LAW AND ECONOMICS SEMINAR **Professor Polinsky**
Autumn Quarter 2018

Thursday, October 4, 2018
4:15 - 5:45 p.m.
Stanford Law School
Room 185

"Data Pollution"

by

Omri Ben-Shahar

(University of Chicago Law School)

Note: It is expected that you will have reviewed the speaker's paper before the seminar.

DATA POLLUTION

Omri Ben-Shahar*

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, diminishing the privacy of the people whose information is used. It claims that a central problem in the digital economy has been largely ignored: how the information individuals give 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 external social harms from industrial pollution. The data pollution framework also opens up a rich roadmap for new regulatory devices—an environmental law for data protection—that focuses on controlling these external effects. The article examines how the tools used to control environmental pollution—production restrictions, carbon tax, and emissions liability—could be adapted to govern data pollution.

* University of Chicago. I am grateful to Ronen Avraham, Oren Bar-Gill, Karen Bradshaw, Daniel Hemel, Jaime Hine, William Hubbard, Florencial Marrota-Wurgler, Jennifer Nou, Ariel Porat, Eric Posner, Ricky Revesz, Lior Strahilevitz, Mark Templeton, and workshop participants at the University of Chicago Law School for helpful discussions, and to Jason Grover for research assistance.

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

INTRODUCTION

Digital information is the fuel of the new economy. It is the resource that creates new companies and products, new markets and currencies, and endless new opportunities.[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 rethink the ways the data economy has to be regulated and its harms controlled.

Digital information embraces everything—history, geography, literature, physics—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 they can be influenced. The widespread aggregation of such personal data is creating personalized environments with enormous private and social benefits. But they also produce potential harm. The external harm is less concrete and, until recently, far less noticed. Understanding the scope of this potential harm and reducing its magnitude is 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 elections have dramatically illustrated. Facebook's database of personal information was used by others to spread false political ads.[2] Political lies are not new, but their effect is magnified when propelled and pinpointed by a data-rich process. 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

---

[1] *Data Is Giving Rise To A New Economy*, Economist (May 6, 2017), available at https://www.economist.com/news/briefing/21721634-how-it-shaping-up-data-giving-rise-new-economy. See also Will Thomas DeVries, *Protecting Privacy in the Digital Age*, 18 BERKELEY TECH. L.J. 283, 291 (2003) (the digital technological change on a scale matching or exceeding the industrial revolution); Howard Isenberg, *The Second Industrial Revolution: The Impact of the Information Explosion*, 27 INDUS. ENG'G. 14, 15 (1995).

[2] Kevin Granville, *Facebook and Cambridge Analytica: What You Need to Know as Fallout Widens*, THE NEW YORK TIMES (Mar. 19, 2018).

dossiers of 143 million consumers were stolen, is a prominent exemplar of this data emission problem.[3]

Societies are searching for paradigms to understand, and techniques to address, the actual harms and the 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 privacy paradigm, the collection of personal data creates various harms to the *individuals* whose data is collected, used, shared, or lost. 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.[4] The privacy paradigm is founded on the premise that the injury from the personal data enterprise is private in nature—to the "core self"[5]—although by sheer aggregation (or by more nuanced channels) these deeply private injuries have a derivative social impact.[6]

The privacy paradigm is disturbingly incomplete because the harms from data misuse are often far greater than the private injuries to the individuals whose information gets released. 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, separate from the impact felt by private people who use polluting products. 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 the collection of personal data affect institutions and groups of people—beyond those whose data is taken. Facebooks data

---

[3] *Massive Equifax data breach hits 143 million*, BBC NEWS (Sept. 8, 2017).

[4] See, e.g., Paul M. Schwartz and Karl-Nikolaus Peifer, *Transatlantic Data Privacy Law*, 106 Geo. L. J. 115, 126 (2017);

[5] See ALAN WESTIN, PRIVACY AND FREEDOM 32 (1967). See also Jeffrey H. Reiman, *Privacy, Intimacy, and Personhood,* in PHILOSOPHICAL DIMENSIONS OF PRIVACY: AN ANTHOLOGY 300 (Shoeman ed., 1982).

[6] See, e.g., Paul M. Schwartz, *Privacy and Democracy in Cyberspace,*52 VAND. L.Rev. 1609, 1653 (1999) (database privacy is necessary for democratic deliberation); Julie E. Cohen, *Examined Lives: Informational Privacy and the Subject As Object*, 52 Stan. L. Rev. 1373(2000) (privacy is necessary for a thriving civil society, free expression, and collective comfort); James P. Nehf, *Recognizing the Societal Value in Information Privacy*, 78 Wash. L. Rev. 1, 69-71 (2003) (privacy is necessary to the proper functioning of a democratic political system); George Ashenmacher, *Indignity: Redefining the Harm Caused by Data Breaches*, 51 Wake Forest L. Rev. 1 (2016) (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, Daniel J. Solove, *Conceptualizing Privacy*, 90 Cal. L. Rev. 1087 (2002) (discussing the private and public domains of privacy protection).

practices lucidly illustrated the need for a data pollution perspective. When Facebook lent political groups access to its personal database in a way that potentially distorted voting decisions, the negative effect was not sufficiently measured by any real or hypothetical injury to the specific individuals whose data was used and whose voting was affected. The critical negative effect was far broader, measured by the harm to the entire electoral and political ecosystem and to the people affected by the voting outcome.

The concept of data pollution helps organize three distinct contributions this article makes. The first contribution is to identify and characterize the scope of data's social harm problem. A vast literature has combed through every aspect of the private harms from data collection—the potential privacy injuries to the people whose data is collected. The externality problem, however, has been entirely neglected: how the participation of people in data-harvesting services affects others, and the entire public. Part I exposes the various faces of this external, societal, effect. It distinguishes data's previously unrecognized social harm from its widely notice private harm, thus beginning to build a new and complementary justification of its regulation. The discussion in part I also helps solve a profound puzzle—how to reconcile the widely shared unease about personal data collection with the widely exhibited indifference by people who continue to "pay with their data." Until now, this misalignment has largely been regarded as a "privacy paradox."[7] Data pollution solves the paradox: people care about data's social harm, about how it affects society as a whole—but not so much about the potential private harm. Privately, they find data's private benefits irresistible.

The second contribution of this article is to recognize and explain the failure of existing legal tools in addressing the problems of data pollution. Part II 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 are primarily due the public nature of the harm. Pollution is an externality; it affects an entire environment, not merely the individuals with whom the polluter transacted, or whose data it emitted. Data pollution, like its industrial ancestor, creates

[7] Benjamin Wittes and Jodie Liu, The Privacy Paradox: The Privacy Benefits of Privacy Threats; Athey et al., The Digital Privacy Paradox: Small Money, Small Costs, Small Talk; Yoan Hermstruwer, *Contracting around Privacy: The (Behavioral) Law and Economics of Consent and Big Data*, 8 J. Intell. Prop. Info. Tech. & Elec. Com. L. 9, 17 (2017). For attempts to explain the privacy paradox as a problem of asymmetric information, see A. Michael Froomkin, *Regulating Mass Surveillance as Privacy Pollution: Learning om Environmental Impact Statements*, 2015 U. Ill. L. Rev. 1713, 1732-35 (2015); Alessandro Acquisti et al., *Privacy and human behavior in the age of information*, 347 Science 509 (2015) (discussing the uncertainty and complexity of privacy decisions as the cause for behavior unprotective of privacy).

harms to the public, and by the time specific individuals are injured it is hard to identify the cause or the full magnitude of harm. For the same reasons that, historically, plaintiffs had difficulties attributing their pollution-caused diseases to specific industrial emissions,[8] data-misuse victims are unable to prove the injurious causal links. And even when causation is established, the magnitude of the harm suffered by specific plaintiffs and its population-wide scale are too speculative for the administration of private law remedies.

I further argue in Part II 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 failed to reduce industrial emissions, markets for digital products are failing to give meaningful attention to reduction of data pollution. People are *not* contracting over data pollution for a variety of reasons, but primarily because it is an externality—a public good—and private contracts are the wrong institution to solve the depletion of a commons. The pervasive hope that contracts and behaviorally-informed choice architecture would help people manage the dissemination of their personal data to reduce the risks of data pollution is fundamentally misguided. It is not people who need to be protected from a web of data-predatory contracts; rather, *it is the ecosystem that needs to be protected from the data sharing contracts that people casually enter*.

If Part I offers a new diagnosis of data's harm, and Part II explains the failure of existing approaches to address this harm, Part III presents the third, and most important contribution of this article: developing an alternative regulatory paradigm 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 eclectically justified.[9]

---

[8] Donald N. Dewees, The Role of Tort Law in Controlling Environmental Pollution, Canadian Public Policy XVIII 425, 429 (1992).

[9] 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, their focus was on privacy harms and privacy law—how data collection causes private injuries to the people whose data is collected. Closest to the analysis of this article are Dennis D. Hirsch, *Protecting the Inner Environment: What Privacy Regulation Can Learn from Environmental Law*, 41 Ga. L. Rev. 1; A. Michael Froomkin, *Regulating Mass Surveillance as Privacy Pollution: Learning om Environmental Impact Statements,* 2015 U. Ill. L. Rev. 1713 (2015); and Nehf, supra note 6. Like the analysis here, Hirsch and Froomkin each 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, at 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, at 28 ("companies benefit from the information they collect, but

The primary regulatory method for pollution control is 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 production 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 transferred, when must it be deleted, and more. Such quantity limitations are increasingly favored by European privacy regulators,[10] and in some narrow bands of U.S privacy law (for example, in dealings with children).[11] Quantity restrictions are the archetypical command-and-control regulation, and they are generally 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.

Recognizing the dilemma of quantity restrictions, Part III then turn to examine another central technique to control pollution: pricing the social cost. It is widely thought that "Pigouvian taxes" on the activity, on the fuels that make it run, or on the output products it generates, could correct the distortion produced by a negative externality. In industrial production, the most common application of this approach is a carbon tax, and in the digital economy it would be a data tax. The social cost of private data collection could be internalized through a tax on the activity of personal data collection, or on the data itself. Part III 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 is important to note, at the outset, that the data tax approach varies dramatically from recent proposals to require firms to pay people for their personal data.[12] Pay-for-data is a zero-sum transfer

---

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." Both Hirsch and Froomkin look to command-and-control regulatory devices used in environmental law as models for data-harms regulation, but because they view the harms in the digital economy as primarily private and privacy-related (what they call the "inner environment" or "privacy pollution"), their analyses of the regulatory methods lead them to different conclusions than the ones discussed in this article. Nehf, 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. At 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., at 79-80.

[10] European Directive on Data Protection, Council Directive 95/46, art. 25, 1995 O.J. (L 281) 31, 56-57 (EC), and the second is the General Data Protection Regulation (GDPR), Commission Regulation 2016/679, 2016 O.J. (L 119) 1, 60--62 (EU).

[11] Children's Online Privacy Protection Act of 1998 (COPPA), 15 U.S.C. § 6501.

[12] Eric Posner and Glen Weyl, Radical Markets __ (2018); Brittany Kaiser, Facebook

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 utilize ex-post devices, to be triggered in the aftermath of harmful data emissions. Like toxic waste releases from industrial production, data spills are rapidly becoming a major social problem of the digital era. Environmental law uses various tools to shift the harm from 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 can be reduced by a proper system of deterrence. Liability equal to the social cost of spills (punctuated by compulsory liability insurance) would lead to better precautions and self-regulation.

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 is harvested. Part III begins to correct his oversight, exploring a regulatory design for data pollution law that if founded on the perspective of social cost. Environmental law and regulations were born in the industrial era because private law dealing with private harms failed to protect public goods and the environment.[13] We now need a twenty-first century version of environmental law—data pollution regulation—to expand the focus and begin to address the *public* harms from the personal data enterprise. This article offers a roadmap for such transformation.

I. 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

should pay its 2bn users for their personal data, Financial Times (April 9, 2018).

[13] KENNETH S. ABRAHAM, THE LIABILITY CENTURY 149 (2008).

privacy harm. At times, the emotional injury to victims is clear and present. For example, when the database of a website used by people to find partners for extramarital affairs is hacked, the potential privacy harm to the millions of people whose information leaked is significant.[14] Such episodes of clear and present emotional harm have helped sustain a ubiquitous premise—that the injury arising from the assembly of databases containing loads of personal information is personal and private in nature. While some privacy theorists 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 undermining a "thriving civil society"[15]—the public harms they identify are still derivatives of the personal damage to the individuals whose private information is taken.

The ideas that the data's problem is privacy, and that the solution is privacy protection, are alluring because the databases firms deploy consist primarily of private information that people don't always 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."[16] If the problem is that smart devices and apps are "spying on you (even in your own home),"[17] then the first harm that comes to mind is personal in nature and the obvious redress to this threat 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."[18] Despite the vast attention lawmakers and advocates lavish on

---

[14] See, e.g. Simon Thomsen, *Extramarital affair website Ashley Madison has been hacked and attackers are threatening to leak data online*, Business Insider, 21 July 2015.

[15] See, e.g., Paul M. Schwartz, *Privacy and Democracy in Cyberspace,*52 Vnd. L. Rev. 1609, 1653 (1999) (database privacy is necessary for democratic deliberation); Julie E. Cohen, *Examined Lives: Informational Privacy and the Subject As Object* , 52 Stan. L. Rev. 1373(2000) (privacy is necessary for a thriving civil society, free expression, and collective comfort); James P. Nehf, *Recognizing the Societal Value in Information Privacy*, 78 Wash. L. Rev. 1, 69-71 (2003) (privacy is necessary to the proper functioning of a democratic political system); George Ashenmacher, *Indignity: Redefining the Harm Caused by Data Breaches*, 51 Wake Forest L. Rev. 1 (2016) (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, Daniel J. Solove, *Conceptualizing Privacy*, 90 Cal. L. Rev. 1087 (2002) (discussing the private and public domains of privacy protection).

[16] Jacob Silverman*, Just How 'Smart' Do You Want Your Blender to Be?,* New York Times Magazine, (June 14, 2016), https://www.nytimes.com/2016/06/19/magazine/just-how-smart-do-you-want-your-blender-to-be.html?_r=1.

[17] Joseph Steinberg, *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/#2f4fbcecb859.

[18] Benjamin Wittes and Jodie Liu, The Privacy Paradox: The Privacy Benefits of Privacy Threats; Athey et al., The Digital Privacy Paradox: Small Money, Small Costs, Small Talk;

privacy risks and privacy protection, and despite widespread popular sentiment—documented through survey evidence[19]—that data privacy matters, people largely behave as if it does not. They say that they greatly value their personal data, but they turn around and give it up for meager quid pro quo.[20] The revealed preference for data privacy is distinctly lower than the declared valuation. There is not much evidence, in short, that the privacy concerns commonly articulated—emotional health, personal dignity, autonomy, participation, or expression—are impaired by the digital data enterprise in a discernable, measurable, manner.

And yet, the concerns over the harm from data technology are 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. The American consumer financial system has been jolted by the massive leakage of consumers' personal and financial data and its potential fraudulent misuses. And major jurisdictions around the world are enacting widely popular deep reforms intended to make data collection more difficult.[21] The the longstanding concerns with 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 among people to sharing their own private data? This is perhaps the most fundamental question haunting the field of data privacy law, and a variety of explanations have been proposed.[22] An explanation not yet

---

Yoan Hermstruwer, *Contracting around Privacy: The (Behavioral) Law and Economics of Consent and Big Data*, 8 J. Intell. Prop. Info. Tech. & Elec. Com. L. 9, 17 (2017).

[19] Wendy Pollack and Mike Sullivan, *The Information Subscribers Most Likely to Pay for Google Among Tech Services*, The Information, April 20, 2018; *EMC Privacy Index*, Dell (2014) https://www.emc.com/campaign/privacy-index/global.htm (cross-country survey of consumer perceptions and attitudes and valuation of data privacy); *New Survey Finds Deep Consumer Anxiety over Data Privacy and Security*, IBM News Room (Apr. 16, 2018) http://newsroom.ibm.com/2018-04-16-New-Survey-Finds-Deep-Consumer-Anxiety-over-Data-Privacy-and-Security ("85 Percent of Consumers Say Businesses Should Be Doing More to Actively Protect Their Data"); Timothy Morey et al., *Customer Data: Designing for Transparency and Trust,* Harvard Business Review (May 2015).

[20] Susan Athey et al., *The Digital Privacy Paradox: Small Money, Small Costs, Small Talk*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2916489: Alessandro Acquisti, Leslie K. John & George Loewenstein, "What is Privacy Worth?" The 42 Journal of Legal Studies 249 (2013); Lior J. Strahilevitz and Matthew B. Kugler, *Is Privacy Policy Language Irrelevant to Consumers?*, 45 J. Legal Stud. (2016).

[21] The California Consumer Privacy Act of 2018, A.B. 375.

[22] See A. Michael Froomkin, *Regulating Mass Surveillance as Privacy Pollution: Learning om Environmental Impact Statements*, 2015 U. Ill. L. Rev. 1713, 1732-35 (2015); Alessandro Acquisti et al., *Privacy and human behavior in the age of information*, 347 Science 509 (2015) (discussing the uncertainty and complexity of privacy decisions as the

proposed, and the 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 continue to share their data.

Thus, separate from the thoroughly discussed question whether data's private injuries are significant, it is necessary to investigate data's external harm. The effects of a database consisting of personal information could be felt by an entire ecosystem, not merely by those whose data is misused or emitted. Accordingly, the remainder of this section examines the patterns by which collection of personal data affects the public—how data pollutes.

A. Effects on Public Interests

Industrial pollution degrades a public good. It is the quintessential negative externality, afflicting many who are not part to the polluting activity. It impacts an ecosystem as a whole, as well as the health of many individuals. The primary method to measure the latter impact is to add up how many individuals would be affected and how severely, but this summation of private effects is merely an accounting technique to capture a concrete component of the social impact.

Emissions of data are like emissions of pollutants: the costs are often external, degrading social interests. A digital database is not like the library card catalog of generations past, which at the time was merely the simple indexed sum of individual items of information. A digital data base is super-additive: new things can be learned that were not known when the information was atomized, which affect other parties. The database can reveal previously unknown *qualitative properties* that implicates the public interest. Or it may reveal information about individuals, other than those the data base, in a way that harms 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 news more effectively, the infected public interest was the integrity of the voting process. This effect reaches far beyond the private

---

cause for behavior unprotective of privacy); Alex Matthews and Catherine Tucker, *Government Surveillance and Internet Search Behavior*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2412564 (finding a chilling effect on search behavior from salient reports about government surveillance on the Internet, in the aftermath of Snowden's discoveries.

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 vote, not experiencing any personal harm.)

A second example illustrating how a database can reveal information affecting public interests other than the user's 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, 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.[23] What else can a cluster of physical workouts in the Sahara Desert 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.

The concern about public harm from databases is reflected in governments' efforts to limit cross-border transfers of commercial databases by establishing data exit controls and requiring "data localization."[24] The concerns driving such policies range from national security and law enforcement to trade protection and domestic industry prop up.[25] 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.[26] 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.[27]

The potential for databases to be used in ways that harm public goods and publicly shared values is illustrated in data's ability to personalize treatment and enable new forms of harmful discrimination. In general, the correlations in the database teach things about people that their individual

---

[23] Richard Perez-Pena and Matthew Rosenberg, *Strava Fitness App Can Reveal Military Sites, Analysts Say*, New York Times, (January 29, 2018); Daniel Brown, *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) (showing heat maps).

[24] Bret Cohen et al*, Data Localization Laws and Their Impact on Privacy, Data Security and the Global Economy*, 32(1) Antitrust 107 (2017).

[25] Id., at 108.

[26] The Cybersecurity Law, Art. 37 ("personal information and important data collected and generated by critical information infrastructure operators operating within the borders of the People's Republic of China should be stored within China.").

[27] See Hong Yanqing, *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) ("For instance, hostile forces could combine the data with other datasets and use various analysis methods such as data mining to gain information that could threaten national security.")

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.[28]

But when the correlations inferred from the digital database allow for personalized treatments that harm groups of people, the discriminatory impact is socially undesirable. For example, when online ads promoting STEM careers are shown less often to women than to men,[29] or when online ads suggestive of arrest records appear more often along search results of black-sounding names,[30] 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 restricting their placement to black-sounding names. Such campaigns are made possible by personalized data analytics, and they merely respond to people's demand for information.[31] 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 helps optimize discrimination.

It is not always easy to distinguish discrimination from its close relative personalization—another form of tailored treatment but one that often creates much good. Personalized medicine, education, and nutrition helps 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

---

[28] See, e.g., Amalia R. Miller and Catherine Tucker, *Frontiers of Health Policy: Digital Data and Personalized Medicine*, 17 Innovation Policy and the Economy 49, 51-54 (2017) (Digital medical records reduce neonatal mortality).

[29] Anja Lambrecht and Catherine Tucker, *Algorithmic Bias? An Empirical Study into Apparent Gender-Based Discrimination in the Display of STEM Career Ads* (https://www.ssrn.com/abstract=2852260). See also Amit Datta et al., *Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination,* In Proceedings on Privacy Enhancing Technologies, (PoPETs), 2015.

[30] Latanya Sweeney, *Discrimination in Online Ad Delivery*, 11(3) ACMQueue 10 (2013).

[31] See, Generally Amit Datta et al., *Discrimination in Online Advertising A Multidisciplinary Inquiry*, 81 Proceedings of Machine Learning Research 1 (2018) (surveying the context and causes for online discriminatory advertising).

incentives and technical tools to commercialize and monetize them.[32] The negative externalities, in contrast, remain orphan. And the negatively affected groups are too broad and dispersed, suffering injuries that are too abstract for private remedies to be effective. Besides, the toxicity from discriminatory treatment degrades the environment also for those not discriminated.

B. Effects on Other People

Pollution could have negative external effects not just on an ecology, but also on identified victims. An oil spill, for example, can affect the livelihood of a specific group of fishermen and businesses in the vicinity. Similarly, the externalization of costs associated with the digital data enterprise could occur through mechanisms that affect not an entire system but only identified individuals—different than the data givers.

The most common way such external effect occurs is when users give specific information about others. 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 internal, and governed by their user agreements. 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 user? The content of their messages too is viewed and collected by Google, as byproduct of the authorization Gmail users give. 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 external effect of the Gmail transaction. If non-Gmail users were able to reduce contact with Gmail users, the external effect would be internalized. But such "Coasian" selective contracting is defeated by a host of transactions costs.

Another example of data affecting others 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 within the users' circles of biological relationships, who never agreed to give to participate in such personal-origins discovery. Possibly, that information could be life-saving to the third-party relatives. It could also be socially desirable when "genetic informants" help solve crimes or when the data help reunite families.[33] But the information could also affect others

---

[32] Important footnote on how competition can dissipate these profits and lead to under-investment.

[33] See *Took an ancestry DNA test? You might be a 'genetic informant' unleashing secrets about your relatives*, USA Today (April 27, 2018), at https://www.usatoday.com/story/tech/nation-now/2018/04/27/ancestry-genealogy-dna-test-privacy-golden-state-killer/557263002/ (cases of solved crimes); *After 60 years of*

negatively, especially in circumstances where genetic anonymity is crucial.[34]

Finally, consider a social network that gains authorized access to its users' data, which includes valuable information about these users' "friends"—including to those who try to limit their exposure. Notwithstanding the effort to anonymize, this penumbra of contacts becomes target to the various targeted services that the social network disseminates.[35] Short of exiting the networks altogether, there is little that the affected third-parties could do. Their efforts are undermined once the data is harvested through the portals of their friends. 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 matters for the design of the regulatory response. Public law remedies, like quantity restrictions or taxes, may work for both categories of externalities, as discussed in Part III 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.

C. Precaution and Insurance Externalities

Data's external effects also arise when the prevention and reduction of the harm, rather than the harm itself, is a public good. Victims of pollution are often part of a large pool who share common exposure that depends on the aggregate level of prevention, which in turn depends on the activity of each member of the pool. In the environmental context, people who could

---

friendship, they learned they're biological brothers, CNN (December 27, 2017), at https://www.cnn.com/2017/12/27/health/friends-brothers-dna-discovery-hawaii-trnd/index.html (reunited family).

[34] Kiley Crossland, *The hidden risks of at-home DNA testing*, World (January 5, 2018), at https://world.wng.org/content/the_hidden_risks_of_at_home_dna_testing

[35] Often, third-party apps allowed users to log in using their Facebook account. 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 like their name, location, email or friends list. Facebook enabled this practice, and then suspended it in 2015, but the third parties did not have to delete the previously collected data. *See* Josh Constine, "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/#.8fhohr:byBc; Paul Lewis, "'Utterly horrifying': ex-Facebook insider says covert data harvesting was routine," The Guardian (Mar. 20, 2018) https://www.theguardian.com/news/2018/mar/20/facebook-data-cambridge-analytica-sandy-parakilas.

deploy private emission-prevention measures fail to engage in the optimal levels, discounting the impact their effort on other people.

A public good problem could arise even when prevention measures have 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).[36] 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 the insurance externality 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 post-breach remediation efforts. But they are largely insured against these private fraud-related losses, through various statutory insurance programs[37] and covered for the residual loss through typical homeowners insurance policies.[38] The economic costs of data spills are significant,[39] but only a small fraction of it is borne by the consumers whose data is stolen.

The divergence between the private and social cost of data spills

---

[36] Insurance contract can mitigate and even overcome the moral hazard problem by creating incentives for care. See, Generally, Omri Ben-Shahar and Kyle Logue, *Outsourcing Regulation: How Insurance Reduces Moral Hazard*, 111 Michigan Law Review 197 (2012). I discuss below how some form of insurance can substitute for public regulation of data pollution. See infra, text accompanying note 161.

[37] Justin C. Pierce, *Shifting Data Breach Liability: A Congressional Approach*, 57 WM. & MARY L. REV. 975, 982 (2016) (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* N. Eric Weiss & Rena S. Miller, CONG. RESEARCH SERV. [CRS], R43496, THE TARGET AND OTHER FINANCIAL DATA BREACHES: FREQUENTLY ASKED QUESTIONS *(2014), https://www.fas.org/sgp/crs/misc/R43496.pdf*; "Lost or Stolen Credit, ATM, and Debit Cards," Consumer Information, FEDERAL TRADE COMMISSION, https://www.consumer.ftc.gov/articles/0213-lost-or-stolen-credit-atm-and-debit-cards#Limit (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).

[38] Standard home owners insurance policies cover unauthorized use of credit card or fund transfer, including forgery. See HO3, Section I.E.6 (https://www.iii.org/sites/default/files/docs/pdf/HO3_sample.pdf). Identity theft insurance is an optional endorsement available under a homeowner's policy See, e.g., "Identity Fraud Expense Coverage," Liberty Mutual Insurance (https://www.libertymutual.com/identity-theft-insurance).

[39] Infra, at __

means that consumers pay indirectly for the protection they receive. They are insured, for example, against credit card fraud,[40] but if banks and merchants have to bear these costs when the data are emitted, they charge their customers higher fees for credit cards services and higher prices for products. Ultimately, the cost is borne by consumers. 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 it pays. The incentive to pay for anti-spill precautions 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.[41] 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, I began to assemble the argument that data's harm is public. It is public not in the derivative, secondary, sense that the privacy literature suggested—the deeply personal privacy injuries that are thought to demoralize and degenerate the civic functioning of individuals and possibly impoverish public spheres and institutions.[42] Instead, the harm is directly affecting public ecosystems, and it is often unrelated to any impact on the specific individuals whose data is used. The digital economy creates digital smog, and the question is what to do about it.

II. THE FAILURE OF PRIVATE LAW

Part I identified a problem—data pollution—that Parts II and III will now try to solve. I begin in Part II by explaining what NOT to do—what regulatory approaches are not suited to deal with data pollution. It is a

---

[40] 15 U.S.C. § 1643(a) (limiting cardholder liability for unauthorized use to $50). See also *Lost or Stolen Credit, ATM, and Debit Cards*, Federal Trade Commission, https://www.consumer.ftc.gov/articles/0213-lost-or-stolen-credit-atm-and-debit-cards#Limit.

[41] Timothy Morey et al., *Customer Data: Designing for Transparency and Trust,* Harv. Bus. Rev. (May 2015), (survey data demonstrating consumers "deeply anxious about how their personal information may be used").

[42] Supra note 6.

necessary first step, because it appears that much of the current regulatory response falls into this category of ineffective law. Specifically, I argue that optimal control of data pollution cannot be guaranteed by private law and by enactments intended to facilitate private precaution. This discussion will subsequently be followed in Part III by a set of ideas on more effective solutions.

The failure of private law in the data pollution area is remarkable, because personal rights in data 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. When the baseline rights are so crystal clear, and when contracting over them is so explicit and rampant, why is private law failing? In this section, I show that private law's shortcoming in the data pollution realm is an exact replay of its failure to regulate industrial pollution.

A. Failure of Contracting

People care about their data ecosystem.[43] Usually, when consumers care about an attribute of a product, firms compete to provide it. We face a puzzle, then: whey is data pollution not subject to preference-satisfying contracts? Why is the prevention of data emissions not bargained for?

Some firms offer a menu of data control options to their customers. They offer, for example, "premium" services in which customers may pay with money instead of data.[44] Indeed, 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.[45] And firm do, indeed, write contracts that give themselves such licenses. Every website, app, or store has a "data policy" that explains to consumers what data is collected and how it is used. The market environment is sizzling with intensive contracting over data and with endless opportunity

---

[43] Wendy Pollack & Mike Sullivan, "The Information Subscribers Most Likely to Pay for Google Among Tech Services," The Information (Apr. 20, 2018).

[44] See "AT&T Charges Steep Premium for Privacy, Calls it a 'Discount'", DSL Reports (March 17, 2016), available at https://www.dslreports.com/shownews/ATT-Charges-Steep-Premium-for-Privacy-Calls-it-a-Discount-136511

[45] *See, e.g.,* Electronic Communications Privacy Act, 18 U.S.C.S. §§ 2510–2522 (2018); Health Coverage Availability and Affordability Act of 1996., 110 STAT. 1936 or 45 C.F.R. § 164.502 (2018); Personal Information Protection Act, 815 ILL. COMP. STAT. 530/1 (2006); California Financial Information Privacy Act, 2003 CAL. ADV. LEGIS. SERV. 241 (2003) (prohibiting financial institutions from sharing or selling personally identifiable nonpublic information without obtaining a consumer's consent); California Online Privacy Protection Act Cal. BUS. & PROF. CODE § 2275 (Deering 2018).

to protect personal data. Why, then, is there so much data pollution?

For the same reason that people do not contract over environmental 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 environmental emissions (and over public goods more generally), my focus below is to show how these factors apply to the data pollution context.

1. Externalities

Pollution harms people not side to the transaction. The production of meat, for example, emits toxic waste into the air, water, and ground, 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.[46] 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[47]—the reaction is rarely calibrated to reflect the magnitude of the harm.

I argued in Part I 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 the pollution side of data externalities. True, people are

---

[46] COMMITTEE ON A FRAMEWORK FOR ASSESSING THE HEALTH, ENVIRONMENTAL, AND SOCIAL EFFECTS OF THE FOOD SYSTEM ET. AL, A FRAMEWORK FOR ASSESSING EFFECTS OF THE FOOD SYSTEM, *Chapter 4: Environmental Effects of the U.S. Food System* (Malden C. Nesheim et. al eds., National Academic Press, 2015) *available at,* https://www.ncbi.nlm.nih.gov/books/NBK305182/ (reviewing the environmental effects of food production systems and discussing how agricultural production systems may in many instances deplete natural resources of land and water, disturb ecosystem balance, involve the use of environmental contaminants such as pesticides and nitrogen that pollute the natural environment, and present challenges to human health); Jeff Kohn & Kelsey Kruger, "Understand pollution, environmental impacts from food in 6 charts," GREENBIZ: P2 IMPACT (Nov. 17, 2016), available at, https://www.greenbiz.com/article/understand-pollution-environmental-impacts-food-6-charts (pointing to EPA Toxic Release data to demonstrate the pollution impact of various food manufacturing sectors).

[47] G. Clay Whittaker, "Edible Six-Pack Rings Could Make the Ocean Safe Again," POPULAR SCIENCE (May 19, 2016) https://www.popsci.com/six-packs-could-be-ocean-safe-again; Andrew Menke, "Reusable Water Bottle Market Pushes Forward," GLOBALEDGE (Dec. 5, 2017) https://globaledge.msu.edu/blog/post/54514/reusable-water-bottle-market-pushes-forw; Tatiana Homonoff, "Paper or plastic? How disposable bag bans, fees and taxes affect consumer behavior," THE CONSUMER (Nov. 17, 2015).

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 chose them, and rarely display any affirmative interest to bargain over data pollution. Legal default rules that prohibit companies from harvesting personal data are thoroughly and methodically reversed—because consumers don't seem to care enough about the potential privacy harm, and have no incentive to do much about the public harm.

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 (namely, when 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.[48] Likewise, cloud storage services that invite people to deposit their entire records for remote safekeeping implement tighter data security.[49] Here, contracts go out of their way to provide effective pollution regulation—but only because the harm is purely private.

2. Information

People may fail to contract over optimal pollution emission even when the harm is internal, because of misinformation.[50] 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.[51] Economists sometimes refer to goods that

---

[48] Florencia Marotta Wurgler, *Self-Regulation and Competition in Privacy Policies*, 45 J. Legal Stud. S13, S__ (2016).

[49] Id., at __.

[50] Froomkin, supra note __, at 1732-37 (describing 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).

[51] "Trans Fats**,"** American Heart Association (Apr. 21, 2018) ("Before 1990, very little was known about how *trans* fat can harm your health. In the 1990s, research began identifying the adverse health effects of *trans* fats. Several countries (e.g., Denmark, Switzerland, and Canada) and jurisdictions (California, New York City, Baltimore, and Montgomery County, MD) have reduced or restricted the use of *trans* fats in food service establishments."); "Risks of Breast Implants," U.S. Food and Drug Administration; Anna Rogers, "Breast Implants: The Ticking Time Bomb in Millions of Women's Bodies,"

have such hidden properties as either "experience goods" (hard to observe the harm before purchase), or "credence goods" (hard to observe even after purchase and consumption).[52] Because many harmful effects (or warranted benefits) from products gestate slowly and manifest latently, uncertainty over the true causes leads to contracting failure.[53]

Data security is a credence good. Consumers cannot know how lavish the data sharing and how loose the security practices—until there is a sharing or security crisis. They are not even aware what data is collected and by whom,[54] and what externality it might cause. When their personal data is emitted and a disruption occurs, consumers have difficulty knowing the source of the emission and whether it was caused by sloppy practices. Many emissions turn out to be harmless; and data-related harms can flow from one of many potential causes (perhaps the same personal data was shared with numerous websites, making it difficult to ascertain which one was the emitter)—further blurring the information about the quality of data protection that each business provides. 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 missing information by relying on informed intermediaries. People who care about environmental pollution could seek certifications and rating by ISO,[55] and people who care about data emissions may similarly consult TRUSTe.[56] 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*.[57] (It is

---

Collective Evolution (Oct. 21, 2015).

[52] Uwe Dulleck, Rudolf Kerschbamer, and Matthias Sutter. "The Economics of Credence Goods: An Experiment on the Role of Liability, Verifiability, Reputation, and Competition." *The American Economic Review* 101, no. 2 (2011): 530–32.

[53] See, generally, Karen Bradshaw, *Information Flooding*, 48 Indiana L. Rev. 755, 765 (2015).

[54] A. Michael Froomkin, *The Death of Privacy?*, 52 Stan. L. Rev. 1461, 1501-02 (2000).

[55] https://en.wikipedia.org/wiki/Environmental_certification

[56] TRUST e provides privacy risk assessments and certifications. *See* https://www.trustarc.com/products/enterprise-privacy-certification/.

[57] For example, one of the most objective grading services is PrivacyGrade.org designed by Carnegie Mellon University. It measures "the gap between people's expectations of an app's behavior and the app's actual behavior." 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? See http://privacygrade.org/apps/search?utf8=%E2%9C%93&q=facebook;

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.[58]) 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 quality features that create marketing advantages than to expensively bolster the hidden traits. Besides, firms may 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.[59] 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. 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 to status quos; they are averse to making any inquiry or decision; and more.[60] 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 to 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

---

http://privacygrade.org/apps/com.strava.run.html.

[58] 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, supra note __, at 65.

[59] Cite literature on this effect.

[60] Dominic Johnson & Simon Levin, *The Tragedy of Cognition: Psychological Biases and Environmental Inaction* 97 CURRENT SCIENCE 1593 AT 1597 (2009) (exploring the impact of psychological biases on preferences, perceptions and reactions to environmental change).

are similarly prone to imperfect rationality. Even more than chemical toxicity, digital risks are hard to ascertain.[61] There are no digital illnesses or deaths; the dimensions of the data emissions risks are many and complex, ripe for endless behavioral biases.[62] The manifestations of data-caused harms are sometimes subtle and easy to disregard; and 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[63]), the underlying issues remain abstruse, confusing, and constantly shifting. Ironically, researchers found that the very presence of a "privacy notice" document soothes consumers' privacy worries and causes them to trust a website more.[64] 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 drives consumers to often 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.[65] The overload problem within each transaction, and the accumulation across multiple transactions, are problems too implacable to solve in a world of private contracting. And they are made exponentially more difficult because similar attention is required to address other daily contracting risks, some much more urgent. In environments cluttered with layers upon layers of technical information, who is to say that ignorance and inattention are irrational?

---

[61] See Nehf, supra note __, at 62.

[62] Alessando Acquisti, Laura Brandimarte, and George Loewenstein, *Privacy and Human Behavior in the Age of Information*, 347 Science 509 (2015); Adjerid et al., JLS 2016

[63] See, e.g., Carlos Jensen and Colin Potts, *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 (2004).

[64] Yue Pan and George M. Zinkhan, *Exploring the Impact of Online Privacy Disclosures on Consumer Trust*, 82 J. Retailing 331–38 (2006); Joseph Turow, *The Federal Trade Commission and Consumer Privacy in the Coming Decade*, 3 Journal of Law and Policy for the Information Society 723, 730 (2008).

[65] Ben-Shahar & Schneider, Ch 6 ("The Quantity Problem"). 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 (McDonald and Cranor, 2008).

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 anti-sharing nudge. For example, enacting data-protective prompts in the form of legal default rules has proven futile in many contexts because the default rules are always one click away from deletion—a click that companies are eager to encourage.[66]

Alternatively, enacting data-protective mandatory rules could be effective, but this means (paradoxically) that the only way for contract law to overcome the problems of contract failure to remove the matter from the bounds of permissible contracting. How to design such mandatory rules is the focus of Part III of this article. For private law to continue to have relevance in such environments of mandatory data pollution rules, victims have to be entrusted with enforcement powers. I therefore turn to examine why private enforcement of non-disclaimable anti-pollution rights in data fails.

B. Failure to Tort Law

Part II.A 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 consent, use it in impermissible ways, negligently fail to secure it, or engage in deceptive data practices. All of these violate people's private rights, and should in principle be redressed by tort law.

But tort law fails to address these wrongs, for the same reason that it historically failed to redress many environmental wrongs. In theory, nuisance law was available to deal with pollution. But it is widely recognized that tort liability did not result in reduction of environmental pollution to optimal levels.[67] Here, I focus on three primary reasons why tort law failed to deter and compensate industrial pollution harms: causation, valuation, and societal externalities. I argue that these reasons are equally central in tort law's failure

---

[66] Ben-Shahar and Strahilevitz, JLS 2016. See, generally, Margaret Jane Radin, Boilerplate: The Fine Print, Vanishing Rights, and the Rule of Law (2013); Lauren E. Willis, *When Nudges Fail: Slippery Defaults*, 80 U.of Chi. L. Rev. 1155 (2013).

[67] Abraham, The Liability Century 149.

to control data pollution.

1. Causation

Tort law is effective when harm is immediate and visible. Historically, the threat of tort liability for pollution failed to properly reduce its levels because the harms from industrial emissions are neither immediate nor visible. Neighbors to pollution can show that they are exposed to a new *risk*, but have hard time showing that they suffered actual *harm*.[68] And environmental liability suffers from an acute problem of "long tail"—latent harms that are difficult to causally match with precise wrongs.[69]

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.[70] 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*).[71] It could take years for the misuse of the data to occur, and by then it would be hard in any individual case 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 gestation and the uncertainty over causal links defeat any attempt to apply a negligence-based regime. But they are also blunt some of the more ambitious proposals to expand the reach of tort law into the data pollution area. It is sometimes thought that a shift from negligence to a strict

---

[68] Donald N. Dewees, The Role of Tort Law in Controlling Environmental Pollution, Candadian Public Policy XVIII 425, 429 (1992).

[69] Abraham 139; David Rosenberg, *The Causal Connection in Mass Exposure Cases: A "Public Law"*

Vision of the Tort System, 97 HARV. L. REV. 849, 919 (1984); Daniel C. Esty, *Environmental Protection in the Information Age*, 79 N.Y.U. L. REV. 115, 131 (2004);

[70] 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 negligent in a manner analogous to the inadvertent preventable releases of environmental pollutants.

[71] See, eg., Indep. Cmty. Bankers of Am. v. Equifax, Inc., N.D. Ga.; see also Jimmy H. Koo, *"Equifax Negligent in Data Breach, Community Banks Allege,"* CLASS ACTION LITIGATION REPORT, BLOOMBERG BNA (Dec. 12, 2017).

liability regime would make firms more accountable for data emissions.[72] 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."[73] Unfortunately, strict tort liability still requires proof of causation. If it is difficult to identify the causal chain connecting particular data emission with specific victims, the desired deterrent and activity-regulating effects of strict liability would not occur. Indeed, data breach tort suits sometimes fail despite glaring carelessness of the targeted companies.[74] It is the difficulty of proving the harm, not the negligent conduct, 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 of the latency of harms.[75] If the data were handy—and perhaps in the context of data security breach they are (those whose data were 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 scheme, if based on good actuarial information, could provide optimal deterrence.[76] For example, the Justice Department estimates that an average loss to victims of identity theft is $1500.[77] Adjudicating 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,[78] they are unlikely to succeed in data pollution lawsuits.

---

[72] Danielle Keats Citron, *Reservoirs of Danger: The Evolution of Public and Private Law at the Dawn of the Information Age*, 80 S.Cal L.Rev. 241 (20070.

[73] Citron, at 266.

[74] Cite caselaw.

[75] Viscusi, Forward xi, in Cutting Green Tape; Christopher H. Schroeder, *Lost in the Translation: What EnvironmentalRegulation Does That Tort CannotDuplicate,*41 WASHBURN L.J. 583, 601 (2002) (explaining the difficulty of tracing the harm to its cause); Albert C. Lin, *Beyond Tort: Compensating Victims of Environmental Toxic Injury*, 78 S. CAL. L. REV. 1439, 1452 (2005) ("Ultimately, the judicial struggle with causation reflects the inherent tension between traditional causal analysis and modern science's probabilistic understanding of causation").

[76] Shavell, The Costs of Accidents, pp. &&.

[77] Erika Harrell, Victims of Identity Theft, 2014 (U.S Department of Justice, Revised Nov. 13, 2017), available at https://www.bjs.gov/content/pub/pdf/vit14.pdf

[78] Troyen A. Brennan, *Causal Chains and Statistical Links: The Role of Scientific Uncertainty in Hazardous-Substance Litigation*, 73 CORNELL L. REV. 469, 491–93 (1988) ("Courts are troubled by the probabilistic evidence of causation with regard to hazardous

Plaintiffs have been making such claims in data spill lawsuits—but with little success.[79] A tort remedy based on *expected harm* is exceedingly uncommon in courts,[80] 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 FTC or state AGs. Indeed, developing a scheme of exposure-based remedies injunctive relief for data emissions is a prominent motivation of Part III of this article—focusing on public law solutions to data pollution.

2. Valuation

A second problem with tort liability for pollution is valuation. Even when the impact of the emissions is private and 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 various criteria of remoteness.[81] The different *physical* manifestations of injury made it possible to compensate the ones easier to value. It also helped direct liability for emissions towards *restoration*, rather than compensation, which further eases problems of monetary valuation.[82] Thus, even if some private losses from pollution are quantifiable (e.g., loss of fishermen income due to oil spill), other major losses arise from the deterioration of the

---

substance injury … Courts rely on mechanistic notions of causation and are confused by probabilistic ones."). But see *Norfolk & Western Railway Company v. Ayres*, 538 US 135 (2003).[EXPLAIN]

[79] Despite it being central to whether plaintiffs have 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. Daniel J. Solove & Danielle Keats Citron, *Risk and Anxiety: A Theory of Data Breach Harms*, 96 TEXAS LAW REV. i., ii (Forthcoming). For analysis of courts' treatment of claims alleging increased risk of future harm, see David L. Silverman, *Developments in Data Security Breach Liability*, 73 BUS. LAW. 215, 226 (2017).

[80] Some courts refuse to accept statistical evidence in tort suits. *See* David Rosenberg, *The Causal Connection in Mass Exposure Cases: A "Public Law" Vision of the Tort System*, 97 HARV. L. REV. 849, 857 (1984); *Smith v. Rapid Transit, Inc.*, 58 N.E.2d 754 (Mass. 1945) (statistical evidence alone cannot prove bus company's causal role). See also Marcia R. Gelpe & A. Dan Tarlock, *The Uses of Scientific Information in Environmental Decisionmaking*, 48 S. CAL. L. REV. 371, 374 (1974).

[81] Courts use standards of proof to exclude some harms. See Am. Law Inst., 1 ENTERPRISE RESPONSIBILITY FOR PERSONAL INJURY 319-21 (1991) (surveying the relatively low total damages awarded for environmental injuries). Courts also regard some harms as too remote for compensation. See Don N. Dewees et al., EXPLORING THE DOMAIN OF ACCIDENT LAW: TAKING THE FACTS SERIOUSLY 293-94 (1996).

[82] Ben-Shahar and Porat, The Restoration Remedy in Private Law, 118 Colum. L. Rev. (Forthcoming, 2018).

surrounding ecosystem, to which people exhibit varying sensitivity and from which they suffer speculative loss.

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 now well-known privacy paradox.[83] 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 the harm of identity theft resulting from data emission—the perception of financial harm could be drastically different from its reality.

Data emissions lawsuits are regularly confronting 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 most courts have held that such injury is too speculative to be compensated, and denied standing to sue.[84] Even costs incurred by victims of data breach to monitor their financial information were held 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'."[85]

The difficulty to evaluate individual harms and to distribute 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 do not get to collect. Such decoupling of liability and compensation could be achieved, for example, by *cy pres* settlements, whereby the court directs non-distributable portions of class-action settlements to third party beneficiaries that work to advance the interests of the class.[86] 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.[87] Indeed, it is precisely in the context of a

---

[83] Benjamin Wittes and Jodie Liu, The Privacy Paradox: The Privacy Benefits of Privacy Threats; Athey et al., The Digital Privacy Paradox: Small Money, Small Costs, Small Talk.

[84] Solove and Schwartz, Information Privacy Law 960-62 (6th Ed) ("the majority of courts are reluctant to recognize emotional distress as a harm stemming from data breach"); Beck v. McDonald, 949 F.3d 262 (4th Cir. 2017); *Amburgy v. Express Scripts, Inc*., 671 F.Supp.2d 1046 (E.D. Mo 2009); Spokeo; [RA: check]

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

[86] *Nachshin v. AOL, LLC*, 663 F.3d 1034, 1038 (9th Cir. 2011); Principles of the Law of Aggregate Litig. § 3.07n(Am. Law Inst. 2010); Kerry Barnett, *Equitable Trusts: An Effective Remedy in Consumer Class Actions*, 96 Yale L.J. 1591 (1987) (giving case examples).

[87] The questionable constitutional foundations of *cy pres* distributions was acknowledged by Chief Justice John Roberts, noting "fundamental concerns surrounding the

claim over data pollution that the Supreme Court is about to decide the legality of such private enforcement model.[88]

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 Part I, 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 hard to assess the total harm.

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 was a key factor in my explanation above 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 so widespread that tort law finds difficult to control.[89]

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 nuisances, the public trust doctrine, and *cy pres* settlements permit tort recovery for societal harms.[90] Additionally, scholars have suggested innovative ways to expand tort law's private harm model to societal injuries.[91] These anecdotal expansions notwithstanding, tort law still limits private claimants to sue solely for private harms.[92] To collect recovery for

---

use of such remedies in class action litigation." *See* Marek v. Lane, 134 S.Ct. 8, 9 (2013), *cert. denied* (No. 13–136).

[88] The Supreme Court is examining this question in a pending case, *Frank v. Gaos* (No. 15-15858). In that case, Google is sued for sharing user search terms with third parties.

[89] Kenneth **S.** Abraham, *The Relation Between Civil Liability and Environmental Regulation: An Analytical Overview,* 41 WASHBURN L.J. 379, 379 n.2 (2002) (externalities were the reason for shift to public enforcement of environmental law); Henry N. Butler & Jonathan R. Macey, *Externalities and the Matching Principle: The Case for Reallocating Environmental Regulatory Authority,* 14 YALE L. & POL'Y REv. 23, 29 (1996).

[90] See, e.g., Albert C. Lin, *Public Trust and Public Nuisance: Common Law Peas in a Pod*, 45 U.C.D. L. REV. 1075 (2012) (discussing the history and application of the Public Trust and Public Nuisance doctrines as it relates to environmental protection).

[91] Catherine Sharkey, *Punitive Damages as Societal Damages*, 113 YALE L.J. 347 (2003) (proposing a new measure of damages—"societal damages"—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).

[92] Dewees, supra note __, at 428 ("private individuals may sue only if they suffer damages different in kind from those of the general public, and governments rarely sue, so

public nuisance under the public trust doctrine, for example, public enforcement is still necessary.[93] In environmental contexts, for example, a tort-like remedy of natural resource damages yields large sums of compensation and settlements to restore injured natural resources, but may be pursued solely by the public agencies under the public trust doctrine.[94] And still, it is widely recognized that "the law of nuisance was not up to the task of protecting the environment."[95]

Similar to environmental pollution, data emissions create societal harms. These are the negative externalities discussed in Part I—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 pure public harm—affecting a political ecosystem, not to any single user. What tort remedy could capture it? The harm to users whose financial data is spilled due to security breach is largely an expanded sense of insecurity to all—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 a design of a private property right to be vindicated through private tort action. Like natural resource damages, a compensatory framework for data pollution would have to rely on public enforcement actions.

C. 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 superfluous hope that people would then be able to exercise "informed consent" to the practices. For example, the Video Privacy Protection Act[96] prohibits service providers from sharing customers' personal data without written consent (imposing a

---

this doctrine is of little effect", citing Swanson, Elizabeth J. and Elaine L. Hughes (1990) The Price of Pollution: Environmental Litigation in Canada.)

[93] See Lin, supra note __, at 1093.

[94] See, generally, Karen Bradshaw, *Settling for Natural Resource Damages*, 40 Harv. Env. L. Rev 211 (2016).

[95] Abraham, at 149. See also Donald N. Dewees, The Role of Tort Law in Controlling Environmental Pollution, Candadian Public Policy XVIII 425, 428-29 (1992); Joel Brenner, Nuisance Law and the Industrial Revolution, 3 J. Legal Stud. 403, at __(1974) (Public nuisance doctrine was similarly ineffective in Great Britain in the 19th century).

[96] 18 U.S.C. § 2710

penalty of $2,500 for each violation), with the result that prominent disclosures are meticulously attached to membership contracts.

Similarly, mandated disclosures are the primary response to data security breaches. Once such leakages occur, affected users must be informed, in 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."[97]

Mandated disclosure is without doubt the dominant regulatory approach in American data privacy law.[98] While it is a public form of regulation, mandated disclosure is widely regarded as necessary to allow for private contracting and private controls, and its violation 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 doesn't achieve any of those goals.[99] The notice requirements fail because they 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 people are not entering better contracts. And no matter how expedient the notices they receive post breach, there are few if any precautions they can take 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, is widely criticized for its many costs and few benefits.[100] Public disclosures of toxic releases under the Emergency Planning Community Right to Know Act, that established a Toxic Release

---

[97] Cal. Civ. Code § 1798.29; Cal. Civ. Code § 1798.82. See also leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201320140SB46.

[98] Principles of the Law, Data Privacy, §3-4 (Tentative Draft No. 2, Oct. 24 2014).

[99] Ben-Shahar & Schneider; Ben-Shahar & Chilton; Cite evidence on post-breach notifications

[100] Allison Rosser Barnhill, *The Unraveling of California's Proposition 65*, 24 Wake Forest L. Rev. 367 (1989); Michael Barsa, *California's Proposition 65 and the Limits of Information Economics*, 49 Stan. L. Rev. 1223, 1248 (1997).

Inventory database, may support the clean-up response by public authorities, but the thought that these notices reduce spills or that they are necessary for post-spill mitigation is, in the word of others, "overstated."[101]

A widely held 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 privacy practices.[102] But the results are disappointing.[103] When decisions are complex, simplification of formats cannot have a meaningful effect on people's understanding of the underlying trade offs.

Like the failure of contract and tort law, the futility of disclosures in changing people's behavior is also due to the societal impact of data pollution. If the harmful effects result from the collective behavior of all participants, and they impact an entire ecosystem, why bother read even the simplest of disclosures?

III: PUBLIC REGULATION OF DATA POLLUTION

The pollution model of data emissions is a powerful framework that Part II relied on to explain why private law is not suitable to solve the problem that Part I identified (data's external harm). Can the pollution model of data emissions be equally instructive in pointing to public law solutions? Can it borrow the environmental regulatory framework to begin constructing data pollution law? Part III 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.

---

[101] Linda T. Bui, *Public Disclosure of Private Information as a Tool for Regulating Environmental Emissions: Firm-Level Responses by Petroleum Refineries to the Toxics Release Inventory*, Working paper 05-13, Center for Economic Studies, U.S. Census Bureau; Hyunhoe Bae et al., *Information Disclosure Policy: Do State Data Processing Efforts Help More than the Information Disclosure Itself?* 29 J. Pol'y Anal. & Management 163 (2010).

[102] *Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers,* Federal Trade Commission (2012); *Short Form Notice Code of Conduct to Promote Transparency in Mobile App Practice*s, National Telecommunications and Information Administration (2013), available at https://www.ntia.doc.gov/ les /ntia/publications/july_25_code_draft.pdf; *Consumer Data Privacy in a Networked World: A Frame- work for Protecting Privacy and Promoting Innovation in the Global Digital Economy*, White House (2012).

[103] Ben-Shahar & Chilton, *Simplification of Privacy Disclosures: An Experimental Test*, 45 J. Legal Stud.S41 (2016).

First, physical pollution can often be cleaned up; digital pollution probably not. A "superfund" for data spills might therefore make little sense.[104] Second, the effects of environmental pollutions are always negative (even if the underlying activity causing it is beneficial), whereas data emissions could actually be good—data create enormous positive externalities. Environmental law thus bans substances to toxic for people to use, a technique not likely have a data analogue. Third, environmental impacts can be measured scientifically as a basis for cost-benefit analysis, whereas data externalities are often qualitative and conjectural. What is the cost figure attached to distorted Presidential elections? To 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 ("DPA") and granting it analogous powers to combat data pollution—would not work. A different thinking about the goals and methods of data pollution law might be required. But despite the glaring differences, Part III is dedicated to finding instructive clues in environmental law on how to design social policies in the data emissions area. 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 experience in controlling environmental pollution provides an organizing paradigm for public regulation of digital pollution.[105]

In some ways, what Part III 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 Federal Trade Commission, for example, has long been active in enforcing data rights[106], and it is investigating Facebook's fake-ads data pollution.[107] But its mandate does not go beyond the harms of deception and unfairness, which are largely

---

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

[105] As mentioned in the Introduction, prior work proposed an adaptation of environmental regulatory tools to address data privacy concerns. See Hirsch, supra note __ (examining 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.

[106] See, generally, Chris J. Hoofnagle, FEDERAL TRADE COMMISSION PRIVACY LAW AND POLICY (2017).

[107] *Statement by the Acting Director of FTC's Bureau of Consumer Protection Regarding Reported Concerns about Facebook Privacy Practices*, Federal Trade Commission (Mar. 26, 2018), https://www.ftc.gov/news-events/press-releases/2018/03/statement-acting-director-ftcs-bureau-consumer-protection

muted if polluting companies like Facebook are not breaching 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 about a data breach.[108]

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.[109] The European approach, discussed below, implements a set of principles known as Fair Information Practices (FIPs).[110] 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.[111]

Public enforcement templates exist in data law. But they have been solely 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.

Part III organizes the arsenal of public law' countermeasures to external harms into three distinct families of regulatory devices, mirroring 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.

A. Command-and-Control Regulation

The primary regulatory technique in controlling environmental pollution is to prohibit harmful emitting activity beyond legally set limits—

---

[108] See, e.g., Matt Robinson, *Yahoo to Pay First SEC Penalty Over Its Response to Massive Hack*, Bloomberg BNA, at privacylaw.bna.com/pvrc/7060/split_display.adp?fedfid=132776103&vname=prabulallissues&fcn=1&wsn=1&fn=132776103&split=0 ((April 25, 2018).

[109] The EU enacted two data protection mandates. The first is the European Directive on Data Protection, Council Directive 95/46, art. 25, 1995 O.J. (L 281) 31, 56-57 (EC), and the second is the General Data Protection Regulation (GDPR), Commission Regulation 2016/679, 2016 O.J. (L 119) 1, 60--62 (EU).

[110] Paul M. Schwartz, *The EU-US Privacy Collision,* 126 HARV. L. REv. 1966, 1974-75 (2013).

[111] See Paul M. Schwartz and Karl-Nikolaus Peifer, *Transatlantic Data Privacy Law*, 106 Geo. L. J. 115 (2017).

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 might they 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 the inputs going into production as much as it focuses on the outputs emitted. Factories are generally free to use inputs, as long as they comply with the 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.[112] Data, it might be thought, cannot be sorted along the input/emission divide—the information inputted is the same that potentially gets emitted. Quantity and activity restrictions, therefore, would have to apply to the input side of digital production and limit the data companies may collect.

It may seem that, unlike many environmental pollutants, data are not toxic per se. But upon closer look, the environmental analogy can continue to hold. Even notorious industrial pollutants have ancillary benefits.[113] 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 has proven that the positive externalities can be taken into account in the cost-benefit analysis. Data's externalities are similarly two-directional. Even when emitted (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.[114] Similarly, databases assembled by genetic testing services and both help and harm non-members. Thus, a key challenge for the command-and-control approach to data emissions 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?

---

[112] 42 U.S.C.S. § 7401; 40 CFR 50.

[113] Cite Richard Revesz and Michael Livermore, Rataking Rationality; Graham and Weiner, Risk versus Risk.

[114] S-P Jun, H.S Yoo, and S. Choi, *Ten years of research change using Google Trends: From the perspective of big data utilizations and applications*, 130 Technological Forecasting & Social Change 69 (2018).

The European Union thinks it can. In particular, various quantity restrictions are a key part of the EU's new General Data Protection Regulation (GDPR).[115] Prominent among those are the principles of "data minimization" and "purpose limitation." The Regulation require that data be "processed fairly" and only for "specified, explicit, and legitimate purposes,"[116] 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."[117] 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 deactivated their accounts.[118] And unless used for personalized service, personal data should be anonymized or aggregated.

Quantity restrictions do 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 transfer that could be restricted. The regulation would have to establish categories of circumstances under which data transfer may or may not occur. It could also implement "data localization" standards—the conditions allowing databases to be transferred for storage or use to other countries.[119]

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—obscures the difficulty. Privacy law focuses on the internal harm—loss of individual control over personal information—and the command-and-control rules seek to restore such control. Data pollution law, by contrast, focuses on data's external harm.

---

[115] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

[116] Art. 5, Regulation (EU) 2016/679, 2016 O.J. (L 119) 1 at 35.

[117] Art. 5, Regulation (EU) 2016/679, 2016 O.J. (L 119) 1 at 35 ("collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purpose"

[118] See also 23 NYCRR 500.13 (limitations on data retention) [FULL CITE]

[119] Art. 5, Regulation (EU) 2016/679, 2016 O.J. (L 119) 1 at 61.

The commands must therefore be justified by anticipating the harm that databases cause. This is a staggering challenge: Big Data allows investigations that reveal connections not previously known or anticipated. Who knew that a database of internet searches could provide clues for the detection of major epidemics?[120] Collecting and studying 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. Environmental law focuses its most severe restrictions on the most toxic pollutants. 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 anti-discrimination laws. Accordingly, heightened quantity restrictions may be set to limit 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.[121]

Limitations on how sensitive data is collected and used are, however, a double edge sword: they protect weaker groups against harms, but also deny them important 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.[122] It is not until such discoveries are excavated from the data that their value becomes known. If data regulations jeopardize the creation of new knowledge, then heightened restrictions vis-à-vis protected groups could have the unintended effect of disproportionately harming these groups. Because data produce both positive 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, for example, the National Pollutant Discharge Elimination System prohibits discharge of pollutants to waters unless compliant with specific permits.[123] Permits could be required for particular data activities which pose the greater risks. If a website wants to run an algorithm that collects and uses, for example, race

---

[120] Jeremy Ginsberg et al., *Detecting influenza epidemics using search engine query data*, 457 Nature 1012 (2009).

[121] Art. 9, Regulation (EU) 2016/679, 2016 O.J. (L 119) 1 at 38.

[122] Dhammika Dharmapla and and Aziz Huq, *Imputing Unreported Hate Crimes Using Google Search Data*, (unpublished, 2018)

[123] 33 U.S.C.S. § 1342.

characteristics, 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 ex ante based on each data collector's goals circumstances, as well as the particular potential harms from the applicant's database. It could be administered in conjunction with another regulatory technique used in environmental law—information-forcing requirements. In the same way that would-be environmental pollutants are required to submit environmental impact statements to identify the potential harms and costs,[124] would-be data pollutants seeking particularly sensitive data permit would have to provide more information about their purposes and practices, and the harms they might impose.[125]

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 over-protective, 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.[126] 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 emit air pollution are required to install the "best available control technology" to achieve the "lowest achievable emission rate."[127] In the data realm, regulations could require companies to adopt data processing and security technologies with desired attributes.[128] This could fit well with two central data pollution concerns—transparency and security.

---

[124] The National Environmental Policy Act of 1969, 42 U.S.C.S. §§ 4321-4347.

[125] See Froomkin, supra note 9, at 1745-47 (proposing 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.

[126] Carl E. Schneider, The Practice of Autonomy, Ch. 4 (1998).

[127] 42 U.S.C.S. §§

[128] Hirsch 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." See supra note 9, at 37.

Algorithms used for personalized services could be asked to meet transparency standards. And, 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 methods.[129] 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).[130] Who are the utility power plants of the data economy, and what are the sulfur dioxides emitted by their production? Unlike production of electricity, which is done by a few major plants and 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. If it is the compilation of various layers of data that creates social harm, the regulatory method of quantity regulation would work only if such multi-layer compilation is limited. If the principle of data minimization forbids, for example, a retailer from collecting drivers' license data or amassing 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

---

[129] *What Is Emissions Trading*?, United States Environmental Protection Agency, https://www.epa.gov/emissions-trading-resources/what-emissions-trading; Evaluation of Cap-and-Trade Programs for Reducing U.S. Carbon Emissions, Congressional Budget Office (June 2001) https://www.cbo.gov/publication/13107; Robert N. Stavins, *Experience with Market-Based Environmental Policy Instruments* Nov. 2001 Resources for the Future Discussion Paper 1.

[130] Dallas Burtraw and Sarah Jo Szambelan, *U.S. Emissions Trading Markets for SO2 and NOx,* 40 Resources for the Future Discussion Paper 1 (2009).

data Y because adequate partitioning prevents some perceived social harms. But if A and B can trade the allowances (or merge), a single firm might end up with both data X and data Y, defeating the intended protections. Since massive data compilation is the main concern, a trading system does not address it.

If databases are more likely to pollute the larger they are, then the data pollution paradigm provides a new rationale for legal limits on the size of data giants. Currently, the leading concern surrounding the rise of the internet mega-data companies like Amazon, Facebook, and Google is their market power and the potential anti-competitive behaviors. But if they are also more likely to data-pollute and cause disproportionately greater external harm, size limits become justified even without demonstrating harm to competition.

If large databases are the more likely data polluters, command-and-control restrictions could be selectively imposed on these sources. Such targeted regime would not only address the main pollution harms, but also have the benefit of promoting competition. Otherwise, subjecting both small and large firms to the same set of data restriction could burden disproportionately the small entities, for whom the fixed costs of compliance may be crushing. Indeed, commentators already observe that data restrictions under new privacy laws, like the EU's GDPR, impose compliance costs that large incumbent firms absorb more easily than small firms and entrants.[131]

While there are ways to target quantity restrictions narrowly, the discussion in this section suggests that it would hard to control data in a satisfying manner through ex ante commands defined by the substantive uses of the data.[132] Too much good could be jeopardized by suffocating the most productive activity of our time. And the fact that such approach is being implemented in the realm of privacy law with the goal of protecting people from private harms is hardly reassuring. The restrictions needed to prevent private harms are relatively minor and can be easily satisfied by giving users more "control" over the uses of their data. Data pollution law's restrictions could not be supplanted by users' consent or control, because the harm is to others. Devising substantive data pollution controls is all the more baffling.

B. Data Tax

Could pollution be restricted without the administrative burden and

---

[131] *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;

[132] See Hirsch, supra note 9, at 33-37 (traditional command-and-control methods should not be employed in regulating data emissions).

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 "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, a product responsible for much pollution is carbon, which is present in most fuels, and which emits carbon dioxide when combusted. A carbon tax is thus the best candidate for a Pigouvian tax, widely regarded as an efficient method to regulate pollution.[133] In the digital economy, data are the fuel that generates the activity and all its beneficial value, but also the potential harm. The external harm can thus be internalized through a data tax.[134]

The most natural occasion to levy the tax is at the time of the 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.[135] 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 massively detailed contract can be affixed to the digital transaction, so could a tax.

How might the tax base and 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

---

[133] Gillbert E. Metcalf and David Weisbach, *The Design of a Carbon Tax*, 33 HARV. ENVTL. L. REV. 499, 500 (2009).

[134] 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, supra note 9, at 42-48; Adam Mossoff, *Spam–Oy, What a Nuisance!*, 19 BERKELEY TECH. L.J. 625 (2004); Lily Zhang, *The CAN-SPAM Act: An Insufficient Response to the Growing Spam Problem*, 20 BERKELEY TECH. L.J. 301, 304 (2005).

[135] Walmart.com, 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, whether an email was opened or if ad was effective. See Walmart Privacy Policy," WALMART.COM (updated Nov. 2017), https://corporate.walmart.com/privacy-security/walmart-privacy-policy

measurements.[136] 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.

Moreover, unlike carbon, which creates mostly negative externalities, data has 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 has many unexpected positive benefits, they 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. But not so with the negative effects; emitters have no incentive to "capture" those. Government intervention is needed as a counterweight to that asymmetry. Still, it is possible that even with such one-sided incentive many benefits are too diffuse to capture and a net positive externality remains. Information is a public good and if the owners of databases have to invest in extracting data's benefits they might underinvest. 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 so as to levy some 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 at least a "small" data tax. Even a nominal tax would force firms to carefully rethink the necessity of the data they collect.

In general, firms can assess the potential benefit of the data more accurately than the government, whereas the government is (perhaps) more sensitive and attentive to the potential harm. 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 tax regime with a "small tax" harnesses firms'

---

[136] See *The Social Cost of Carbon: Estimating the Benefits of Reducing Greenhouse Gas Emissions*, United States Environmental Protection Agency, *available at* https://19january2017snapshot.epa.gov/climatechange/social-cost-carbon.html. (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 also Jason Scott Johnston, *The Social Cost of Carbon*, 39 Regulation 36 (2016).

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 marginal tax rate would get higher the more data is collected, to reflect the heightened social concern (now including competitive concerns) with mega databases. It stands to reason that the tax on Amazon need not be equal to the tax paid by a local bookstore on the same 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 data tax 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 cause significant externalities, both positive and negative. If a tax is imposed on the 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'.[137] Ancestry.com clients expose genetic information about their relatives. And Facebook users create portals for data about her friends: a user with 1000 friends is exposing more external data than one who has only 100—and should be taxed more.

---

[137] Complaint at 7, *Daniel Matera v. Google Inc.*, Case No. 5:15-cv-04062 (N.D. Cal. Sept. 4, 2015) https://advance.lexis.com/api/permalink/35a477ee-3758-4a49-8e64-46c9e9f537f2/?context=1000516.

(class action by non-Gmail users).

More generally, giving data is like grazing a common pasture. Others are affected by the database created from individual snippets. The database reveals general attributes about people—not only about the data giver and the 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 think of the people (over)using it as the targets of social intervention. Taxing on the many users who give data that affect others mirrors the typical response to the problem of public goods.

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, they paid upwards of $200 for cars' navigation devices. Then came the free Maps apps, sponsored by collection of 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 behave as if it their data are entirely private and use such currency in disregard for the social impact. But if a social impact exists, a user data tax would 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 the normative shift—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 don't 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 them to the implicit choice they are otherwise making.

In a dramatic way, a data tax would reverse 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 and more personal data, whereas "premium" accounts cost more in money but involve less or no personal data collection.[138] For example, AT&T and Comcast offer

[138] Data discounts are a specific case of a more general "pay-for-data" model, in which

broadband plans that cost more (roughly double) but liberate the bounty-paying users from data collection and from data-driven ads.[139] The great majority of consumers shun these plans—they prefer to pay with data and enjoy the price discount. In a data pollution environment, such choice ignores the negative social impact and should thus trigger the data tax. The appeal of pay-with-data plans would justly decline.

C. Management of Data Spills

Command-and-control rules and data tax are the two primary regulatory techniques targeting the market failure occurring at the time a database is created. They resemble 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 how to address the harm ex post, especially in the aftermath of an unanticipated release.

If release of toxic waste was major problem of the industrial era, data spills are rapidly becoming a major social problem of the digital era.[140]

---

companies would have to pay the users for their personal data. See, e.g., Eric Posner and Glen Weyl, *Want Our Personal Data? Pay For It*, Wall Street Journal (April 20, 2018), https://www.wsj.com/articles/want-our-personal-data-pay-for-it-1524237577.

[139] Cite (see https://www.dslreports.com/shownews/ATT-Charges-Steep-Premium-for-Privacy-Calls-it-a-Discount-136511 for some information).

[140] Estimates of the cost of data breaches vary greatly. A report by New York Attorney General 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 *Information Exposed: Historical Examination of Data Breaches In New York State* 2 (New York Attorney General 2014). See also *Victims of Identity Theft*, 2014," U.S. DEPARTMENT OF JUSTICE OFFICE OF JUSTICE PROGRAMS BUREAU OF JUSTICE STATISTICS, at p. 7 (Sept. 2015, NCJ 248991) https://www.bjs.gov/content/pub/pdf/vit14.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&;

Cybercrime will cost businesses over $2 trillion by 2019, JUNIPER RESEARCH: PRESS RELEASES (Mar. 2017) https://www.juniperresearch.com/press/press-releases/cybercrime-cost-businesses-over-2trillion.

[140] *Grand Theft Data – data exfiltration study: Actors, tactics, and detection*, MCAFEE REPORT (2017) https://www.mcafee.com/us/resources/reports/rp-data-exfiltration.pdf. See also Center for Strategic and International Studies, *Net Losses: Estimating the Global Cost of Cybercrime* (Economic Impact of Cybercrime II), the McAfee Report (June 2014), https://csis-prod.s3.amazonaws.com/ s3fs-public/legacy_files/files/attachments/140609_rp_economic_ impact_cybercrime_report.pdf (accessed 11 Oct. 2017). According to the McAfee Report, 'a conservative estimate would be $375 billion in losses, while the maximum could be as much as $575 billion'.

According to one cybercrime report, half billion people around the world are subject to cybercrime per year, costing them $110 billion.[141] Data spills are often caused by intentional criminal hacking,[142] and they could be prevented, at least in part, by tighter security. Indeed, recent statutory enactments mandate that companies adhere to higher prevention standards.[143] 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 cleanup of hazardous waste sites[144]—which does not have a digital match. In general, data emissions cannot be scrubbed. Digital matter does not exist in a well-confined, excludable and 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.

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.[145] Bolstering these post-release notification schemes is a prominent theme in various proposals to deal with data emissions.[146]

---

[141] Norton, *2012 Norton Cybercrime Report*, http://now-static.norton.com/now/en/pu/images/Promotions/2012/cybercrimeReport/2012_Norton_Cybercrime_Report_Master_FINAL_050912.pdf.

[142] *2017 Annual Data Breach Year-End Review*, Identity Theft Research Center (2017), https://www.idtheftcenter.org/images/breach/2017Breaches/2017AnnualDataBreachYearEndReview.pdf (finding that criminal hacking dwarfs all other methods of data compromise totaling almost 60% of all breaches).

[143] Long Cheng; Fang Liu; Danfeng (Daphne) Yao, *Enterprise data breach: causes, challenges, prevention, and future directions,* 7 WIREs Data Mining and Knowledge Discovery 1 (2017); Yasmine Agelidis*, Protecting the Good, the Bad, and the Ugly: Exposure Data Breaches and Suggestions for Coping with Them,* 31 Berkeley Tech. L.J. 1057 (2016). [Cite provisions from NY data security statute 23 NYCRR 500**]**

[144] Comprehensive Environmental Response, Compensation, and Liability Act of 1980., 94 Stat. 2767. 42 U.S.C.S. §§ 9601–9616.

[145] California Civil Code 1798.29(a) & California Civ. Code 1798.82(a); Consumer Privacy Protection Act of 2017, Sec. 211, H.R. 4081; [CITE other state law].

[146] See Hirsch, supra note 9, at 58 (proposing a new federal "Data Release Inventory

Post breach notifications are not completely useless.[147] 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;[148] regularly check their credit reports; file tax returns early; and more. Yet consumers' response to written security breach notifications is, at best, sluggish.[149] 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.[150] 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 have smartly taught themselves to

---

(DRI)" program that would require companies to report annually how much data they released, both intentionally and unintentionally); Paul M. Schwartz and Edward **J.** Janger, *Notification of Data Security Breaches,* 105 MICH. L. REV. 913 (2006); A**.** Michael Froomkin, *Government Data Breaches,* 24 BERKELEY TECH L. J. 1019 (2009) (data breach notice rules applying to governmental bodies).

[147] Romanovsky et al., Do Data Breach Disclosure Laws Reduce Identity Theft?, 30 J. Pol'y Anal. & Management 256 (2011)(estimating the magnitude of identity theft reduction due to data breach disclosure laws).

[148] Cite congressional act allowing minors to request new ss #.

[149] "The Aftermath of a Data Breach: Consumer Sentiment," Ponemon Institute (April 2014), https://www.ponemon.org/local/upload/file/Consumer%20Study%20on%20Aftermath%20of%20a%20Breach%20FINAL%202.pdf (finding that "the most frequent response to a notification is to ignore it and do nothing").

[150] See Pierce, supra note __, at 982 ("In practice, almost all issuing banks offer fraud monitoring services, identity theft protection, and zero fraud liability as features of a consumer's card. Because of these protections, consumers seldom pay for any fraudulent charges. Aside from the financial liability, consumers have the inconvenience of reporting fraud, obtaining a new card, and perhaps worrying about an increased likelihood of future fraud or identity theft. Compared with the financial effects passed along to issuing banks, these costs are nominal." See also John Kiernan, *2015 Fraud Liability Study: Which Cards Protect You Best?,* CARDHUB, http://www.cardhub.comledul fraud-liability-study/ [https://perma.cc8CEL-CDGV]). As a result of these insurance arrangements, the consequences for victims are greatly moderated. It is estimated that approximately 25% of victims of data breaches subsequently suffer identity theft, and 14% of victims of identity theft experience personal out-of-pocket financial losses of $1 or more, with half suffering less than $100 loss. See 2013 LexisNexis True Cost of Fraud Study: Merchants Struggle Against an Onslaught of High-Cost Identity Fraud and Online Fraud," LexisNexis, September 2013, (www.lexisnexis.com/risk/downloads/assets/ true-cost-fraud-2013.pdf); Erika Harrelland Lynn Langton, "Victims of Identity Theft, 2012," U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics, December 2013 (www.bjs.gov/content/pub/pdf/vit12.pdf).

ignore.[151]

Mitigation of post-emission harm can be done without consumers' active participation, but ordinarily still require 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 monitoring through a program called TrustedID, which required only a minimal effort to enroll.[152] Under the proposed Consumer Privacy Protection Act, spilling companies must 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).[153]

Mitigation can reduce potential private harm to consumers whose data was spilled, but the social cost might still be substantial. For one, identity theft and other violations still occur. Further, 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 ends 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,[154] and some spend much more. Overall, 15% of people experience identity theft at some point in their life, and the risk is associated with non-trivial emotional anxiety.[155]

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

---

[151] See Omri Ben-Shahar and. Carl Schneider, More Than You Wanted To Know: The Failure Of Mandated Disclosure (2014).

[152] See, e.g., *How to Enroll in Equifax TrustedID After the Data Breach*, ConsumerSafety.org (Sept. 11, 2017).

[153] H.R. 4081: Consumer Privacy Protection Act of 2017.

[154] Erika Harrelland Lynn Langton, "Victims of Identity Theft, 2012," p. 10 (U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics, December 2013, at www.bjs.gov/content/pub/pdf/vit12.pdf).

[155] Id. At 13.

2. Ex-post Liability

Environmental law imposes stiff ex-post liability on toxic spilling companies. Exxon's liability for the Valdez oil spill totaled over $1 billion (in addition to a heavily-litigated punitive damages of $507 million);[156] and BP's Deepwater Horizon oil spill in 2010 cost the company well over $40 billion.[157] Can data spills be subject to similar stiff liability?

Part II explained why tort law has failed in holding companies liable for data breach. Problems of uncertainty over causation, societal harm, and valuation make it difficult for potential victims to gain standing in courts 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 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 private class action, it should equal the best estimate of the *risk* that the data release inflicts on society. The long tail problem of data emissions harms would no longer be a bar for 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.[158] Estimates of the likelihood of identity theft to consumers whose social security numbers were exposed vary, perhaps in the range of 14-30%.[159] With such estimates, a fixed charge could be established for each consumer exposed. The fine would then need to be equal this 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 fixed 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

---

[156] *Exxon v. Baker*, 554 U.S. 471 (2008); See https://en.wikipedia.org/wiki/Exxon_Valdez_oil_spill

[157] https://en.wikipedia.org/wiki/Deepwater_Horizon_oil_spill

[158] Supra note __, at __ (https://www.bjs.gov/content/pub/pdf/vit14.pdf). Cite other estimates.

[159] See *Information Exposed: Historical Examination of Data Breaches In New York State* 11 (New York Attorney General 2014).

sensitivity of information, the quantum of affected records, and 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[160]—and the only remaining problem is how to divide the liability among polluters. Various criteria could measure the contribution of each polluter, focusing on the amount and quality of data released and the security shortfalls. While tort law solves similar problems of apportionment in cases with joint tortfeasors, 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.

3. Mandatory Insurance

Ex-post liability can accomplish its deterrence goals, but only if firms can afford to pay the liability and if they 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 threatened to do in the area of environmental liability.[161] 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 firms pay reflect the different external costs they impose.[162] Through mandatory insurance, the equivalent of the data tax discussed above is imposed, not

---

[160] See Al Pascual et al., *2018 Identity Fraud: Fraud Enters a New Era of Complexity*, Javelin Strategy, available at https://www.javelinstrategy.com/coverage-area/2018-identity-fraud-fraud-enters-new-era-complexity.

[161] Kenneth S. Abraham, Distributing Risk __(1986)

[162] Ben-Shahar and Logue, supra note __, at 207; Abraham, id., at 57.

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.[163] Accurate actuarial pricing that accounts for the expected harm given actual prevention efforts invested by the policyholder gives firms incentive to reduce the 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 firm "harden their data security."[164] 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.[165] And they intervene early enough in the aftermath of a security breach to reduce both the magnitude of harm and the legal liability exposure.

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.[166] Environmental law has an elaborate scheme of risk management, requiring sites to implement spill prevention, control, and countermeasures.[167] Even with such robust regulatory background, environmental liability insurance policies often require firms to adhere to stricter private environmental codes than those

---

[163] Steven Shavell, *On Moral Hazard and Insurance*, 93 Q.J. Econ. 541 (1979); Steven Shavell, *On the Social Function and Regulation of Liability Insurance*, 25 Geneva Papers on Risk & Ins. 166, 168–69 (2000).

[164] Shauhin A. Talesh, *Data Breach, Privacy, and Cyber Insurance*, Law and Social Inquiry, 2017.

[165] Id., at __.

[166] *Zurich American Insurance Co. v. Sony Corp. of America*, which settled – see **N***.Y. Court: Zurich Not Obligated to Defend Sony Units in Data Breach Litigation*, Insurance Journal, March 17, 2014.

https://www.insurancejournal.com/news/east/2014/03/17/323551.htm

[167] EPA's spill prevention program includes the Spill Prevention, Control, and Countermeasure (SPCC) and the Facility Response Plan (FRP) rules. See __.

imposed by the EPA.[168] Given the embryonic stage of cyber security 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.

CONCLUSION

Digital data law should not be only about privacy. Too often, the exchange of data between a giver and a taker affects others. The data could contain information about others; or, more importantly, the database can be used or misused in ways that affect public interests other than users' privacy. Data pollution is the name of this problem, and data pollution law is the set if 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, assuming (heroically) that they work, are designed to help people protect themselves, but there is no reason to think that people would cease giving data harming 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 they would throw the baby out with the bathwater, suffocating data's positive externalities.

Perhaps the most promising 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 stopped, pollution only needs to be priced. The design of a rational data tax is extremely challenging, and Part III of this Article made some initial nods towards that mission. Two relatively simple strategies could set us in the right direction. First, let's 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.[169] These subsidies are equivalent to paying people to pollute. Second, a small nominal data tax would go a long way towards stopping the mindless hoarding of unneeded data, without stifling

---

[168] Ben-Shahar and Logue, supra note __, at 225-26; Steven A. Kunzman, *The Insurer as Surrogate Regulator of the Hazardous Waste Industry: Solution or Perversion?* 20 Forum 469, 477 (1985); Benjamin J. Richardson, *Mandating Environmental Liability Insurance*, 12 Duke Envtl. L. & Pol'y F. 293, 316 (2002); *How to Open Pollution Coverage Market— Make Policy Contingent on Obeying Environmental Code*, Ins. Advocate, Apr. 5, 1997, at 10.

[169] Eric Posner and Glen Weyl, Radical Markets __ (2018); Brittany Kaiser, Facebook should pay its 2bn users for their personal data, Financial Times (April 9, 2018).

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 where its previous devices (primarily disclosure) failed. People may 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 distinguising data's external effects from data's privacy harm, and begin to look for the best ways to reduce these social costs.

* * *