pdf).

Liberty Mutual Insurance. 2019. Identity Fraud Expense Coverage. *Liberty Mutual Insurance*. https://www.libertymutual.com/identity-theft-insurance.

Lin, Albert C. 2005. Beyond Tort: Compensating Victims of Environmental Toxic Injury. **78** *S. Cal. L. Rev.* 1439–1528.

——— 2012. Public Trust and Public Nuisance: Common Law Peas in a Pod. **45** *U.C.D. L. Rev.* 1075.

Marotta-Wurgler, Florencia. 2016. Self-Regulation and Competition in Privacy Policies. **45** *J. Legal Stud.* S13–S39.

Matthews, Alex & Catherine Tucker. 2017. *Government Surveillance and Internet Search Behavior*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2412564.

May, Ashley. 2018. Took an Ancestry DNA Test? You Might Be a 'Genetic Informant' Unleashing Secrets about Your Relatives. *USA Today* (April 27, 2018). https://www.usatoday.com/story/tech/nation-now/2018/04/27/ancestry-genealogy-dna-test-privacy-golden-state-killer/557263002/.

McAfee. 2017. Grand Theft Data – Data Exfiltration Study: Actors, Tactics, and Detection. *McAfee Report*. https://www.mcafee.com/us/resources/reports/rp-data-exfiltration.pdf.

McDonald, Alecia M. & Lorrie Faith Cranor. 2008. The Cost of Reading Privacy Policies. **4** *I/S: J.L. & Pol'y for Info. Soc'y*. 540–565.

Metcalf, Gilbert E. & David Weisbach. 2009. The Design of a Carbon Tax. **33** *Harv. Envtl. L. Rev.* 499–556.

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

Miller, Amalia R. & Catherine Tucker. 2017. Frontiers of Health Policy: Digital Data and Personalized Medicine. **17** Innovation Policy and the Economy 49–75.

Mossoff, Adam. 2004. Spam–Oy, What a Nuisance! **19** *Berkeley Tech. L.J.* 625–666.

Morey, Timothy *et al.* 2015. Customer Data: Designing for Transparency and Trust. *Harvard Business Review* (May 2015), 1–11. https://hbr.org/2015/05/customer-data-designing-for-transparency-and-trust.

National Telecommunications and Information Administration. 2013. *Short Form Notice Code of Conduct to Promote Transparency in Mobile App Practices.* https://www.ntia.doc.gov/les/ntia/publications/july_25_code_draft.pdf.

Nehf, James P. 2003. Recognizing the Societal Value in Information Privacy. **78** *Wash. L. Rev.* 1–92.

Nesheim, Malden C. *et al.* eds. 2015. *A Framework for Assessing Effects of the Food System.* Washington, DC: National Academic Press. https://www.ncbi.nlm.nih.gov/books/NBK305182/.

Norton Security. 2012. 2012 Norton Cybercrime Report. *Symantec* (September 5, 2012). https://www.symantec.com/about/newsroom/press-releases/2012/symantec_0905_02.

Office of the New York State Attorney General. 2014. *Information Exposed: Historical Examination of Data Breaches in New York State.* New York State Attorney General (July 7, 2014). https://www.ag.ny.gov/pdfs/data_breach_report071414.pdf.

Pan, Yue & George M. Zinkhan. 2006. Exploring the Impact of Online Privacy Disclosures on Consumer Trust. **82** *J. Retailing* 331–338.

Pascual, Al *et al.* 2018. Identity Fraud: Fraud Enters a New Era of Complexity. *Javelin Strategy* (February 6, 2018). https://www.javelinstrategy.com/coverage-area/2018-identity-fraud-fraud-enters-new-era-complexity.

Perez-Pena, Richard & Matthew Rosenberg. Strava Fitness App Can Reveal Military Sites, Analysts Say. *The New York Times* (January 29, 2018). https://www.nytimes.com/2018/01/29/world/middleeast/strava-heat-map.html.

Pierce, Justin C. 2016. Shifting Data Breach Liability: A Congressional Approach. **57** *Wm. & Mary L Rev.* 975–1017.

Pollack, Wendy & Mike Sullivan. 2014. The Information Subscribers Most Likely to Pay for Google Among Tech Services. *The Information* (April 20, 2018). https://www.theinformation.com/articles/the-information-subscribers-most-likely-to-pay-for-google-among-tech-services.

Ponemon Institute. 2014. *The Aftermath of a Data Breach: Consumer Settlement.* Ponemon Institute (April 2014). https://www.ponemon.org/local/upload/

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

file/Consumer%20Study%20on%20Aftermath%20of%20a%20Breach%20FINAL%202.pdf.

———. 2017. *Cost of Data Breach Study: Global Overview*. Ponemon Institute (June 2017). https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=SEL03130WWEN&.

Posner, Eric & Glen Weyl. 2018. *Radical Markets*. Princeton, NJ: Princeton University Press.

PrivacyGrade.org. 2014. *Search Results for "facebook." Carnegie Mellon University*. http://privacygrade.org/apps/search?utf8=%E2%9C%93&q=facebook.

Radin, Margaret Jane. 2013. *Boilerplate: The Fine Print, Vanishing Rights, and the Rule of Law*. Princeton, NJ: Princeton University Press.

Reiman, Jeffrey H. 1982. Privacy, Intimacy, and Personhood. In F.D. Schoeman, ed., *Philosophical Dimensions of Privacy: An Anthology*, 300–316. Cambridge, UK: Cambridge University Press.

Revesz, Richard L. & Michael A. Livermore. 2011. *Rataking Rationality: How Cost-Benefit Analysis Can Better Protect the Environment and Our Health*. Oxford, UK: Oxford University Press.

Robinson, Matt. 2018. Yahoo to Pay First SEC Penalty over Its Response to Massive Hack. *Bloomberg BNA* (April 25, 2018). https://news.bloomberglaw.com/tech-and-telecom-law/yahoo-to-pay-first-sec-penalty-over-its-response-to-massive-hack.

Richardson, Benjamin J. 2002. Mandating Environmental Liability Insurance. **12** *Duke Envtl. L. & Pol'y F.* 293–330.

Rogers, Anna. 2015. Breast Implants: The Ticking Time Bomb in Millions of Women's Bodies. *Collective Evolution* (October 21, 2015). https://www.collective-evolution.com/2015/10/21/breast-implants-the-ticking-time-bomb-in-millions-of-womens-bodies/.

Romanovsky, Sasha *et al.* 2011. Do Data Breach Disclosure Laws Reduce Identity Theft? **30** *J. Pol'y Anal. & Management* 256–286.

Rosenberg, David. 1984. The Causal Connection in Mass Exposure Cases: A "Public Law" Vision of the Tort System. **97** *Harv. L. Rev.* 849–949.

Schneider, Carl E. 1998. *The Practice of Autonomy: Patients, Doctors, and Medical Decisions*. Oxford, UK: Oxford University Press.

Schroeder, Christopher H. 2002. Lost in the Translation: What Environmental Regulation Does That Tort Cannot Duplicate. **41** *Washburn L.J.* 583–606.

Schwartz, Paul M. 1999. Privacy and Democracy in Cyberspace. **52** *Vand. L. Rev.* 1609–1702.

——— 2013. The EU-US Privacy Collision. **126** *Harv. L. Rev.* 1966.

Schwartz, Paul M. & Edward J. Janger. 2006. Notification of Data Security Breaches. **105** *Mich. L. Rev.* 913–984.

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

Schwartz, Paul M. & Karl-Nikolaus Peifer. 2017. Transatlantic Data Privacy Law. **106** *Geo. LJ.* 115–179.

Sharkey, Catherine. 2003. Punitive Damages as Societal Damages. **113** *Yale L.J.* 347–453.

Shavell, Steven. 1979. On Moral Hazard and Insurance. **93** *Q.J. Econ.* 541–562.

——— 1987. *Economic Analysis of Accident Law*. Cambridge, MA: Harvard University Press.

——— 2000. On the Social Function and Regulation of Liability Insurance. **25** *Geneva Papers on Risk & Ins.* 166–179.

Silverman, David L. 2017. Developments in Data Security Breach Liability. **73** *Bus. L.* 215.

Silverman, Jacob. 2016. Just How 'Smart' Do You Want Your Blender to Be? *The New York Times* (June 14, 2016). https://www.nytimes.com/2016/06/19/magazine/just-how-smart-do-you-want-your-blender-to-be.html.

Solove, Daniel J. 2002. Conceptualizing Privacy. **90** *Cal. L. Rev.* 1087–1155.

Solove, Daniel J. & Danielle Keats Citron. 2018. Risk and Anxiety: A Theory of Data Breach Harms. **96** *Tex. L. Rev.* 737–786.

Solove, Daniel J. & Paul Schwartz. 2017. *Information Privacy Law*, 6th edn. Philadelphia, PA: Wolters & Kluwer.

Strahilevitz, Lior J. & Matthew B. Kugler. 2016. Is Privacy Policy Language Irrelevant to Consumers? **45** *J. Legal Stud.* S69–S95.

Stavins, Robert N. 2003. Experience with Market-Based Environmental Policy Instruments. In Karl-Göran Mäler and Jeffrey Vincent, eds., *Handbook of Environmental Economics*, 355–435. Amsterdam, Netherlands: Elsevier Science.

Steinberg, Joseph. These Devices May Be Spying On You (Even In Your Own Home). *Forbes* (January 27, 2014). https://www.forbes.com/sites/josephsteinberg/2014/01/27/these-devices-may-be-spying-on-you-even-in-your-own-home/#7eb0e320b859.

Sunstein, Cass R. 2017. *#Republic: Divided Democracy in the Age of Social Media*. Princeton, NJ: Princeton University Press.

——— 2018. Is Social Media Good or Bad for Democracy. *Facebook Newsroom* (January 22, 2018). https://newsroom.fb.com/news/2018/01/sunstein-democracy/.

Swanson, Elizabeth J. & Elaine L. Hughes. 1990. *The Price of Pollution: Environmental Litigation in Canada*. Edmonton: Environmental Law Center.

Sweeney, Latanya. 2013. Discrimination in Online Ad Delivery. **11** *ACMQueue* 1–10.

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231

Talesh, Shauhin A. 2017. Data Breach, Privacy, and Cyber Insurance: How Insurance Companies Act as "Compliance Managers" for Business. **43** *Law and Social Inquiry* 417–440.

Thomsen, Simon. 2015. Extramarital Affair Website Ashley Madison Has Been Hacked and Attackers Are Threatening to Leak Data Online. *Business Insider* (July 21, 2015). https://www.businessinsider.com/cheating-affair-website-ashley-madison-hacked-user-data-leaked-2015-7.

TrustArc. 2019. TRUSTe Enterprise Privacy Certification. *TrustArc Privacy Compliance*. https://www.trustarc.com/products/enterprise-privacy-certification/.

Turow, Joseph. 2008. The Federal Trade Commission and Consumer Privacy in the Coming Decade. 3 *J. L. &Pol'y for Info. Soc.* 723–749.

Viscusi, Kip. 2000. Foreword. In Richard L. Stroup and Roger E. Meineres, eds., *Cutting Green Tape: Toxic Pollutants, Environmental Regulation, and the Law*, ix. Piscataway & New Brunswick, NJ: Transaction Publishers.

Walmart. 2017. Walmart Privacy Policy. *Walmart.com* (November 2017). https://corporate.walmart.com/privacy-security/walmart-privacy-policy.

Westin, Alan. 1967. *Privacy and Freedom*. New York, NY: Atheneum Press.

Weiss, N. Eric & Rena S. Miller. 2014. *The Target and Other Financial Breaches: Frequently Asked Questions. Cong. Research Serv., R43496*. https://fas.org/sgp/crs/misc/R43496.pdf.

White House. 2012. *Consumer Data Privacy in a Networked World: A Framework for Protecting Privacy and Promoting Innovation in the Global Digital Economy*. White House (February 2012). https://obamawhitehouse.archives.gov/sites/default/files/privacy-final.pdf.

Wikipedia. 2019a. Environmental Certification. *Wikipedia*. https://en.wikipedia.org/wiki/Environmental_certification.

Willis, Lauren E. 2013. When Nudges Fail: Slippery Defaults. **80** *U. of Chi. L. Rev.* 1155–1229.

Wittes, Benjamin & Jodie C. Liu. 2005. *The Privacy Paradox: The Privacy Benefits of Privacy Threats*. Center for Technology Innovation at Brookings (May 2015). https://www.brookings.edu/wp-content/uploads/2016/06/Wittes-and-Liu_Privacy-paradox_v10.pdf.

Yanqing, Hong. 2017. The Cross-Border Data Flows Security Assessment: An important part of protecting China's basic strategic resources. Yale Law School Paul Tsai China Center. *Working Paper* (June 20, 2017). https://law.yale.edu/system/files/area/center/china/document/dataflowssecurity_final.pdf.

Zetter, Kim. 2009. Do Breach Notification Laws Work? *Wired* (March 9, 2009). https://www.wired.com/2009/03/experts-debate/.

Zhang, Lily. 2005. The CAN-SPAM Act: An Insufficient Response to the Growing Spam Problem. **20** *Berkeley Tech. L.J.* 301–332.

Downloaded from https://academic.oup.com/jla/article-abstract/doi/10.1093/jla/laz005/5578488 by guest on 01 October 2019

Electronic copy available at: https://ssrn.com/abstract=3191231