2-2021

A Unified Theory of Data

William Magnuson

ARTICLE

A UNIFIED THEORY OF DATA

WILLIAM MAGNUSON*

ABSTRACT

How does the proliferation of data in our modern economy affect our legal system? Scholars that have addressed the question have nearly universally agreed that the dramatic increases in the amount of data available to companies, as well as the new uses to which that data is being put, raise fundamental problems for our regulatory structures. But just what those problems might be remains an area of deep disagreement. Some argue that the problem with data is that current uses lead to discriminatory results that harm minority groups. Some argue that the problem with data is that it impinges on the privacy interests of citizens. Still others argue that the problem with data is that its remarkable efficacy as a tool will lead to disruptions in labor markets.

This Article will argue that the disagreements about data and its harms in modern society are the result of overly compartmentalized an alyses of the nature of data itself. Data, after all, is a strikingly broad concept, one that spans everything from where you ate breakfast today to the genetic markers in your DNA to the returns on your 401(k) last year. By focusing narrowly on specific segments of the data industry, both scholars and policymakers have crafted a set of conflicting rules and recommendations that fail to address the core problem of data itself.

This Article aims to correct this gap. First, it provides a taxonomy of the core features of the data economy today and the various behaviors, both positive and negative, that these features make possible. Second, the Article categorizes the types of arguments made about costs and benefits of wider data usage. Finally, the Article argues that the only way to reconcile the varied and overlapping approaches to data in our current regulatory system is to create a more unified law of data. This unified law of data would set forth harmonized and consistent rules for the gathering, storage, and use of data, and it would establish rules to incentivize beneficial data practices and sanction harmful ones. Ultimately, the Article concludes, governing data will require a more comprehensive approach than the limited and piecemeal efforts that have ruled to date.

TABLE OF CONTENTS

* Associate Professor, Texas A&M University School of Law; J.D., Harvard Law School, 2009; M.A., Università di Padova, 2007; A.B., Princeton University, 2004.

I. Introduction

The data economy has arrived, and it is sending shockwaves throughout the nation. Large technology companies like Facebook, Google, Apple, and Amazon have emerged as massive repositories of personal data, collecting information about everything from who our friends are, to where we go, to what we buy.[1] Financial institutions are using data to improve their investment decisions and allocate capital.[2] Health care companies are aggressively seeking out data in order to study disease and create medicines.[3] "Data ag-

[1] *See* Jack Nicas et al., *How Each Big Tech Company May Be Targeted by Regulators*, N.Y. Times (Sept. 8, 2019), https://www.nytimes.com/2019/09/08/technology/antitrust-amazon-apple-facebook-google.html [https://perma.cc/W28E-8C2H]; Jennifer Valentino-DeVries et al., *Your Apps Know Where You Were Last Night, and They're Not Keeping It Secret*, N.Y. Times (Dec. 10, 2018), https://www.nytimes.com/interactive/2018/12/10/business/location-data-privacy-apps.html [https://perma.cc/J44M-Z3D5]; Stuart A. Thompson & Charlie Warzel, *Twelve Million Phones, One Dataset, Zero Privacy*, N.Y. Times (Dec. 19, 2019), https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html [https://perma.cc/TZ3Y-Z2F4]; Sam Schechner & Mark Secada, *You Give Apps Sensitive Personal Information. Then They Tell Facebook*, Wall St. J. (Feb. 22, 2019), https://www.wsj.com/articles/you-give-apps-sensitive-personal-information-then-they-tell-facebook-11550851636 [https://perma.cc/SVR3-SJS8]; Geoffrey A. Fowler, *I Found Your Data. It's For Sale*, Wash. Post (July 18, 2019), https://www.washingtonpost.com/technology/2019/07/18/i-found-your-data-its-sale/ [https://perma.cc/9EWF-RASL]; Carly Minsky, *Is Consumer Protection Legislation Fit for Purpose?*, Fin. Times (Nov. 19, 2019), https://www.ft.com/content/3901dd14-ca55-11e9-af46-b09e8bfe60c0 [https://perma.cc/T4AJ-3BYH].

[2] *See The Fintech Revolution*, Economist (May 9, 2015), https://www.economist.com/leaders/2015/05/09/the-fintech-revolution [https://perma.cc/SK3R-8MZL]; Nathaniel Popper, *Where Finance and Technology Come Together*, N.Y. Times (Nov. 14, 2016), https://www.nytimes.com/2016/11/15/business/dealbook/where-finance-and-technology-come-together.html [https://perma.cc/MMS5-NVQS]; *Machine Learning Promises to Shake Up Large Swathes of Finance*, Economist (May 25, 2017), https://www.economist.com/finance-and-economics/2017/05/25/machine-learning-promises-to-shake-up-large-swathes-of-finance [https://perma.cc/3ESL-QXFR] [hereinafter *Machine Learning in Finance*].

[3] *See Data and Medicine: A Revolution in Health Care Is Coming*, Economist (Feb. 1, 2018), https://www.economist.com/leaders/2018/02/01/a-revolution-in-health-care-is-coming [https://perma.cc/HB6B-JW87]; Jason Millman, *What Big Data Could Do for Health Care*, Wash. Post (July 9, 2014), https://www.washingtonpost.com/news/wonk/wp/2014/07/09/what-big-data-could-do-for-health-care/ [https://perma.cc/4ZTG-KXBX]; Madhumita Murgia & Max Harlow, *How Top Health Websites Are Sharing Sensitive Data with Advertisers*, Fin. Times (Nov. 12, 2019), https://www.ft.com/content/0fbf4d8e-022b-11ea-be59-e49b2a136b8d [https://perma.cc/8QK7-LWJG].

gregators," whose sole *raison d'être* is to collect and monetize data, have emerged as major players across industry sectors.[4]

The rise of the data economy has generated fierce controversy, including a fair deal of hand-wringing from participants in the industry itself. Tim Cook, the CEO of Apple, has stated that "stockpiles of personal data serve only to enrich the companies that collect them" and has called for a federal data law that minimizes the collection of personal data.[5] Mark Zuckerberg, the founder of Facebook, has similarly called for "robust conversations" about the relationship between social media platforms and data.[6] Congress, for its part, has held hearings and launched investigations into the data practices of technology companies.[7] The state of California has gone further, in 2018 enacting the wide-ranging California Consumer Privacy Act ("CCPA") aimed at enhancing privacy protections for consumer data.[8]

The scholarship on data and its discontents has accelerated as well. Scholars have argued that the data economy will lead to systematic discrimination against minorities.[9] They have argued that it will lead to major intru-

---

[4] *See* Natasha Singer, *Mapping, and Sharing, the Consumer Genome*, N.Y. TIMES (June 15, 2012), https://www.nytimes.com/2012/06/17/technology/acxiom-the-quiet-giant-of-consumer-database-marketing.html [https://perma.cc/NG4N-ES8R]; Nathaniel Popper, *Banks and Tech Firms Battle over Something Akin to Gold: Your Data*, N.Y. TIMES (Mar. 23, 2017), https://www.nytimes.com/2017/03/23/business/dealbook/banks-and-tech-firms-battle-over-something-akin-to-gold-your-data.html [https://perma.cc/PT9U-2TE5]; *Crunching the Numbers*, ECONOMIST (May 19, 2012), https://www.economist.com/special-report/2012/05/19/crunching-the-numbers [https://perma.cc/AJ8N-QJ5N]; Ryan Tracy, *Lawmakers Call for Investigation of Fintech Firm Yodlee's Data Selling*, WALL ST. J. (Jan. 17, 2020), https://www.wsj.com/articles/lawmakers-call-for-investigation-of-fintech-firm-yodlees-data-selling-11579269600?st=ii27zllwxudumfm [https://perma.cc/BW5P-MVHQ]; *For AI, Data Are Harder to Come By than You Think*, ECONOMIST (June 11, 2020), https://www.economist.com/technology-quarterly/2020/06/11/for-ai-data-are-harder-to-come-by-than-you-think [https://perma.cc/73LG-T2Y7].

[5] *See* Sara Salinas & Sam Meredith, *Tim Cook: Personal Data Collection Is Being "Weaponized Against Us with Military Efficiency,"* CNBC (Oct. 24, 2018), https://www.cnbc.com/2018/10/24/apples-tim-cook-warns-silicon-valley-it-would-be-destructive-to-block-strong-privacy-laws.html [https://perma.cc/ZK27-UE33].

[6] *See* Jeff Horwitz & Deepa Seetharaman, *Facebook's Zuckerberg Backs Privacy Legislation*, WALL ST. J. (June 26, 2019), https://www.wsj.com/articles/facebooks-zuckerberg-backs-privacy-legislation-11561589798 [https://perma.cc/W4Q2-27H3].

[7] *See* John D. McKinnon, *Big Tech Companies to Appear Before Senate to Discuss Privacy*, WALL ST. J. (Sept. 12, 2018), https://www.wsj.com/articles/big-tech-companies-to-appear-before-senate-to-discuss-privacy-1536750001 [https://perma.cc/334F-2CR8]; David McCabe, *Lawmakers Urge Aggressive Action from Regulators on Big Tech*, N.Y. TIMES (Sept. 17, 2019), https://www.nytimes.com/2019/09/17/technology/senate-antitrust-tech-hearing.html [https://perma.cc/3QQE-6E3H].

[8] *See* California Consumer Privacy Act of 2018 ("CCPA"), CAL. CIV. CODE §§ 1798.100-1798.198 (West 2020); *see also* Natasha Singer, *What Does California's New Data Privacy Law Mean? Nobody Agrees*, N.Y. TIMES (Dec. 29, 2019), https://www.nytimes.com/2019/12/29/technology/california-privacy-law.html [https://perma.cc/5RAP-KV5R].

[9] *See* Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 673–77 (2016); Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L.J. 2218, 2224–25 (2019) (arguing that the use by police, prosecutors, and judges of algorithmic risk assessments leads to racial inequality, not because of input data or code, but because of the nature of prediction itself); Matthew Adam Bruckner, *The Promise and Perils of Algorithmic Lenders' Use of Big Data*, 93 CHI.-KENT L. REV. 3, 27–28 (2018) (arguing that the use of big data and

sions into the privacy of citizens.[10] They have argued that it will destabilize our markets and concentrate power in the hands of a few giant companies.[11] This literature raises serious questions about the harmful effects of the new data economy on important societal values.

On the other hand, accompanying these dire predictions about data's harms is a separate, and seemingly irreconcilable, perception that data can provide solutions to society's most pressing problems. Here are just a few examples of recent entries in this category. In March 2020, in an effort to stem the spread of coronavirus within the country, South Korea launched a massive data collection platform to track the location and activities of its citizens, and this granular data was widely regarded as having allowed the country's response to be so effective.[12] In May 2020, a group of college admissions experts penned an article in the *Wall Street Journal* arguing that "data-rich, comprehensive information on high schools based on their socioeconomic status and academic profile" might fill a gap in college admissions decisions in the wake of SAT cancellations during the coronavirus pandemic.[13] In May 2020, the city of Austin, Texas released a report on community policing calling for more and better data to be generated in order for the police department to improve its policing strategies.[14]

---

machine learning in credit decisions could exacerbate financial services discrimination); Anupam Chander, *The Racist Algorithm?*, 115 MICH. L. REV. 1023, 1025 (2017) (arguing that the possibility of racist or sexist algorithms, based on facts that are suffused with discrimination, calls for algorithmic affirmative action, in which algorithms are designed to take into account race and gender); Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 633 (2017) (arguing that algorithmic decision-making could lead to "incorrect, unjustified, or unfair results" that cannot be solved through transparency alone); Sonia K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54, 105–06 (2019) (arguing that artificial intelligence presents a variety of civil rights problems that must be addressed, not just by governments, but also by corporations); Talia B. Gillis & Jann L. Spiess, *Big Data and Discrimination*, 86 U. CHI. L. REV. 459, 459 (2019) (arguing that algorithms that extract information from big data could exacerbate discrimination and impede the application of anti-discrimination laws).

[10] *See* Jack M. Balkin, *Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation*, 51 U.C. DAVIS L. REV. 1149, 1184 (2018) (arguing that governments have legitimate interests in imposing rules on how information fiduciaries collect, use, and distribute user data).

[11] Ryan Calo, *Digital Market Manipulation*, 82 GEO. WASH. L. REV. 995, 1041 (2014) (arguing that technology companies can use data to manipulate consumers in pernicious ways, including uncovering and triggering consumer frailties); Cynthia Estlund, *What Should We Do After Work? Automation and Employment Law*, 128 YALE L.J. 254, 267–70 (2018) (arguing that advances in robotics and artificial intelligence enable firms to circumvent labor rules aimed at protecting workers).

[12] *See* Eun A Jo, *South Korea's Experiment in Pandemic Surveillance*, DIPLOMAT (Apr. 13, 2020), https://thediplomat.com/2020/04/south-koreas-experiment-in-pandemic-surveillance/ [https://perma.cc/X294-F3TJ].

[13] Brennan Barnard et al., *Will the Pandemic Revolutionize College Admissions?*, WALL ST. J. (May 29, 2020), https://wsj.com/articles/will-the-pandemic-revolutionize-college-admissions-11590763879 [https://perma.cc/C3QF-P5ME].

[14] *See* Kelsey Bradshaw, *Audit: Austin Police Need More Time, Better Data in Community Policing*, AUSTIN-AMERICAN STATESMAN (May 27, 2020), https://www.statesman.com/news/20200527/audit-austin-police-need-more-time-better-data-in-community-policing [https://perma.cc/FJ6F-YR9M].

These conflicting narratives are driven not just by different values, but also by different focuses. It is only natural that if we ask what the primary threats to privacy are today, we will find that data ranks high on the list. It is similarly natural that if we ask how to develop a vaccine for a deadly virus as quickly as possible, data will also rank high on the list. But these conflicting narratives can also cause real harm. When we focus on one feature of data without also focusing on other features, we risk creating legal structures that overweight some values at the expense of others. Just as worryingly, we may create overlapping or inconsistent regulations that either benefit no one or favor powerful political constituencies.

This Article attempts to remedy this gap by providing a unified theory of data. It begins, in Part II, by describing what makes data today different from data in the past. It argues that three key features of data in the modern world have enabled the emergence of the data economy. First, there is simply more data than ever before. The sheer magnitude of the data that is being produced on a daily basis dwarfs any previous period and continues to accelerate as mobile phones, computers, and other internet-connected gadgets proliferate. Second, data today lasts longer than ever. The permanence of data—that is, its ability to be stored for long periods—has been driven both by increased capacity and by increased interest. Third, data is more easily accessed than ever. The portability of data has allowed markets to develop for the production, sale, and analysis of datasets, as well as an explosion in theft and hacking.

Part III turns to the question of consequences: how should we assess this new data ecosystem? In recent years, controversies have emerged over a wide range of issues, on everything from data privacy to data discrimination to data manipulation. It is easy to forget that these issues are all related to the same thing: the magnitude, permanence, and portability of data. And, just as importantly, they all tend to revolve around three core axes. The first axis addresses how the spread of data in the modern world will affect the fairness of our societal institutions. Will it impinge on fundamental values like equality, dignity, and freedom, or will it, instead, preserve or enhance them? The second axis turns on how data will affect the efficiency of our decisions. Will companies and governments use data to exploit or oppress their consumers and citizens, or will they instead use it to empower or inform them? The third axis addresses how data will affect the stability of our broader systems. Will hackers undermine our democratic elections and threaten our markets, or will governments harness data to become more responsive to citizen preference and stabilize our financial institutions? In sum, a review of the literature on data reveals a remarkably symmetric quality, with every risk having an equal and opposite opportunity.

Having identified the dichotomies that typify the scholarly literature on data, Part IV next considers whether there are unifying legal principles that can simultaneously address both the concerns and the opportunities of data. It answers this question in the affirmative. Specifically, it argues that a uni-

fied law of data could help harmonize the patchwork of varied and conflicting data regulations currently existing in the United States and also encourage better data practices across industries. This unified law of data would be animated by three core principles. First, consumers must have clear rights of ownership over the data they generate, allowing them both to hide and to share their data as they see fit. Second, governments must have the right to access data for legitimate purposes, giving them a type of eminent domain power over necessary data. Third, data possessors must protect their data with adequate cybersecurity measures, paired with speedy and effective mechanisms for liability in the event of breaches. Together, these principles of a unified law of data would go a long way towards ensuring a fairer, safer, and more efficient data economy.

II. The Nature of Data

Data, at its heart, is simply information.[15] Information about people, relationships, transactions, and many, many other things. As such, there is an element of data that is timeless and unchanged. We have always had information, and we have always sought to use it. But there are elements of data today that are truly unique, if not in category, then at least in degree. These new elements of data have paved the way for new uses and strategies that would have been difficult or impossible in the past. In order to understand the controversies and conflicts that surround data today, then, we must be careful in distinguishing the unique, new features of data from the unoriginal, old features of it. It may well be true that companies are using data in ways that discriminate against certain groups, but they have also used other information to discriminate as well.[16] It may well be true that governments have impinged on important privacy rights by processing data, but they have also historically violated privacy rights using many other types of information.[17] It may well be true that data-driven investment algorithms could destabilize financial markets, but many other kinds of information have destabilized markets as well.[18] Thus, the question is not whether data raises legal problems—it certainly does. The question is whether data raises legal problems that are different in kind from those that existed before. In order to

---

[15] 1 Oxford English Dictionary (2d ed. 1989) ("An item of (chiefly numerical) information, esp. one obtained by scientific work, a number of which are typically collected together for reference, analysis, or calculation.").

[16] *See, e.g.*, Katzenbach v. McClung, 379 U.S. 294, 296–97 (1964) (concerning a restaurant in Birmingham, Alabama that refused to serve African-Americans in its dining accommodations).

[17] *See, e.g.*, Mapp v. Ohio, 367 U.S. 643, 644–45 (1961) (concerning police officers that arrested the defendant for possession of pornographic materials after forcing open the door to her house without a search warrant).

[18] *See generally* John Carswell, The South Sea Bubble (1960) (concerning the notorious financial crisis brought on by the collapse of the South Sea Company).

answer that question, we must first identify the aspects of data that are different from information more generally.

This Part will argue that the data economy is driven by three unique features of data today—magnitude, permanence, and portability. First, the sheer magnitude or volume of data that is being produced and recorded on a daily basis has increased enormously. Second, the permanence of data has increased as well, allowing data to be stored and used for longer and longer periods of time. And third, the portability of data has risen, with data now easily accessible and transferable around the globe at the click of a button. In combination, these three features of data—magnitude, permanence, and portability—have paved the way for entirely new uses for data. These new uses, in turn, raise new risks.

A. Magnitude

Never before has data been produced and stored at such a remarkable rate.[19] Between 2010 and 2018, the amount of data generated in the United States increased at an annual rate of 31.9 percent.[20] In China, it increased at a rate of 41.9 percent a year.[21] It is estimated that between 2020 and 2021, the world will produce around ninety zettabytes of data, more than all the data that has been produced since the creation of the computer.[22] Accompanying this rapid acceleration in data production is a concomitant acceleration in storage capacity.[23] In 2015, data centers stored 171 exabytes of data worldwide.[24] In 2020, they stored 985 exabytes of data.[25] Another measure of aggregate data storage—the amount of electricity that data centers consume—shows the enormous amount of resources devoted to data: a recent study found that the world's data centers consume around 30 billion watts of electricity, the same amount as 30 nuclear power plants.[26] Google's data centers alone use 300 million watts of electricity;[27] Facebook's use 60 million.[28] To

---

[19] *See* Jeff Desjardins, *How Much Data Is Generated Each Day?*, WORLD ECONOMIC FORUM (Apr. 17, 2019), https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/ [https://perma.cc/HQX4-KKKN].

[20] *See* Ludwig Siegele, *A Deluge of Data Is Giving Rise to a New Economy*, ECONOMIST (Feb. 20, 2020), https://www.economist.com/special-report/2020/02/20/a-deluge-of-data-is-giving-rise-to-a-new-economy [https://perma.cc/KY8W-GRMK].

[21] *See id.*

[22] *Id.*

[23] *See* Angus Loten, *Data Center Market Powers Up to Meet Cloud Demand*, WALL ST. J. (Aug. 27, 2019), https://www.wsj.com/articles/data-center-market-powers-up-to-meet-cloud-demand-11566898200 [https://perma.cc/8XBS-ZF7K].

[24] Arne Holst, *Amount of Data Actually Stored in Data Centers Worldwide from 2015 to 2021*, STATISTA (Mar. 2, 2020), https://www.statista.com/statistics/638613/worldwide-data-center-storage-used-capacity/ [https://perma.cc/QCM3-Q9E7].

[25] *Id.*

[26] *See* James Glanz, *Power, Pollution and the Internet*, N.Y. TIMES (Sept. 22, 2012), https://www.nytimes.com/2012/09/23/technology/data-centers-waste-vast-amounts-of-energy-belying-industry-image.html [https://perma.cc/N2PY-3A4P].

[27] *Id.*

put this in perspective, in 2015, Google's electricity use was roughly equal to that of the entire city of San Francisco.[29]

The rise in data production and storage has been driven by several related features of modern life. First, the increase in mobile phone usage has meant that people bring computing power with them wherever they go, generating new information about themselves and their activities that can then be stored and shared by other companies.[30] In 2011, 35 percent of U.S. adults owned a smartphone.[31] In 2019, 81 percent did.[32] And of course, every time an individual engages in activity on a computer, data is generated, which can then be stored either locally or, increasingly, in the cloud.[33] Often, data is collected even without user input.[34] Telephone companies,[35] mapping apps,[36] and social media companies[37] all have the capacity to track smartphone location data even when users are not actively interacting with their phones. These companies can use this data to generate remarkably intricate portraits of an individual's day-to-day routines. Facebook, for example, has 1.4 billion daily users, who spend, on average, an hour per day on the service.[38] And it is not just smartphones that gather data about individuals and their activities. Today, a growing array of gadgets, from smart appliances to security cameras to self-driving cars, collect and store data about

---

[28] *Id.*

[29] *See* Adam Brinklow, *Google Consumes as Much Electricity as San Francisco*, CURBED (Dec. 7, 2016), https://sf.curbed.com/2016/12/7/13875996/google-san-francisco-electricity-power [https://perma.cc/5FTF-P7NF].

[30] *See* Orly Lobel, *The Law of the Platform*, 101 MINN. L. REV. 87, 94 (2016); Jane E. Brody, *Hooked on Our Smartphones*, N.Y. TIMES (Jan. 9, 2017), https://www.nytimes.com/2017/01/09/well/live/hooked-on-our-smartphones.html [https://perma.cc/C8TU-8PPR].

[31] *See Mobile Fact Sheet,* PEW RSCH. CTR. (June 12, 2019), https://www.pewresearch.org/internet/fact-sheet/mobile/ [https://perma.cc/958T-XJ6D].

[32] *Id.*

[33] *See* Ross Douthat, *Your Smartphone Is Watching You*, N.Y. TIMES (June 8, 2013), https://www.nytimes.com/2013/06/09/opinion/sunday/douthat-your-smartphone-is-watching-you.html [https://perma.cc/47RM-Q2CS].

[34] *See* Jennifer Valentino-DeVries et al., *Your Apps Know Where You Were Last Night, and They're Not Keeping It Secret*, N.Y. TIMES (Dec. 10, 2018), https://www.nytimes.com/interactive/2018/12/10/business/location-data-privacy-apps.html [https://perma.cc/J44M-Z3D5].

[35] *See* Ephrat Livni, *Your Cellphone Location Data Is Now the Last Vestige of Your Privacy*, QUARTZ (June 22, 2018), https://qz.com/1312339/this-supreme-court-ruling-means-cellphone-location-data-is-now-the-last-vestige-of-your-privacy/ [https://perma.cc/2BDX-8MKM].

[36] *See* Andrew Griffin, *Google Stores Location Data "Even When Users Have Told It Not To,"* INDEPENDENT (Aug. 14, 2018), https://www.independent.co.uk/news/world/americas/google-location-data-privacy-android-sundar-pichai-a8490636.html [https://perma.cc/8ET6-ZE5L].

[37] *See* Tony Romm et al., *U.S. Government, Tech Industry Discussing Ways to Use Smartphone Location Data to Combat Coronavirus*, WASH. POST (Mar. 17, 2020), https://www.washingtonpost.com/technology/2020/03/17/white-house-location-data-coronavirus/ [https://perma.cc/PRJ9-7H3Z].

[38] *See* Eduardo Porter, *Your Data Is Crucial to a Robotic Age. Shouldn't You Be Paid for It?*, N.Y. TIMES (Mar. 6, 2018), https://www.nytimes.com/2018/03/06/business/economy/user-data-pay.html [https://perma.cc/XY23-V4Y5].

their environments.[39] These gadgets, too, create massive amounts of data. Self-driving cars are estimated to produce around 100 gigabytes of data *per second*.[40] Together, these developments, bringing computing power into rapidly expanding corners of our lives, mean that an ever-greater portion of our lives is being recorded and stored in the form of computer data.

Second, and relatedly, rapid increases in computing power have enabled all this new data to be processed and stored. The well-known Moore's Law, which posits that the number of components per integrated circuit, roughly equivalent to a computer's processing power, will double every eighteen to twenty-four months, held true for decades and contributed to rapidly accelerating computational capacity every year.[41] While Moore's Law may have lost some of its predictive power in recent years, the collective increase in the power of computing to date has already been enough to transform our very conception of what a computer can do.[42] The ability of computers to process more data has opened pathways to new techniques in computer science, such as machine learning and neural networks, that have made major breakthroughs by processing massive amounts of data and detecting patterns in it.[43] In 2011, for example, IBM's artificial intelligence-based Watson program beat the world's best *Jeopardy!* contestants by a sizeable margin.[44] In 2017, a machine learning-based program known as AlphaGo beat the world's best Go player.[45] Commercial applications have shown promise as well, notching remarkable achievements in health care,[46] image recognition,[47] and fraud detection.[48] Optimism about the potential of

---

[39] *See* Stacy-Ann Elvy, *Commodifying Consumer Data in the Era of the Internet of Things*, 59 B.C. L. Rev. 423, 435–39 (2018).

[40] *See Regulating the Internet Giants: The World's Most Valuable Resource Is No Longer Oil, but Data*, Economist (May 6, 2017), https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data [https://perma.cc/DEC5-2NQD].

[41] *See* John O. McGinnis, *Accelerating AI*, 104 Nw. U. L. Rev. 366, 370 (2010).

[42] *See* Shara Tibken, *CES 2019: Moore's Law Is Dead, Says Nvidia's CEO*, CNet (Jan. 9, 2019), https://www.cnet.com/news/moores-law-is-dead-nvidias-ceo-jensen-huang-says-at-ces-2019/ [https://perma.cc/4FCA-V6MW].

[43] *See* William Magnuson, *Artificial Financial Intelligence*, 9 Harv. Bus. L. Rev. 337, 339–49 (2020).

[44] *See* Stephen Baker, Final Jeopardy: The Story of Watson, the Computer that Will Transform Our World 251 (2012).

[45] *See* Larry Greenemeier, *AI versus AI: Self-Taught AlphaGo Zero Vanquishes Its Predecessor*, Scientific Am. (Oct. 18, 2017), https://www.scientificamerican.com/article/ai-versus-ai-self-taught-alphago-zero-vanquishes-its-predecessor/ [https://perma.cc/DMK4-8UTS].

[46] *See* Emma Hinchliffe, *IBM's Watson Supercomputer Discovers 5 New Genes Linked to ALS*, Mashable (Dec. 14, 2016), http://mashable.com/2016/12/14/ibm-watson-als-research/#oKfRVPG3C8qI [https://perma.cc/4RY3-QXR6].

[47] *See, e.g.*, *Large Scale Visual Recognition Challenge 2017 (ILSVRC2017),* Imagenet (2017), http://image-net.org/challenges/LSVRC/2017/ [https://perma.cc/3X7R-ZMP5].

[48] *See* Daniel Chatfield, *Fighting Fraud with Machine Learning*, Monzo (Feb. 3, 2017), https://monzo.com/blog/2017/02/03/fighting-fraud-with-machine-learning/ [https://perma.cc/HV64-RBB8]; *Machine Learning in Finance*, *supra* note 2.

artificial intelligence and other big data techniques has, in turn, generated new demands for data, in a kind of data-focused feedback loop.[49]

Finally, as companies have increasingly come to understand and appreciate the value of data, new producers of data have emerged to respond to the demand for it. One aspect of this new source of data production is the rise of "data brokers" or "data aggregators," corporate producers of data that specialize in monetizing data.[50] Take, for example, Acxiom, a data broker in the advertising space. Acxiom's business is based on collecting data about consumer behavior and then selling it to clients in order to improve their advertising results.[51] The information it possesses about individuals is enormous—as early as 2012, its database stored information on 500 million active consumers, with around 1,500 data points per consumer.[52] Data brokers like Acxiom have become a major economic force in recent years. One study found that the data brokerage industry generates around $200 billion of revenue a year.[53] Acxiom itself sold its marketing business for $2.3 billion in 2018.[54] But it is not just specialized data brokers that are responding to the demand for more data. Another aspect of the supply side of data is internal to companies themselves. Recognizing that many features of business today require large amounts of data, companies are increasingly employing (directly or indirectly) their own workers to produce the data. Scale AI, a start-up focusing on machine-learning applications for everything from drones to self-driving cars to robots, has more than 30,000 workers entirely devoted to "tagging," or labeling, images from things like self-driving cars in order to improve the company's software's ability to identify objects in the real world.[55] Another company, iMerit, employs more than 2,000 workers around

---

[49] *See* WILL MARKOW ET AL., IBM, THE QUANT CRUNCH: HOW THE DEMAND FOR DATA SCIENCE SKILLS IS DISRUPTING THE JOB MARKET 3–4 (2017), https://www.ibm.com/downloads/cas/3RL3VXGA [https://perma.cc/S92C-6AJN].

[50] *See* Amy J. Schmitz, *Secret Consumer Scores and Segmentations Separating "Haves" from "Have-Nots,"* 2014 MICH. ST. L. REV. 1411, 1419–25 (describing the emergence of the data broker industry); Daniel J. Solove & Chris Jay Hoofnagle, *A Model Regime of Privacy Protection*, 2006 U. ILL. L. REV. 357, 364–68 (highlighting the unregulated nature of data brokers).

[51] *See* Julie Brill, *Demanding Transparency from Data Brokers*, WASH. POST (Aug. 15, 2013), https://www.washingtonpost.com/opinions/demanding-transparency-from-data-brokers/2013/08/15/00609680-0382-11e3-9259-e2aafe5a5f84_story.html [https://perma.cc/5PTA-BS5F].

[52] *See* Singer, *supra* note 4.

[53] *See* David A. Hoffman, *Intel Executive: Rein in Data Brokers*, N.Y. TIMES (July 15, 2019), http://nyti.ms/2XLaDO2 [https://perma.cc/97VP-LFD5].

[54] *See* Alexandra Bruell & Suzanne Vranica, *IPG to Acquire Acxiom Division for $2.3 Billion*, WALL ST. J. (July 2, 2018), https://www.wsj.com/articles/ipg-in-advanced-talks-to-acquire-acxiom-division-for-over-2-2-billion-1530561951?st=xndus119srg84vn [https://perma.cc/DY6X-P65N].

[55] *See* Lucas Matney, *Scale AI and its 22-Year-Old CEO Lock Down $100 Million to Label Silicon Valley's Data*, TECHCRUNCH (Aug. 5, 2019), https://techcrunch.com/2019/08/05/scale-ai-and-its-22-year-old-ceo-lock-down-100-million-to-help-label-silicon-valleys-data/ [https://perma.cc/D3BN-FNWD]; *Are Data More Like Oil or Sunlight?*, ECONOMIST (Feb. 20, 2020), https://www.economist.com/special-report/2020/02/20/are-data-more-like-oil-or-sunlight [https://perma.cc/K5YM-WXLF].

the world to help Amazon create its online data-labeling service.[56] Investment banks, too, have gotten in on the act, hiring machine learning scholars from academia to develop their own data-intensive applications in finance.[57] Put together, these developments have contributed to a massive rise in the creation and propagation of data, covering everything from the highly personal (such as where people are located and what they are doing) to the highly impersonal (such as images of objects and financial results).[58]

The magnitude and volume of data being created and stored in digital form today, driven by mobile phones, the internet, and processing power, have meant that there is more data *potentially* available to individuals, companies, and governments than ever before. This does not, of course, mean that data is *freely* available. Indeed, one of the major concerns of the new data economy is that data is a resource that can be weaponized for commercial advantage.[59] Large companies might refrain from sharing their data with others, as they can use it to develop their competitive strategies.[60] Governments might refrain from sharing data with the public out of a concern that doing so would harm their criminal enforcement priorities or their efforts to fight tax fraud.[61] So even if there is more data than ever before, the landscape of the data economy is also fragmented and siloed in important ways.

B. Permanence

Data today is not just more prolific than it has ever been. It is also more permanent.[62] Once data has been created, it is increasingly stored and acces-

---

[56] *See* Cade Metz, *A.I. Is Learning from Humans. Many Humans*, N.Y. Times (Aug. 16, 2019), https://www.nytimes.com/2019/08/16/technology/ai-humans.html [https://perma.cc/WC7H-43KD].

[57] *See* Sarah Butcher, *The Top Machine Learning Teams in Investment Banks*, eFinancial Careers (May 23, 2018), https://news.efinancialcareers.com/uk-en/315969/top-machine-learning-teams-banks [https://perma.cc/L85P-3HMS].

[58] On the dangers of even anonymized and de-personalized datasets, *see* Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. Rev. 1701, 1703–06, 1716–22 (2010); Jane Yakowitz, *Tragedy of the Data Commons*, 25 Harv. J.L. & Tech. 1, 3–5 (2011).

[59] *See* Cesare Fracassi & William Magnuson, *Data Autonomy*, 74 Vand. L. Rev. (forthcoming 2021).

[60] *See id.*

[61] *See* Sarah Brayne, *The Criminal Law and Law Enforcement Implications of Big Data*, 14 Ann. Rev. L. & Soc. Sci. 293, 305 (2018).

[62] *See* Viktor Mayer-Schonberger, Delete: The Virtue of Forgetting in the Digital Age 6–7 (2011); Daniel J. Solove, The Future of Reputation: Gossip, Rumor, and Privacy on the Internet 33 (2007); Jonathan Zittrain, The Future of the Internet and How to Stop It 229 (2008); Jean-François Blanchette & Deborah G. Johnson, *Data Retention and the Panoptic Society: The Social Benefits of Forgetfulness*, 18 Info. Soc'y 33 (1998); H. Brian Holland, *Inherently Dangerous: The Potential for an Internet-Specific Standard Restricting Speech that Performs a Teaching Function*, 39 U. S.F. L. Rev. 353, 403–04 (2005); Michael L. Rustad & Sanna Kulevska, *Reconceptualizing the Right to be Forgotten to Enable Transatlantic Data Flow*, 28 Harv. J.L. & Tech. 349, 352, 355 (2015); Alexander Tsesis, *The Right to Erasure: Privacy, Data Brokers, and the Indefinite Retention of Data*, 49 Wake Forest L. Rev. 433, 444–45 (2014); *but see* Agnieszka McPeak, *Disappearing Data*,

sible for long periods of time, and potentially forever.[63] Information stored in digital form is not subject to fading memory or decaying paper.[64] It is permanent in a way that was hardly conceivable before the dawn of the computer age, or even just a few years ago. The life span of a CD, after all, is only five to ten years.[65] Today, with easy and affordable cloud storage, data is increasingly being stored on servers owned and operated by large technology firms, which regularly maintain and update them.[66] This applies to wide ranges of data. Company records are increasingly backed up on the cloud.[67] Individuals back up photos and documents to cloud services.[68] Social media posts can be copied or downloaded and then shared with others.[69] Police departments that download video from Ring doorbell cameras can keep the video forever.[70] Data has thus acquired a degree of permanence.

Part of the shift to permanence in the data landscape has been driven by purely technological change. Companies and individuals now have much of their information stored on computers and servers, and the rise of cloud storage has meant that they can store more information for longer and for cheaper.[71] Thus, they simply *can* store data permanently. Another part of the

---

2018 WIS. L. REV. 17, 19 (discussing the spread of "ephemeral" data platforms like Snapchat where data is deleted after a brief period of time).

[63] *See* Farhad Manjoo, *Do We Want an Erasable Internet?*, WALL ST. J. (Dec. 22, 2013, 4:46 PM), https://www.wsj.com/articles/do-we-want-an-erasable-internet-1387748729 [https://perma.cc/8X9E-YF4D].

[64] *But see* David Talbot, *The Fading Memory of the State*, MIT TECH. REV. (July 1, 2005), https://www.technologyreview.com/2005/07/01/39714/the-fading-memory-of-the-state/ [https://perma.cc/P56V-V39A] (detailing the problems associated with storing data for long periods of time in increasingly incompatible formats).

[65] *See* Laura Sydell, *How Long Do CDs Last? It Depends, But Definitely Not Forever*, NAT'L PUB. RADIO (Aug. 18, 2014, 5:21 PM), https://www.npr.org/sections/alltechconsidered/2014/08/18/340716269/how-long-do-cds-last-it-depends-but-definitely-not-forever [https://perma.cc/93EL-JG73] (describing the struggle of archivists to maintain information stored on degrading CDs).

[66] *See* Quentin Hardy, '*Where Does Cloud Storage Really Reside? And Is It Secure?*,' N.Y. TIMES (Jan. 23, 2017), https://www.nytimes.com/2017/01/23/insider/where-does-cloud-storage-really-reside-and-is-it-secure.html [https://perma.cc/G6F4-VUSV].

[67] *See* IBM Cloud Education, *Backup and Disaster Recovery*, IBM CLOUD LEARN HUB (Dec. 6, 2018), https://www.ibm.com/cloud/learn/backup-disaster-recovery [https://perma.cc/QG2W-635W].

[68] *See* David Pierce, *Cluttered Phone and Computer? Put Your Files in the Cloud*, WALL ST. J. (Sept. 9, 2018, 9:00 AM), https://www.wsj.com/articles/cluttered-phone-and-computer-put-your-files-in-the-cloud-1536498000 [https://perma.cc/EBZ5-D7RZ]; Tim Bradshaw, *Dropbox Faces Growing Competition in Cloud Storage Wars*, FIN. TIMES (Aug. 18, 2013), https://www.ft.com/content/88be965e-edd8-11e2-816e-00144feabdc0 [https://perma.cc/W59H-HMZW].

[69] *See* Thomas H. Koenig & Michael L. Rustad, *Digital Scarlet Letters: Social Media Stigmatization of the Poor and What Can Be Done*, 93 NEB. L. REV. 592, 609 (2015).

[70] *See* Drew Harwell, *Police Can Keep Ring Camera Video Forever and Share with Whomever They'd Like, Amazon Tells Senator*, WASH. POST (Nov. 19, 2019, 3:32 PM), https://www.washingtonpost.com/technology/2019/11/19/police-can-keep-ring-camera-video-forever-share-with-whomever-theyd-like-company-tells-senator/ [https://perma.cc/2G42-HUZX].

[71] For an example of one downside of all this cheap storage, see David E. Sanger, et al., *Attack Gave Chinese Hackers Privileged Access to U.S. Systems*, N.Y. TIMES (June 20, 2015),

shift has been driven by incentives. As it has become increasingly clear that data has substantial value to businesses, governments, and researchers, actors have recognized the incentives to collect troves of data for future use.[72] Thus, they also *desire* to store data permanently. But yet another part of the increasing permanence of data is due to the fact that data is now often stored in many different places at once. It is not just that individuals are storing more data on secure cloud servers, or that companies are collecting more data on consumer behavior, or that social media sites maintain vast records of individual actions and photos and messages. It is that, increasingly, the same data is stored in many different places by many different actors.[73] A post on Twitter might, for example, be tweeted out by an individual. It might then be retweeted by others. Still others might take a screenshot of it and share it on Facebook. Google might store a cached version of the post on its servers. Yet another version might be stored on archived web pages stored by the *Wayback Machine* (a site that preserves copies of defunct webpages). If it was a particularly popular tweet, it might even be written about in a newspaper or blog.[74] This means that, long after the post has been deleted (either intentionally or unintentionally) by the original writer, copies of it will still exist and potentially be widely available. Data, once created, has become surprisingly durable—it is hard for it to simply be erased or forgotten.

In many ways, the permanence of data is a welcome development. It is convenient for individuals to store treasured memories and old emails on cloud servers that ensure they will always be able to retrieve it. It is helpful to researchers when they can access data on health outcomes, or demographic trends, or financial results. Some scholars have argued that data should be made *more* permanent, at least when it comes to government data. These scholars argue that where governments possess data that the public has a right to access, they should make the data available on the internet with a permanent address so that researchers and others can view and analyze the data in the future.[75] Doing so would make government decision-making more transparent and open opportunities for "innovation and dynamism" by allowing private parties to find and leverage public data for a variety of uses.[76]

---

https://www.nytimes.com/2015/06/21/us/attack-gave-chinese-hackers-privileged-access-to-us-systems.html [https://perma.cc/HLX8-37AB].

[72] *See* Rana Foroohar, *How Much Is Your Data Worth?*, FIN. TIMES (Apr. 8, 2019), https://www.ft.com/content/3f2b0f0e-57cc-11e9-91f9-b6515a54c5b1 [https://perma.cc/V82B-TSLS].

[73] *See* David W. Opderbeck, *Cybersecurity and Executive Power*, 89 WASH. U. L. REV. 795, 798 n.10 (2012) (discussing the advantages and disadvantages of data redundancy).

[74] *But see* Jill Lepore, *The Cobweb: Can the Internet Be Archived?*, NEW YORKER (Jan. 19, 2015), https://www.newyorker.com/magazine/2015/01/26/cobweb [https://perma.cc/5XRU-3N8N] (discussing the ways in which data from websites disappears).

[75] *See* David Robinson, et al., *Government Data and the Invisible Hand*, 11 YALE J.L. & TECH. 160, 167–68 (2009).

[76] *Id.* at 161–62.

But data permanence is also a cause of growing concern for many observers. An embarrassing video might stick around on the internet for years, causing psychological harm to an individual long after its initial post.[77] A defamatory blog post might be easily discoverable by future employers or family members.[78] Or, more broadly, vast data on an individual's purchase history, website visits, and physical movement might remain on a company's servers indefinitely, creating myriad privacy and cybersecurity issues.[79] All of these are troubling possibilities and, all too often, have caused serious damage.[80] As Justice Benjamin Cardozo wrote in 1931, "[w]hat gives the sting to writing is its permanence in form. The spoken word dissolves, but the written one abides and perpetuates the scandal."[81] What was true then is only more true today.

There are, of course, important limitations to the permanence of data. One is simply cost. The process of backing up data is not free. Institutions must either store the data locally on their own hard drives or pay other companies to store it for them.[82] Both of these can be expensive. In 2018, Australian telecom companies spent over $22 million on data retention services in order to comply with their country's stringent metadata storage laws.[83] As mentioned earlier, data centers require large amounts of electricity to maintain.[84] Thus, in the absence of strong reasons to maintain data, one might expect governments, companies, and individuals to erase data that they no longer need. And it is certainly true that companies are constantly seeking

---

[77] *See* Gabriel Snyder, *One Day, All of This Will Be Embarrassing*, ATLANTIC (June 11, 2019), https://www.theatlantic.com/family/archive/2019/06/sharenting-its-way-becoming-old-fashioned/591361/ [https://perma.cc/E7YY-RRBB].

[78] *See* Paul Sullivan, *Negative Online Data Can Be Challenged, at a Price*, N.Y. TIMES (June 10, 2011), https://www.nytimes.com/2011/06/11/your-money/11wealth.html [https://perma.cc/37CL-36NT].

[79] *See* Emily Price, *How to Download Your Entire Amazon Purchase History*, LIFEHACKER (Apr. 27, 2019, 1:41 PM), https://lifehacker.com/how-to-download-your-entire-amazon-purchase-history-1834353979 [https://perma.cc/N6HL-FT4Q].

[80] *See* Benjamin P. Edwards, *Cybersecurity Oversight Liability*, 35 GA. ST. U. L. REV. 663, 664 (2019).

[81] Ostrowe v. Lee, 175 N.E. 505, 506 (N.Y. 1931).

[82] *See Information Storage: Tape Rescues Big Data*, ECONOMIST (Sept. 26, 2013) https://www.economist.com/babbage/2013/09/26/tape-rescues-big-data [https://perma.cc/CB36-WRJL]; Richard Waters, *Amazon Pushes Cloud Revolution Into Its Second Phase*, FIN. TIMES (Dec. 5, 2019), https://www.ft.com/content/6699593c-1780-11ea-9ee4-11f260415385 [https://perma.cc/KEG2-R78D].

[83] *See* Chris Duckett, *Data Retention Costs Australian Telcos Upwards of AU$210 Million to Date*, ZDNET (July 23, 2019, 7:15 AM), https://www.zdnet.com/article/data-retention-costs-australian-telcos-upwards-of-au210-million-to-date/ [https://perma.cc/TDF2-7M4Q].

[84] *See* Anupam Chander & Uyen P. Le, *Data Nationalism*, 64 EMORY L.J. 677, 723–24 (2015) ("Operating such a data center remains expensive because of enormous energy and other expenses—averaging $950,000 in Brazil, $710,000 in Chile, and $510,000 in the United States each month."); Sara Castellanos, *Cisco CIO Says Shift to Cloud Will Cut Energy Use, Costs*, WALL ST. J. (Aug. 15, 2019, 5:51 PM), https://www.wsj.com/articles/cisco-cio-says-shift-to-cloud-will-cut-energy-use-costs-11565905879 [https://perma.cc/T5NQ-NVF6].

ways to cut down on data retention costs.[85] At the same time, the actual cost of storing a given amount of data has dropped enormously. In 1967, a hard drive that stored a single megabyte of data cost around $1 million.[86] Today, a gigabyte of storage capacity would cost around two cents.[87] Thus, while cost certainly is a limiting factor on how much data governments, companies, and individuals are willing to store, the actual substantive amount of such data that is able to be stored at reasonable prices is significantly larger than it has ever been.

Another important limitation on the permanence of data is digital incompatibility.[88] Sometimes referred to as "digital obsolescence" or "bit rot," the problem of maintaining data in readable formats over long periods of time, when new software and hardware is invented and adopted and old software and hardware is abandoned, is only beginning to be understood.[89] Accessing stored data requires a number of interlinked parts, all of which must function in order for the data to be maintained: data formats, software programs, operating systems, hardware components.[90] If any of these break down, the data can be lost forever. NASA has already lost data on its first missions to the moon after the original machines used to read the data were discarded.[91] An estimated seventy-seven percent of datasets for research papers published between 1991 and 2011 have been lost.[92] In order to avoid data loss, some experts have started recommending that data be migrated into different formats every five to ten years in order to ensure compatibility.[93] And even storing data on the cloud is not perfect. Companies providing these services can suffer data losses or go out of business, potentially leading to massive amounts of data being lost forever. Policy changes might lead

---

[85] *See* Bob Violino, *Can the Cloud Save You Money? These Companies Say Yes*, InfoWorld (Oct. 21, 2019, 3:00 AM), https://www.infoworld.com/article/3445206/can-the-cloud-save-you-money-these-companies-say-yes.html [https://perma.cc/4MFR-GZAK].

[86] *See* Lucas Mearian, *CW@50: Data Storage Goes from $1M to 2 Cents Per Gigabyte*, Computerworld (Mar. 23, 2017, 3:00 AM), https://www.computerworld.com/article/3182207/cw50-data-storage-goes-from-1m-to-2-cents-per-gigabyte.html [https://perma.cc/X6WL-YC7X].

[87] *Id.*

[88] *See* Richard Whitt, *"Through a Glass, Darkly": Technical, Policy, and Financial Actions to Avert the Coming Digital Dark Ages*, 33 Santa Clara High Tech. L.J. 117, 127 (2017); *Digital Data: Bit Rot*, Economist (Apr. 28, 2012), https://www.economist.com/leaders/2012/04/28/bit-rot [https://perma.cc/HS9W-274S]; Pritam Roy, *All the Data We Lost to Incompatible APIs*, Medium (Feb. 4, 2018), https://medium.com/swlh/a-moment-for-all-of-our-data-lost-in-time-260a54fef86 [https://perma.cc/YSE7-23CN].

[89] *See* Kari Kraus, *When Data Disappears*, N.Y. Times (Aug. 6, 2011), https://www.nytimes.com/2011/08/07/opinion/sunday/when-data-disappears.html [https://perma.cc/B4XX-3XZN]; *Digital Archiving: History Flushed*, Economist (Apr. 28, 2012), https://www.economist.com/international/2012/04/28/history-flushed [https://perma.cc/EQL7-445M].

[90] *See* Whitt, *supra* note 88, at 117.

[91] *See Digital Data: Bit Rot*, *supra* note 88.

[92] *See* Timothy H. Vines et al., *The Availability of Research Data Declines Rapidly with Article Age*, 24 Current Biology 94, 95 (2014).

[93] *See* David Pogue, *Should You Worry About Data Rot?*, N.Y. Times (Mar. 26, 2009), https://www.nytimes.com/2009/03/26/technology/personaltech/26pogue-email.html [https://perma.cc/8J5T-VPCR].

cloud providers to delete data as well, even when companies and individuals want to keep it. In 2009, for example, Amazon remotely erased copies of George Orwell's novels *1984* and *Animal Farm* from users' Kindle e-readers after Amazon learned that the novels had been uploaded by a company that did not have rights to them.[94] In 2019, Microsoft shut down its e-book program and announced that it would be deleting any books that users had bought from their libraries.[95]

A final limitation on data permanence comes from law. A variety of rules and regulations in jurisdictions around the world constrain the ability of companies to store data for long periods. The European Union's General Data Protection Regulation ("GDPR") famously includes a so-called "right to be forgotten," under which individuals may force companies to erase personal data related to them.[96] California's recently enacted CCPA similarly gives individuals the right to request businesses to delete their personal information in certain circumstances.[97] Other special rules apply to particular industries. The Health Insurance Portability and Accountability Act (or "HIPAA") imposes a number of requirements on health care institutions to properly delete data containing protected health information.[98] The GDPR goes even further, forbidding companies from retaining personal data if it is no longer necessary for them to keep it.[99]

But despite these important limitations on data permanence, the basic observation remains the same—vast amounts of data about our world, our

---

[94] *See* Brad Stone, *Amazon Erases Orwell Books from Kindle*, N.Y. TIMES (July 18, 2009), https://www.nytimes.com/2009/07/18/technology/companies/18amazon.html [https://perma.cc/JFN9-5YFS].

[95] *See* Brian Barrett, *Microsoft's Ebook Apocalypse Shows the Dark Side of DRM*, WIRED (June 30, 2019), https://www.wired.com/story/microsoft-ebook-apocalypse-drm/ [https://perma.cc/4DYE-FLJC].

[96] *See* Commission Regulation 2016/679 of Apr. 27, 2016, General Data Protection Regulation (GDPR), 2016 O.J. (L119); Robert C. Post, *Data Privacy and Dignitary Privacy: Google Spain, The Right to be Forgotten, and the Construction of the Public Sphere*, 67 DUKE L.J. 981, 986–87 (2018); Giancarlo Frosio, *The Right to be Forgotten: Much Ado About Nothing*, 15 COLO. TECH. L.J. 307, 309 (2017).

[97] Cal. Civ. Code § 1798.105 (West 2020).

[98] *See* DEP'T OF HEALTH AND HUM. SERV'S, DISPOSAL OF PROTECTED HEALTH INFORMATION, https://www.hhs.gov/hipaa/for-professionals/faq/disposal-of-protected-health-information/index.html [https://perma.cc/D5QY-QX7L].

[99] *See* GDPR, *supra* note 96, at Art. 25(2) ("The controller shall implement appropriate technical and organisational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed. That obligation applies to the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility."); GDPR, *supra* note 96 at recital 39 ("The personal data should be adequate, relevant and limited to what is necessary for the purposes for which they are processed. This requires, in particular, ensuring that the period for which the personal data are stored is limited to a strict minimum . . . In order to ensure that the personal data are not kept longer than necessary, time limits should be established by the controller for erasure or for a periodic review."); *see also* Ari Ezra Waldman, *Privacy's Law of Design*, 9 U.C. IRVINE L. REV. 1239, 1246 n.32 (2019); Clare Sullivan, *GDPR Regulation of AI and Deep Learning in the Context of IOT Data Processing—A Risky Strategy*, 22 J. INTERNET L. 1, 11 (2018).

thoughts, and our actions are being recorded and stored for indefinite periods of time. How we deal with this data is thus tremendously important.

C. Portability

A final unique feature of data today is its astonishing portability. Data can be copied, transferred, sold, accessed, moved, and viewed more easily than ever.[100] It can be emailed, downloaded, Bluetoothed, placed on digital storage devices, saved on shared cloud drives, and moved in any number of other convenient and instantaneous ways.[101] Indeed, the entirety of an individual's genomic sequence could be sent over a typical email platform and be shared with hundreds or thousands at the click of a button.[102]

The portability of data has been driven by many of the same phenomena driving data's magnitude and permanence. The spread of the internet, the mass adoption of smartphones, the improvement of broadband and cell phone networks, the increasing speed of computer chips—all of these developments have meant that data is more easily accessed and transferred, and in larger amounts.[103] Sometimes referred to as the "digital communications revolution," the ease and speed of data transfer today is remarkable and is largely responsible for many of the transformative technologies of recent years, from smartphones to videocalls to virtual assistants.[104] All of these would not have been possible without fast and reliable data transfers.[105]

One consequence of data portability is that data can be transformed and repurposed in any manner of ways, by any number of parties, and for any number of purposes.[106] Take, for example, the fact that I went to the grocery store this morning. Google Maps might use that piece of information to make recommendations about nearby restaurants or gas stations.[107] It might also use it to estimate traffic patterns in the city and share that with other

---

[100] *See* Siegele, *supra* note 20.

[101] *See* Edmund M. Hart et al., *Ten Simple Rules for Digital Data Storage*, 12 PLoS Comput. Biol. 10, 10–11, 17 (2016).

[102] *See* Scott Christley et al., *Human Genomes as Email Attachments*, 25 Bioinformatics 274, 274 (2009).

[103] *See* Siegele, *supra* note 20; Colin Jeffrey, *World Record Internet Data Transfer Rate Almost 50,000 Times Faster Than Broadband*, New Atlas (Feb. 12, 2016), https://newatlas.com/fastest-internet-data-rate-optical-ucl/41797/ [https://perma.cc/D7W9-C8G5].

[104] *See* Mark Cooper, *The Long History and Increasing Importance of Public-Service Principles for 21st Century Public Digital Communications Networks*, 12 J. Telecomm. & High Tech. L. 1, 19–20 (2014).

[105] *See* Eric Eckel, *Apple's Siri: A Cheat Sheet*, TechRepublic (June 22, 2020, 11:12 AM), https://www.techrepublic.com/article/apples-siri-the-smart-persons-guide/ [https://perma.cc/DT8P-C7QD].

[106] *See* Stuart A. Thompson & Charlie Warzel, *Smartphones Are Spies. Here's Whom They Report To.*, N.Y. Times (Dec. 20, 2019), https://www.nytimes.com/interactive/2019/12/20/opinion/location-tracking-smartphone-marketing.html [https://perma.cc/GU5D-NNP3].

[107] *See* Frederic Lardinois, *The New Google Maps with Personalized Recommendations Is Now Live*, TechCrunch (June 26, 2018, 12:52 PM), https://techcrunch.com/2018/06/26/the-new-google-maps-with-personalized-recommendations-is-now-live/ [https://perma.cc/7SAU-TNWQ].

Google Maps users.[108] My credit card company might use the information to learn my purchase patterns and load it into its fraud-detection software to detect scam purchases in the future.[109] If any apps on my phone share location data with advertising companies (and many of them do), advertisers might use that information to show me ads that are tied to my location.[110] If I use a budgeting app, the budgeting app might use the data to prepare spending reports and recommend ways to save money.[111] Government health officials might use the data to detect movement patterns, assess social distancing efforts, and prevent the spread of infectious diseases by performing contact tracing on my precise location.[112] The uses for this single piece of data are almost limitless. They are also not constrained by the fact that I did not take any active decision to share that data in particular (we will return to the question of whether and when prior consent to data sharing should be construed as affirmative agreement to future uses). As a result, a single payment can be included in many different databases, and thus used to study and analyze many different phenomena.[113] This means that companies can package and repackage data into discrete datasets and sell or transfer those datasets to others.[114] Indeed, the portability of data has been an essential driver of the Big Data revolution. Once the possibility of large-scale commercial databases arose, a data ecosystem emerged to exploit it: producers of data shared their information with data brokers, data brokers sold that information to companies, and companies used the data to run their businesses better.[115] And once a data ecosystem emerged, data-intensive research methods, most importantly machine learning and its various iterations today, became more useful.[116] It is no coincidence that many of the most important

---

[108] *See* Brendan Hesse, *How Google Recognizes Traffic Jams in Maps*, LIFEHACKER (Feb. 4, 2020, 4:30 PM) https://lifehacker.com/how-google-recognizes-traffic-jams-in-maps-1841455880 [https://perma.cc/PD25-LP79].

[109] *See* Randy Macaraeg, *Credit Card Fraud Detection*, TOWARDS DATA SCIENCE (Sept. 5, 2019), https://towardsdatascience.com/credit-card-fraud-detection-a1c7e1b75f59 [https://perma.cc/Y544-PL95].

[110] *See* Thompson & Warzel, *supra* note 1.

[111] *See* Fracassi & Magnuson, *supra* note 59.

[112] *See* Sara Morrison, *Apple and Google Look Like Problematic Heroes in the Pandemic*, VOX (Apr. 16, 2020, 3:10 PM), https://www.vox.com/recode/2020/4/16/21221458/apple-google-contact-tracing-app-coronavirus-covid-privacy [https://perma.cc/9KP9-LKTE].

[113] *See Cross-Border Credit Reporting is Becoming a Reality*, ECONOMIST (Apr. 11, 2019), https://www.economist.com/finance-and-economics/2019/04/11/cross-border-credit-reporting-is-at-last-becoming-a-reality [https://perma.cc/7YEK-EQVC].

[114] *See* Douglas McMillan, *Data Brokers Are Selling Your Secrets. How States Are Trying to Stop Them*, WASH. POST (June 24, 2019) https://www.washingtonpost.com/business/2019/06/24/data-brokers-are-getting-rich-by-selling-your-secrets-how-states-are-trying-stop-them/ [https://perma.cc/8A4Q-GLH8].

[115] *See* Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 22 (2014); Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C. L. REV. 93, 96–99 (2014); Omer Tene & Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 NW. J. TECH. & INTELL. PROP. 239, 253–54 (2013).

[116] *See* Tom Young et al., *Recent Trends in Deep Learning Based Natural Language Processing No. 1708-02709*, ARXIV LABS (2018), https://arxiv.org/abs/1708.02709 [https://

breakthroughs in machine learning have occurred in the last decade or so, precisely as the data economy was coming into existence.

But the portability of data has also raised new risks. The ease of transferring data, after all, does not just extend to consensual transfers. Hackers, thieves, and spies can also take advantage of the portability of data for their own purposes. Combined with data's magnitude and permanence, the harm from these attacks can be particularly severe and, just as problematically, difficult to prevent. In 2020, for example, it was revealed that a flaw in Apple's email software for iPhones allowed hackers to send an email to recipients and, without any action whatsoever from users, gain access to the contents of their iPhones, and thereby download and copy personal data in messages, photos, and other formats.[117] If data is easily copied, it is also easily stolen.

Data portability, of course, has important limitations. One of these is that most companies go to great lengths to prevent their data from being used or seen by others.[118] Some of these restrictions are obvious. Companies want to keep their trade secrets from being exposed.[119] Individuals want to ensure that the personal information they give to banks or other service providers is not leaked to outsiders.[120] Intelligence agencies want to keep their sources and methods secret.[121] To protect their data from unauthorized use, many entities encrypt their data, or store it in offline locations.[122] But other restrictions are less obvious. One major concern in recent years has been the difficulties that consumers face when attempting to move their data from one company to another.[123] In the financial world, consumers may face limits on whether and how their other providers can view their banking information.[124] In the social media world, consumers may face limits on how

---

perma.cc/R8EA-QLFF]; NAT'L SCI. & TECH. COUNCIL, PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE (2016), at 5.

[117] *See* Robert McMillan, *Apple iPhone May Be Vulnerable to Email Hack*, WALL ST. J. (Apr. 22, 2020, 11:29 AM), https://www.wsj.com/articles/apple-iphone-may-be-vulnerable-to-email-hack-11587556802 [https://perma.cc/P8RY-GYWT].

[118] *See* Raphael Gellert, *Understanding Data Protection as Risk Regulation*, 18 J. INTERNET L. 3, 3, 6–7 (2015).

[119] *See* Elizabeth A. Rowe, *Rats, Traps, and Trade Secrets*, 57 B.C. L. REV. 381, 381 (2016); Scott J. Shackelford & Scott Russell, *Above the Cloud: Enhancing Cybersecurity in the Aerospace Sector*, 10 FIU L. REV. 635, 643 (2015).

[120] *See* Omri Ben-Shahar, *Data Pollution*, 11 J. LEGAL ANALYSIS 104, 110 (2019); McKay Smith & Garrett Mulrain, *Equi-Failure: The National Security Implications of the Equifax Hack and a Critical Proposal for Reform*, 9 J. NAT'L SECURITY L. & POL'Y 549, 550 (2018).

[121] *See* Kristen E. Eichensehr, *The Cyber-Law of Nations*, 103 GEO. L.J. 317, 356 (2015); Melanie J. Teplinsky, *Fiddling on the Roof: Recent Developments in Cybersecurity*, 2 AM. U. BUS. L. REV. 225, 252-65 (2013).

[122] *See* Sara Castellanos, *Why Don't Companies Just Encrypt All Their Data? It Isn't So Simple*, WALL ST. J. (May 29, 2018, 10:05 PM), https://www.wsj.com/articles/why-dont-companies-just-encrypt-all-their-data-it-isnt-so-simple-1527645900?mod=rss_Technology [https://perma.cc/Q6T7-Z6V5].

[123] *See* Christopher S. Yoo, *When Antitrust Met Facebook*, 19 GEO. MASON L. REV. 1147, 1154 (2012).

[124] *Id.*

they can download their friend groups or messages or photos.[125] Sometimes, these restrictions are driven by legitimate social reasons, such as cybersecurity or privacy.[126] In others, they may be driven by more suspect rationales, such as impeding competition or reducing choice.[127]

Another important limitation on data portability is data transfer regulation. The European Union's GDPR includes many such rules, as does California's Consumer Privacy Act. The GDPR, for example, includes provisions governing data transfers across borders and, in some cases, requires such transfers to take place only after the European Union's Commission has decided that the recipient country ensures an adequate level of data protection under its domestic laws.[128] The CCPA prohibits companies from transferring for consideration the personal data of consumers who are under sixteen years of age unless the consumer has affirmatively authorized the transfer.[129] These types of legal restrictions on data portability prohibit companies from transferring certain types of data in certain cases, and companies expend substantial resources in complying with them, as it is not always easy to separate restricted from unrestricted data.[130]

But again, the limitations on data portability do not negate the fact that most data, most of the time, is remarkably easy to transfer, and substantially easier than in periods predating the spread of the internet and the proliferation of smartphones. The speed and ease of sharing data have thus become an essential feature of the modern data economy.

---

[125] *See* Dina Srinivasan, *The Antitrust Case Against Facebook: A Monopolist's Journey Towards Pervasive Surveillance in Spite of Consumers' Preference for Privacy*, 16 BERKELEY BUS. L.J. 39, 89 (2019); Jeff Horwitz, *Facebook Lays Out Challenges of Letting Users Take Their Data to Other Platforms*, WALL ST. J. (Sept. 4, 2019, 12:30 PM), https://www.wsj.com/articles/facebook-lays-out-challenges-of-letting-users-take-their-data-to-other-platforms-11567614600 [https://perma.cc/TR5M-WUD7].

[126] *Id.*

[127] *See* Lina M. Khan, *Amazon's Antitrust Paradox*, 126 YALE L.J. 710, 782–83 (2017); Nathan Newman, *Search, Antitrust, and the Economics of the Control of User Data*, 31 YALE J. ON REG. 401, 441–42 (2014); *Competition in the Digital Age: How to Tame the Tech Titans*, ECONOMIST (Jan. 18, 2018), https://www.economist.com/leaders/2018/01/18/how-to-tame-the-tech-titans [https://perma.cc/B7ZQ-9PHL].

[128] *See* Detlev Gabel & Tim Hickman, *Chapter 13: Cross-Border Data Transfers – Unlocking the EU General Data Protection Regulation*, WHITE & CASE (Apr. 5, 2019), https://www.whitecase.com/publications/article/chapter-13-cross-border-data-transfers-unlocking-eu-general-data-protection [https://perma.cc/R8Q8-5UP5]; *see generally* Matthew Humerick, *The Tortoise and the Hare of International Data Privacy Law: Can the United States Catch Up to Rising Global Standards?*, 27 CATH. U. J.L. & TECH. 77 (2018).

[129] *See* Sara H. Jodka *California's Data Privacy Law: What It Is and How to Comply (A Step-By-Step Guide)*, DICKINSON WRIGHT (July 12, 2018), https://www.dickinson-wright.com/-/media/files/news/2018/07/2californias-data-privacy-law-what-it-is-and-how-t.pdf [https://perma.cc/SK7Q-4GZS].

[130] *See* Nicole Lindsey, *Understanding the GDPR Cost of Continuous Compliance*, CPO MAG. (May 31, 2019), https://www.cpomagazine.com/data-protection/understanding-the-gdpr-cost-of-continuous-compliance/ [https://perma.cc/2Z9U-P57C].

III. Data's Dichotomies

The data economy today is driven by three unique features of data: its magnitude, its permanence, and its portability. There is more data than ever, and it is being produced at an ever-accelerating rate; the data that has already been produced is, for most intents and purposes, permanent, and can be stored for an indefinite period of time; and data can be copied, transferred, shared, and stolen with greater ease than ever. Combined, these features have transformed data into a resource with substantial value to individuals, companies, governments, and researchers. But the rise of the data economy has also generated controversy.[131] Some worry that individual privacy has been sacrificed to the insatiable demand of technology companies for data. Others worry that data is being used to discriminate against disfavored minorities. Still others worry that algorithms driven by Big Data will destabilize our markets or distort our decisions.

This Part will argue that the controversies over data revolve around three core axes: fairness, efficiency, and stability. While these axes overlap and interact in various ways, they are surprisingly robust and resilient categories that show up throughout the literature on data and its discontents. Data may lead to unfair or immoral outcomes; it may lead to inefficient or

---

[131] *See* Jane Bambauer, *Is Data Speech?*, 66 Stan. L. Rev. 57, 60–61 (2014) (arguing that data constitutes speech and thus should receive protection under the First Amendment); Emily Berman, *A Government of Laws and Not of Machines*, 98 B.U. L. Rev. 1277, 1278 (2018) (arguing that the use of machine learning algorithms by government actors undermines the rule of law in areas where government discretion is closely constrained by constitutional, statutory, or regulatory rules); Hannah Bloch-Wehba, *Access to Algorithms*, 88 Fordham L. Rev. 1265, 1265 (2020) (arguing that laws granting citizens access to government records and proceedings could reduce the threat that government algorithms present to civil rights and liberties); Dan L. Burk, *Algorithmic Fair Use*, 86 U. Chi. L. Rev. 283, 285 (2019) (arguing that algorithmic mediation of copyright protection and its exceptions could alter fair use standards); Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 Geo. L.J. 1147, 1147–48 (2017) (arguing that governmental use of machine learning can "comfortably fit within . . . conventional legal parameters"); Ashley S. Deeks, *Predicting Enemies*, 104 Va. L. Rev. 1529, 1573 (2018) (arguing that the military should clearly identify the laws and policies governing how it uses algorithms in its military operations); Lina M. Khan & David E. Pozen, *A Skeptical View of Information Fiduciaries*, 133 Harv. L. Rev. 497, 497–98 (2019) (arguing that the concept of information fiduciaries contains "a number of lurking tensions and ambiguities" and has a limited "capacity to resolve them satisfactorily"); Andrew Keane Woods, *Litigating Data Sovereignty*, 128 Yale L.J. 328, 371–74 (2018) (arguing that disputes over state regulation of the internet are best resolved through the doctrine of comity); Andrew Keane Woods, *Against Data Exceptionalism*, 68 Stan. L. Rev. 729 , 755–56 (2016) (arguing that the issues surrounding government jurisdiction to access personal data are not novel or unprecedented); Omer Tene & Jules Polonetsky, *Taming the Golem: Challenges of Ethical Algorithmic Decision-making*, 19 N.C. J.L. & Tech. 125, 125–26 (2017) (arguing that the law should distinguish between algorithms that lack "an active editorial hand" and algorithms that are "intentionally framed to further a designer's policy agenda"); Jonathan Zittrain, *Facebook Could Decide an Election Without Anyone Ever Finding Out*, New Republic (June 1, 2014), https://newrepublic.com/article/117878/information-fiduciary-solution-facebook-digital-gerrymandering [https://perma.cc/2JLT-8ABF] (arguing that web companies should act as information fiduciaries, in which they agree not to use data to put their own interests above those of their users).

wasteful outcomes; and it may destabilize or undermine important systemic structures. These critiques of the data economy are powerful and persuasive, but they also have important counterpoints. Data may also lead to fairer or more moral outcomes; it may lead to more efficient or optimal outcomes; and it may buttress or support important systemic structures. The dichotomies—that is, data's tendency to simultaneously support and undermine important values—are essential to understanding the ways in which data affects our legal regimes. This Part will explore these dichotomies with an eye towards developing a more comprehensive view of the types of arguments made about data.

A. Fairness

Arguably the most powerful and prevalent critiques of the data economy have been fairness-oriented, asserting that data generates a wide range of harms to important moral values, such as dignity, freedom, equality, and privacy. At the same time, some of the strongest defenses of greater data usage have also focused on fairness, arguing instead that data can help preserve those values in the modern world. What should we make of these competing claims?

Let us begin by examining the fairness-oriented critiques of data. In general, they fall into one of three categories: first, that the data economy has raised serious privacy concerns for individuals; second, that the data economy has led to invidious discrimination; and third, that the data economy fails to treat individuals as morally responsible actors. Let me say at the outset that, while I will treat these critiques as predominantly morality-oriented, they also have important efficiency and stability aspects. After all, morally reprehensible actions can also be inefficient ones, and may well undermine systemic values.

So how does the rise of the data economy harm our interests in privacy? It can do so in a number of ways.[132] At the most basic level, both governments and companies today collect and store vast troves of data about our behavior, data that provides them intimate views into our lives. But the privacy critique of data would not be especially pernicious if it were simply an observation that others know much about us. Every time I walk on the street or go to work, people can observe large amounts of information about me: what I choose to wear, my approximate age, my hair color, my gender, my accent, my race. Arguably, some of these facts are more integral to my conception of myself than what I bought on Amazon yesterday. But there are elements of modern data collection practices that make it uniquely suited to undermining our interests in keeping information about ourselves hidden

[132] *See* ZITTRAIN, *supra* note 62, at 200–234.

from others—our "right 'to be let alone,' " as Warren and Brandeis put it.[133] While the concept of privacy is wide-ranging, and can encompass many different theoretically distinct values, most scholars writing on data tend to use "privacy" to describe the interest that individuals have in choosing what to reveal about themselves to others, and also what not to reveal.[134] Data raises a number of problems on this front. For one, current data practices may simply not respect the choices of individuals about what they choose to hide and to reveal.[135] Companies may collect information about users without user consent, or they may share information they rightfully received with others without requesting consent for the additional sharing.[136] Or they may receive our consent to store data, but then fail to protect it, unwittingly allowing hackers or unauthorized actors to access it.[137] All of these scenarios raise the very real possibility (indeed, given what we know today, the near certainty) that we are more or less forced to share information about ourselves that we would prefer to keep secret.[138] Another major privacy problem is that individuals may not fully understand what they are consenting to when they "choose" to share data with companies or governments, or they may fail to understand the consequences of data collection.[139] The massive amounts of information being collected from us often are provided by us without our full knowledge—at most, we may have clicked "accept" on a terms-and-conditions page, but we probably don't remember the terms, and even if we do, we probably aren't actively thinking about them as we go about our data-generating days.[140] Similarly, the magnitude of data today has meant that even "anonymized" data—that is, databases that have removed names and other identifying information—can be used in conjunction with

---

[133] *See* Samuel D. Warren & Louis D. Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193, 195 (1890) (quoting THOMAS M. COOLEY, A TREATISE ON THE LAW OF TORTS OR THE WRONGS WHICH ARISE INDEPENDENT OF CONTRACT 29 (2d ed. 1879)).

[134] *See, e.g.*, A. Michael Froomkin, *The Death of Privacy*?, 52 STAN. L. REV. 1461, 1464 (2000) (defining "informational privacy" as "the ability to control the acquisition or release of information about oneself").

[135] *See generally* Daniel J. Solove, *Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1880 (2013).

[136] *See id.* at 1881.

[137] *See* Jonathan Mayer, *Government Hacking*, 127 YALE L.J. 570, 586–89 (2018).

[138] *See* Cameron F. Kerry, *Why Protecting Privacy Is a Losing Game Today—and How to Change the Game*, BROOKINGS INST. (Jul. 12, 2018), https://www.brookings.edu/research/why-protecting-privacy-is-a-losing-game-today-and-how-to-change-the-game/ [https://perma.cc/7AW5-VU9Y].

[139] *See* Anita L. Allen, *Protecting One's Own Privacy in a Big Data Economy*, 130 HARV. L. REV. F. 71, 71–72 (2016).

[140] *See* Lawrence Lessig, *The Law of the Horse: What Cyber Law Might Teach*, 113 HARV. L. REV. 501, 510 (1999) ("We wander through cyberspace, unaware of the technologies that gather and track our behavior. We cannot function in life if we assume that everywhere we go such information is collected. Collection practices differ, depending on the site and its objectives. To consent to being tracked, we must know that data is being collected. But the architecture disables (relative to real space) our ability to know when we are being monitored, and to take steps to limit that monitoring.").

other databases to identify individuals.[141] When in the hands of governments, this data could be used to chill the exercise of civil liberties and free speech rights or target politically unpopular groups.[142] When in the hands of hackers and robbers, it could be used to steal and harass.

Another major fairness-oriented critique of the modern data economy is its potential to lead to discrimination against minority groups. It is a fundamental principle of the law that we should not use factors such as race, religion, or sex to make important decisions about whom to hire, rent to, and make a loan to. But the widespread shift to data-driven decision-making might make such discrimination simultaneously more likely and more difficult to root out.[143] Scholars have identified a number of mechanisms by which this might work. First, governments or companies, or officials inside of these entities, might intentionally create algorithms that discriminate against disfavored groups.[144] Banks might, for example, use a credit-scoring algorithm that makes mortgage decisions based on factors that they know to be correlated with race, such as ZIP Code or first name.[145] It would be difficult for an outside observer to know, based on reviewing the code for the algorithm, whether the coder was motivated by racial animus when they assigned weights to particular factors.[146] And given the massive amount of data available about individuals today, and thus the wide variety of seemingly objective factors that might be used to make decisions, data might be used to hide invidious discrimination.[147] Second, governments or companies might unintentionally rely on algorithms that they believe to be unbiased and accu-

---

[141] *See generally* Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. Rev. 1701 (2010).

[142] *See* Neil M. Richards, *The Dangers of Surveillance*, 126 Harv. L. Rev. 1934, 1962–64 (2013).

[143] *See* Jack M. Balkin, *The Three Laws of Robotics in the Age of Big Data*, 78 Ohio St. L.J. 1217, 1238–40 (2017); Barocas & Selbst, *supra* note 9, at 694–714; Stephanie Bornstein, *Antidiscriminatory Algorithms*, 70 Ala. L. Rev. 519, 523–24 (2018); Bruckner, *supra* note 9, at 25–27; Chander, *supra* note 9, at 1026; Gillis & Spiess, *supra* note 9, at 459; Margaret Hu, *Algorithmic Jim Crow*, 86 Fordham L. Rev. 633, 658–60 (2017); Aziz Z. Huq, *Racial Equity in Algorithmic Criminal Justice*, 68 Duke L.J. 1043, 1115–23 (2019); Mikella Hurley & Julius Adebayo, *Credit Scoring in the Era of Big Data*, 18 Yale J.L. & Tech. 148, 149 (2016); Katyal, *supra* note 9, at 56; Jon Kleinberg et al., *Discrimination in the Age of Algorithms*, 10 J. Legal Analysis 113, 114 (2018); Kroll et al., *supra* note 9, at 678; David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. Davis L. Rev. 653, 703–05 (2017); Ziad Obermeyer et al., *Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations*, 366 Science 447, 450 (2019); Tene & Polonetsky, *supra* note 131, at 135.

[144] *See* Lehr & Ohm, *supra* note 143, at 703–04.

[145] *See* Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 Wash. L. Rev. 1, 13–14 (2014) ("Credit scores are only as free from bias as the software and data behind them. Software engineers construct the datasets mined by scoring systems; they define the parameters of data-mining analyses; they create the clusters, links, and decision trees applied; they generate the predictive models applied. The biases and values of system developers and software programmers are embedded into each and every step of development.").

[146] *See* Lehr & Ohm, *supra* note 143, at 656.

[147] *See* Lehr & Ohm, *supra* note 143, at 656.

rate, but that, in reality, incorporate biased data or assumptions.[148] They might, for example, use an algorithm to determine the likelihood that a felon will re-offend based on factors such as prior offenses and personality disorders, but if it turns out that juries are more likely to convict individuals of certain racial groups with committing crimes, or if psychologists are more likely to diagnose people of certain genders with personality disorders, then the algorithm itself may lead to results that systematically discriminate against these groups, even if it appears to be based on objective factors.[149] And it turns out that an enormous number of "objective" factors are in fact correlated with race, sex, religion, or national origin.[150] Even more problematically, one of the most popular methods of analyzing large datasets is machine learning—and, to be more precise, the sub-category of machine learning known as deep learning or neural networks—which tends to provide analyses that are particularly complex and difficult to interpret, making the actual identification of bias in decision-making even more opaque.[151] Thus, many observers argue, the proliferation of the data economy has raised concerns about its contribution to systemic discrimination.

Finally, another fairness-based critique of the data economy has been that data collection and usage practices fail to respect the dignity and equality of individuals as moral actors. According to this view, in some contexts, it is simply inappropriate to treat individuals as an agglomeration of statistical data points, rather than autonomous moral beings deserving of recognition as such.[152] Making important decisions about their welfare based on computer algorithms and datasets inflicts deep-rooted harm to our conceptions of individuality and participation.[153] These issues are particularly salient in the area of criminal justice. Take, for example, parole decisions. A number of states utilize a commercial software program, COMPAS, in mak-

---

[148] *See* Barocas & Selbst, *supra* note 9, at 672.

[149] *Cf.* Barocas & Selbst, *supra* note 9, at 677–94.

[150] *See* Gillis & Spiess, *supra* note 9, at 464–72.

[151] *See* Finale Doshi-Velez et al., *Accountability of AI Under the Law: The Role of Explanation*, Berkman Klein Ctr. For Internet & Soc'y 6–9 (2017), https://cyber.harvard.edu/publications/2017/11/AIExplanation [https://perma.cc/5KW2-JD4U]; Jenna Burrell, *How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms*, 3 Big Data & Soc'y 1, 5–8 (2016); Aaron M. Bornstein, *Is Artificial Intelligence Permanently Inscrutable?*, Nautilus (Sept. 1, 2016), http://nautil.us/issue/40/learning/is-artificial-intelligence-permanently-inscrutable [https://perma.cc/2U79-JN5Q].

[152] Tal Zarsky, *The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making*, 41 Sci. Tech. & Human Values 118, 118–19, 129 (2016); Sofia Grafanaki, *Autonomy Challenges in the Age of Big Data*, 27 Fordham Intell. Prop. Media & Ent. L.J. 803, 843–44 (2017).

[153] *See* Ric Simmons, *Big Data, Machine Judges, and the Legitimacy of the Criminal Justice System*, 52 U.C. Davis L. Rev. 1067, 1076–85 (2018); *but see* Aziz Z. Huq, *A Right to a Human Decision*, 106 Va. L. Rev. 611, 612 (2020) (arguing that "concerns about due process, privacy, and discrimination in machine decisions are typically best addressed through a justiciable 'right to a well-calibrated machine decision'").

ing parole decisions.[154] The COMPAS algorithm uses data about prior reoffense levels to create an individualized risk assessment, and its recommendations have been used on numerous occasions to support decisions to deny individuals parole.[155] Some scholars have argued that this kind of algorithm-based treatment, which reduces individuals to a collection of data points, violates the rights of individuals to due process and, more generally, fair and respectful treatment.[156] We simply should not make decisions about freedom and imprisonment based on these kinds of simplifying and reductive factors.[157] But it is not just in the realm of criminal justice that data raises dignity and equality concerns. Similar arguments have been made in the context of genetic information and targeted advertising and online speech.[158] While these types of fairness critiques are broad and varied, they all involve some wider concern about treating humans as statistics, not moral actors.

But if some scholars believe that the data economy is, in one way or another, leading to unfair results or processes, many others believe that it in fact leads to the opposite outcome—more fair, more just decisions. Let us start with the question of privacy. It is assuredly true that, for the vast majority of people, the proliferation of data collection, storage, and use in the modern world has led to more entities having more information about them. This reduces the ability of people to *hide* information about themselves from others. But the flipside is that the data economy has also given people signif-

---

[154] *See* Niraj Chokshi, *Can Software Predict Crime? Maybe So, But No Better Than a Human*, N.Y. TIMES (Jan. 19, 2018), https://www.nytimes.com/2018/01/19/us/computer-software-human-decisions.html [https://perma.cc/RB4V-LG83].

[155] *See* Aziz Z. Huq, *Racial Equity in Algorithmic Criminal Justice*, 68 DUKE L.J. 1043, 1047–48 (2019); BERNARD E. HARCOURT, AGAINST PREDICTION: PROFILING, POLICING, AND PUNISHING IN AN ACTUARIAL AGE 31–34 (2007).

[156] *See* J.C. Oleson, *Risk in Sentencing: Constitutionally Suspect Variables and Evidence-Based Sentencing*, 64 SMU L. REV. 1329, 1388–93 (2011); John Monahan, *A Jurisprudence of Risk Assessment: Forecasting Harm Among Prisoners, Predators, and Patients*, 92 VA. L. REV. 391, 427–28 (2006).

[157] *See* Emily Berman, *A Government of Laws and Not of Machines*, 98 B.U. L. REV. 1277, 1331 (2018) ("In each of these contexts, questions that deserve serious democratic, deliberative consideration are instead decided by law enforcement and national security officials or the vendors who supply them with analytic products. This is a troubling short-circuit of the democratic process around important policy debates."); Daniel J. Steinbock, *Data Matching, Data Mining, and Due Process*, 40 GA. L. REV. 1, 78 (2005) ("Challenges to the algorithms used in data matching or data mining . . . may not fit well with the kind of individualized hearings that are the due process paradigm.").

[158] *See* Julie E. Cohen, *Examined Lives: Informational Privacy and the Subject as Object*, 52 STAN. L. REV. 1373, 1377 (2000) ("[O]ne must, if one values the individual as an agent of self-determination and community-building, take seriously a conception of data privacy that returns control over much personal data to the individual. We must carve out protected zones of personal autonomy, so that productive expression and development can have room to flourish."); Joseph W. Jerome, *Buying and Selling Privacy: Big Data's Different Burdens and Benefits*, 66 STAN. L. REV. ONLINE 47, 50 (2013) ("Ever-increasing data collection and analysis have the potential to exacerbate class disparities. They will improve market efficiency, and market efficiency favors the wealthy, established classes."); Natalie Ram, *Genetic Privacy After Carpenter*, 105 VA. L. REV. 1357, 1381 (2019) ("[G]enetic information is 'deeply revealing,' and so it is presumptively private in nature.").

icantly more ability to *reveal* information about themselves to others.[159] This, after all, is the foundation of modern social media platforms: they provide people with a truly unique and powerful way to reveal important facts about themselves with others.[160] And it is not just social media that benefits from the magnitude, permanence, and portability of data. Individuals can now share their data for many other purposes, such as proving that they have the means to pay back a mortgage,[161] or to show that they really do have a degree from a particular educational institution,[162] or to improve their investment portfolio.[163] Many important features of the data economy today are only possible because individuals have the capacity to reveal more information about themselves more broadly. Uber and Lyft, after all, would never have existed if there weren't convenient ways for people to share their locations with others. Thus, while the prevalence, portability, and permanence of data have generally reduced people's ability to hide information from others, they have also increased people's ability to reveal information to others. Whether, on balance, this is more or less fair for private citizens is, thus, not so straightforward.

Similarly, with respect to discrimination, many scholars have argued that wider use of data can in fact reduce, not exacerbate, concerns about decisions being made based on improper considerations such as race or sex. The arguments here are relatively straightforward. First, while algorithms may be biased against certain groups (either because of faulty programming, imperfect data, or some other reason), so are human beings.[164] The question

---

[159] *See* Susan Park & Patricia Sanchez Abril, *Digital Self-Ownership: A Publicity-Rights Framework for Determining Employee Social Media Rights*, 53 AM. BUS. L.J. 537, 568–70 (2016); Michael Bromley, et al., *The Social Reality of Blogging and Empowerment Among Malaysian Bloggers*, 23 J. ASIAN PAC. COMMISSION 210, 214 (2013). *See generally* MARTHA MCCAUGHEY ET AL., CYBERACTIVISM: ONLINE ACTIVISM IN THEORY AND PRACTICE (Martha McCaughey & Michael D. Ayers eds., 2003).

[160] Facebook's Mission Statement provides that the company's mission is to "give people the power to build community and bring the world closer together. People use Facebook to stay connected with friends and family, to discover what's going on in the world, and to share and express what matters to them." *See FAQs,* FACEBOOK, https://investor.fb.com/resources/default.aspx. [https://perma.cc/N936-N9T5]. Twitter's Mission Statement provides that its mission is to "give everyone the power to create and share ideas and information instantly without barriers." *See FAQ,* TWITTER, https://investor.twitterinc.com/contact/faq/default.aspx [https://perma.cc/3AJ8-R25C].

[161] *See* AITE GROUP, ALTERNATIVE DATA ACROSS THE LOAN LIFE CYCLE: HOW FINTECH AND OTHER LENDERS USE IT AND WHY 7-8 (2018), https://www.experian.com/assets/consumer-information/reports/Experian_Aite_AltDataReport_Final_120418.pdf [https://perma.cc/LE4G-Z2Y3].

[162] *See* Sean Gallagher, *How the Value of Educational Credentials Is and Isn't Changing*, HARV BUS. REV., (Sept. 20, 2019), https://hbr.org/2019/09/how-the-value-of-educational-credentials-is-and-isnt-changing, [https://perma.cc/4F4D-SLQY].

[163] *See* Victoria Guida, *Banks, Fintech Startups Clash Over 'The New Oil' — Your Data*, POLITICO (Feb. 7, 2020, 6:24 PM), https://www.politico.com/news/2020/02/07/banks-fintech-startups-clash-over-the-new-oil-your-data-112188 [https://perma.cc/Q4LT-QUPQ].

[164] For one of the classic studies on racial bias in hiring, see Marianne Bertrand & Sendhil Mullainathan, *Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination*, 94 AM. ECON. REV. 991, 992–93, 1011 (2004).

we need to ask is not whether basing our decisions on data will lead to perfectly unbiased decisions, but rather whether doing so will lead to decisions that are *less biased* than having human beings make the decisions.[165] And there are strong reasons to believe that, in many contexts, human decision-makers are more prone to bias than well-written, well-scrutinized algorithms.[166] Second, even in the contexts where algorithmic decision-making based on large datasets will be subject to strong bias, it is relatively simple to detect that bias.[167] Through back-testing, variable adjustment, and zero-knowledge proofs, policymakers can review whether the results of the data analysis are leading to results that systematically disfavor certain groups.[168] Even in machine learning contexts, where opacity is a concern, there are ways to determine how much of a role particular factors, such as race, sex, or religion, are playing in the result.[169] Thus, even if discrimination is a problem in the Big Data world, it is easier to identify that discrimination when it occurs. And finally, with respect to remedies, while it is extraordinarily difficult to eliminate biases in human decision-making, it is much easier to do so with data-based algorithms.[170] One can simply adjust variables, change data, or eliminate certain factors that the algorithm uses so that discrimination based on certain features is removed.[171] In one study focused on health care, in which the researchers found that a widely used algorithm for assessing risk was systematically underestimating the health needs of black patients, the researchers found that if they simply reformulated the algorithm to eliminate costs as a proxy for needs, the racial bias was removed.[172] Thus, wider use of data may well be the best tool we have for reducing systemic discrimination.[173]

Finally, and on a related note, scholars have argued that greater and wider use of data is the best way to promote dignity and equality in the world.[174] It is not just that it might lead to less discrimination, although this is assuredly an important part of the argument. It is also that, if we take seriously Platonic arguments about self-knowledge as a route to empowerment, then creating and storing more information in more easily accessible

---

[165] *See* Huq, *supra* note 153, at 638–39.

[166] *See* Huq, *supra* note 153, at 638–39.

[167] *See* Jon Kleinberg et al., *Discrimination in the Age of Algorithms*, 10 J. Legal Analysis 113, 113 (2019); Barocas & Selbst, *supra* note 9, at 700–01; Kroll et al., *supra* note 9, at 682–90.

[168] *See* Kroll et al., *supra* note 9, at 668–69.

[169] *See* Kroll et al., *supra* note 9, at 668–69.

[170] *See* Edward H. Chang et al., *The Mixed Effects of Online Diversity Training*, 116 PNAS 7778, 7778 (2019).

[171] *See* Obermeyer et al., *supra* note 143, at 447. *But see* Devin Desai and Joshua A. Kroll, *Trust but Verify: A Guide to Algorithms and the Law*, 31 Harv. J.L. & Tech. 1, 11 (2017) (discussing the limits of monitoring and regulating algorithms).

[172] *See* Obermeyer et al., *supra* note 143, at 447.

[173] *See* Kleinberg et al., *supra* note 167, at 113.

[174] *See, e.g.*, Omer Tene & Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 Nw. J. Tech. & Intell. Prop. 239, 241–42 (2013).

and analyzable forms might well be the best way to promote individual well-being.[175] Just as importantly, respecting the voluntary decisions of individuals about how and what to disclose and share with others might well require us to accept that people simply desire to share data with companies in order to receive their services.[176]

B. Efficiency

But just as there are serious questions about the fairness of the data economy, there are also questions about its efficiency. Setting aside whether we believe the data economy is desirable from a moral or ethical perspective, we might still want to know whether it is leading to more informed, speedier, or cheaper transactions. And just as we saw on the question of fairness, we will also see that scholars have sharply opposing viewpoints on this question.[177]

Let us begin with the arguments that data usage today is leading to inefficient results. One of the most common concerns about the data economy has been that companies are using the vast troves of data they have on consumers to manipulate and exploit them.[178] We know, for example, that companies use large databases on consumer behavior to narrowly target advertising to particular users.[179] A number of scholars have argued that this provides companies with a powerful tool to take advantage of the cognitive

---

[175] For an analysis of Plato's arguments about knowledge and philosophy, see MALCOLM SCHOFIELD, PLATO: POLITICAL PHILOSOPHY 136–94 (2006). For a discussion of the welfare-enhancing effects of data, see VIKTOR MAYER-SCHONBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK 17–18 (2014).

[176] *But see* Spencer Williams, *Predictive Contracting*, 2019 COLUM. BUS. L. REV. 621, 689–91 (2019).

[177] The reader might well ask why discrimination is not included in this Part on the efficiency concerns of data. After all, if algorithms used to determine who receives a loan or who receives health care treatment systematically discriminate against minorities, this must assuredly be an inefficient outcome. The author does not disagree. The primary reason for categorizing discrimination as a fairness-oriented problem of data, not an efficiency-oriented one, is simply that the arguments about discrimination tend to focus more on issues of autonomy and equality, and less on transaction costs and market failures. But the overlapping nature of these problems highlights just how contentious the debates over categorization can be.

[178] *See* Jonathan Zittrain, *How to Exercise the Power You Didn't Ask For*, HARV. BUS. REV. (Sept. 19, 2018), https://hbr.org/2018/09/how-to-exercise-the-power-you-didnt-ask-for [https://perma.cc/B7J9-C9TC]; Ryan Calo, *Digital Market Manipulation*, 82 GEO. WASH. L. REV. 995, 1022 (2014); Sandra Wachter & Brent Mittelstadt, *A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI*, 2019 COLUM. BUS. L. REV. 494, 505–15 (2019); Ryan Calo & Alex Rosenblat, *The Taking Economy: Uber, Information, and Power*, 117 COLUM. L. REV. 1623, 1651 (2017); Daniel Susser, et al., *Online Manipulation: Hidden Influences in a Digital World*, 4 GEO. L. TECH. REV. 1, 3–4 (2019); Chris Jay Hoofnagle et al., *Behavioral Advertising: The Offer You Cannot Refuse*, 6 HARV. L. & POL'Y REV. 273, 291–95 (2012).

[179] *See* Stuart A. Thompson, *These Ads Think They Know You*, N.Y. TIMES (Apr. 30, 2019), https://www.nytimes.com/interactive/2019/04/30/opinion/privacy-targeted-advertising.html [https://perma.cc/5H2H-PCAA].

biases of consumers.[180] This might well lead to a unique kind of market failure, where consumers are regularly duped into purchasing products or services that they do not truly want or at prices that are unreasonably high.[181] Of course, this is a feature of all advertising—the very premise of advertising is that it leads consumers to purchase things that they would not have purchased otherwise.[182] But the nature of data today makes it easier for companies to acquire information about consumers and act on it in intrusive ways—by, for example, seeing that you visited a Nike website this morning to look for running shoes and then running Nike ads on your future website visits.[183] This opens up entirely new ways of manipulating consumer behavior.

A second major efficiency-related critique of the data economy is that large companies may use data to create or strengthen dominant competitive positions.[184] Data, after all, has become an ever more valuable resource, one that allows companies to sell services, develop strategies, and understand market behaviors in ever more sophisticated ways.[185] But if a few large companies—such as Google and Facebook, or JP Morgan and Goldman Sachs, or Amazon and Walmart—possess vastly more data than other actors, they may well be able to develop impregnable positions that exclude other en-

---

[180] *See* Ryan Calo, *Digital Market Manipulation*, 82 GEO. WASH. L. REV. 995, 995 (2014).

[181] *See* Tal Zarsky, *Online Privacy, Tailoring, and Persuasion*, *in* PRIVACY AND TECHNOLOGIES OF IDENTITY: A CROSS-DISCIPLINARY CONVERSATION 219–20 (Katherine J. Strandburg & Daniela Stan Raicu eds., 2006).

[182] *See* Shmuel I. Becher & Yuval Feldman, *Manipulating, Fast and Slow: The Law of Non-Verbal Market Manipulations*, 38 CARDOZO L. REV. 459, 461 (2016) ("The basic premise of the science of marketing is that consumers' purchasing decisions are highly influenced by sellers' manipulation and selling tactics.").

[183] *See* Calo, *supra* note 180, at 1010 ("Emerging methods of big data present a new and vastly more efficient way to identify cognitive bias by attempting to pinpoint profitable anomalies. Rather than hypothesize and then test a promising deviation, as a lab experimenter would, firms can work backward from raw data.").

[184] *See* Lina M. Khan, *The Separation of Platforms and Commerce*, 119 COLUM. L. REV. 973, 1025 (2019); Kai-Uwe Kühn & Steve Tadelis, Algorithmic Collusion, Address before the Competition and Regulation European Summer School and Conference (CRESSE) (2017), https://www.cresse.info/wp-content/uploads/2020/02/2017_sps5_pr2_Algorithmic-Collusion.pdf [https://perma.cc/Y2G5-5A4K]; FED. TRADE COMM'N, THE COMPETITION AND CONSUMER PROTECTION ISSUES OF ALGORITHMS, ARTIFICIAL INTELLIGENCE, AND PREDICTIVE ANALYTICS (Nov. 14, 2018), https://www.ftc.gov/news-events/audio-video/audio/algorithmic-collusion [https://perma.cc/8LMM-F3QT]; Antonio Capobianco, *Algorithms and Collusion*, ORG. ECON. COOP. & DEV. 5, 16–22 (2017); Azriel Ezrachi & Maurice E. Stucke, *Artificial Intelligence & Collusion: When Computers Inhibit Competition*, 2017 U. ILL. L. REV. 1775, 1777 (2017); Kira Radinsky, *Data Monopolists Like Google Are Threatening the Economy*, HARV. BUS. REV. (Mar. 2, 2015), https://hbr.org/2015/03/data-monopolists-like-google-are-threatening-the-economy [https://perma.cc/6NBA-TKFM]; Nathan Newman, *Search, Antitrust, and the Economics of the Control of User Data*, 31 YALE J. ON REG. 401, 421–23 (2014); Rory Van Loo, *Making Innovation More Competitive: The Case of Fintech*, 65 UCLA L. REV. 232, 243 (2018); Yafit Lev-Aretz & Katherine J. Strandburg, *Privacy Regulation and Innovation Policy*, 22 YALE J.L. & TECH. 256, 304 (2020).

[185] *See* Van Loo, *supra* note 184, at 242–43.

trants in the market.[186] The resulting monopolies could lead to less useful or more costly services being offered to consumers.[187] As access to data becomes an ever more crucial component in business strategies, the data economy could lead to markets being dominated by fewer, but larger, companies.[188]

A third, and somewhat conflicting, efficiency-related critique of the modern data economy is that widespread reliance on large datasets might well lead to worse decision-making by the users of data themselves.[189] The arguments here are multifold, but their basic thrust is that algorithmic analyses possess certain characteristics that tend to skew or constrain the decision-making processes.[190] For example, it is a known problem in machine learning that deep learning algorithms—which analyze historical datasets for patterns—tend to struggle with problems that exhibit "non-stationary" behavior, that is, where the performance of the observed target changes over time.[191] If an airline company had trained its algorithm on data on consumer patterns *before* the coronavirus pandemic, it might well have done a bad job at predicting consumer behavior *after* the pandemic, given the fundamental changes in psychology, society, and law that the pandemic ushered in. Other biases in machine learning exist as well, from difficulties handling low-prevalence subsets of populations to problems with overfitting data to training sets.[192] Another, perhaps even more troubling efficiency problem is the interaction between human decision-makers and data.[193] Data is a tool, and it can be used to help identify unexpected relationships between variables, to spot patterns, and to analyze information faster than humans could ever do. This can obviously be tremendously helpful to policymakers, but it also creates the risk that humans will put *too much* weight on the recommendations that

---

[186] *See* Daniel L. Rubinfeld & Michal S. Gal, *Access Barriers to Big Data*, 59 ARIZ. L. REV. 339, 370 (2017).

[187] *See* Simon Loertscher & Leslie M. Marx, *Digital Monopolies: Privacy Protection or Price Regulation?*, 71 INT'L J. INDUS. ORG. 1, 12 (2020).

[188] *See* Maurice E. Stucke, *Should We Be Concerned About Data-Opolies?*, 2 GEO. L. TECH. REV. 275, 323–24 (2018).

[189] *See* Yesha Yadav, *How Algorithmic Trading Undermines Efficiency in Capital Markets*, 68 VAND. L. REV. 1607, 1617–31 (2015); Rory Van Loo, *The Rise of the Digital Regulator*, 66 DUKE L.J. 1267, 1294 (2017); William Magnuson, *Financial Regulation in the Bitcoin Era*, 23 STAN. J.L. BUS. & FIN. 159, 202 (2018).

[190] *See* Magnuson, *supra* note 43, at 355–65.

[191] *See generally* MASASHI SUGIYAMA & MOTOAKI KAWANABE, MACHINE LEARNING IN NON-STATIONARY ENVIRONMENTS: INTRODUCTION TO COVARIATE SHIFT ADAPTATION (2012).

[192] *See* STUART J. RUSSELL & PETER NORVIG, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH 705 (3d ed. 2010); RICHARD BERK, STATISTICAL LEARNING FROM A REGRESSION PERSPECTIVE 142 (2008); Lehr & Ohm, *supra* note 143, at 678–84; Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 PROC. MACHINE LEARNING RES. 1, 1–2 (2018), http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf [https://perma.cc/G8VD-LZG7].

[193] *See, e.g.*, Nizan Geslevich Packin, *Consumer Finance and AI: The Death of Second Opinions?*, 22 N.Y.U. J. LEGIS. & PUB. POL'Y 319, 346–49 (2020); A. Michael Froomkin, et al., *When AIs Outperform Doctors: Confronting the Challenges of a Tort-Induced Over-Reliance on Machine Learning*, 61 ARIZ. L. REV. 33, 72–81 (2019).

algorithms generate.[194] In other words, as governments, companies, and individuals gain access to large datasets and algorithms capable of analyzing them, they may well be tempted, not just to use the data to improve their decisions, but also to entirely replace their own judgment with the judgment of the algorithm. They might simply accept the algorithm's recommendation that an accused criminal represents a flight risk, or its conclusion that a borrower does not deserve a loan, or its advice that a patient should receive a particular treatment, without further reflection or consideration. A tool might quickly become a crutch.

But other scholars have argued that, far from creating efficiency losses, data will lead to enormous efficiency gains. Indeed, of all the arguments in favor of wider data creation, storage, and use, greater efficiency is likely the most common. Data can be used in innumerable ways to empower and inform individuals.[195] Data about school outcomes and test scores empowers parents to decide where to educate their children.[196] Data about hospital performance empowers patients to decide where to receive medical treatment.[197] Data about spending patterns and financial outcomes empowers people to save for houses, retirements, and other goals.[198] Consumers need data in order to make these decisions. And in order for that data to be provided, they often need either governments or companies to gather and disseminate it.

The data economy may also lead to more competitive markets with fewer dominant players.[199] Far from creating monopolies and erecting barriers to entry, some scholars argue, data has become a tool for small start-ups and enterprising tech companies to compete with large incumbents like

---

[194] *See* Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1271–72 (2008); Linda J. Skitka et al., *Automation Bias and Errors: Are Crews Better than Individuals?*, 10 INT'L J. AVIATION PSYCHOL. 85, 85 (2000); Packin, *supra* note 193, at 322.

[195] *See* Van Loo, *supra* note 184, at 238–42.

[196] *See, e.g.*, DEP'T OF EDUC., EVERY STUDENT SUCCEEDS ACT STATE AND LOCAL REPORT CARDS NON-REGULATORY GUIDANCE (2017), https://www2.ed.gov/policy/elsec/leg/essa/essastatereportcard.pdf [https://perma.cc/4QXY-QGNR]; Jenny Abamu, *How Transparent Is School Data When Parents Can't Find or Understand It?*, EDSURGE (June 26, 2018), https://www.edsurge.com/news/2018-06-26-how-transparent-is-school-data-when-parents-can-t-find-or-understand-it [https://perma.cc/F7RD-LF99].

[197] *See generally* Jeffrey H. Silber, *When Public Reporting Misleads the Public: The Case of Medicare's Hospital Compare Mortality Model*, 68 DEPAUL L. REV. 407 (2019); Nathan Cortez, *Regulation by Database*, 89 U. COLO. L. REV. 1 (2018).

[198] *See generally* Rory Van Loo, *Digital Market Perfection*, 117 MICH. L. REV. 815 (2019); Benjamin P. Edwards, *The Rise of Automated Investment Advice: Can Robo-Advisers Rescue the Retail Market?*, 93 CHI.-KENT L. REV. 97 (2018); Christopher G. Bradley, *Fintech's Double Edges*, 93 CHI.-KENT L. REV. 61 (2018); Chris Brummer & Yesha Yadav, *Fintech and the Innovation Trilemma*, 107 GEO. L.J. 235 (2019); William Magnuson, *Regulating Fintech*, 71 VAND. L. REV. 1167 (2018); William Magnuson, *Financial Regulation in the Bitcoin Era*, 23 STAN. J. L. BUS. & FIN. 159 (2018).

[199] *See* Richard A. Posner, *Antitrust in the New Economy*, 68 ANTITRUST L.J. 925, 938 (2001); Michael Del Priore, *The Trope of Parity*, 36 CARDOZO ARTS & ENT. L.J. 181, 206–08 (2018); Ilene Knable Gotts & Joseph G. Krauss, *Antitrust Review of New Economy Acquisitions*, 15 ANTITRUST 59, 59 (2000); D. Daniel Sokol & Roisin Comerford, *Antitrust and Regulating Big Data*, 23 GEO. MASON L. REV. 1129, 1136 (2016); Deborah T. Tate, *Net Neutrality 10 Years Later: A Still Unconvinced Commissioner*, 66 FED. COMM. L.J. 509, 518 (2014).

never before.[200] Fintech companies are competing for market share with large Wall Street banks.[201] Health technology, or "healthtech," startups, with just tens of employees, are competing with large pharmaceutical companies with thousands.[202] And despite the dominance of social media companies like Facebook and Twitter, it is worth remembering that these companies were themselves startups just fifteen years ago.[203] And their dominance has been due to a large extent by the fact that they have acquired new entrants, not that new entrants have failed to emerge at all.[204] Data, if anything, has led to significantly lower costs of entry.

More generally, many scholars have argued that digital data is the key to better decision-making in the internet era.[205] Pharmaceutical companies can develop vaccines to defeat pandemics because of their ability to gather and analyze large amounts of data.[206] Self-driving car companies can design safer automobiles because of their ability to fine-tune their algorithms.[207] And while machine learning algorithms have weaknesses—what method of analysis does not?—they also have tremendous strengths, and have led to remarkable breakthroughs in areas as diverse as image and voice recognition, mapping, and fraud detection.[208] Data can be used in all these scenarios to augment or improve human decision-making. Thus, to many, the data

---

[200] *See* Sokol & Comerford, *supra* note 199, at 1136 ("The data requirements of new competitors are far more modest and qualitatively different than those of more established firms. Little, if any, user data is required as a starting point for most online services. Instead, firms may enter with innovative new products that skillfully address customer needs, and quickly collect data from users, which they can then use for further product improvement and success. As such, new entrants are unlikely to be at a significant competitive disadvantage relative to incumbents in terms of data collection or analysis.").

[201] *See* Magnuson, *Regulating Fintech*, *supra* note 198, at 1173–87.

[202] *See* Jorge L. Contreras, *The False Promise of Health Data Ownership*, 94 N.Y.U. L. Rev. 625, 627–30 (2019).

[203] Ashlee Vance, *Facebook: The Making of 1 Billion Users*, Bloomberg (Oct. 4, 2012), http://www.bloomberg.com/bw/articles/2012-10-04/facebook-the-making-of-1-billion-users [https://perma.cc/6ZQN-44PV]; Matthew Braga, *Twitter's Road to IPO: Grow First, Monetize Later*, Fin. Post (Sept. 13, 2013), http://business.financialpost.com/fp-tech-desk/twitters-road-to-ipo-grow-first-monetize-later [https://perma.cc/3LZ9-CWXR].

[204] *See* John F. Coyle & Gregg D. Polsky, *Acqui-Hiring*, 63 Duke L.J. 281, 283–84 (2013); Kara Swisher, *Big Tech's Takeovers Finally Get Scrutiny*, N.Y. Times (Feb. 14, 2020), https://www.nytimes.com/2020/02/14/opinion/ftc-investigation-google-facebook.html [https://perma.cc/W2B4-TQJJ].

[205] Susan Athey, *Beyond Prediction: Using Big Data for Policy Problems*, 355 Science 483, 483 (2017); Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 Geo. L.J. 1147, 1223 (2017); Huq, *supra* note 153, at 638–39 ("Even when there is a human substitute for a decision, moreover, studies in a variety of fields suggest that large gains in human well-being can be attained by using a machine-learning tool rather than a person.").

[206] *See* Cade Metz, *Making New Drugs with a Dose of Artificial Intelligence*, N.Y. Times (Feb. 5, 2019), https://www.nytimes.com/2019/02/05/technology/artificial-intelligence-drug-research-deepmind.html [https://perma.cc/2SY3-HC3N].

[207] *See* Nidhi Kalra & David G. Groves, *The Enemy of Good: Estimating the Cost of Waiting for Nearly Perfect Automated Vehicles*, Rand Co. 3 (2017), https://www.rand.org/pubs/research_reports/RR2150.html [https://perma.cc/V2ZG-MHYQ].

[208] *See* Magnuson, *supra* note 43, at 339.

economy will lead to more, not less, efficiency in markets and decision-making.

C. Stability

Finally, another important strand of data scholarship focuses on how the data economy will affect the broader stability of existing systems. Will data change the fundamental nature of government? Will it render markets more prone to crashes? Will it disrupt the relationship between labor and capital? Scholars operating in a wide variety of fields have argued that data raises systemic stability concerns and have called for reforms to reduce the risks. At the same time, another group of scholars, sometimes in dialogue with the first group, and sometimes not, assert that data will make these systems more resilient and less prone to disruption.

One of the central stability-focused critiques of the data economy is that it will undermine our democratic system of government. It might do so in a variety of ways. First, the wide availability of copious amounts of information about citizens might lead governments to grow more oppressive as they ramp up surveillance of their populations.[209] Media reports of recent years have raised serious alarms about governments around the world using data collection to crack down on civil dissent. With the capacity to track citizens' movements, actions, and words with ever-greater precision, governments can punish citizens for behaviors that they would never have even been able to observe before. But it is not just surveillance that scholars are worried about. Another major concern is that companies, political groups, or foreign governments will use data to manipulate political outcomes.[210] In the 2016 election, Russian government hackers penetrated the accounts and computers of Hillary Clinton's campaign for president and leaked embarrassing emails.[211] In 2019, Facebook announced that it had identified (and removed) four different government-backed disinformation campaigns on the social media site.[212] The possibility that bad actors might use or alter data to achieve nefarious political goals, such as tilting an election in favor of a

---

[209] *See* Neil M. Richards, *The Dangers of Surveillance*, 126 HARV. L. REV. 1934, 1945–57 (2013). *See generally* SURVEILLANCE AND DEMOCRACY (Kevin D. Haggerty & Minas Samatas eds., 2010); REBECCA MACKINNON, CONSENT OF THE NETWORKED (2012); EVGENY MOROZOV, THE NET DELUSION (2011).

[210] *See* Ido Kilovaty, *Legally Cognizable Manipulation*, 34 BERKELEY TECH. L.J. 449, 452 (2019) ("[P]erpetrators may use personal information not only for direct financial gain, as they did for more than two decades, but also for the largely unanticipated political manipulation and direct microtargeting of the data subjects.").

[211] *See* Ellen Nakashima & Shane Harris, *How the Russians Hacked the DNC and Passed Its Emails to Wikipedia*, WASH. POST (July 13, 2019), https://www.washingtonpost.com/world/national-security/how-the-russians-hacked-the-dnc-and-passed-its-emails-to-wikileaks/2018/07/13/af19a828-86c3-11e8-8553-a3ce89036c78_story.html [https://perma.cc/FQM5-4EML].

[212] *See* Mike Isaac, *Facebook Finds New Disinformation Campaigns and Braces for 2020 Torrent*, N.Y. TIMES (Oct. 21, 2019), https://www.nytimes.com/2019/10/21/technology/facebook-disinformation-russia-iran.html [https://perma.cc/8664-Y97X].

candidate or fomenting division within society, has become a major focus of intelligence agencies in recent years.[213] The nature of data today—its magnitude, its permanence, and its portability—provides determined adversaries a number of avenues for affecting wider political systems.

Another major critique of the data economy is that it has rendered markets more volatile and prone to crashes.[214] The arguments here tend to focus on how the accelerating trend towards algorithmic trading, based on analyses of financial data, could lead to system-wide effects that alter the nature of stock markets.[215] High-frequency traders that specialize in rapid purchases and sales of securities, for example, might increase the speed of market shifts, and thereby increase panic among other investors.[216] The complexity of data-based financial algorithms might make the process of identifying risk

---

[213] *See* Press Release, Nat'l Sec. Agency, Joint Statement from DOJ, DOD, DHS, DNI, FBI, NSA, and CISA on Ensuring Security of 2020 Elections (Nov. 5, 2019), https://www.nsa.gov/news-features/press-room/Article/2009338/joint-statement-from-doj-dod-dhs-dni-fbi-nsa-and-cisa-on-ensuring-security-of-2/ [https://perma.cc/WHG4-Z2QY] ("Russia, China, Iran, and other foreign malicious actors all will seek to interfere in the voting process or influence voter perceptions. Adversaries may try to accomplish their goals through a variety of means, including social media campaigns, directing disinformation operations or conducting disruptive or destructive cyber-attacks on state and local infrastructure."). *See generally* S. SELECT COMM. ON INTEL., 116TH CONG., REP. ON RUSSIAN ACTIVE MEASURES CAMPAIGNS AND INTERFERENCE IN THE 2016 ELECTION (2019), https://www.intelligence.senate.gov/publications/report-select-committee-intelligence-united-states-senate-russian-active-measures [https://perma.cc/U7TV-LUUX].

[214] *See* Saule T. Omarova, *New Tech v. New Deal: Fintech as a Systemic Phenomenon*, 36 YALE J. ON REG. 735, 790 (2019) ("Potential systemic risk amplifiers, on the other hand, include the heightened tendency toward herding behavior and procyclicality, greater vulnerability to technical glitches and operational failures, and the rise of the systemic importance of non-financial firm."); Gregory Scopino, *Preparing Financial Regulation for the Second Machine Age: The Need for Oversight of Digital Intermediaries in the Futures Markets*, 2015 COLUM. BUS. L. REV. 439, 446–49 (2015); Hilary J. Allen, *Driverless Finance*, 10 HARV. BUS. L. REV. 157, 179 (2020); Van Loo, *supra* note 198, at 861; Kristin N. Johnson, *Regulating Innovation: High Frequency Trading in Dark Pools*, 42 J. CORP. L. 833, 837 (2017); Tom C.W. Lin, *Artificial Intelligence, Finance, and the Law*, 88 FORDHAM L. REV. 531, 541–43 (2019); Yadav, *supra* note 189, at 1612 ("Algorithmic markets are characterized by a systemic degree of "model risk" caused by widespread reliance on stylized models and programming to capture messy real world behavior."); Christopher K. Odinet, *Securitizing Digital Debts*, 52 ARIZ. ST. L.J. 477, 496 (2020) ("[T]he [fintech] securitization process creates a great deal of opacity . . . and that opacity combined with the increasingly complex deep learning underwriting techniques of fintech lending creates systemic risk concerns."). *See generally* Dirk A. Zetzsche et al., *From Fintech to Techfin: The Regulatory Challenges of Data-Driven Finance*, 14 N.Y.U. J.L. & BUS. 393 (2018).

[215] *See* Allen, *supra* note 214, at 160 ("[T]he propensity for increased delegation of decision-making to a few algorithms may lead to destabilizing correlation that undermines financial stability [and] increased use of algorithms could undercut existing financial stability regulation, including regulatory attempts to instill a more stability-oriented financial culture in financial institutions.").

[216] Andrei A. Kirilenko & Andrew W. Lo, *Moore's Law versus Murphy's Law: Algorithmic Trading and Its Discontents*, 27 J. ECON. PERSPECTIVES 51, 52 (2013); Pankaj Jain, et al., *Does High-Frequency Trading Increase Systemic Risk?*, 31 J. FIN. MKTS. 1, 20–22 (2016).

in the market more difficult, both for firms and for regulators.[217] Correlations between Big Data strategies might reduce liquidity in times of stress.[218] All of these problems suggest that data-driven markets could be fragile and volatile.

On the other hand, a number of scholars have argued that the data economy will lead to more resilient, more stable systems. With respect to democratic government, data collection and analysis provides an important tool for states to protect themselves.[219] One of the primary efforts in fighting the coronavirus pandemic was to develop a system for tracking the movements of citizens in order to contact-trace and reduce the risk of spread of the virus.[220] This was made possible by the widespread use of mobile phones and the disclosure of location data.[221] Similarly, national security agencies rely on collecting data about emails and website visits in order to prevent terror attacks and other threats to the nation.[222] An entire new field of legal scholarship (sometimes referred to as "RegTech") is devoted to exploring the ways in which governments can harness data to improve regulatory structures.[223] The underlying thesis of this literature is that more data can lead to better, more responsive government.

Similarly, a significant body of literature has emerged describing the ways in which expanded data use in finance can lead to more stable markets.[224] If financial institutions have more data about firms, consumers, and

---

[217] *See* Allen, *supra* note 214, at 192 ("Regulators . . . have an important role to play in addressing the increasing automation of financial services [but] innovation in financial algorithms will undoubtedly make their jobs more challenging.").

[218] *See* Allen, *supra* note 214, at 182–87.

[219] *See* Sara Binzer Hobolt & Robert Klemmemsen, *Responsive Government? Public Opinion and Government Policy Preferences in Britain and Denmark*, 53 POL. STUDIES 379, 391–96 (2005); Julie Freeman & Sharna Quirke, *Understanding E-Democracy: Government-Led Initiatives for Democratic Reform*, 5 J. E-DEMOCRACY & OPEN GOV. 141, 149–150 (2013); Marijn Janssen et al., *Driving Public Sector Innovation Using Big and Open Linked Data*, 19 INF. SYS. FRONTIERS 189, 192–93 (2017).

[220] *See* Jack Nicas & Daisuke Wakabayashi, *Apple and Google Team Up to 'Contact Trace' the Coronavirus*, N.Y. TIMES (Apr. 10, 2020), https://www.nytimes.com/2020/04/10/technology/apple-google-coronavirus-contact-tracing.html [https://perma.cc/D28Y-HFDL].

[221] Byron Tau, *Government Tracking How People Move Around in Coronavirus Pandemic*, WALL ST. J., Mar. 28, 2020, https://www.wsj.com/articles/government-tracking-how-people-move-around-in-coronavirus-pandemic-11585393202 [https://perma.cc/TT7H-Y26F].

[222] *See* Rajeev Syal, *Bulk Data Collection Vital to Prevent Terrorism in UK, Report Finds*, GUARDIAN (Aug. 19, 2016), https://www.theguardian.com/world/2016/aug/19/bulk-data-collection-vital-to-prevent-terrorism-in-uk-report-finds [https://perma.cc/7MTF-VRLD].

[223] *See* Saule T. Omarova, *Dealing with Disruption: Emerging Approaches to Fintech Regulation*, 61 WASH. U. J.L. & POL'Y 25, 48–52 (2020); Douglas W. Arner et al., *FinTech, RegTech, and the Reconceptualization of Financial Regulation*, 37 NW. J. INT'L L. & BUS. 371, 373 (2017); William Boyd, *Environmental Law, Big Data, and the Torrent of Singularities*, 64 UCLA L. REV. DISCOURSE 544, 548 (2016).

[224] *See* Terrence Hendershott et al., *Does Algorithmic Trading Improve Liquidity?*, 66 J. FIN. 1, 28–30 (2011); Bjorn Hagstromer & Lars Norden, *The Diversity of High-Frequency Traders*, 16 J. FIN. MKTS. 741, 768–69 (2013); Jennifer Conrad et al., *High-Frequency Quoting, Trading, and the Efficiency of Prices*, 116 J. FIN. ECON. 271, 289–90 (2016); Terrence Hendershott & Pamela C. Moulton, *Automation, Speed, and Stock Market Quality: The NYSE's Hybrid*, 14 J. FIN. MKTS. 568, 601 (2011).

trends, they can better allocate their capital to deserving companies.[225] This can reduce the size and length of price asymmetries in markets and thereby make them better reflect the value of assets.[226] Greater data can also lead to less volatility as new information is constantly reflected in stock prices.[227] As there is more information in the markets, the value of any single piece of information declines, thereby reducing the likelihood that a news item will lead to dramatic shifts in the market.[228] And as the likelihood of shocks decreases, markets should become less prone to crashes and, thus, systemic risks that reverberate throughout the economy.[229] From the regulatory side, greater data access also means that regulators can identify risks sooner and more accurately, thereby reducing the chance that a systemically important financial institution has unknown vulnerabilities.[230] This literature thus argues that the data economy is contributing to greater, not lesser, market stability.

IV. A Unified Law of Data

The data economy today is driven by data's unique ability to provide information with magnitude, permanence, and portability. But these features have also created sharp debates about the risks and rewards of data. The debates revolve around three key axes: fairness, efficiency, and stability. As the last Part demonstrated, along each of these dimensions, there are strong arguments on either side. Data might be discriminatory, or it might be egalitarian. Data might be wasteful, or it might be efficient. Data might be destabilizing, or it might be reinforcing. The dichotomies are stark and often irreconcilable. Intuitions about the proper use of data tend to depend on whether we favor the particular result, not on the actual type of data or the decision whether to collect and store it in the first place. We might be happy for a company to access our location data to give us better mapping directions. We would be less happy if they allowed a hacker to steal that data. The difficulty of fashioning answers to these problems has led data regulation to be narrowly targeted and circumscribed in scope—a Health Insurance Portability and Accountability Act to regulate health data, a Financial Services Modernization Act to regulate financial data, a Federal Information Security

---

[225] *See* Jennie Bai et al., *Have Financial Markets Become More Informative?*, 122 J. Fin. Econ. 625, 62 (2016).

[226] *See* Conrad et al., *supra* note 224, at 271.

[227] *See* Hamid Mohtadi & Stefan Ruediger, *Does Greater Transparency Reduce Financial Volatility?*, (Working Paper), https://cpb-us-w2.wpmucdn.com/sites.uwm.edu/dist/0/252/files/2016/07/does-transparency-reduce-financial-volatility_5_14_2013-1fz624i.pdf [https://perma.cc/X8DH-HLPS].

[228] *See* Hedi Benamar et al., *Demand for Information, Uncertainty, and the Response of U.S. Treasury Securities to News*, (Working Paper, Feb. 19, 2019), https://conference.nber.org/conf_papers/f117462.pdf [https://perma.cc/H4CW-JGNE].

[229] *See* Magnuson, *Regulating Fintech*, *supra* note 198, at 1188–93.

[230] *See* Omarova, *supra* note 223, at 48–52; Arner et al., *supra* note 223, at 371.

Management Act to regulate data held by the government—or left to states to develop in ad hoc and often conflicting legislation—the CCPA, Massachusetts' Act Relative to Consumer Protection from Security Breaches, and Maryland's Online Consumer Protection Act.[231]

But, as this Part will argue, there is value in creating a more cohesive set of principles to govern data. Data cannot be easily cabined into industry or area, and it crosses borders instantaneously. A single piece of data can be used in a multiplicity of ways, and a unified law of data would bring needed clarity and uniformity to a tumultuous area of law. It would go a long way towards ensuring that data is fairer, more efficient, and more stable. This Part will explore three key principles of what such a unified law might look like—related to ownership, access, and security—as well as ways in which these principles might be enshrined to favor particular values over others, and where flash points of conflict are most likely to arise.[232]

A. Private Ownership

First and foremost, a unified law of data would need to establish clear property rights over data.[233] The default here would be to grant consumers, users, and individuals substantial control over any data related to them. A corollary of this proposition is that data owners would have rights to possess, control, exclude, and dispose of their data as they see fit. Their location data, their email data, their phone call data, their health data, and their financial

---

[231] *See generally* Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-191, 110 Stat. 1936 (1996); Financial Services Modernization Act, Pub. L. 106-102, 113 Stat. 1338 (1999); Federal Information Security Management Act of 2002, Pub. L. 107-347, 116 Stat. 2899 (2002); California Consumer Privacy Act of 2018 ("CCPA"), CAL. CIV. CODE §§ 1798.100-1798.198 (West 2020); An Act Relative to Consumer Protection from Security Breaches, 2018 Mass. Acts 444; Maryland Online Consumer Protection Act SB0613, 2020 Leg., 441st Sess. (Md. 2020).

[232] It should be noted that the European Union's General Data Protection Regulation and California's Consumer Privacy Act both come close to being comprehensive regulations of data and its treatment, at least within their respective jurisdictions. So, the proposal here is not quite as radical as it might at first glance appear to be. It would, however, mark a sharp departure from current approaches under U.S. federal law.

[233] For analyses of legal status of data ownership rights, see Jeffrey Ritter & Anna Mayer, *Regulating Data as Property: A New Construct for Moving Forward*, 16 DUKE L. & TECH. REV. 220, 223 (2018) (arguing that an "explicit, legal mechanism to establish, claim and transfer property rights in data must be adopted"); Vera Bergelson, *It's Personal but Is It Mine? Toward Property Rights in Personal Information*, 37 U.C. DAVIS L. REV. 379, 383 (2003) (arguing that "in order to protect privacy, individuals must secure control over their personal information by becoming its real owners"); Stacy-Ann Elvy, *Commodifying Consumer Data in the Era of the Internet of Things*, 59 B.C. L. REV. 423, 463 (2018) (noting that "[c]ommentators have adopted contrasting positions on questions of data ownership, rights in data and the legal regimes that should govern related issues"); Nancy S. Kim, *Contract's Adaptation and the Online Bargain*, 79 U. CIN. L. REV. 1327, 1356 (2011) ("It is unclear what legal right or interest, if any, consumers have in their personal information."); Andreas Boerding et al., *Data Ownership: A Property Rights Approach from a European Perspective*, 11 J. CIV. L. STUD. 323, 325 (2018).

data would all be owned by the individuals themselves. They could keep it or sell it when and how they wanted.

This might seem like an obvious point that could not be seriously debated. But in fact, ownership rights over data are both legally ambiguous and deeply contentious.[234] In a well-publicized case against the file sharing site Megaupload, the U.S. government argued that users who had stored data on the cloud did not retain ownership over the data uploaded.[235] The Brookings Institution issued a report in 2019 arguing that individuals do not, and should not, have property rights over data, even if it is related to their personal lives.[236] Indeed, the issue is so unclear that in 2019, Senator John Kennedy introduced a bill, the Own Your Own Data Act, that would have provided that "each individual owns and has an exclusive property right in the data that individual generates on the internet."[237] The bill has not passed.

Defining just what counts as personal data worthy of property recognition would, of course, be a crucial element of the right. After all, data is simply information, and thus raises a number of issues not presented by, say, property rights in a house or a car. Some areas would be easy. Data ownership rights would clearly extend to medical records, meaning that patients would own their X-rays and vaccination histories and could take them with them wherever they go. Data ownership rights would also extend to location data, meaning that cellphone users could access and transfer records of where they have been. Ownership rights would apply to bank records, as well, meaning that consumers could store and share information about where and how they have spent money. But other areas are harder. Does the recipient or the sender own the data in an email message, or do they both? If a borrower defaults on a mortgage, does the bank own the data related to that fact, or does the borrower? If a shop has a security camera that records customers who enter, who owns the data on the camera? These are difficult problems, and their seriousness should not be minimized. At the same time, there are ways to deal with them. The GDPR, for example, starts with a broad right (all "personal data" must be processed lawfully and fairly) and

---

[234] In the healthcare space, see, e.g., Barbara J. Evans, *Would Patient Ownership of Health Data Improve Confidentiality?*, 14 Am. Med. Ass'n J. Ethics 724, 728 (2012); I. Glenn Cohen, *Is There a Duty to Share Healthcare Data?*, *in* Big Data, Health Law, & Bioethics 209 (I. Glenn Cohen et al., eds., 2018).

[235] Brief of the United States Regarding the Breadth and Format of a Hearing to Determine the Applicability of Federal Rule of Criminal Procedure 41(g), United States v. Kim, No. 1:12–cr–3, 2012 WL 5474807 (E.D.Va. 2012).

[236] *See* Cameron F. Kerry & John B. Morris, *Why Data Ownership Is the Wrong Approach to Protecting Privacy*, Brookings Inst. (June 26, 2019), https://www.brookings.edu/blog/techtank/2019/06/26/why-data-ownership-is-the-wrong-approach-to-protecting-privacy/ [https://perma.cc/FT44-Q9WD].

[237] Own Your Own Data Act, S. 806, 116th Cong. § 2(a) (2019); *see* Daniel R. Stoller, *Lawmakers Question Need for Data Ownership at Senate Hearing*, Bloomberg (Oct. 24, 2019, 3:36 PM), https://news.bloomberglaw.com/privacy-and-data-security/lawmakers-question-need-for-data-ownership-at-senate-hearing [https://perma.cc/WHX2-ZFKY].

then proceeds to delineate a lengthy series of exceptions to the right.[238] A similar approach could be used in developing data ownership rights.

One of the advantages of a data ownership model of data regulation is that it creates room for individuals to make their own decisions about the tradeoffs of privacy, convenience, and other values. Rather than imposing a single regime about when and where firms can use consumer data, data ownership would allow more individualized results. One of the dangers of current regulations of data privacy, for example, is that they tend to create barriers even to consensual data sharing practices. It is remarkably complex for consumers to share a comprehensive picture of their health data, requiring them to go to numerous health care providers, fill out duplicative authorization forms, and pay large fees.[239] The process is so complex that more than seventy percent of Americans have not seen even a single health record in the last year.[240] Similarly, the European Union's regulations about strong customer authentication for banks have raised the barriers that consumers must overcome in order to share their financial data.[241] A data ownership regime would help reduce these sorts of costly overweighting problems by giving individuals the right to decide for themselves how to use, store, and share their data.

One important point of tension in a data ownership regime would surround the question of consent. Property rights are premised on the ability of owners to do as they wish with their property, and thus, we must delve into their intent when deciding proper legal outcomes. But, of course, consent is tricky on the internet. Companies might, for example, attempt to skirt around data ownership rights by simply including a provision in their terms and conditions forcing owners to waive their rights. Take, for example, the current terms and conditions in Apple's iCloud service for storing data (which, to be clear, is far from an extreme example).[242] Section V.H.1 of the terms and conditions provides that, by submitting or posting content on areas of iCloud that are accessible to the public or other users, users "grant Apple a

---

[238] The GDPR defines "personal data" to mean "'any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person." Regulation (EU) 2016/679, of the European Parliament and the Council of 27 April 2016 on the protection of natural persons with regards 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 at arts. (4)(1), (5)(1)(a), 6(1).

[239] *See* Harlan Krumholz, *It's Your Right to See Your Medical Records. It Shouldn't Be This Hard to Do*, NPR (Aug. 28, 2019, 5:01 AM), https://www.npr.org/sections/health-shots/2019/08/28/754725843/opinion-its-your-right-to-see-your-medical-records-it-shouldn-t-be-this-hard-to- [https://perma.cc/8ZTC-STJD].

[240] *See* Picnic Health, Data Ownership, https://picnichealth.com/data-ownership [https://perma.cc/69JR-P6TH].

[241] *See* Council Directive 2015/2366, art. 4(30), 2015 O.J. (L 337/35).

[242] *iCloud Terms and Conditions*, Section V.H.1, https://www.apple.com/legal/internet-services/icloud/en/terms.html [https://perma.cc/V8AY-2Q4J].

worldwide, royalty-free, non-exclusive license to use, distribute, reproduce, modify, adapt, publish, translate, publicly perform and publicly display such Content on the Service solely for the purpose for which such Content was submitted or made available, without any compensation or obligation to you."[243] The somewhat confusing provision appears to suggest that if you upload a photo to the cloud and share it with others, Apple could publicly display it, or even Photoshop it, so long as Apple had the same general purpose as the user initially had. It should be noted that Apple expressly disclaims ownership of information stored on its servers, but such a claim could potentially be made by other companies and might well defeat the purpose of establishing data ownership rights in the first place. Given the problems inherent in online contracting—including whether people actually read terms and conditions—default rules requiring strong consent procedures for the transfer of data ownership must be established.

B. Public Access

Strong private data ownership rights, however, must be paired with equally strong public access rights. A type of eminent domain for data would give governmental entities the right to collect, access, and use data—even privately held data—so long as they have legitimate purposes for doing so. Government data access could help promote essential societal goals, such as responding to healthcare crises, preventing crime and terrorism, and stabilizing financial markets.

Again, public access rights might appear to be a basic and incontrovertible proposition—how could government not have the right to access the data it needs for its functions? But again, the question is contentious and legally ambiguous. In a well-publicized case in 2019, Apple refused to provide assistance to government efforts to unlock the iPhone of a gunman who had attacked the Pensacola Naval Air Station and who law enforcement agents believed had connections with Al Qaeda.[244] Apple argued that creating a backdoor for law enforcement agencies would cripple the company's cybersecurity mechanisms and open up avenues for other, less well-intentioned actors to exploit.[245] Similarly, during the coronavirus pandemic, the federal government was forced to rely on mobile advertising companies to gain access to location data of citizens to track the spread of the virus, after

---

[243] *Id.*

[244] *See* Chris Welch, *The FBI Successfully Broke Into a Gunman's iPhone, But It's Still Very Angry at Apple*, VERGE (May 18, 2020, 1:05 PM), https://www.theverge.com/2020/5/18/21262347/attorney-general-barr-fbi-director-wray-apple-encryption-pensacola [https://perma.cc/K95Y-8SPM].

[245] *Id.*

uncertainties arose about whether they could access the data directly from cellphone companies.[246]

But a thriving data economy that is impervious to government scrutiny is a problem. While claims about the dangers of granting backdoors into data storage systems are valid and important, they should not overcome the basic fact that individuals store more and more of their information on such systems. To prevent governments from accessing that data for their operations would be to deprive them of the most important tool they have to promote the public good. Governments should have a right to access data. And as a corollary, companies should have an obligation to give governments the tools to do so.

This does not mean that the power of eminent domain over data would be unlimited. Even if governments have the right and the ability to access data created and stored by private entities for legitimate reasons, the scope of this right would need to be conscribed by prudential principles, particularly transparency and reviewability. In recent years, a drumbeat of discoveries have highlighted the extent of government data collection, and, in particular, the access that police departments can gain to private data.[247] In 2019, Amazon disclosed that video footage from Ring doorbell cameras could be downloaded and stored by police departments, potentially forever, and that police departments could share the footage with others.[248] Police departments can also access audio recordings captured by Amazon's virtual assistant device Alexa.[249] Police officers have also received warrants to search consumer DNA sites for genetic profiles to solve crimes.[250] In order to reduce the risk of abuse, these programs must be transparent to citizens and reviewable by the judiciary.

C. Security

A final principle of a unified law of data would be that possessors of data must protect it with adequate cybersecurity measures. Of course, saying that companies should prevent hacks is a bit like saying that people should be healthy: it's a fine idea, but devilishly difficult to accomplish. At the same

---

[246] *See* Byron Tau, *Government Tracking How People Move Around in Coronavirus Pandemic*, WALL ST. J. (Mar. 28, 2020), https://www.wsj.com/articles/government-tracking-how-people-move-around-in-coronavirus-pandemic-11585393202 [https://perma.cc/XGY7-M3HC].

[247] *See* Jack Goldsmith & Andrew Keane Woods, *Internet Speech Will Never Go Back to Normal*, ATLANTIC (Apr. 25, 2020), https://www.theatlantic.com/ideas/archive/2020/04/what-covid-revealed-about-internet/610549/ [https://perma.cc/9UHU-KU4B].

[248] *See* Harwell, *supra* note 70.

[249] *See* Jon Fingas, *Florida Police Obtain Alexa Recordings in Murder Investigation*, ENGADGET (Nov. 2, 2019), https://www.engadget.com/2019-11-02-florida-police-obtain-alexa-recordings-in-murder-case.html [https://perma.cc/Q72R-JYQZ].

[250] *See* Kashmir Hill & Heather Murphy, *Your DNA Profile is Private? A Florida Judge Just Said Otherwise*, N.Y. TIMES (Nov. 5, 2019), https://www.nytimes.com/2019/11/05/business/dna-database-search-warrant.html [https://perma.cc/5PXK-8VHY].

time, regulation can play an important role in improving cybersecurity practices, both through ex ante monitoring structures and ex post liability requirements.

First and foremost, data regulations must set out clear, specific, and affirmative obligations on possessors of data to protect it from unauthorized access. Data security rules would need to establish the basic obligation of cybersecurity (which is currently shoehorned in through a variety of disparate rules and regulations), but they should also go further. Companies should be required to establish not just "reasonable" cybersecurity procedures, but "best-in-class" procedures.[251] Too often, cybersecurity procedures at companies amount to little more than "check-the-box" exercises.[252] Regulators could issue rules specifically setting out the particular tools and mechanisms that companies in particular industries needed to adopt.[253] And furthermore, regulators should increasingly offer "cyberhygiene" scans, in which experts review companies' cybersecurity mechanisms to detect flaws or vulnerabilities.[254]

Second, there need to be strong and broad liability rules requiring possessors of data to compensate individuals if their data is hacked. Too often, possessors of data can deflect or delay liability by arguing that someone else was at fault or that the consumer cannot prove damages or even that the consumer lacks standing to bring a claim in the first place.[255] Instead of placing the burden on consumers to identify the methods used by the hacker or the precise harms generated to them, data regulation should establish a clear rule that possessors of data must compensate consumers when hackers have

---

[251] *See* NAT'L CONF. STATE LEG., DATA SECURITY LAWS—PRIVATE SECTOR (2019), https://www.ncsl.org/research/telecommunications-and-information-technology/data-security-laws.aspx [https://perma.cc/RQQ6-AJPQ] (noting that most state data security laws "require businesses that own, license, or maintain personal information about a resident of that state to implement and maintain 'reasonable security procedures and practices' appropriate to the nature of the information and to protect the personal information from unauthorized access, destruction, use, modification, or disclosure"); Jeff Kosseff, *Congress Is Finally Tackling Privacy! Now Let's Do Cybersecurity*, SLATE (Dec. 3, 2019, 3:00 PM), https://slate.com/technology/2019/12/congress-national-privacy-law-cybersecurity.html [https://perma.cc/E9VH-8C4Z].

[252] *See* Roger A. Grimes, *2 Critical Ways Regulations and Frameworks Weaken Cybersecurity*, CSO (Jan. 10, 2019, 3:00 AM), https://www.csoonline.com/article/3332139/2-critical-ways-regulations-and-frameworks-weaken-cybersecurity.html [https://perma.cc/DWF8-XCPE].

[253] New York, for example, has an extensive list of specific cybersecurity procedures that financial institutions must abide by. *See* N.Y. COMP. CODES R. & REGS. tit. 23, § 500 (2017).

[254] *See* Michael Wines, *Wary of Hackers, States Move to Upgrade Voting Systems*, N.Y. TIMES (Oct. 14, 2017), https://www.nytimes.com/2017/10/14/us/voting-russians-hacking-states-.html [https://perma.cc/GC33-ULSM].

[255] *See generally* Seth D. Rothman & Dennis S. Klein, *Defending a Data Breach Class Action*, 255 N.Y.L.J. S6 (Jun. 6, 2016), https://www.hugheshubbard.com/news/defending-a-data-breach-class-action [https://perma.cc/CBH8-66G6]. *See also* Jon R. Knight, *The New Normal: Easier Data Breach Standing Is Here to Stay*, CYBERSEC. LAW REP. (Feb. 6, 2019), https://www.privacysecurityacademy.com/wp-content/uploads/2019/06/SF1608-REVISED_The-New-Normal-Easier-Data-Breach-Standing-Is-Here-to-....pdf [https://perma.cc/5JA7-5K2W].

gained access to their data. If it turns out that a third-party vendor or other party is at fault, then the primary data possessor should have the right to receive indemnification from those parties ex post. Such broad liability rules would help incentivize companies to devote greater resources to cybersecurity and help compensate victims in a speedy and efficient way. They would also provide an efficient way to force companies that profit from storing data to "internalize" the cost of breaches. To the extent that such rules would make data collection and storage more costly, this would merely correct for the current deflated costs companies expend while externalizing the risk to others.

These rules would also need to be mandatory and non-waivable in order to prevent companies from contracting around them. To return to Apple's iCloud service, the terms and conditions provide the following with regard to hacking and cybersecurity:

- "You are solely responsible for maintaining the confidentiality and security of your Account and for all activities that occur on or through your Account . . . . Provided we have exercised reasonable skill and due care, Apple shall not be responsible for any losses arising out of the unauthorized use of your Account resulting from you not following these rules."[256]
- "Apple does not represent or guarantee that the service will be free from loss, corruption, attack, viruses, interference, hacking, or other security intrusion, and Apple disclaims any liability relating thereto."[257]
- "You expressly understand and agree that Apple . . . shall not be liable to you for any direct, indirect, incidental, special, consequential or exemplary damages, including, but not limited to, damages for loss of profits, goodwill, use, data, cost of procurement of substitute goods or services, or other intangible losses (even if Apple has been advised of the possibility of such damages), resulting from: (i) the use or inability to use the service . . . (iii) the unauthorized access to or alteration of your transmissions or data; . . . and (vi) any other matter relating to the service."[258]
- "You agree to comply with this Agreement and to defend, indemnify and hold harmless Apple from and against any and all claims and demands arising from usage of your Account, whether or not such usage is expressly authorized by you."[259]

These sorts of broad waivers might well deter consumers from bringing even valid claims related to hacks, and, if broadly adopted, could defeat the essential purpose of cybersecurity rules. Instead, a unified data regulation must clarify that data possessors have non-waivable obligations to establish best-

---

[256] *See iCloud Terms and Conditions*, *supra* note 242.
[257] *Id.* at Section IX.
[258] *Id.*
[259] *Id.*

in-class cybersecurity procedures and to compensate users for breaches when they happen.

V. Conclusion

The explosion of the data economy has raised a host of thorny legal, political, and social problems. Some of these are related to data's drawbacks, some are related to its opportunities, and many are somewhere in between. This Article has attempted to provide a unified treatment of the field, highlighting both areas of agreement and disagreement. It has also attempted to sketch out principles that should apply broadly to the industry as a whole. Too often, the conversations about the future of data take place in isolation and without dialogue. It is hoped that this Article will bring some unity to the field.