

RULE BY DATA: THE END OF MARKETS?

KATHARINA PISTOR*

I

INTRODUCTION

This Article explores data as a source and, in their processed variant, as a means of governance that will likely replace both markets and the law. Discussing data not as an object of transactions or an object of governance, but as a tool for governing others on a scale that rivals that of nation states with their law, seems a fitting topic for a special issue that is devoted to the legal construction of markets. Here, I argue that while it may well be the case that law constitutes markets, markets are not the only way in which economic relations may be organized, and law is not the only feasible mode of governing these relations. Central planning under socialism posed an alternative, which proved ultimately non-viable.¹ The rise of big tech companies (Big Tech) and their accumulation of vast amounts of data offers yet another possibility: the rule by data.

In his path-breaking article on the nature of the firm, published in 1937, Ronald Coase famously posed the following question: if markets are the optimal form of economic organization, why do firms exist?² His answer to this question was that some transactions are better governed inside hierarchically organized firms than in open markets.³ The explanation he gave focused on the inevitable transaction costs that beset markets, where markets are defined by “price movements [that] direct production, which is co-ordinated through a series of exchange transactions on the market.”⁴ Implied in this definition (which I will embrace for this Article) is that goods or services are traded for a unit of account, or money, and that a series of transactions reveals the prevailing market price. And as Coase added in another seminal paper on the “problem of social costs,” the allocation of property rights is key for market exchange.⁵ Yet, as I will discuss

Copyright © 2020 by Katharina Pistor.

This Article is also available online at <http://lcp.law.duke.edu/>.

* Edwin B. Parker Professor of Comparative Law, Columbia Law School. I would like to thank participants of the workshop on “The Legal Construction of Markets” at Yale Law School on September 20–21, 2019. I am particularly grateful to Amy Kapczynski for her insightful comments. All remaining errors are mine.

1. For a comprehensive analysis of the weaknesses of the socialist system, see JANOS KORNAI, *THE SOCIALIST SYSTEM: THE POLITICAL ECONOMY OF COMMUNISM* (1992).

2. Ronald H. Coase, *The Nature of the Firm*, 4 *ECONOMICA* 386, 388 (1937).

3. *Id.*

4. *Id.*

5. See Ronald H. Coase, *The Problem of Social Cost*, 3 *J.L. & ECON.* 1, 15–19 (1960) (“Once the costs of carrying out market transactions are taken into account it is clear that such a rearrangement of

in this Article, Big Tech has found a way to retain ownership over data even as it sells the data again and again, and on terms that Big Tech controls. Money changes hands, yet only *access* to data and their predictive power is granted in return. If this is a market in the original sense of the word, it is a rather peculiar one.

The Coasean framework also suggests that the enormous reduction of transaction costs would lessen the need for law. “[I]f market transactions were costless, all that matters (questions of equity apart) is that the rights of the various parties should be well-defined and the results of legal actions easy to forecast.”⁶ In other words, even in a hypothetical world with zero transaction costs, there is still room for law to allocate the initial rights. How exactly these rights are allocated does not matter much for efficiency (in contrast to distributional effects) because zero transaction costs will ensure that these rights will eventually be allocated to whoever can make the most efficient use of them.

Today, over eighty years after Coase’s 1937 article, transaction costs have arguably been brought under control. This has given us the possibility to communicate with almost anybody anywhere in the world within seconds; to conduct high frequency trading; and to surveil millions of people, second by second.⁷ If Coase was right that firms are primarily a product of (or response to) transaction costs, we should be witnessing the disappearance of firms and the flattening out of markets. If markets are the optimal organization of economic exchange, then this is true even if information technology also reduces intra-firm transaction costs.

Despite the fact that information technology has brought down transaction costs dramatically, it has not eliminated firms. Rather, the enormous amount of data produced by information technology has given us the rise of Big Tech—global corporate behemoths. Big Tech surveils individuals (the data producers) who use its services or have become objects of its data surveillance schemes. It harvests the data, and it turns the data into governance tools capable of predicting and shaping the behavior of individuals and groups (the data targets) in the interests of Big Tech’s own clients, who purchase predictive power from it. Big Tech is not fostering efficient markets by collecting and sharing information with its customers or its clients; rather, it is inserting itself into the exchange by extracting information from consumers and selling processed information to various clients. These clients then possess sufficient predictive power to make the behavior of their customers highly predictable. Big Tech and its clients gain from this asymmetry of predictive power at the expense of the consumers.

rights will only be undertaken when the increase in the value of production consequent upon the rearrangement is greater than the costs which would be involved in bringing it about.”).

6. *Id.* at 19.

7. See 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/H6WA-UXJY>] (noting the treasure trove of data transmitted by mobile phones).

Moreover, in the world of big data, information costs may not be zero, but they certainly approximate this lower bound. Yet, buyers and sellers do not transact on equal footing with knowledge only of their respective legal rights and some sense about how a dispute might be resolved. Instead, one of the two parties, typically the seller, is granted access to the predictive power of data that have been collected from data producers, whereas the buyer on the other end of this chain is left in the dark. In short, (close to) zero transaction costs have not eliminated firms; nor have they foregrounded legal rights and perfect forecast for all parties to a transaction. Instead, Big Tech firms have monopolized the informational content and governance power of data and monetized it by granting only selective access to it.

Indeed, although the raw data individuals produce may be non-rivalrous, Big Tech companies are transforming them such that their *use* becomes rivalrous. The choice between markets and firms then is not, or is not primarily, a response to transaction costs—but rather a response to the differential ability to create asymmetries of power. Coase might agree with this; after all, he realized that the problem of social costs is at least in part the result of existing power relations. However, those who cling to the Coase theorem—that is, the notion that under conditions of zero transaction costs markets allocate resources efficiently—may be less inclined to agree.⁸

Characterizing the business of Big Tech as creating asymmetries of power sheds new light on their relation to the consumers whose behavior they wish to influence. Consumers access digital platforms for a variety of purposes, and in doing so are turning themselves into unpaid data producers. For example, they may wish to purchase goods on Amazon; communicate via their Apple or Android smart phones or Gmail or Microsoft Outlook email accounts; build online social networks on Facebook and Twitter; or simply surf the web for all kinds of information on Google or Bing.⁹ Consumers may have more than one platform to choose from; however, powerful network effects imply that once networks have reached a certain size, they are likely to tip, leaving the winner to take all.¹⁰

Entry barriers and search costs are down for all these services, but this does not mean that consumers are in fact free to choose which services they use, or which goods they buy. Big Tech has taken over the process of selecting and choosing. It shares with paying clients data about the behavior of consumers that it has collected itself, often supplemented by additional data it acquires from data brokers. Big Tech companies also occupy a dominant market position vis-à-vis their clients, who buy processed information that allows them to target

8. For a critique by Coase himself of the focus on the Coase theorem in the literature, see RONALD H. COASE, *THE FIRM, THE MARKET, AND THE LAW* 159 (1990).

9. Note that not every search engine operates as a data harvester. Firefox, for example, does not and is therefore better described as an access provider.

10. See DIGITAL COMPETITION EXPERT PANEL, *UNLOCKING DIGITAL COMPETITION* 4 (2019).

consumers with ads and information to enhance the probability that their products will be purchased. In short, Big Tech uses its surveillance and power to serve two “markets”: the market for consumers and the market for sellers.

One might compare the role of Big Tech with brokers who operate as matchmakers in commodities markets or stock exchanges, for example.¹¹ Unlike classical brokers, however, Big Tech does not match buyers and sellers. Instead, companies in this sector extract data from data producers without pay and build a database by cleansing, classifying, and processing the vast amounts of raw information they have harvested. Once they have refined the analytics, they then sell the predictive power of this information to their clients—the sellers of goods and services to customers, many of whom have earlier served as data producers. This is how Big Tech monetizes the information it controls. Given the vast amounts of data these companies have amassed, they face little competition.

In sum, in the world of big data controlled by Big Tech, data are not primarily objects of exchange transactions; rather, they are both the source for and the means of control by Big Tech and their clients over others: consumers of goods and services, workers, voters, members in organizations, or whatever other targets they might choose. Indeed, in this environment, the contracts between the sellers of goods and services and their customers have not been bargained for in the Coasean sense; rather, they have been preordained by an algorithm.

The remainder of this Article proceeds as follows. In Part II, I discuss the value of data, and compare data with other ordinary goods or property claims. I also discuss how the power data confer on data controllers leads away from property rights and from markets. Part III rethinks the world of data as hierarchy, not markets, and I suggest that the world of Big Tech and big data poses new and fundamental challenges to the formal equality of market participants because of their one-sided control over predictive and manipulative data power. In Part IV, I argue that there is a need for a third wave of accountability, namely accountability of the data harvesters and their clients to the data producers, many of whom are also the consumers of products and services at the other end of the data value chain. In Part V, I sketch a collective governance structure for databases, and, in Part VI, I conclude.

II

THE VALUE OF DATA

Most dictionaries define data as facts, numbers, or just information that is used as the foundation for reasoning and decision-making.¹² Every person, simply

11. For a similar comparison, see generally Anish Agarwal et al., *A Marketplace of Data: An Algorithmic Solution*, 2019 ACM CONF. ON ECON. & COMPUTATION 4–5.

12. See, e.g., *Data*, MERRIAM-WEBSTER ONLINE DICTIONARY, <https://www.merriam-webster.com/dictionary/data> [<https://perma.cc/G257-A4DV>]; *Data*, CAMBRIDGE DICTIONARY ONLINE, <https://dictionary.cambridge.org/us/dictionary/english/data> [<https://perma.cc/74R8-Q3MC>].

by being and behaving, constantly produces raw information that can be turned into data and then organized into databases that can generate the power to predict subsequent behavior. People collect information about others and their environment, process these data, and act on the results of that information processing. Humans would not be able to function without processing information—not even on Robinson Crusoe’s lonely island, where, given the absence of other human beings, there was little need for social norms or for law. Still, without observing and processing information about plants, animals, and water sources, and acting upon that information, Robinson would have perished in no time. In short, data are the stuff from which *relations* among human beings and between them and their environment are made. Surveilling a single individual generates data not only about her, but also about many others with whom she interacts. Controlling these data bestows enormous power on whoever can get their hands on them.

In fact, power seems a better explanation for the rise of Big Tech than the standard transaction cost arguments. It may well be that firms have an advantage over markets in terms of transaction costs in general. The rise of information technology, however, makes this argument far less compelling than it was in the 1930s, when Coase wrote his famous article. Today, controlling vast amounts of data and determining who can access them and at what price is an expression of economic power. And by making access to digital platforms contingent on individual data producers releasing their data, platform owners have effectively created a self-replenishing well for the source of their power.¹³

The worth of data does not lie in their exchange value but in the power they confer on data controllers. Data controllers harvest data from millions of often unsuspecting individuals, and process and analyze them with the goal of perfecting predictability about the future behavior of their targets. This also empowers the controllers of large datasets to make predictions at the macro-level about aggregate behavior, including GDP forecasts. The classified and processed data and their predictive power, rather than the raw information, can be monetized. Other companies are buying access to these data or to the predictive algorithms. The elimination of future uncertainty for only one of the parties to the “transaction,” the seller, tilts the playing field in its favor. This one-sided exploitation of data leads to the elimination of markets (as a series of exchange transactions for money that reveal value), and, in its place, rule by data.

To demonstrate this, I compare data in this Part with ordinary goods or property that can be turned into wealth generating, or capital, assets, with the

13. For more information on data surveillance, see generally SHOSHANA ZUBOFF, *THE AGE OF SURVEILLANCE CAPITALISM* (2019). And to learn more about “nudging,” see generally RICHARD H. THALER & CASS R. SUNSTEIN, *NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS* (2009).

right legal coding.¹⁴ Data are often labeled the “new oil,” which earlier replaced gold as an important, if not the primary, source of wealth. In fact, oil was given the telling name “black gold.”¹⁵ These metaphors are misleading, however, in that they treat data as just another asset that is tangible and finite, and where its value is derived from using or exchanging it. Gold has been cherished for these characteristics. It has been used for jewelry, artwork, furniture, and buildings. Moreover, gold coins have served as a means of payment (money), as a storage of value, and as a unit of account. Gold was eventually complemented by and later substituted with paper money. And, at times, gold was used to back paper money, which was used as the primary means of exchange. In this role, gold retained its image as an asset of intrinsic value: its precious material, shiny surface, and scarcity. Oil is certainly less aesthetically attractive than gold and has rarely, if ever, been used as money. Its value derives from the fact that it is a powerful source of energy for which there has been growing demand for over a century as more and more countries developed economically and relied heavily on oil to quite literally fuel their growth.

Data are different. Data are intangible, not tangible; they are ubiquitous, not scarce; they have neither use nor (much) exchange value in the form of small amounts of unprocessed, raw information. Indeed, data are rarely exchanged in markets that offer price discovery. Instead buyers have to rely on sellers’ representations of the data’s worth. Most of data’s value is derived from economies of scale: from the sheer volume of data harvested and the capacity to process, analyze, and use them in order to induce behavior in others. What counts is neither their use nor their exchange value, but rather their predictive power. It is this power that is monetized by the data harvesters or controllers when selling access to the raw data or the results of algorithms that generate the desired predictions.

The shift from expected returns on assets to the predictive power of data to estimate consumer behavior might be interpreted as a new turn in the evolution of property rights. John Commons discovered the legal foundations of capitalism when he was on the lookout for the sources of economic value.¹⁶ He detailed the shift from protecting property rights as use value to expected value in the case law of the U.S. Supreme Court between the late nineteenth and the early

14. For details, see generally KATHARINA PISTOR, *THE CODE OF CAPITAL: HOW THE LAW CREATES WEALTH AND INEQUALITY* (2019).

15. On the origin of this analogy and for a critical assessment, see Samuel Flender, *Data is Not the New Oil*, TOWARDS NEW DATA SCI. (Feb. 10, 2015), <https://towardsdatascience.com/data-is-not-the-new-oil-bdb31f61bc2d> [<https://perma.cc/Y6AA-6J7D>] (“The thing with oil is, once an oil company find[s] it in the ground, they know more or less which steps they have to follow to turn that oil into profits: drill, extract, refine, sell. This is far from the reality that you face when dealing with data: *when dealing with data, it is far from clear how exactly to turn that data into profits.*”).

16. See JOHN R. COMMONS, *THE LEGAL FOUNDATIONS OF CAPITALISM* 304–05 (1924) (noting the relationship between the law and economic coercion).

twentieth centuries. This finding led him to the conclusion that economic value is ultimately derived from law and that law is the foundation of capitalism.

In contrast, the turn to predictive power of consumer behavior leads away from property rights and from markets. First, the predictive power of data can be generated without much help from the state in the creation and initial allocation of property rights.¹⁷ This distinguishes data from the notion of expected returns and other intangible assets, such as intellectual property rights or financial assets, which owe their existence to law. The value of these legal constructions depends on the willingness of states to back them with their coercive powers and render them enforceable. In contrast, data controllers rely mostly on self-help by way of technological barriers to accessing their databases. To be sure, data controllers have lobbied states to protect the data they have harvested and placed on some physical device against hacking, or theft, and more recently have championed the tightening of trade secrecy law, including the addition of criminal sanctions for breach. However, Big Tech has stopped short of claiming full-throttled property rights protection from the state. Not only do tech companies not need this kind of support because they have technological means at their disposal to govern access to the data they have amassed, but they have also benefited from the ambiguity that has surrounded data ownership as they have moved to enclose and extract data from billions of individuals. In effect, Big Tech treated the data it captured as *res nullius*, or wild animals: things that belong to no one but can be claimed by whoever catches them first.¹⁸

Data are, of course, not the first or the only assets that have benefited from legal protection that was granted after the fact, or after de facto control had already been established. The French anarchist, Pierre-Joseph Proudhon, famously asserted that “all property is theft.”¹⁹ There is some truth to it in that most state sanctioned property rights begin that way. Rarely, if ever, are de facto control rights altered when property rights are formalized and sanctioned by law; instead, the allocation of property rights typically entrenches de facto control rights.²⁰ This pattern can be observed in the enclosure of land in early modern Europe, as well as in the enclosure of intellectual property rights on a global scale.²¹

17. This is not to deny that Big Tech has physical footprints on territory controlled by nation states, nor that they use contracts or other legal devices.

18. Incidentally, this rule also governs the capture of underground resources in many jurisdictions, including oil. On this concept and for a comparison with *res communis* (or public goods), see generally Richard A. Epstein, *Property Rights and Governance Strategies: How Best to Deal with Land, Water, Intellectual Property, and Spectrum*, 14 COLO. TECH. L.J. 181 (2016).

19. PIERRE-JOSEPH PROUDHON, *WHAT IS PROPERTY? AN INQUIRY INTO THE PRINCIPLE OF RIGHT AND GOVERNMENT* ch. I (Benjamin R. Tucker trans., Benjamin R. Tucker 1876) (1840).

20. A good example is De Soto's suggestion that de facto property rights should be identified by listening to the barking dogs. See HERNANDO DE SOTO, *THE MYSTERY OF CAPITAL: WHY CAPITALISM TRIUMPHS IN THE WEST AND FAILS EVERYWHERE ELSE* 189 (2003).

21. For the case of land in England, see generally Briony McDonagh, *Making and Breaking Property: Negotiating Enclosure and Common Rights in Sixteenth-Century England*, 76 HIST. WORKSHOP

The major difference between land and oil on the one hand and data on the other is that data can be harvested, amassed, and turned into monetary assets without allocating property rights at the outset. In fact, data harvesters have exploited the legal ambiguity around data ownership, captured raw information, and claimed property rights to the aggregate databases they created. An important policy question is whether allocating property rights to individual data producers would make much of a difference. Because the value of data rests on aggregation and scaling of data and data analytics, individual data producers might not be able to take advantage of individual property rights even if the legal system assigned property rights to them.

Economic property rights theories rest on the notion of scarcity. Owners wish to capture the expected returns from an asset they own, but by the same token, they also must bear the costs of using it. Property rights are meant to ensure that owners will internalize most of the costs of using their assets.²² A “tragedy of the commons” occurs whenever too many individuals can use a resource without having to account for the costs of doing so.²³ Data, however, are non-rivalrous goods and they are available in infinite amounts. The data harvesters only need to capture them. They might encounter some costs for capturing and processing data or for running algorithms through their dataset, but this has no effect on the raw data or the ability of others to use them. Not the raw data, but only the processed data and the algorithms that allow data controllers to extract predictive value from them, are rivalrous.

The limits of property rights theories in relation to data is also apparent when applying the flipside of the tragedy of the commons story: the tragedy of the anti-commons.²⁴ As Michael Heller has explained, parsing rights to an asset into too many claims creates too many veto players, which makes it impossible to put an asset to its most efficient use.²⁵ There is some reason to believe that if every individual held the property rights to her own data, this might create a gridlock

J. 32 (2013). For recent trends on a global scale, see generally Charles Geisler & Fouad Makki, *People, Power, and Land: New Enclosures on a Global Scale*, 79 RURAL SOC. 28 (2014). For a comparison of the enclosure of land with the enclosure of intellectual property rights, see generally Ford C. Runge & Edi Defrancesco, *Exclusion, Inclusion, and Enclosure: Historical Commons and Modern Intellectual Property*, 34 WORLD DEV. 1713 (2006). And for information specifically on intellectual property rights, see generally Susan Sell & Christopher May, *Moments in Law: Contestation and Settlement in the History of Intellectual Property*, 8 REV. INT'L POL. ECON. 468 (2001).

22. See Harold Demsetz, *Toward A Theory of Property Rights*, 57 AM. ECON. REV. 347, 350 (1967). Demsetz realized that not all costs would be internalized in this fashion; however, he suggested that the much-reduced number of owners post-privatization would allow them to bargain with one another over how to best allocate the remaining costs.

23. See Garrett Hardin, *The Tragedy of the Commons*, 162 SCI. 1243, 1244–45 (1968) (noting that, in such a situation, rational people will continue to use a resource to the point of exhausting it).

24. See Michael Heller, *The Tragedy of the Anti-Commons*, 111 HARV. L. REV. 621, 623–26 (1998) (“When there are too many owners holding rights of exclusion, the resource is prone to underuse.”). For further development of this argument, see MICHAEL HELLER, GRIDLOCK ECONOMICS: HOW TOO MUCH OWNERSHIP WRECKS MARKETS, STOPS INNOVATION, AND COSTS LIVES (2008).

25. See Heller, *supra* note 24, at 633–40; see also HELLER, *supra* note 24, at 2.

effect, because harvesting the data would require prior consent. In practice, this might not be much of a problem given the ease with which consent can be acquired. The European data regulation law, the General Data Protection Regulation (GDPR), is a good example. It requires prior consent from individuals, or an alternative legal basis, for harvesting any *personal* data.²⁶ This requirement has been translated into an “agree” button for internet users to click whenever they enter a site that uses cookies and similar devices that track their behavior. As it turns out, most internet users do so without hesitation. The consent requirement is a nuisance, but it does not create gridlock.

Acquiring consent for data surveillance of real-world activities that are captured by sensors is more difficult. This is why notifying data producers that their data will be captured is often deemed sufficient, and consent is implied once notice has been given.²⁷ Either way, data producers cannot effectively prevent their data from being collected without their consent if collected in violation of existing regulations. They may challenge the data harvesting *ex post*, but this places the burden on them as well as on regulators in a world in which *quintillions* of data bytes are produced every day, ready to be captured.²⁸

Thus, the economic case for property rights in data is not clear cut. Yet, the data *harvesters* have lobbied hard to obtain legal protection on par with property rights for the aggregate data they have collected. The United States enacted the Computer Fraud and Abuse Act in 1986.²⁹ Under this Act, hacking is treated as theft, thereby sanctioning the exclusive use rights of the data controller, whether or not such protection is warranted as a matter of economic theory or justice.³⁰ Similarly, in the Data Base Directive in 1996, the European Union sought to attract tech companies by granting copyright protection to original data or data on a device if the compilation of the data required substantial investments.³¹ In either case, legal protection was granted to owners of physical devices that held large amounts of data—that is, to the data harvesters, not the data producers.

26. See *What Are the GDPR Consent Requirements?*, GDPR.COM, <https://gdpr.eu/gdpr-consent-requirements/> [<https://perma.cc/9XCD-LED3>] (“Consent is one of the easiest [legal bases] to satisfy because it allows you to do just about anything with the data—provided you clearly explain what you’re going to do and obtain explicit permission from the data subject.”).

27. Existing laws in the United States rely heavily on notice rather than requiring explicit consent as does the GDPR in the European Union. See Cameron F. Kerry, *Filling the Gaps in US Data Privacy Laws*, BROOKINGS (July 12, 2018), <https://www.brookings.edu/blog/techtank/2018/07/12/filling-the-gaps-in-u-s-data-privacy-laws/> [<https://perma.cc/GF6F-V7HE>].

28. See Cameron F. Kerry, *Why Protecting Privacy is a Losing Game Today—and How to Change the Game*, BROOKINGS (July 12, 2018), <https://www.brookings.edu/research/why-protecting-privacy-is-a-losing-game-today-and-how-to-change-the-game/> [<https://perma.cc/2ERA-27MD>] (“More and more data about each of us is being generated faster and faster from more and more devices, and we can’t keep up. It’s a losing game both for individuals and for our legal system.”).

29. 18 U.S.C. § 1030 (2018).

30. See *id.*

31. The directive was recently revised. For details, see *Policy for the Protection of Databases in the Digital Single Market*, EUR. COMM’N, <https://ec.europa.eu/digital-single-market/en/protection-databases> [<https://perma.cc/F9XZ-3W2G>].

If economic theory fails to give guidance on allocating property rights to data, are more powerful justifications found in theories of justice? Hanoch Dagan and Avihay Dorfman have recast property rights as a human right.³² Specifically, according to them, property rights encompass a relation between human beings, each respecting each other and his or her quest for self-authorship.³³ While this may not apply to all claims, property rights should be recognized as human rights “to the extent that it is *crucial* to our self-determination and insofar as it is made equally available to us all.”³⁴ This perspective creates certain demands for the property rights regime. Not just any regime will do, but only one that “provides all individuals alike entitlement to the authority over others with respect to certain resources when, and to the extent that, this authority secures the possibility of developing *their own life-plans, rather than the plans imposed on them* by other persons or by society at large.”³⁵

At first glance, this framework makes a powerful case for individual property rights in data. The act of capturing data by Big Tech (or governments) gives them unchecked power to impose their economic or political preferences on the data producers. While not every data point, in isolation, that is captured raises human rights concerns, amassing enough information to constitute digital identities of individual data producers and using that information to influence their behavior to the benefit of the data controller does.

The test Dagan and Dorman propose is “self-imposability by an end.”³⁶ Applied to data this raises the question whether “substantively free and equal” data producers would choose to impose on themselves the obligations demanded by the data controllers.³⁷ In principle, the answer to this question should be a resounding “no.” In practice, however, individual consent is tied to services that most data producers want and for which there are increasingly few, if any, viable alternatives.³⁸ Individual consent by dispersed data producers who are faced with a take-it-or-leave-it option is not very meaningful, especially in a situation where they face oligopolies or even monopolies that control access to digital platforms. This hardly amounts to “self-imposability *to an end*,” which would require an agreement based on mutual respect for self-authorship, when, in fact, the very purpose of the data harvesting is to undermine self-authorship.

As discussed, property rights are a form of self-ordering given a set of legal rights that are backed by the coercive powers of a state. If self-ordering violates fundamental norms of justice, an alternative solution might be to deny legal

32. Hanoch Dagan & Avihay Dorfman, *The Human Right to Private Property*, 18 THEORETICAL INQUIRES L. 391, 393 (2017).

33. *Id.* at 405.

34. *Id.* at 397.

35. *Id.* (emphasis added).

36. *Id.* at 401 (citing ALAN BRUNDER & JENNIFER M. NADLER, *THE UNITY OF THE COMMON LAW* 142 (2d ed. 2013)).

37. *Id.* at 401–04.

38. See generally Kerry, *supra* note 28.

support for certain activities. Just as legal systems prohibit the alienability of body parts, they might and perhaps should prohibit the alienability of data. This has not happened, and for the reasons given above about the ineffectiveness of property rights regimes, may not be workable. Still, some attempts have been made to rein in the use of data companies have harvested. The German antitrust authority (Bundeskartellamt), for example, ruled in early 2019 that Facebook may *not combine* data it obtained from various services offered by different apps, such as Facebook, Instagram, and WhatsApp.³⁹ Still, it will hardly take more than an amendment of the consent declaration for Big Tech to overcome these legal restrictions. This only reinforces the point that governing data in the interest of data producers cannot be accomplished with the standard legal toolkits. Vesting *individual* data producers with property rights to their data, whether for economic purposes or for principles of justice, does not protect individuals against the power that data harvesters exert by controlling entry to digital platforms on which more and more economic transactions and social interactions take place. Effective protection of individuals in their data would instead require a collective mechanism.

III

RETHINKING THE WORLD OF DATA AS HIERARCHY, NOT MARKETS

The value of data does not depend on trade. Harvesting data occurs without exchange; it is a unilateral action that does not require the producer's consent. Indeed, it can occur without the producer even noticing it. The producer may have been put on notice in the agreement that she clicks when accessing a platform, but the act of surveillance is not felt or heard. Whereas for ordinary goods an appropriation without consent is easily qualified as theft, the data harvester is not breaching physical control over an asset. It is simply assembling data points that invariably result from even the most innocuous actions, and even inactions, that individuals take in everyday life. The imposition of consent requirements for personal data is an attempt by lawmakers to introduce an exchange element into the harvesting process. But unless harvesting without consent can be actively policed and sanctioned, and consent itself becomes more than an automatic click, the consent requirement itself is a sham.

Not only is there no market transaction at the source—that is, between the data producers and the data harvesters—but calling the other relation between data controllers and their own clients (who buy access to the data or the algorithms) a market is also a stretch of the imagination. Data harvesters extract the economic value from data not by reselling the aggregate data at a higher price, but by selling the data's predictive power. This can take the form of a simple

39. Bundeskartellamt [Federal Cartel Office], Facebook Decision, 6th Decision Division B6-22/16, Dec. 22, 2015 (Ger.), https://www.bundeskartellamt.de/SharedDocs/Entscheidung/EN/Entscheidungen/Missbrauchsaufsicht/2019/B6-22-16.pdf?__blob=publicationFile&v=5 [<https://perma.cc/N5LK-RCJW>].

service contract or an auction—mechanisms which, of course, are also widely used in markets in the Coasean sense. The difference is that the clients purchase this predictive power from oligopolies and have no way of discovering market prices through repeat transactions. Every data application is unique and allows the seller to set prices without a market test. Finally, in the relation between these clients, who are the sellers of goods and services, and their consumers, one might also question the existence of a market relation. Goods and services are not sold to the buyer who values them the most, but to customers whose future behavior has become most predictable to the seller.

In short, the business of data is first and foremost about control, because only control over massive amounts of data can be effectively monetized. The value of data is not revealed by the price others are willing to pay for data points or even the sum of all these data points, but by the processing and analytical capacity of the data controller. This capacity depends on scale; that is, on the volume of data already captured and on future access to data.

Given the purpose data serve in the hands of data harvesters and future users (the data controllers), data are better characterized as means for exerting control than as tradable goods. Controlling this resource and future access to it confers power over the behavior of consumers as data targets, which include data producers as well as other “like” individuals. The business of data is not about markets. It is about hierarchy. Going beyond Coase’s theory of the firm, hierarchy is not employed to address transaction costs, but to exert control over future behavior and future economic outcomes. The data controllers use the data for their own purposes or share them with paying clients; in contrast, neither their clients’ customers nor the data producers have access to these aggregate data as a means of governance. This uneven access to data as a source of wealth, as well as a means of governance, results in distributional inequities.

Markets that are unperturbed by information asymmetries or power have always been more of an ideal than a reality. Economic might and differential access to law, which function as institutionalized coercive powers, have not rendered flat markets, but rather deeply structured relations of power. Only with the help of antitrust law and mandatory disclosure requirements has the illusion of a free and efficient market—in which both sellers and buyers can rely on the price signal even when transacting with bigger or more powerful parties—been maintained.

Big Tech is distorting this less-than-perfect market even further. First, data harvesters are monetizing their access to raw information by processing these data and giving select access to them to well-paying clients. These clients use their access to the data to nudge future consumers to behave as their models predict. Neither the data harvesters nor their clients share the data or their predictive power with the customers; nor do they share any of the superior returns these data produce with the original data producers. When consumers and data

producers overlap, they lose twice: first, they lose the value of their data; second, they lose their ability to choose transactions without manipulation.

This picture contrasts with modern contract theory, which holds that efficient contracting is all about two parties achieving a mutually beneficial outcome. Both parties face fundamental uncertainty, which renders their contracts incomplete, and both therefore need to find common solutions.⁴⁰ Big Tech will not conquer fundamental uncertainty entirely, but big data combined with rapidly developing analytical power can substantially reduce the scope of unknowable unknowns. In theory, big data could help *both* sides of a transaction by reducing uncertainty and thereby achieving optimal mutually beneficial outcomes. In practice, monopolizing data in the hands of Big Tech and its clients is key for the success of the data-driven *business* model. Who needs to bargain for mutually beneficial results when, for one party, the behavior of the other party is not only highly predictable, but manipulatable?

The rise of markets as a central mode for organizing economic relations is associated with the breaking down of social structures such as guilds and social classes, with their legally recognized differentiation between the nobility, peasants, and merchants.⁴¹ Henry Maine coined the phrase “from status to contracts” to capture this transition.⁴² It entailed the elimination of special legal regimes for different social classes, as well as the institutionalization of the state’s coercive powers such that private agents could harness it for ordering their private affairs with one another. The transition from status to contract is well illustrated in the organization of the French Civil Code of 1804: its first book is devoted to persons, both natural and legal, and their role as subjects of civil legal rights; the second book covers the objects persons might own; and the third specifies the means by which objects or rights to objects can be transferred from one person to another. This is the law of obligations, or contract law. The legal organization of persons, property rights, and contracts are the foundation for markets.

The fact that not all market participants are equal has long been used to critique the ideal of free markets. Still, formal equality has been largely preserved, if only with the helping hand of the state in the form of strong antitrust and disclosure rules. I therefore suggest that the world of Big Tech and big data poses new and fundamental challenges to the formal equality of market participants because of their one-sided control over predictive and manipulative data power. This belies the illusion of formal equality and of voluntary exchange

40. For foundational work on incomplete contracts, see generally Oliver Hart & John Moore, *Foundations of Incomplete Contracts*, 66 REV. ECON. STUD. 115 (1999). For a transposition of this concept into the law of contracts, see Robert E. Scott, *The Law and Economics of Incomplete Contracts*, 2 ANN. REV. L. & SOC. SCI. 279, 280–83 (2006).

41. See KARL POLANYI, *THE GREAT TRANSFORMATION: THE POLITICAL AND ECONOMIC ORIGINS OF OUR TIME* 116 (1944).

42. HENRY MAINE, *ANCIENT LAW: ITS CONNECTION WITH THE EARLY HISTORY OF SOCIETY AND ITS RELATION TO MODERN IDEAS* 101 (J.H. Morgan ed., J.M. Dent & Sons Ltd. 1917) (1861).

transactions worthy of their name. Most legal systems will not enforce a contract that has been entered into under duress. Using control over data to ensure that a target will enter into a transaction may not be the same as holding a gun to her head; however, it is arguably closer to it than the idea of a voluntary transaction between parties that contract to maximize their mutual benefits in the face of future uncertainty.

IV

DIAGNOSIS AND REMEDIES

The purpose of diagnosing a problem is to understand the root cause of a malaise and find the best remedy to treat it. Much of the literature on data to date has focused on the question of whether the market for data exhibits problems of market failure on the one hand and privacy concerns on the other. This has led researchers to inquire about the need for legally sanctioned property or privacy rights, respectively.⁴³ Proponents of market failure theories reason that state intervention is justified, perhaps even necessary, (only) in cases of market failure.⁴⁴ Absent any serious incentive problems that might affect the collection, aggregation, and monetization of data, there is no reason for conferring property rights on data.⁴⁵ Beyond personal data, for which a separate regime has been created, data harvesters have, and according to free marketeers should have, free rein. Further, since data harvesters are turning their freedom into something of economic value, the state should stay out and leave the evolution of this asset class and its use for exchange or other purposes to private actors.⁴⁶ Privacy advocates, for their part, have stressed the importance of privacy as a human right, a right that protects self-authorship and dignity.⁴⁷ However, as discussed earlier, the protection of privacy has been turned into a rather meaningless notice and consent ritual.

More recently, observers have shifted their attention to *market* dominance. The standard approach for dealing with market dominance is to subject it to the

43. Nadeshda Purtova, *The Illusion of Personal Data as No One's Property*, 7 LAW, INNOVATION, & TECH. 83 (2015).

44. For an intellectual history of economic theories of market failure (externalities and public goods) and the need for regulation, see generally Marianna Johnson, *Public Goods, Market Failure, and Voluntary Exchange*, 47 HIST. POL. ECON. 174 (2015).

45. See Wolfgang Kerber, *A New (Intellectual) Property Right for Non-Personal Data? An Economic Analysis*, 65 GEWERBLICHER RECHTSSCHUTZ UND URHEBERRECHT, INTERNATIONALER TEIL [GRUR INT] 989, 992 (2016) (Ger.) (“As regards traditional intellectual property rights such as patents and copyrights, the main argument for justifying the grant of temporary monopolistic positions is that innovations might suffer from a public good problem leading to insufficient incentives for investing in innovation or creative works In the discussion about data property so far, nobody has claimed that we have a general incentive problem.”).

46. See Lothar Determann, *No One Owns Data*, 70 HASTINGS L.J. 1, 9 n.43 (2018) (noting the potential alignment of social and private incentives).

47. Adam Moore, *Towards Informational Privacy Rights*, 44 SAN DIEGO L. REV. 809 (2007) (arguing that individuals have a moral right to their data).

scrutiny of antitrust law. Not just any form of market dominance will suffice to justify intervention. After all, domination may result from success in the marketplace, or from features of a particular market or product that give rise to natural monopolies. In most jurisdictions, including the European Union, only the abuse of market power will be sanctioned.⁴⁸ Remedies include fines, the obligation to unbundle certain products or services, or, on occasion, the breaking up of a dominant company into separate parts. In addition, dominant players may be prevented from expanding their empire by acquiring other companies and thereby further entrenching their market dominance.

Some of the less drastic measures have already been invoked against select Big Tech companies. The European Union, for example, imposed fines on Google for abusing its market dominance by giving illegal advantage to its own shopping service.⁴⁹ In the United States, antitrust cases against Google, Amazon, Facebook, and Uber are under consideration, but details have not been made public yet.⁵⁰ Amazon has been singled out as a relatively easy target for antitrust policy because of an only too obvious conflict of interests. It operates a giant digital market platform that is tied to a storage and transportation infrastructure to ensure delivery. Amazon is on its way to create a payment infrastructure as well.⁵¹ Critically, Amazon not only operates a digital retail platform, it also offers its own products on the same platform and competes directly with its paying clients.⁵² In fact, Amazon seems to have developed a pattern of developing its own product line after carefully monitoring successful retailers on its own platform, offering to buy them out, and, if they refuse, undercutting their prices and, allegedly, employing its control over the digital infrastructure that powers the marketplace to prioritize its own products.⁵³ To address this inherent conflict,

48. In the United States, like in the European Union, abuse is not a constitutive element of antitrust law, but antitrust enforcement rarely proceeds in the absence of it. *See* Lina Khan, *Amazon's Antitrust Paradox*, 126 *YALE L.J.* 710, 710 (2017) (arguing that the purpose of the law was to limit high concentration and abuse of power).

49. Commission Decision of June 27, 2017 (Case AT. 39740—Google Search (Shopping)), 2018 O.J. (C 009) ¶¶ 11–14, https://ec.europa.eu/competition/antitrust/cases/dec_docs/39740/39740_14996_3.pdf [<https://perma.cc/P7ZM-8PG8>].

50. The Department of Justice opened anti-trust investigations into several Big Tech companies, including Google and Facebook, in July of 2019, and the Federal Trade Commission recently announced that it is taking a 10-year back review of mergers in the sector. *See* Jack Crowe, *DOJ Opens Sweeping Antitrust Probe of Big Tech Firms*, *NAT'L REV.* (July 23, 2019), <https://www.nationalreview.com/news/doj-opens-sweeping-anti-trust-probe-into-big-tech-firms/> [<https://perma.cc/7HKY-VQ53>]; Morgan Rubin, *FTC Taking a 10 Year Look-Back at Big Tech*, *NAT'L L. REV.* (Mar. 3, 2020), <https://www.natllawreview.com/article/ftc-taking-10-year-antitrust-look-back-big-tech> [<https://perma.cc/26LE-LSRJ>].

51. *See* Ron Shevlin, *Amazon's Impending Invasion of Banking*, *FORBES* (July 8, 2019), <https://www.forbes.com/sites/ronshevlin/2019/07/08/amazon-invasion/#4296ed679216> [<https://perma.cc/5GJ4-LDZ5>].

52. For details on Amazon's business model, see Kahn, *supra* note 48, at 746–55, which notes that Amazon's business model includes “a willingness to sustain losses and invest aggressively at the expense of profits, and integration across multiple business lines.”

53. *See id.* at 768–70 (discussing the case of Quidsi, an e-commerce company in the business of selling baby products such as diapers, household essentials, and beauty products).

India has prohibited digital platform providers from selling their own products on said platform.⁵⁴

A recent report commissioned by the United Kingdom government and chaired by Jason Furman on “unlocking digital competition” offers several policy proposals for improving competition in digital markets.⁵⁵ The report notes that some digital marketplaces are prone to “tipping,” subjecting them to a winner-takes-all structure.⁵⁶ In response, the report offers preventive measures, such as enhancing individual users’ “data mobility.”⁵⁷ If data producers had the right to easily transfer all of their data from one data platform to another (say from Facebook to Twitter, or from Google to Yahoo), this would allegedly increase competition between platform providers and might even avoid tipping altogether.⁵⁸

Yet, data mobility at the individual level would affect competition only if consumers shifted from one platform to another in a concerted action—a move that, if coordinated, might well spark its own antitrust investigations.⁵⁹ Another problem with this solution is that it fails to address free data harvesting and usage. It is not clear, for example, whether (and, if so, how), the data producer might prevent the first digital platform from continuing to rely on his data for its analytics, or from digitally stalking him as he moves to a different platform. From the perspective of individual data producers, it does not really make much of a difference who harvests and controls the data; what matters is harvesting and controlling the data itself. In addition, a consumer who is the target of the data’s predictive power does not necessarily benefit from greater competition among data controllers. Either way, consumers will be shortchanged.⁶⁰

The same indifference to the problem of data domination and rule by algorithms can be found in the “Utah Statement,” a joint recommendation by antitrust academics and other experts who are calling for a more robust anti-monopoly regime for Big Tech.⁶¹ They seek to reverse several important doctrinal shifts over the past few decades. Their recommendations include, among others, a call for overturning legal doctrines that hinder the sanctioning

54. See Aftab Ahmet & Sankap Phartyal, *India Tightens E-commerce Rules; Likely to Hit Amazon, Flipkart*, REUTERS (Dec. 26, 2018), <https://www.reuters.com/article/us-india-ecommerce/india-tightens-e-commerce-rules-likely-to-hit-amazon-flipkart-idUSKCN1OP14M> [<https://perma.cc/C342-8W9E>].

55. See generally DIGITAL COMPETITION EXPERT PANEL, *supra* note 10.

56. *Id.* at 4.

57. *Id.* at 5.

58. *Id.* at 9.

59. In the United States, antitrust law has been used against labor unions and other coordinated actions in the past. See Sanjukta Paul, *The Enduring Ambiguities of Antitrust Liability for Worker Collective Action*, 47 LOY. UNIV. CHI. L.J. 969, 990 (2016).

60. To read foundational work on the power of algorithms, see generally FRANK PASQUALE, *THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION* (2015).

61. For details, see Tim Wu, *The Utah Statement: Reviving Antimonopoly Traditions for the Era of Big Tech*, ONEZERO (Nov. 18, 2019), <https://onezero.medium.com/the-utah-statement-reviving-antimonopoly-traditions-for-the-era-of-big-tech-e6be198012d7> [<https://perma.cc/B4AU-D4AE>].

of monopolies, such as the presumption of a public benefit for vertical mergers,⁶² and for reviving doctrines that require dominant players to offer their services to anybody, such as the structural presumption in merger reviews or the structural facilities doctrine.⁶³ Like the Furman report sponsored by the United Kingdom government,⁶⁴ the Utah Statement tackles Big Tech as a problem of market dominance that is in principle susceptible to antitrust policies.

While the concentration of large amounts of data in the hands of very few companies exacerbates this problem, even smaller players can engage in data-sourced governance that reduces real choice on the part of consumers and thereby contributes to the marginalization of markets. Preventing Big Tech from expanding and from combining data sourced from different platforms is, of course, important, but it will not be a game changer for the practice of harvesting consumer data and turning data into lucrative tools of governance over consumers.

The real threat that emanates from Big Tech using big data is not just market dominance. Rather, it is the power to transform free contracting and markets into a controlled space that gives a huge advantage to sellers over buyers. Only the sellers get to employ the predictive power of data to their advantage against their own customers, who are often also the data producers. If data harvesting hurts as much as paying one's taxes does, it would not take long for the slogan "no taxation without representation" to be translated into the information age as "no data harvesting without participation." Unfortunately, the act of data harvesting has little impact; consumers do not feel it and they often obtain access to something that they want in return. But this should not distract from the fact that the problem is not market dominance but power, and that the remedy therefore should be accountability, not just market structure or antitrust law.

Nobody questions that the massive data sweeps by government agencies are matters of power and therefore of accountability. And nobody would suggest that we should alleviate concerns about accountability by empowering more government agencies to follow suit and sweep their own data. Nor would anyone be satisfied if obtaining consent to access individual data was conditioned on the threat to withhold government service unless consent was given. Such solutions equally miss the point when applied to Big Tech. Instead, what is needed is accountability of the data harvesters and controllers to the data producers, many of whom also serve as the targeted customers.

In calling for accountability, one should distinguish between at least two types: first, accountability for data harvesting itself, as well as data analyses that might hardwire existing social biases. Many cited examples include the failure of facial recognition software to properly identify minorities or females. As Frank

62. *See id.* ("Vertical coercion, vertical restraints, and vertical mergers should enjoy no presumption of benefit to the public.").

63. *Id.*

64. *See generally* DIGITAL COMPETITION EXPERT PANEL, *supra* note 10.

Pasquale has noted, this kind of accountability takes data harvesting and the use of algorithms to affect social outcomes as a given. It only addresses certain aspects of their use.⁶⁵ Furthermore, there is what Pasquale calls the “second wave of accountability” that addresses the more fundamental questions about when and for what purposes digital empires should be built.⁶⁶ This second wave seems to have helped slow down (at least the visible parts of) the development of artificial intelligence and has further advanced the inclusiveness of digital technology, especially with regards to disadvantaged communities.

The argument advanced in this Article suggests that there is need for a third wave of accountability, namely accountability of the data harvesters and their clients to the data producers, many of whom are also the consumers of products and services at the other end of the data value chain. The importance of this third wave of accountability follows from the fundamental questions that second wavers have asked but have not fully answered: when and for what purposes shall the power of data be deployed? Neither transparency of algorithms nor greater inclusion address these questions. As elaborated in the following Part, what is needed is a governance regime that ensures that data producers have a share in the economic benefits of their data and some say in the data’s future use.

V

HOLDING BIG DATA ACCOUNTABLE

The question about when, and for what purposes, big data shall be amassed is inextricably tied to the question of who gets to decide when, and for what purposes, data is amassed. If data were like ordinary goods and services, the answer to this question would follow from who holds property rights to data. That person can determine how to use the data, whether or when to alienate the data, or even whether or when to destroy the data. If individual data producers held property rights, they could freely transfer them to the data controllers, subject to any conditions the producers might impose. If, however, allocating property rights to raw information is difficult or creates too high transaction costs, the data harvesters can claim property rights to the aggregate data they have amassed—either by claiming capture of a *res nullius* or by invoking property rights protection of data on devices they own.

The difficulty of protecting raw data as property lies in the fact that it is next to impossible to exclude data scavengers from access to data in their raw form. Legal restrictions have been imposed on “personal” data in some jurisdictions, but even these restrictions are weak; simple consent legalizes access and any

65. See Frank Pasquale, *The Second Wave of Algorithmic Accountability*, *Law and Political Economy*, LAW & POL. ECON. (Nov. 25, 2019), <https://lpeblog.org/2019/11/25/the-second-wave-of-algorithmic-accountability/> [<https://perma.cc/88HR-SSJK>] (“While the first wave of algorithmic accountability focuses on improving existing systems, a second wave of research has asked whether they should be used at all—and, if so, who gets to govern them.”).

66. *Id.*

“legitimate” use beyond the one that has been explicitly consented to is deemed legal as well.⁶⁷ The burden for challenging the legitimacy falls squarely on the data producers, who face substantial costs and uncertainties in proving their claims.

As a result, the question as to whether data shall be amassed at all and for what purpose falls to the data harvesters, both de facto and arguably even de jure. Some digital platform operators have imposed voluntary restrictions on themselves about the kind of data they will harvest and for what uses. In theory, this should allow data producers to select from among different providers and shun the most aggressive data harvesters among them. This, however, assumes that data producers have full knowledge about how their data will be used by different harvesters, and also that the services these harvesters offer are largely equivalent, neither of which can be taken for granted.

In fact, leaving the decision to individual data producers makes little sense. They are dispersed and cannot easily exclude others from their data, not even by foregoing online services altogether, since their behavior might be detected by sensors others have installed. Data producers are facing a problem of a mismatch of scale. Each data producer is better off trading her data in return for access to a digital platform or service; yet, all data producers would be better off if they collectively set the terms for harvesting their data and for the data’s future use.

The raw data might also be characterized as public goods, at least in some respects: they are non-excludable and non-rivalrous, which defies their status as private goods.⁶⁸ Yet, unlike standard public goods, such as public safety, the major problem is not “free riding,” that is, failure to pay one’s due share for a good that is provided for everyone. Rather, the problem is the unhindered and free appropriation of raw data by Big Tech and their transformation into monetizable assets followed by technological and legal enclosure. As self-anointed owners, Big Tech will now determine to what ends aggregate data shall be used, by whom, and for what purposes. If data producers themselves wish to have a say, they need to create a collective governance structure that prevents the appropriation of their data. In the alternative, they should at least get a fair share of the future cash flow that the database generates and a say in the purposes for which data will be used.

67. Processing of the data is lawful without the consent of the data subject (that is, the data producer) if it is necessary for the “performance of a task carried out in the public interest or in the exercise of official authority vested in the controller” or “the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject.” Commission Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data; Repealing Directive 95/46/EC (General Data Protection Regulation), art. 6(1)(e), 2016 O.J. (L 119) 1.

68. For a foundational work on public goods, see generally MANCUR OLSON, *THE LOGIC OF COLLECTIVE ACTION: PUBLIC GOODS AND THE THEORY OF GROUPS* (1971). See also Elinor Ostrom, *How Types of Goods and Property Rights Jointly Affect Collective Action*, 15 J. THEORETICAL POL. 239 (2003) (discussing different types of goods and bundles of property rights).

Attempts to prevent the harvesting of data, including restrictions on the kinds of data that may be harvested, are likely to be futile for the reasons already discussed. This leaves rights to the database itself as the target for a collective governance regime. Below I sketch such a structure; working out the details requires further research.

First, data producers should be given a claim to the economic returns not on their individual data, but on the database in a prorated fashion. Allocating individual data producers property rights over their raw information so that they can sell them would give them only a pittance. The U.S. Supreme Court recently raised the question whether data producers have standing to bring a suit against a Big Tech company (Google, in this case) under the Stored Communications Act.⁶⁹ Standing requires specific, or concrete, injury. In a case in which the original data a person had produced were passed on to another server without permission, concrete injury required careful examination.⁷⁰ The same argument could be applied to the data harvesting process itself.

Instead of seeking to control the use of data *ex post* through tort actions, treating data producers as co-owners of the database would give them a claim to future profits the database produces on a *pro rata* basis as well as governance over its use. Just as shareholders pool their resources to invest in the expectation of future returns, so would data producers be treated as joint-venturers who could generate higher returns by pooling their resources. The Big Tech companies would be treated as agents. They may be owed compensation for harvesting, cleansing, and processing the data, but not the exclusive right to monetize the returns on these data. Importantly, as agents, they would also owe fiduciary duties to the data owners, the ultimate owners of the database. In this role, the data producers could also control both the harvesting and use of the data.

Technically, this solution could be implemented by creating a trust fund that owns the data. A possible model is the Depository Trust Company (DTC), which was created by an act of the U.S. Congress and operates as the formal owner of shares in publicly listed companies.⁷¹ It holds these shares on behalf of the ultimate beneficial owners. The DTC was established in 1968 to reduce transaction costs in an age when every transaction in shares, bonds, and other securities left a lengthy paper trail.⁷² As markets grew, the New York Stock

69. Frank v. Gaos, 139 S. Ct. 1041, 1044 (2019) (per curiam).

70. *Id.* at 1046 (remanding for consideration of plaintiffs' standing in light of the Court's decision in *Spokeo, Inc. v. Robbins*, 136 S. Ct. 1540 (2016)); see also Noah Feldman, *Supreme Court Isn't Sold on the Harms of Big Tech*, BLOOMBERG OPINION, (Mar. 20, 2019, 3:22 PM), <https://www.bloomberg.com/opinion/articles/2019-03-20/supreme-court-ruling-in-google-class-action-suit-is-good-for-tech> [<https://perma.cc/D6N5-T2QG>] (analyzing *Gaos* and noting that "[t]he justices' action strongly hinted that a majority thinks the suit should never have been allowed to go forward in the first place").

71. *The Depository Trust Company (DTC)*, DTCC, <https://www.dtcc.com/about/businesses-and-subsidiaries/dtc> [<https://perma.cc/UW3X-T3VX>].

72. *Id.*

Exchange suffered a heart-attack, which prompted the U.S. Congress to establish the DTC.⁷³

The DTC's basic model could be repurposed for data: the interests in data could be transferred to a (public) trust. This trust would have a right to a share in the earnings the company derived from the data and would have the task of channeling these earnings to the data producers. The trust would also exercise voting rights on behalf of the data producers. A separate trust might be established for each major company, or possibly one for all combined. The latter would have the advantage of economies of scale with regards to the operation costs of the trust, but also in terms of the bargaining power vis-à-vis individual data harvesters. However, there are also reasons to believe that tailoring trust companies to each harvester might enhance monitoring and might better reflect the specific interests of data producers in different firms. Either way, the qualifications of the trustees, their appointment and dismissal, as well as their independence from the shareholders, directors, and officers of the data harvesters, would have to be carefully regulated and fortified with liability rules.⁷⁴

Implementing these ideas would hardly go unchallenged. Big Tech companies and their existing shareholders will claim that a law that would turn data producers into equity holders would amount to expropriation. Yet, whether or not Big Tech companies actually own the raw data they have harvested remains ambiguous. So far, this has worked in favor of Big Tech companies, which have proceeded to harvest data without asking for consent or offering a consideration of equal value in return. They might invoke the Roman law principle that assets that are not owned by anyone can be captured and that the act of capture creates private property rights in a *res nullius*. But not everything that is not owned by someone can be appropriated. Roman law distinguished between *res nullius* and *res communis*, the latter encompassing assets, which according to the collective will, shall not be privately appropriated.⁷⁵ Legal intervention that clarifies the status of data as *res communis* and allocates claims to future cash flows the aggregate databases generate, as well as the right to decide their use, to data producers is an act of allocating the initial property rights, not an act of expropriation. And the power to allocate these rights is a matter for legislatures, not for private data appropriators to decide.

How to ensure that this fix survives future change is yet a different matter. As the history of corporate law suggests, tying the legislature to the mast is not easy.

73. Adam J. Levitin, *The Paper Chase: Securitization, Foreclosure, and the Uncertainty of Mortgage Title*, 63 DUKE L.J. 637, 681 (2013) (discussing the establishment of the DTC in 1968 in response to increased trading volumes that could no longer be processed effectively).

74. This may require some changes to existing trust law, or specific legislative requirements for data trustees. On the “stripping” of core features of common law trust law in recent decades, see Adam S. Hofri-Winogradow, *The Stripping of the Trust: A Study in Legal Evolution*, 65 UNIV. TORONTO L.J. 1, 2 (2015) (“Legislatures worldwide have been eliminating traditional rules of trust law designed to impose liability on negligent trustees, protect vulnerable beneficiaries, and prevent the use of trust law to evade the claims of settlors’ and beneficiaries’ creditors.”).

75. See Epstein, *supra* note 18.

Just as the DTC, this new Data Trust would have to be introduced by federal law to eliminate the threat that regulatory competition between the states could quickly erode its efficacy. In addition, techniques that have been used in other contexts might be employed to protect data producers, such as grandfathering their rights and thereby immunizing them against future legislative change. Only databases that are created anew from future data would side-step the original constraints. Given the importance of scale for the value of databases, attempts to sidestep these constraints may not be productive.

Lastly, the new Data Trust must have effective control rights over data harvesters, which may be difficult given the existing control structure of several Big Tech companies. Facebook, Google, and Amazon all have chosen governance structures that perpetuate the founders or owners in office irrespective of any change in the future composition of shareholders or their desire to put new management in their place. They have created dual-class ownership structures that allocate to the founder(s) multiple voting rights or transfer a significant number of shares to a foundation, which the founder controls. As a result, Mark Zuckerberg controls close to 58 percent of the votes in Facebook,⁷⁶ and the founders of Google control over 50 percent of the votes in Alphabet, Google's holding company.⁷⁷ Big Tech still lives off the image of the gig economy, in which individual talent counts and anything is possible, but the founders of these companies have walled themselves in behind legal protections and can thwart any proxy fight or takeover attempt, the means for changing control over a corporation. Far from subjecting themselves to shareholder democracy, they have created a governance structure that resembles autocracies. Still, if all data producers obtained common stock in return for the data they produced, this could potentially dwarf the voting control the founding shareholders of these companies have given themselves.

76. Facebook, Inc., Proxy Statement (Form DEF 14A) 40–41 (Apr. 12, 2019).

77. See Alphabet Inc., Proxy Statement (Form DEF 14A) 34 (Apr. 22, 2019) (revealing that Larry Page owned 26.1 percent and Sergei Brin owned 25.2 percent of all voting shares, and Eric Schmidt another 5.6 percent).

VI

CONCLUSION

This Article identifies a puzzle concerning markets in the age of information capitalism. According to Ronald Coase, in a world with zero or close to zero information costs, there should be no firms; instead markets should reign freely.⁷⁸ This follows from his argument that firms emerge (only) in response to transaction costs; when information is abundant they should therefore not exist. Yet, the age of information technology, and the dramatic lowering of information costs that has gone along with it, has not seen the flattening of hierarchies into markets, but instead the rise of Big Tech companies, such as Google, Facebook, Amazon, Apple, and others. These companies have taken command over data that individuals leave as they make use of Big Tech's services or are captured by sensors that are installed in more and more devices. Big Tech has been able to monetize its control over these data by erecting technological barriers around the data and selling their predictive power to well-paying clients.

Hierarchy, it turns out, not only reduces the costs of doing business inside the firm, it also creates a comparative advantage in securing dominance in the "real" world, where, according to economic theory, in the absence of transaction costs, markets should reign. The first-mover advantage of Big Tech has come at the expense of data producers as well as the final consumers, and to some extent even of the clients Big Tech services. Data producers have been unable to mount an effective counter-offensive and secure broad access to data because they are hampered by collective action problems. And Big Tech clients have been co-opted into a scheme that is costly for them, but allows them to acquire dominance over both data producers and consumers.

This analysis renders important insights for the relation of both markets and hierarchy to law. Markets, it turns out, are not the natural default mode of economic ordering when transaction costs approximate zero. They emerge only if *all* market participants have equal access to information. When some actors can monopolize access to information (raw data) and the predictive powers of processed data, hierarchy, rather than markets, will dominate.

Moreover, law privileges hierarchy over markets. It is not a coincidence that Big Tech companies are all organized as corporations, with their hierarchical governance structure and built-in legal privileges such as limited liability for shareholders and protection of the corporations from claims on their assets by their own shareholders or their personal creditors. These features have allowed Big Tech companies to raise capital in the early stages of their operation when they still depended on outside capital and to incubate the returns they made behind a corporate veil.⁷⁹

78. See Coase, *supra* note 2, at 388.

79. On the critical function of corporate law in providing both owner and entity shielding, see generally Henry Hansmann & Reinier Kraakman, *The Essential Role of Organizational Law*, 110 YALE L.J. 387 (2000). See also Henry Hansmann et al., *Law and the Rise of the Firm*, 119 HARV. L. REV. 1335,

Given that both markets and firms are legal constructs, and that information costs alone cannot explain when one prevails over the other, what or who determines when either markets or firms shall reign? The case of Big Tech suggests that private agents will often prefer hierarchy over markets, because it greatly increases their control rights and creates economies of scale from which they can benefit disproportionately, especially when they have free rein to design the governance structure of these firms. If hierarchy is the “natural” outcome, proactive intervention is needed to recreate the resemblance of markets in which parties can bargain at least on formally equal footing.

Viewed in this light, the rise of private entities that monopolized data and monetized their predictive power can be described as a massive governance failure: a failure by the state to intervene in a timely fashion to protect data from private appropriation by declaring them a *res communis*.⁸⁰ This governance failure is difficult to rectify *ex post*. Still, the legal grey zone in which data still reside offers an opportunity to reassert collective rights over them. Contract law and antitrust law are unlikely to accomplish this re-ordering on their own. The antidote to the dominance by Big Tech can be found only in organizational law, in the legal tools for governing firms, not markets.

1337 (2006) (discussing “entity shielding”—the “rules that protect a firm’s assets from the personal creditors of its owners”).

80. The choice between *res nullius*, goods that can be appropriated and become private property with their appropriation, and *res communis*, goods that may not be privately appropriated, is a collective choice of fundamental importance. See Epstein, *supra* note 18, at 182.