ICLE International Center for Law & Economics

FTC Hearings on Competition & Consumer Protection in the 21st Century

FTC Project No. P181201

Comments of International Center for Law & Economics:

Understanding Competition in Markets Involving Data or Personal or Commercial Information

Hearing # 6 (Nov. 6-8, 2018)

Submitted January 7, 2019

Authored by:

Dirk Auer (Senior Fellow, International Center for Law & Economics)

Geoffrey A. Manne (President & Founder, International Center for Law & Economics)

Julian Morris (Executive Director, International Center for Law & Economics)

Kristian Stout (Associate Director, International Center for Law & Economics

I. Introduction and Executive Summary

We thank the Commission for the opportunity to comment on "Competition and Consumer Protection in the 21st Century Hearings."

The International Center for Law and Economics (ICLE) is a nonprofit, nonpartisan research center whose work promotes the use of law & economics methodologies to inform public policy debates. We believe that intellectually rigorous, data-driven analysis will lead to efficient policy solutions that promote consumer welfare and global economic growth.[1]

ICLE's scholars have written extensively on competition and consumer protection policy. Some of our writings are included as references in the comment below. Additional materials may be found at our website: www.laweconcenter.org.

In this comment, we primarily address the fifth topic raised by the Commission ("Are there policy recommendations that would facilitate competition in markets involving data or personal or commercial information that the FTC should consider?").

Our comment addresses several pressing issues regarding competition in markets that rely heavily on data to operate. For a start, commonly repeated analogies between data and oil are highly misleading (Section II). Oil is physical commodity that is highly rivalrous (a user cannot use oil without impairing others' ability to use the same oil) and readily excludable (it can easily be stored in ways that prevent use by non-authorized parties). By contrast, data is simply information that bears some of the traits of a public good: it is often non-rivalrous in consumption (the same information may be used by multiple parties without any degradation) and difficult to appropriate because it is difficult to prevent others' use of the same data, it is difficult to ensure optimal investment in its creation). Moreover, in most instances, it is not data that is scarce, but the expertise required to generate and analyze it. In any case, most successful internet companies started life with little to no data. This suggests that data is more a *byproduct* of the ongoing operation of internet platforms than it is a critical input for their creation.

Crucially for antitrust enforcers, data is unlikely to constitute a barrier to entry, and even less likely to amount to an essential facility (Section III). As George Stigler famously argued, a barrier to entry is "[a] cost of producing that must be borne by a firm which seeks to enter an industry but is not borne by firms already in the industry."[2] There is no reason that the cost of obtaining data for a new

[1] ICLE is supported entirely by voluntary contributions from individuals, foundations, and corporations. The ideas expressed here are the authors' own and do not necessarily reflect the views of ICLE's advisors, affiliates, or supporters. Please contact us with questions or comments at icle@laweconcenter.org.

[2] GEORGE STIGLER, THE ORGANIZATION OF INDUSTRY 67 (1968).

entrant should be any higher than it was for an incumbent. In fact, the opposite will often turn out to be true.

Other ills that allegedly plague data-rich markets (and the merits of proposed solutions) are equally dubious (Section IV). This is notably the case for the relationship between mandated data portability and competition. Contrary to what some scholars have advanced, it is far from clear that mandated data portability will increase consumer welfare in data-reliant markets. Not only is this type of portability unlikely to significantly affect switching costs for consumers but, even if it did, this would have ambiguous consumer welfare consequences (as is generally the case for consumer lock-in and regulatory interventions to overcome it). To make matters worse, mandated data portability is not without its risks. Most notably, data portability poses data security and user privacy risks.

Likewise (also Section IV) fears of costly price discrimination and widespread algorithmic collusion are greatly overblown. While it is true that big data may have a transformative effect on firms' ability to price discriminate, there is no strong reason to believe that this would have a detrimental effect on consumer welfare. Instead, as with all forms of price discrimination, it may potentially expand output and allow less well-off consumers to participate in markets they might otherwise be priced out of. Similarly, the idea that big data and algorithms will lead to collusion is deeply flawed. Fears of collusion rest on the faulty premise that online marketplaces and the use of big data will dramatically increase transparency, thus facilitating collusion. In fact, the opposite is just as likely (and, in any case, the manifest benefits of increased transparency, likely outweigh the speculative costs).

In short, we argue that the advent of data-enabled markets does not support the calls for a significant expansion of antitrust tools and antitrust enforcement being made in its name. Contrary to what has sometimes been claimed, data does not present unique (and uniquely large) anticompetitive risks. Data is not irrelevant, of course, but it is just one amongst a plethora of factors that enforcement authorities and courts should consider when they analyze firms' behavior.

II. The economics of data: Data is not the "new oil"

"Data is the new oil" has been a catchphrase for policymakers[3], business investors[4] and reporters[5] for the best part of a decade. Behind the slogan lies an unspoken fear for antitrust policy makers and enforcers: left to their own devices, today's dominant digital platforms will become all powerful—like the industrial giants of the gilded age and specifically the Standard Oil company. These comparisons are not just implicit. The Economist and other press outlets have routinely used Standard Oil Company-related imagery to depict the rise of digital platforms (see Figure 1).[6]

In this Section, we consider the elements of this analogy, beginning, in Part A, with an analysis of the degree to which data may be excluded and appropriated. Part B discusses the scarcity of data and Part C looks at ways in which data may be monetized. The section concludes by returning to the question: is data the new oil?

---

[3] *See* Meglena Kuneva, Keynote Speech: Roundtable on Online Data Collection, Targeting and Profiling, March 31, 2009, *available at* http://europa.eu/rapid/press-release_SPEECH-09-156_en.htm. ("Personal data is the new oil of the internet and the new currency of the digital world.").

[4] *See* Martin Pelletier, *Why data is the new oil and what it means for your investment portfolio*, FINANCIAL POST, Aug. 29, 2017, *available at* https://business.financialpost.com/investing/why-data-is-the-new-oil-and-what-it-means-for-your-investment-portfolio.

[5] *See* Leaders, *The world's most valuable resource is no longer oil, but data*, THE ECONOMIST, May 6, 2017, *available at* https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data.

[6] *See* Leaders, *Battle of the internet giants Survival of the biggest*, THE ECONOMIST, May 6, 2017, *available at* https://www.economist.com/leaders/2012/12/01/survival-of-the-biggest. *See also*, Scott Galloway, *Silicon Valley's Tax-Avoiding, Job-Killing, Soul Sucking Machine*, ESQUIRE (Feb. 8, 2018), *available at* https://www.esquire.com/news-politics/a15895746/bust-big-tech-silicon-valley/. *For a more detailed discussion*, *see* Dirk Auer & Nicolas Petit, *Two Systems of Belief About Monopoly: The Press Versus Antitrust*, 39 CATO JOURNAL (forthcoming 2018), working paper available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3112150.

Figure 1: Top left, top right and bottom left: depictions of Google Amazon and Facebook (contemporary). Bottom right, depiction of the Standard Oil Company (1904).

A. Data is information

One of the most salient features of the data that digital platforms create and consume is that, jargon aside, it is just information. As with other types of information, it thus tends to have at least some traits that are usually associated with public goods (i.e. goods that are non-rivalrous in consumption and not readily excludable).[7] The marginal cost of collecting and employing data is usually close to zero, making it close to non-excludable. Meanwhile, multiple economic agents can simultaneously use the same data, making it non-rivalrous in consumption.

This is not to say that data requires some special protection to be provided by the market. Far from it. As Ronald Coase famously showed, public goods are a theoretical construct – like perfect competition or monopoly – that rarely exists outside of economic textbooks.[8] Instead, the public good analogy shows that data bears some traits which make it almost irreconcilable with the alleged hoarding and dominance that came to be associated with the oil industry of the late nineteenth and early twentieth centuries.

[7] *See* CARL S. SHAPIRO & HAL R. VARIAN INFORMATION RULES: A STRATEGIC GUIDE TO THE NETWORK ECONOMY 3 (1998).

[8] *See* Ronald H Coase, *The Lighthouse in Economics,* 17 J.L. & ECON. 15 (1974).

Moreover, data, broadly speaking, is useful to *all* industries. Collecting data on consumers is not a new phenomenon restricted to online companies. The market for data, even if narrowly described as data for targeted advertising, is much broader than the online world. Offline retailers have long used data about consumers to better serve them. Through devices like coupons and loyalty cards (to say nothing of targeted mailing lists and the age-old practice of data mining check-out receipts), brick-and-mortar retailers have long tracked purchase data and used it to better serve consumers.[9] Not only do consumers receive better deals as a result, but retailers know better what products to stock and advertise and when and on what products to run sales.

1. *Access to data is not exclusive*

Data tends to be non-rivalrous (or at least, the cost of producing a marginal copy of some piece of data is usually close to zero).[10] For this reason, one agent's use of a given piece of information does not automatically preclude its rivals from using the same information.

The non-rivalrous nature of information seriously undermines the views of critics who have compared digital platforms to Standard Oil and argued that government authorities need to step in to limit the platforms' control over data.[11] To say that data is like oil betrays a serious misunderstanding. Google knowing my birthday doesn't limit the ability of Facebook to know my birthday, as well. While databases may be proprietary, the underlying data usually is not.

In other words, most data are non-exclusive. Not only can the same data be used by many different economic agents, but there are also numerous ways in which it can be obtained through different platforms. As we discuss in more detail below (see *infra* Section III), antitrust authorities should thus be highly skeptical about claims that rivals will be unable to independently generate equivalent data to that which is held by dominant platform.

2. *Data is hard to appropriate*

> [W]e expect a free enterprise economy to underinvest in invention and research (as compared with an ideal) because it is risky, because the product can be appropriated only to a limited extent, and because of increasing returns in use.[12]

---

[9] *See, e.g.*, Nancy Kross, *Big Data Analytics Revolutionizing The Way Retailers Think*, BIDNESS ETC, Jun. 26, 2014, http://www.bidnessetc.com/business/big-data-analytics-revolutionizing-the-way-retailers-think/; Dianne Heath, *How Panera Uses Rewards Card to Increase Customer Loyalty & Attract Customers*, ANALYST DISTRICT (Nov. 4, 2011), http://www.analystdistrict.com/2011/11/panera-increase-customer-loyalty.html.

[10] *See* SHAPIRO & VARIAN, *supra* note 7, at 3.

[11] *See* Nathan Newman, *Taking on Google's Monopoly Means Regulating Its Control of User Data*, HUFFINGTON POST (Sept. 24, 2013), *available at* http://www.huffingtonpost.com/nathan-newman/taking-on-googles-monopol_b_3980799.html.

[12] Kenneth Arrow, *Economic welfare and the allocation of resources for invention*, *in* THE RATE AND DIRECTION OF INVENTIVE ACTIVITY: ECONOMIC AND SOCIAL FACTORS 619 (1962). *But see* Jack Hirshleifer, *The Private and Social Value of Information and*

The second key feature of information is that is it hard to appropriate. In practice, this means that companies that have acquired a valuable piece of data will struggle both to prevent their rivals from obtaining the same data as well as to derive competitive advantage from the data. For these reasons, it also means that firms may well be more reluctant to invest in data generation than is socially optimal.[13] In fact, to the extent this is true there is arguably more risk of companies *under*-investing in data generation than of firms *over*-investing in order to create data troves with which to monopolize a market. This contrasts with oil, where complete excludability is the norm. The fact that appropriating data is a complicated task can be seen in a number of instances.

First, specific pieces of data can usually be obtained through a variety of channels. This undermines oft-repeated claims that large online platforms such as Google and Facebook have acquired an insurmountable data advantage over their competitors.[14] In other words, it is almost impossible to build an insurmountable data advantage because there will generally be an alternative way (or, more likely, a multitude of ways) to amass the same data. To take just one example, mobile ISPs like Verizon have access to considerable data about their users, likely at least comparable to what Google and Facebook have. What's more, mobile ISPs have uniquely good access to location data, increasingly the coin of the realm in a world where the most important and valuable consumer interactions are shifting to mobile. This may not be the identical information, and even where it overlaps it is certainly a somewhat different dataset. Yet there can be no doubt that Verizon's data can be used by advertisers (among others) for the same purposes as is data from Google and Facebook.

Another important example concerns the ubiquity of data scraping on the internet. Contrary to popular belief, numerous firms in data-heavy industries do not rely solely on proprietary data to improve and market their products. Instead, these firms routinely "scrape" the internet in order to obtain the data they require.[15] This practice has led to a blossoming industry. Critically, this is one space where dominant firms arguably have little advantage over more nimble rivals. Indeed, stories abound of startups going head to head with large incumbents and generating more useful insights form the same publicly accessible data.[16]

---

the Reward to Inventive Activity, 61 AM. ECON. REV. 561 (1971); Harold Demsetz, *Information and efficiency*, 12 J. L. & ECON. 14 (1969).

[13] *See* Arrow, *id.* at 617.

[14] *See, e.g.,* Leaders, *supra* note 5.

[15] *See* Klint Finley, *Scraper' Bots and the Secret Internet Arms Race*, WIRED (Jul. 23, 2018), *available at* https://www.wired.com/story/scraper-bots-and-the-secret-internet-arms-race/.

[16] *See, e.g.,* Miranda Katz, *A Lone Data Whiz Is Fighting Airbnb – and Winning*, WIRED, Feb. 10, 2017, *available at* https://www.wired.com/2017/02/a-lone-data-whiz-is-fighting-airbnb-and-winning/. *See also*, Julia Angwin & Steve Stecklow, *Scrapers' Dig Deep for Data on Web*, THE WALL STR. J. (Oct. 12, 2010), *available at* https://www.wsj.com/articles/SB10001424052748703358504575544381288117888.

The upshot is that the ease with which data can be obtained — notably by identifying or creating new sources of information or by using publicly accessible information — suggests that it is an unlikely tool for firms to perpetuate monopoly power over lasting periods of time. A monopoly that relies on data to cement its position is thus built on sand, because any data-related advantage can be eroded the moment rivals come up with an alternative way of attaining comparable information.

It is important not to overstate the fungibility of data. While for many types and uses of data fungibility is the norm, it certainly will not always be. But, properly understood, the uniqueness of data is not a strong argument for antitrust enforcement against firms successfully using big data. First, unique agglomerations of data for which comparable substitutes do not (yet) exist inevitably reflect unique entrepreneurial foresight into the value of certain data, superior data processing abilities, and/or a particularly innovative mechanism for generating unique data.[17] In all of these cases, there are potentially considerable consumer advantages from the underlying conduct that enables the unique appropriation of data, and penalizing the successful use of data means also penalizing broader innovative activities. Indeed, the inseparability of data from the product or services that generate or use it is one of the key problems of calls for antitrust intervention against big data: We do not use our antitrust laws (in the US, at least) against effective competition, but only against abuse of market power.

Second, data use by multi-sided platforms may often appear competitively unique when looking at only one side of the platform. But any anticompetitive significance may also be mitigated or undermined by the fungibility of the data on the *other* side of the platform. To take one obvious example, the data used and generated by Google Search is significantly different than that used and generated by Facebook. And, not coincidentally, on the user side of the platform Google and Facebook offer substantially different products, used primarily for divergent purposes. But on the advertising side, of course, the distinctions are substantially less relevant. Both Google and Facebook collect, generate, and process data to help advertisers identify and reach likely customers. The mechanisms by which they do this are quite different, but the purpose and aggregate content of the data is not likely very different at all.[18] The lack of advertising-side differentiation is no doubt

[17] *See, e.g.,* Ufuk Akcigit and Qingmin Liu, *The Role of Information in Innovation and Competition*, 14 J. EUR. ECON. ASS'N 4 (2015) (Studying how firms pursue different innovation strategies owing to imperfect information distribution across firms) *available at* https://onlinelibrary.wiley.com/doi/abs/10.1111/jeea.12153.

[18] It has to be mentioned, as well, that the difference between the sets of specific users advertisers might access on each platform approaches zero as each platform approaches ubiquity. For advertisers, the substitutability of Facebook for Google (and vice-versa) increases as each increases in size. Whether this increase in competition offsets any (alleged) competitive problems resulting from their size is an empirical question (but one that advocates for antitrust action against these firms because of their size never address).

bolstered by user multi-homing and the increasing ability of users to transfer data between platforms.[19]

B. Data is not scarce; expertise is

Another important feature of data is that it is ubiquitous. Contrary to oil, the challenge for firms is not so much obtaining data but is rather drawing useful insights from it. This has two important implications as far as antitrust policy is concerned. First, although data does not have the same self-reinforcing characteristics as network effects, there is a sense that acquiring a certain amount of data and expertise is necessary to compete in data-heavy industries. However, it is equally apparent that this "learning by doing" advantage rapidly reaches a point of diminishing returns. Second, it is firms' capabilities, rather than the data they own, that lead to success in the marketplace. Critics who argue that firms such as Amazon, Google, and Facebook are successful because of their superior access to data thus have causality in reverse. It is because these firms have come up with successful industry-defining paradigms that they have amassed so much data, and not the other way around.

1. Learning by doing

It is sometimes claimed that data-intensive industries naturally lead to winner-take-all markets. The argument goes that superior access to data allows firms to improve their products and gain more users. This then leads to even more data, thereby creating a self-reinforcing circle that eventually causes one firm to dominate the market. In other words, these industries exhibit "data network effects."[20] Though the intuition is appealing, it has neither been translated into a rigorous economic model, nor has it been established empirically. In fact, the anecdotal evidence that has been used to support this naïve assertion merely shows that learning-by-doing plays an important role in the tech industry, just as it does in the rest of the economy.

Take Google, which has become the poster child for unsophisticated "data network effects" arguments. In the words of Nathan Newman:

> While there are a number of network effects that come into play with Google, ["its intimate knowledge of its users contained in its vast databases of user personal data"] is

[19] *See* the Data Transfer Project at https://datatransferproject.dev/ (last visited Jan. 6, 2019). The DTP is an initiative, begun in 2017, of Google, Facebook, Microsoft, Twitter, and a number of other data platforms to make data portability between platforms more efficient and user-friendly. An overview of the project is available at Data Transfer Project Overview and Fundamentals (Jul. 20, 2018), *available at* https://datatransferproject.dev/dtp-overview.pdf.

[20] *See, e.g.*, Maurice E. Stucke & Allen P. Grunes, *Debunking the myths over big data and antitrust*, 5 CPI ANTITRUST CHRONICLE 2015 6 (2015). ("Data-driven industries can be subject to several network effects, including: Traditional network effects, such as social networks like Facebook; Network effects involving the scale of data; Network effects involving the scope of data…). The authors provide no evidence to support the existence of these purported data-related network effects.

> likely the most important one in terms of entrenching the company's monopoly in search advertising.
>
> * * *
>
> Google's overwhelming control of user data… might make its dominance nearly unchallengeable.[21]

There are numerous problems with this claim. At the most basic level, it misapplies the theory of network effects. Network effects occur when a consumer's utility for a good is, at least in part, a function of the expected number (and quality) of other agents using the same product.[22] These valuable users may be located on the same side (direct network effects) or on the opposite side (indirect network effects) of a platform.[23] In both cases, the bottom line is that consumers place a premium on utilizing a product whose network contains a large number of users (or higher quality ones). To a first approximation, however, this means that network effects are a *benefit* to users, not a cost.[24]

Telling a story of problematic "data network effects" for a company like Google is difficult. Direct network effects are not an issue for Google: Search users don't interact with other search users, and they do not benefit directly from there being more of them. And search users receive no benefit from indirect network effects; in fact, the more advertisers the more likely the value of the platform is *reduced* for search users.

What proponents of a data network effects theory propose is that "[t]he gain for Google from its network of users is not just data on each individual user, but the cumulative data that can reveal how similar users behave."[25] For proponents of this view, Google's access to this crucial data is unparalleled and unsurmountable.

But the relevant information is as available to Bing as it is to Google: observable patterns of users' interactions with readily indexable, web-connected content. Google certainly makes observations about its greater number of users' behaviors that it uses to improve its product. But Bing also has that capability, as well as the support structure of one of the most valuable companies in the world (Microsoft) and teams of talented programmers. Roughly a quarter of all US searches were

---

[21] Nathan Newman, *Search, Antitrust, and the Economics of the Control of User Data*, 31 YALE J. ON REG. 401, 420 & 423 (2014)

[22] *See, e.g.,* Michael L. Katz & Carl Shapiro, *Systems Competition and Network Effects*, 8 J. ECON. PERSP. 93, 96 (1994).

[23] *See, e.g.,* Jean-Charles Rochet & Jean Tirole, *Platform Competition in Two-Sided Markets*, 1 J. EUR. ECON. ASS'N 990 (2003).

[24] *See* Geoffrey A. Manne & Joshua D. Wright, *Google and the Limits of Antitrust: The Case Against the Case Against Google*, 34 HARV. J. L. & PUB. POL'Y 171, 224-25 (2011).

[25] Newman, *supra* note 21, at 421.

performed on Bing in 2018.[26] Surely, given its resources and teams of programmers, Bing (or another well-funded and technologically savvy competitor) is *capable* of competing away Google's gains. There's no indication that any more than a *significantly* smaller volume of data is required to train a search algorithm.[27] After all, the power of machine learning is that it can make useful inferences about user behavior based on small sample sizes — one doesn't need "all the data" to make useful machine learning algorithms. Bing has "enough" data — it indexes the same public web, and has access to a very large share of user activity — it just so happens that users prefer Google's results and the other features of its product.

The conclusion, from Newman and others, that "Microsoft, with nearly half of Google's user base, still generated $2.6 billion in losses compared to its costs shows the height of the competitive barrier"[28] is tellingly misguided. It is theoretically possible that data barriers have prevented Microsoft's success relative to Google — but it is far more likely that Microsoft simply offered an inferior product. That proponents of a data network effects story ignore relative product quality along multiple dimensions and assume that the quantity of data alone is outcome determinative highlights the paucity of the argument. Data matters to the extent that it is used to provide value to users, within a product or service that is also attractive, functional, and usable. The quality of the underlying algorithm, informed in part by data derived from users, certainly contributes to that, but it is far from the only factor.

Network effects and user inertia might theoretically prevent rivals from successfully competing against a dominant network.[29] And yet, in practice, this intuition often turns out to be false. For instance, Stan Liebowitz and Stephen Margolis show that one of the most commonly cited examples of "excess inertia" — the failure of Dvorak keyboards to displace the allegedly less-efficient QWERTY layout — did not withstand empirical scrutiny.[30] The authors conclude that:

> The trap constituted by an obsolete standard may be quite fragile. Because real-world situations present opportunites for agents to profit from changing to a superior standard, we cannot simply rely on an abstract model to conclude that an inferior standard has persisted. Such a claim demands empirical examination.[31]

---

[26] *See Share of search queries handled by leading U.S. search engine providers as of October 2018*, STATISTA (Oct. 2018), *available at* https://www.statista.com/statistics/267161/market-share-of-search-engines-in-the-united-states/.

[27] Manne & Wright, *supra* note 24, at 212.

[28] Newman, *supra* note 21, at 419.

[29] *See* Joseph Farrell & Garth Saloner, *Installed base and compatibility: Innovation, product preannouncements, and predation*, 76 AM. ECON. REV. 940 (1986).

[30] *See, e.g.* Stan J Liebowitz & Stephen E Margolis, *The fable of the keys*, 33 J.L. & ECON. 21 (1990).

[31] *Id.*

The upshot is that there is a theoretical, though empirically dubious, case to be made for network effects leading to potential market failures. In the case of "data networks effects," however, even the theoretical model is weak, at best. Because users do not attach any standalone value to platforms with more data,[32] there is literally an infinite number of ways in which firms may offer a superior product without having the same or as much data as their rivals. Firms can notably differentiate themselves on a variety of features, ranging from price, to quantity and invasiveness of ads to which users are exposed, to the degree of privacy protection afforded to users. This has notably been the case for search engines.[33]

There is also a mathematical difference between a conventional network effect and the kinds of effects seen when data is utilized on a platform. In its most extreme form, the network effects story holds that the value of a network increases quadratically, at the rate $n(n-1)/2$, where n is the number of users (or "nodes" in a network).[34] Even for communications networks, the relationship — in value terms — is likely closer to $n.\ln(n)$, a much less radical rate of growth.[35] Nonetheless, many commentators have argued that platform markets tend to exhibit "winner-take-all" characteristics, because at some point the huge benefits generated by a larger network allegedly make its position unassailable.[36]

But when it comes to data, it is more appropriate to consider the growth of firms and the size of their networks as a function of "learning by doing."[37] Learning by doing is the idea that a firm's productivity improves with experience,[38] which is usually found to be much less pronounced than network effects. For instance, in his seminal paper about learning-by-doing, Arrow cites empirical literature indicating that "to produce the Nth airframe of a given type, counting from the inception of production, the amount of labor required is proportional to $N^{-1/3}$." Contrary to network effects,

---

[32] Data collection and use is merely a tool that a platform uses to customize user experience, not the experience itself. Firms can offer the same end-user experience (which is, logically, what consumers actually value) using different data in different amounts.

[33] Duck Duck Go, For instance, has experienced significantly increased traffic in recent years (though it still lags very far behind Google in terms of users). *See* Duck Duck Go, https://duckduckgo.com/traffic (last visited Jan. 6, 2019). Crucially, competition between Google Search and Duck Duck Go does not seem to be primarily dependent on the data these firms hold. On the one hand, Google offers much lower default levels of privacy protection but proposes a full suite of online applications free of charge. In contrast, Duck Duck Go differentiates itself by offering a search engine with a higher levels of privacy protection. It is not clear how much the data owned by these companies influences consumer choices.

[34] *See* SHAPIRO & VARIAN, *supra* note 7, at 185.

[35] *See, e.g.,* Bob Briscoe, Andrew Odlyzko & Benjamin Tilly, *Metcalfe's law is wrong-communications networks increase in value as they add members-but by how much?*, 43(7) IEEE SPECTRUM 34 (2006).

[36] *Contra* E. Glen Weyl & Alexander White, *Let the Best "One" Win: Policy Lessons from the New Economics of Platforms*, 10 COMPETITION POL'Y INT'L 28-51 (2014).

[37] *See* Hal R Varian, *The economics of internet search*, 96 RIVISTA DI POLITICA ECONOMICA 179 (2006).

[38] *See* Kenneth J Arrow, *The economic implications of learning by doing*, 29 REV. ECON. STUD. 155 (1962). *See also*, Armen Alchian, *Reliability of progress curves in airframe production*, 31 ECONOMETRICA 679 (1963).

learning by doing is thus generally assumed to involve decreasing marginal benefits, and to become almost irrelevant beyond a point.[39] In other words, learning by doing generates significant advantages in the early stages of improving a process, but these incremental advantages drop off sharply after a certain point because firms have picked all the low hanging fruit and because knowledge spills over to rival firms that can imitate the learned process improvements. For this reason, for data-driven platforms, growth more commonly follows a "learning curve" and is not subject to the winner-takes-all effect implied by the conventional network effect assumption.[40]

Another important difference is that, in the case of learning by doing, success is, by definition, a function of superior capabilities (and/or efficiency because of increased productivity). Large returns can (and do) exist in industries in which learning by doing is important (arguably in proportion to the technological complexity of the industry[41]). But it makes no sense to attack such firms even where they may enjoy large profits and market power as a result of their superior skill; this is precisely the type of benefit that the antitrust laws were designed to promote.[42] And there is even less of an argument that learning by doing constitutes a barrier to entry than do network effects because incumbents and entrants must bear roughly the same costs to move down their respective learning curves. Moreover, initial advantages are typically dissipated over time as information spills over. This contrasts with the widely accepted definition of barriers to entry, which holds that a barrier to entry is any cost that must be borne by entrants but not incumbents.[43]

Although these may seem like abstract distinctions, they have very real consequences. Take a recent presentation given by the Chief Economist of the European Commission. In a nutshell, the Commission official held that a positive feedback loop allowed dominant platforms to extract ever more data from its users. The intuition is that a platform with more users generates more data, and this allegedly leads to superior targeted advertisements. These, in turn, allegedly lead to more users

---

[39] *See* Arrow, *id.* at 680.

[40] *See* D'Arcy Coolican and Li Jin, *The Dynamics of Network Effects*, ANDREESSEN-HOROWITZ (Dec. 13, 2018), *available at* https://a16z.com/2018/12/13/network-effects-dynamics-in-practice/.

[41] *See* Phillip E. Auerswald, *Entry and Schumpeterian Profits: How Technological Complexity Affects Industry Evolution*, 20 J. EVOLUTIONARY ECON. 553 (2010) ("In industries where production processes are simple, I find that profits rapidly converge on the norm, particularly when imitation is possible. In industries where production processes are more complex, persistent profits accrue to surviving firms. Such profits are greatest in the early stages of industries where technology is of intermediate complexity — that is, where learning is rapid enough to confer a competitive advantage, but imitation is sufficiently uncertain to deter later entry.").

[42] *See* R.H. BORK, ANTITRUST PARADOX 105 (1993).

[43] *See* STIGLER, *supra* note 2, at 67 (1968).

because the platform can reinvest the added revenue they generate, etc.[44] But this is precisely the conceptual trap that competition authorities should avoid.

This flawed reasoning implies that there is a linear, or even superlinear (e.g. quadratic) relationship between the data owned by a firm and the money it can extract from targeted advertisements. Putting aside the fact that the revenue required to fund platform growth can come from *any* source, not just advertising itself,[45] this leaves out consideration of two crucial questions: (1) when does additional data cease to markedly improve ad targeting, and (ii) at what point does superior ad targeting no longer significantly increases revenues? It is clear that data used for ad targeting exhibit diminishing returns to scale, and that it does so at a fairly modest threshold.[46] Moreover, although this area of research is still in its infancy, there is at least some evidence that highly targeted advertisements might not always be effective because consumers perceive them to be overly intrusive.[47] This suggests that there may indeed be a point at which more data used to improve ad targeting no longer provides any meaningful benefits.

The bottom line is that so-called data "network effects" are in reality a form of learning by doing. They thus raise little antitrust concern and should be embraced by policymakers because they ultimately lead to superior efficiency, the very goal of antitrust law.

2. *Dynamic capabilities*

This leads us to an important second point. The challenge for firms in data-reliant industries is multidimensional. Not only must they acquire data (and this is not merely a matter of "data network effects") but, just as importantly, they must also develop the expertise to analyze this data, draw useful insights from it, and turn these insights into successful products. In doing so, acquiring the *right* data and getting the best out of a firm's engineers is at least as important as controlling a large amount of data or engineering expertise. In other words, there is no single ingredient that mechanically leads to success. Instead, it is up to firms to identify and seize upon emerging business opportunities.

---

[44] *See* Tommaso Valletti, *Après moi, le déluge! Tech giants in the digital age*, CRA Conference (Dec. 5, 2018), *available at* https://ecp.crai.com/wp-content/uploads/2018/12/Tommaso-Valletti-2018.pdf.

[45] *See* Manne & Wright, *supra* note 24, at 210-11 & fn. 137 ("[T]hough Google perhaps generates the funds for its continued product development through its successful business, the same business model need not be adopted by competitors. In fact, Microsoft, one of Google's primary competitors, has a market capitalization substantially larger than Google's, and higher profits generated by its other businesses to invest in search engine functionality improvements. There is no reason why it matters if this investment comes from advertising revenue, the sale of operating systems, or outside capital sources.").

[46] *See* William Terdoslavich, *Big Data and the Law of Diminishing Returns*, INFORMATIONWEEK (Nov. 27, 2015), http://www.informationweek.com/big-data/big-data-analytics/big-data-and-the-law-of-diminishing-returns/d/d-id/1323310; Hal Varian, *Is there a data barrier to entry?* (presentation, Jun. 2015), *available at* http://www.learconference2015.com/wp-content/uploads/2014/11/Varian-slides.pdf.

[47] *See* Avi Goldfarb & Catherine Tucker, *Online display advertising: Targeting and obtrusiveness*, 30 MARKETING SCIENCE 398 (2011).

Under this light, the resounding success of certain technology platforms appears to be down to their respective "dynamic capabilities" rather than the operation of positive feedback loops.

Dynamic capabilities can be defined as:

> [T]he particular (nonimitability) capacity business enterprises possess to shape, reshape, configure, and reconfigure assets so as to respond to changing technologies and markets and escape the zero-profit condition.[48]

Critically, David Teece adds that "[t] he dynamic capabilities framework recognizes that the business enterprise is shaped but not necessarily trapped by its past."[49]

This is of great importance for antitrust authorities. Though it may seem obvious, not all firms will possess the requisite capabilities to compete and flourish in these dynamic marketplaces. And evolving market realities imply that some prosperous firms will fall out of favor with consumers for no other reason than the firms' failure to adapt to new market realities (these firms will often find themselves in situations where it is too late to turn the ship and opt for another business strategy). Antitrust enforcers may often be tempted to try and prop-up these failing firms – under the faulty premise that their demise is due to anticompetitive behavior rather than a mix of poor decisions and bad luck. But forcing successful firms to share their assets will often only delay the inevitable.

These factors can notably be seen at play in the early days of the search engine market. In 2013, the Atlantic ran a piece titled "what the web looked like before Google."[50] By comparing the websites of Google and its rivals in 1998 (when Google Search was launched), the article shows how the current champion of search marked a radical departure from the status quo:

---

[48] *See* D.J. TEECE, DYNAMIC CAPABILITIES AND STRATEGIC MANAGEMENT: ORGANIZING FOR INNOVATION AND GROWTH 48 (2009).

[49] *Id.* at 50.

[50] *See* Rebecca Greenfield, *What the Web Looked Like Before Google*, THE ATLANTIC (Sep. 27, 2013), , *available at* https://www.theatlantic.com/technology/archive/2013/09/what-web-looked-google/310544/.

Figure 2: Above, Homepages of Yahoo, AltaVista and AOL, 1998. Bellow, Homepage and first result page of Google, 1998.

These images reveal critical differences between Google and its rivals. Even if it stumbled upon it by chance (and although it was not necessarily apparent at the time), Google immediately identified a winning formula for the search engine market. It ditched the complicated classification schemes favored by its rivals and opted, instead, for a clean page with a single search box. This notably ensured that users could access the information they desired in the shortest possible amount of time (thanks in part to Google's PageRank algorithm).[51]

It is hardly surprising that Google's rivals struggled to keep up with this shift in the search engine market. The theory of dynamic capabilities tells us that firms who have achieved success by indexing

[51] *See* Steven Levy, *Exclusive: How Google's Algorithm Rules the Web*, WIRED (Feb. 22, 2010), *available at* https://www.wired.com/2010/02/ff_google_algorithm/.

the web will struggle when the market rapidly moves towards a new paradigm (in this case, Google's single search box and ten blue links). During the time it took these rivals to identify their weaknesses and repurpose their assets, Google kept on making successful decisions (most notably, the introduction of Gmail,[52] its acquisitions and of YouTube and Android,[53] the introduction of Google Maps,[54] etc.). All these products tied-in with one of its key capabilities, which is to provide users with information through whatever platform they are using (desktop or mobile) and regardless of the medium in which it stored (be it web pages, online videos, maps, emails, etc.). Seen from this evolutionary perspective, Google thrived because its capabilities were perfect for the market at that time, while rivals were ill-adapted.

If this interpretation is to be believed, then Google's meteoric rise had nothing to do with "data network effects" and everything to do with its specific capabilities and the strategy it deployed, over many years, to capture latent consumer demand in the search engine market. In fact, it overcame a tremendous data *disadvantage* to catch-up with—and overtake—firms such as Yahoo and AltaVista (who had entered the search engine market long before Google).[55]

This should, at the very least, give pause to proponents of the "data network effects" theory. Indeed, it is hard to take such claims seriously when they completely ignore one of the most significant competitive events in the history of the search engine. If the data network effects fable were true, and search engines with more data inevitably prosper compared to data-poor rivals, then Yahoo and AltaVista should have obliterated Google. The reality is that competition in the search engine market (and probably all other online markets) is about far more than data.

---

[52] *See* Annys Shin, *When Google introduced Gmail*, THE WASHINGTON POST (March 30, 2017), *available at* https://www.washingtonpost.com/lifestyle/magazine/google-introduces-gmail/2017/03/27/8a210e9c-034a-11e7-b1e9-a05d3c21f7cf_story.html?utm_term=.e8f2239b75ee.

[53] *See* Matt Reynolds, *If you can't build it, buy it: Google's biggest acquisitions mapped*, WIRED (Nov. 25, 2017), *available at* https://www.wired.co.uk/article/google-acquisitions-data-visualisation-infoporn-waze-youtube-android.

[54] *See* Brett Taylor, *Mapping your way*, GOOGLE OFFICIAL BLOG (Feb. 8, 2005), https://googleblog.blogspot.com/2005/02/mapping-your-way.html.

[55] Whereas the early search engine giants like Lycos scrambled to index the quickly growing set of web pages (one billion by the year 2000. *See* Stephen Goehlerab, Masud Cadera, and Harold Szua, *Smart Internet search engine through 6W* at 3 in PROCEEDINGS OF SPIE - THE INTERNATIONAL SOCIETY FOR OPTICAL ENGINEERING (2006) *available at* https://www.researchgate.net/publication/271498959, Google took a different tack, eschewing the collation of large data sets (as was the habit of its competitors) and focusing on the relevance of pages to given queries:

> Google embraced the philosophy of quality over quantity. They didn't try to index every page in existence. Instead, Google focused on trying to retrieve the best possible results to meet the user's query. Google tried to display the few highly relevant results before the thousands of slightly relevant results that plagued the older search engines. Google also introduced the concept of page ranking to help move towards their goal. Google's quick popularity forced other major search engines to redesign their own algorithms to keep pace.

Id.

The theory of dynamic capabilities also sheds light on the European Union's recent *Google Search* and *Google Android* decisions.[56] In these cases, the European Commission concluded that Google had excluded its rivals from the search engine market. On closer inspection, however, it seems at least plausible that these rivals simply failed because of poor business judgement.

In the *Search* case, Foundem (one of the complainants) based its entire business on comparison shopping services.[57] In so doing, it took no steps to protect itself from potential changes to the Google Search engine on which it depended, despite a clear industry trend towards single search boxes leading to all results. As Geoffrey Manne put it:

> Google's purpose is not to send traffic away from its site; it's "to bring all the world's information to users seeking answers." It just happens that sending users away from its site was the best and quickest way to provide answers on the Web in, say, 1999. But as Google's technological abilities and resources grew, and as users sought even quicker answers — especially ones provided by voice or on mobile devices — its mechanisms for serving its users evolved.[58]

Much of the same can be said about Yandex, a Russian search engine that was a complainant in the *Android* case.[59] Yandex notably argued that it was being excluded from the search market as a result of Google's dominance of the Android mobile platform. Regardless of the merits of the underlying case, two facts are particularly relevant: (i) Yandex never attempted to launch its own mobile OS, and (ii) with the rise of virtual assistants, the market for search will likely become less and less distinct from the mobile OS market. Though Yandex has not been excluded from the Russian market (proof that the theory of data network effects is greatly exaggerated), its market share has slowly declined (it nevertheless remains the first search engine in Russia).[60]

The upshot is that, in both of these cases, enforcers struggled to distinguish exclusion resulting from anticompetitive conduct from shifts in the marketplace that incidentally caused some firms to fall by the wayside (due to their failure to adapt to new circumstances). When deciding on such matters, it is crucial that authorities do not ignore the important role that dynamic capabilities may play

---

[56] Case AT.39740 *(Google Search (Shopping)),* Commission Decision of 27 June 2017, ¶ 379, *available at* http://ec.europa.eu/competition/antitrust/cases/dec_docs/39740/39740_14996_3.pdf. *See also*, Commission Decision in Case AT.40099 (Google Android), slip op. (Jul. 18, 2018).

[57] *Google Search (Shopping)*, *id.* at ¶ 379.

[58] *See* Geoffrey A Manne, *The Real Reason Foundem Foundered*, ICLE ANTITRUST & CONSUMER PROTECTION RESEARCH PROGRAM WHITE PAPER 2018-02 (2018).

[59] *See Russia's Yandex says complained to EU over Google's Android*, REUTERS, Nov. 13, 2015, *available at* https://www.reuters.com/article/us-eu-antitrust-google-yandex/russias-yandex-says-complained-to-eu-over-googles-android-idUSKCN0T21L420151113.

[60] *See* Katerina Rubinova, *Battle of the Titans: Yandex vs Google*, THE FUTURE MEDIA BLOG, Mar. 11, 2016, http://blogs.bcu.ac.uk/futuremedia/2016/03/11/battle-of-the-titans-yandex-vs-google/. *See also*, *Search Engine Market Share Russian Federation: Dec. 2017-2018*, STATCOUNTER GLOBALSTATS, http://gs.statcounter.com/search-engine-market-share/all/russian-federation.

during these industry transition periods. In contrast to the claims made by those who allege that "data network effects" account for the success of some firms and failure of others, these capabilities actually do appear to be a key predictor of a firm's success or failure. In short, business model competition necessarily implies that some firms will be left out, not because they don't have data, but because they have chosen a strategy that either left them with too few users to generate relevant data or collecting the wrong type(s) of data.

C. The path to data monetization

Policymakers should also bear in mind that platforms must often go to great lengths in order to create data about their users — data which these same users often do not know about themselves. Under this framing, data is a by-product of firms' activity rather than an input that is necessary for rivals to launch a business. This is especially clear when one looks at the formative years of numerous online platforms. Most of the time, these businesses were started by entrepreneurs who did not own much data but, instead, had a brilliant idea for a service that consumers would value. Even if data ultimately plays a large role in the monetization of these platforms, it does not appear to be necessary for their creation.

> While data collected from users can be important to online providers in improving the services offered and their ability to monetize, user data is only one of many inputs into providing online services. The quality of services offered by online providers, and the ability to monetize effectively, is driven by much more than user data. There are many other sources of data, inputs into providing high quality services, dimensions of quality, and means of attracting users(such as distribution arrangements). Online providers can make investments in quality and distribution that are independent of its scale of users. And, through these investments, a provider can attain scale. Thus, it is incorrect to assert that an online platform lacking scale today can never attain scale. The fact that online providers can gain user scale in ways that do not involve user data weakens the claimed user data-service quality feedback-loop.[61]

1. Platforms create data

Possessors of information are assumed to benefit from the private use of information. But, while this is undoubtedly true for some data, it is often the case that information has no realizable value unless and until a mechanism is created for *using* it. At the extreme, for example, there is no intrinsic value to a consumer in the knowledge that she likes music by the Grateful Dead. There is value to *her*, however, in *others* knowing and using this information — most obviously, music recommendation services and music sellers (but also, in the non-commercial sense, friends and social communities).

[61] Andres V. Lerner, *The Role of "Big Data" In Online Platform Competition* (unpublished manuscript, Aug. 24, 2014) at 28, *available at* http://awa2015-concurrences.nursit.com/IMG/pdf/big.pdf.

There is thus also value to the consumer from making sure that others know this information about her.

That applies to information that is known to the consumer. But there is also information that does not even "exist" in any real sense (or at least is not known) until the mechanism is created to *elicit* it. Indeed, "[i]t is questionable whether wants, as conscious motives to conduct, ever exist unless we are in a position of having to choose, to adopt one line of conduct and renounce another."[62] Whether or not someone likes her brother's latest photo of her dog isn't "information" in any meaningful sense until the photo exists, is shared with her, and she considers her reaction to it. In this sense, the vast majority of (actionable) information exists only because of some activity that creates the mechanism for the information to be created (or coalesced).[63]

This type of information is extremely important, but routinely overlooked, in discussions of big data. It is, in fact, arguably the most important sort of data employed by these platforms, and it does not exist absent the platforms on which it is created. Crucially, data of this sort is most obviously the manifestation of users' preferences. A user's *preferences* may be, in some philosophical sense, pre-existing. But the user may not even know what they are until asked, and certainly external users of that information cannot know it without it being communicated either directly (e.g., "I like the Grateful Dead") or indirectly (e.g., through a user's music purchase history).

Thus, information about users' preferences can perhaps be known, but, more to the point, it must typically be *elicited*. "[The consumer] does not know what he will want, and how much, and how badly; consequently he leaves it to producers to create goods and hold them ready for his decision when the time comes."[64]

And users have an interest in that information being elicited and shared. Moreover, users have no particular comparative advantage in the eliciting or interpreting of that information: as noted, it may not even be known *ex ante*, and, even if it is, in many cases it is virtually useless. As a result, the mechanisms that elicit and share that information with others who *do* have a comparative advantage in using it are of great value to users — not only because the information, once processed, may be used by others in ways that ultimately impart value to the user, but also because the very act of eliciting and sharing the information imparts knowledge to the user directly.

---

[62] FRANK H. KNIGHT, RISK, UNCERTAINTY, AND PROFIT 60 (1964).

[63] This sort of information must be distinguished from statistical knowledge, which consists of making inferences based on past experience. Although This is related to the distinction between "information" and "news" in John M. Marshall, *Private Incentives and Public Information*, 64 AM. ECON REV. 373, 373-74 (1974) ("In common usage the word information is ambiguous. It means either information that is known, as it is after being delivered, or unknown, as it is when it is purchased. The word 'information' will be used here only in the latter sense, while the former meaning will be conveyed by the term 'news. Thus, the purchase of information eventually results in news.").

[64] KNIGHT, *supra* note 62, at 241.

This is a crucial and overlooked aspect of policy discussions surrounding data-intensive platforms. The value of a user's interactions with Facebook and Google, for example, is not, as commonly assumed, only in the platform's aggregation and use of the data generated through those interactions, but also in the user's own, immediate access to information that either didn't exist or wasn't known to her beforehand. An enormous quantity of the data at issue in these policy discussions is of this sort: it is non-existent, unknown, and/or useless, *even for personal use by the user*, until it is made manifest through some activity by which the user interacts with the platform. Thus the value of those activities is not just in the sharing of information with others, but in the creation of information in the first place.

Why is this so important? Because, as we discuss below,[65] it turns the generally assumed "platform information asymmetry" on its head. To begin with, there is information that a user *does* often know about herself that the platform does not: that is, her preferences. Any information that the platform gleans about her preferences is necessarily incomplete and indeterminate, and the platform can make only inferences — inferences that, even when accurate, can quickly become obsolete. Information asymmetry in this regard runs in favor of the individual user, not the platform.

In addition there is information that even the user does not know about herself and that becomes known only *because of* the platform. Even though the user does not (unlike the platform) know the aggregate information from many users of which her data is only a minuscule part, the private use value of that information is better known to the user than to the platform. While the user knows whether the information is accurate and valuable, and while there is no limitation on what the user can do with that information, the platform is able to use it only to make inferences about its relevance and importance to the user, with limited accuracy.

To be sure, a platform can also combine this information with other data to create yet more information and to derive value inaccessible to the individual user. But the relative magnitudes of these different types of information and their value to different users, and to the platform itself, is uncertain. It cannot simply be assumed that there is "asymmetry" or that it flows in only one direction.

Indeed, this puts paid to the canard that "if you're not paying for it, you are the product" or "the price for 'free' services is your data." In truth, much of the information we share is shared because it is only by doing so that its value can be realized. More important, much of the data we share with platforms does not even exist (or is not known) separately from our interactions with these platforms. **In this sense it is not data that is the "price" users pay for platform services; it is platform services that are the "price" platforms pay for data.**

[65] *See infra* Section III.A.2.

Of course, none of this is especially new; it is simply overlooked. The great UCLA economist, Jack Hirshleifer, noted many of these dynamics as long ago as the 1970s.

> The possessor can in general benefit simply by private use of the information for his own productive or consumptive decisions. But in a market context it might also be possible for him to profit from sale of the information to others. The information-seeker might correspondingly find it advantageous to produce socially "new" information by direct inquiry of Nature (research) or to purchase "secondhand" information in the market. Viewed as a tradeable commodity, information has (as we shall see) a number of special features.... In the market process information can be regarded as "pulled" from the possessor by purchase, i.e., by payment of an explicit price. **But what is surprising, the possessor may find it preferable to give away this valuable commodity, to disseminate it without pull of compensation. Indeed it may be highly profitable for him to incur costs so as to gratuitously "push" information to potential recipients!** As for the information-seeker, his knowing that the possessors are so motivated may lead to adoption of a monitoring or listening mode of learning behavior.[66]

2. *Most platform businesses started without any data*

Another important point is that data often becomes significant only at a relatively late stage in these businesses' development. A quick glance at the digital economy is particularly revealing in this regard. Google and Facebook, in particular, both launched their platforms under the assumption that building a successful product would eventually lead to significant revenues. It took five years from its launch (and 300 million users) for Facebook to start making profits. But even then, it was not entirely clear whether the social network would generate most of its income from app sales or online advertisements.[67] It was another three years before Facebook started to cement its position as one of the world's leading providers of online ads.[68] During this eight-year timespan, it seems that Facebook's first concern was not so much the monetization of its platform, but user growth.

Facebook thus appears to have concluded (rightly, it turns out), that once its platform attracted enough users, it would surely find a way to make it highly profitable. This suggests that data might not have been of critical importance during the formative years of the Facebook platform (or at least not for its monetization). This might explain how Facebook managed to build a highly successful

---

[66] Jack Hirshleifer, *Where Are We in the Theory of Information?*, 63 AM. ECON. REV. 31, 32 (1973) (emphasis added).

[67] *See* Derek Thompson, *Facebook Turns a Profit, Users Hits 300 Million*, THE ATLANTIC (Sep. 17, 2009), *available at* https://www.theatlantic.com/business/archive/2009/09/facebook-turns-a-profit-users-hits-300-million/26721/.

[68] *See* Rebecca Greenfield, *2012: The Year Facebook Finally Tried to Make Some Money*, THE ATLANTIC (Dec. 14, 2012), *available at* https://www.theatlantic.com/technology/archive/2012/12/2012-year-facebook-finally-tried-make-some-money/320493/.

platform, despite a large data disadvantage over rivals like MySpace.[69] The upshot is that, in the case of Facebook, data does not seem to have been a prerequisite for building a successful platform.

And Facebook is no outlier. Other successful technology firms have similar origins. For instance, Snapchat managed to build a successful platform that has 187 million daily active users. Snap Inc. achieved this feat without much, if any, user data, and despite entering the market later than numerous high-profile rivals,[70] including Facebook,[71] Instagram,[72] and WhatsApp.[73] Like Facebook, Snapchat chose to build its network without a clear monetization strategy, deferring this question to a later stage, when it would have an established user base.[74] Granted, Snapchat may yet succumb to larger rivals (at the time of writing, Instagram seems to be winning the battle and may ultimately drive Snapchat out of the market). But these rivals' success does not appear to have anything to do with superior access to data.[75] Instead, the Snapchat's possible decline appears to be down to Instagram having introduced more attractive features to its app.[76] Far from being suggestive of data-related market failures, Snapchat's decline at the hands of Instagram appears to be a sign of healthy competition. It thus shows that competition between digital platforms is about much more than data, and that it is perfectly feasible for innovative companies to enter these markets despite significant data disadvantages.

And consider companies like Uber, Lyft and Sidecar that have taken over the personal transport sector. They too had no customer data when they began to challenge established cab companies that *did* possess such data. If data were really so significant, they could never have competed successfully. But Uber, Lyft and Sidecar have been able to effectively compete because they built products that

---

[69] *See* Harrison Jacobs, *Former MySpace CEO explains why Facebook was able to dominate social media despite coming second*, BUSINESS INSIDER (May 9, 2015), *available at* https://www.businessinsider.fr/us/former-myspace-ceo-explains-why-facebook-was-able-to-dominate-social-media-despite-coming-second-2015-5.

[70] *See* J.J. Colao, *Snapchat: The Biggest No-Revenue Mobile App Since Instagram*, FORBES, Nov. 27, 2018, *available at* https://www.forbes.com/sites/jjcolao/2012/11/27/snapchat-the-biggest-no-revenue-mobile-app-since-instagram/#75c31ad97200.

[71] *See A Faster Way to Message on Mobile*, FACEBOOK, Oct. 19, 2011, *available at* https://www.facebook.com/notes/facebook/a-faster-way-to-message-on-mobile/10150249543542131

[72] *See* MG. Siegler, *Instagram Launches with the Hope of Igniting Communication Through Images*, TECHCRUNCH, OCT. 6, 2010, *available at* https://techcrunch.com/2010/10/06/instagram-launch/.

[73] *See* Parmy Olson, *Exclusive: The Rags-To-Riches Tale Of How Jan Koum Built WhatsApp Into Facebook's New $19 Billion Baby*, FORBES, Feb. 19, 2014, *available at* https://www.forbes.com/sites/parmyolson/2014/02/19/exclusive-inside-story-how-jan-koum-built-whatsapp-into-facebooks-new-19-billion-baby/#18e781f82fa1.

[74] *See* Colao, *supra* note 57.

[75] *See* Kurt Wagner & Rani Molla, *Why Snapchat is shrinking*, RECODE, Aug. 7, 2018, *available at* https://www.recode.net/2018/8/7/17661756/snap-earnings-snapchat-q2-instagram-user-growth.

[76] *See* Sara Salinas, *Instagram Stories has twice as many daily users as Snapchat's service – and it now has background music*, CNBC, June 28, 2018, *available at* https://www.cnbc.com/2018/06/28/instagram-stories-daily-active-users-double-snapchats.html.

users wanted to use[77] — they came up with an idea for a better mousetrap. The data they have accrued came *after* they innovated, entered the market and mounted their successful challenges — not before.

The list of companies that prevailed despite starting with little to no data, and before they implemented (or even identified) a data-dependent monetization strategy is vast. Other examples include AirBnb, Amazon, Twitter, PayPal, etc. These abundant illustrations severely undermine ideas that data constitutes a barrier to entry, that "data network effects" inevitably lead to tech platform tipping, or that data constitutes an essential facility.

A more apt economic parallel can be made with regard with the economic literature on two-sided markets. In these markets, it is well established that firms face a "chicken and egg problem". Because the success of their business hinges on attracting two complementary groups of users, these platforms must often decide which group of users to favor early on in the hope that this will then kickstart any positive feedback loops that may exist between users on both sides of the platform.[78] One particularly relevant strategy for ad-supported business is what David Evans refers to as "sequential entry":[79]

> In some cases it is possible to get one group of agents on board over time and then make these agents available to the other group of agents later in time. That is the situation with advertising-supported media. One can use content to attract viewers and then bring advertisers on board later. This dynamic works because there are non-positive indirect network effects between the two sides: viewers do not care about advertisers (and may dislike advertising) but come to platform for the content.[80]

The ubiquity of the sequential entry strategy, which is relevant for many internet firms (including, Google, Facebook and their rivals), contradicts arguments that access to data is necessary to compete in these industries. Granted, at some point firms may need data to earn profits and reinvest in their platforms. But saying that this dynamic somehow interferes with competition is merely a repeat of the "deep pocket" fallacy that plagued early predatory pricing theory.[81] In the case of online

---

[77] *See* Karen Mathews & Verena Dobnick, *Uber Cars in New York Now Outnumber Yellow Cabs*, HUFFINGTON POST, Mar. 19, 2015, *available at* http://www.huffingtonpost.ca/2015/03/19/new-york-citys-storied-y_n_6900980.html.

[78] *See* Bernard Caillaud & Bruno Jullien, *Chicken & egg: Competition among intermediation service providers*, RAND J. OF ECON. 310 (2003). *See also*, Geoffrey G Parker & Marshall W Van Alstyne, *Two-sided network effects: A theory of information product design*, 51 MANAGEMENT SCIENCE 1496 (2005). *See also*, Jean-Charles Rochet & Jean Tirole, *Platform competition in two-sided markets*, 1 J. OF THE EUR. ECON.ASS'N 990 (2003).

[79] *See* David S Evans, *How catalysts ignite: the economics of platform-based start-ups*, *in* PLATFORMS, MARKETS AND INNOVATION 109 (2011).

[80] *Id.*

[81] The intuition is that firms with significant financial resources can sustain losses for longer periods of time and thus evict smaller rivals through predatory pricing. *See* Corwin D Edwards, *Conglomerate bigness as a source of power*, *in* BUSINESS CONCENTRATION AND PRICE POLICY 334-335 (1955). This notion was severly exposed by Chicago-school scholars, notably

platforms, there is no reason to believe that firms who earn less profits will invest less in their products. Instead, all these firms need to do is convince investors that they will ultimately have the best product and the most users. If capital markets work properly – and there are literally billions of dollars flowing to Silicon Valley tech startups every year[82] – then being able to immediately monetize data offers firms little to no advantage over their rivals that must call upon capital markets.

The inevitable conclusion is that, in reality, those who complain about data facilitating unassailable competitive advantages have it exactly backwards. Companies need to innovate to attract consumer data, otherwise consumers will switch to competitors (including both new entrants and established incumbents). As a result, the desire to make use of more and better data *drives* competitive innovation, with manifestly impressive results: the continued explosion of new products, services and apps is evidence that data is not a bottleneck to competition but a spur to drive it.

D. Is data the new oil?

While the metaphor "data is the new oil" may be rhetorically appealing, the comparison could hardly be any less apt. As we have shown, unlike oil, data is ultimately a form of information and as such is non-rivalrous and in many cases non-exclusive. Moreover, the value of a given dataset hinges critically on the expertise that firms can bring to bear in order to analyze the data. Unfortunately, this combination of learning by doing and firmwide capabilities in data-intensive markets has often been mislabeled as a "data network" effect.

Finally, unlike an oil company that must first drill and refine oil before it can make sales, large amounts of data often become important only in later stages of a digital platform's development. At the same time, much of the data used by platforms does not, in any meaningful sense, *pre-exist* the platforms' interactions with their users; rather, it is created by those interactions. As we have discussed, firms routinely build successful businesses without having access to pre-existing data. Instead, they hope that a strong product on the user side of the market, will eventually translate into substantial revenues, notably by leveraging the data that is eventually generated on the platform.

III. The faulty logic of "data barriers to entry"

In antitrust policy circles, it is often claimed that data constitutes a barrier to entry that prevents or makes it more difficult for competitors to develop alternative products in the marketplace — the so-called "data barrier to entry." The argument is that upstarts do not have sufficient data to compete with established players like Google and Facebook, which in turn employ their data to attract online

---

because it assumes capital market imperfections. *See* Lester G Telser, *Cutthroat competition and the long purse*, 9 J.L. & ECON. 270 (1966).

[82] *See* Kate Clark, *Venture capital investment in US companies to hit $100B in 2018*, TECHCRUNCH, Oct. 8, 2018, *available at* https://techcrunch.com/2018/10/09/venture-capital-investment-in-us-companies-to-hit-100b-in-2018/.

advertisers and to foreclose their competitors from this crucial source of revenue. It is thus sometimes argued that firms with large agglomerations of data are inherently protected from competition, and that antitrust enforcement must be used to restructure firms to overcome their data barrier.[83]

There are at least two important reasons to be dubious of such arguments. First, superior competition, notably of the sort that generates large amounts of data, is not a barrier to entry. Second, and related, tearing down supposed data barriers to entry entails punishing firms for success and fails to properly account for the long-run effect on innovation.

A. Superior competition, notably through data, is not a barrier to entry

The logic of entry barriers implied by many who assert *data* as an entry barrier is a curious one. Because data (in the context relevant here) can be used to improve the quality of products and/or to subsidize their use, the idea of data as an entry barrier suggests that any product improvement or price reduction made by an incumbent may amount to an entry barrier to any new entrant. This effect may be magnified with network effects. And if the product improvement itself is in the product's processing of data or its ability to access data, it is doubly so.

With this in mind, any assessment of the data barrier to entry argument should thus consider whether it is tantamount to an argument that competition itself is a cognizable barrier to entry. Without more, the concept of barriers to entry has no intrinsic antitrust relevance; it's merely a statement that the better the incumbent is (or the cheaper its product), the harder it is for new entrants to compete. It would be a curious approach to antitrust if this were treated as a problem, as it would imply that firms should under-compete — should forego consumer-welfare enhancements — in order to bring about a greater number of firms in a given market simply for its own sake.

Even if, somehow, we thought that this expansive view of barriers to entry were correct, how would it be implemented? At what point would competition, product improvements, and price reductions become anticompetitive? It can't be at *any* point at which they make entry more difficult, because on the margin that must happen at every single point in the product lifecycle. But otherwise the dividing line may be essentially arbitrary. And, of course, any approach here that impedes competition on the merits for incumbents would also, on the margin, make new entrants compete less vigorously, invest less in their own products, etc., where competition is for the market (as it often seems to be

[83] *See* Newman, *Search, Antitrust, and the Economics of the Control of User Data*, *supra* note 21, at 404 ("This Article is largely a case for reorienting many antitrust investigations in the technology sphere — and more generally regulatory approaches — to focus far more on the issue of how control of user data can entrench monopoly power and harm consumer welfare in an economy shaped increasingly by the power of companies collecting personal data.").

in high-tech, platform markets) — which could easily undermine the entire "entry over innovation" rationale.[84]

And there is a fundamental underlying error in the entire barriers to entry enterprise: It is rooted in the idea that barriers tend to determine the number of firms, and the number of firms determines competitiveness. But this is a far too simplistic view.

> For example, firms can compete against each other by investing in the development of new products, in the promotion of the product, or in the reduction of costs. All these features are determined in equilibrium together with industry concentration. One can show in these models that as markets grow in size, the industry structure that can emerge is not one of atomistic competition with constant quality but rather one where concentration remains high but product quality increases. Therefore, competition along nonprice dimensions can explain why concentration does not necessarily diminish as industries grow. The significance of this point cannot be overstated. Models that focus on only price competition may fail miserably to correctly predict industry concentration and consumer welfare when there are other product dimensions along which competition occurs. This is likely to be particularly true in industries requiring investment and creation of new products. It is no coincidence that many of the most controversial antitrust and regulatory cases have arisen in high- technology industries (e.g., computers and telecommunications) where competition in research and development and new products is paramount.[85]

The confusion surrounding the meaning of "barriers to entry" often results because the precise consequence of having an entry barrier is unclear. If there are such "barriers," is anticompetitive conduct facilitated by them? The proper analysis doesn't end with entry barriers; it starts with analysis of what would happen without barriers, and then assesses whether barriers change anything. In so doing, it must also account for the benefits of existing conduct, including the benefits of market conduct and structures that may operate to impede new entry ("barriers"), but also facilitate new investment and innovation by incumbents. Where it does not, it again tends the assessment toward protection of the *status quo*.

A key problem in the analysis of entry barriers is the assumption of essentiality of inputs or other relationships created by early movers.

---

[84] For more on this, see the discussion, *infra*, at notes 97-101 and accompanying text.

[85] Dennis W. Carlton, *Barriers to Entry*, in 1 ISSUES IN COMPETITION LAW AND POLICY 601, 603-04 (ABA Section of Antitrust Law, 2008).

1. Microsoft and the applications pathway to entry

Consider this error in the *Microsoft* court's analysis of entry barriers: The court pointed out that new entrants face a barrier that Microsoft didn't face, in that Microsoft didn't have to contend with a powerful incumbent impeding its entry by tying up application developers.[86]

But while this may be true, Microsoft did face the *absence* of any developers at all and had to essentially create (or encourage the creation of) businesses that didn't previously exist. In fact, although the court dismissed this argument (in a different context), it noted that, "[a]ccording to Microsoft, it had to make major investments to convince software developers to write for its new operating system, and it continues to 'evangelize' the Windows platform today."[87] Yet the court also notes:

> Because the applications barrier to entry protects a dominant operating system irrespective of quality, it gives Microsoft power to stave off even superior new rivals. The barrier is thus a characteristic of the operating system market, not of Microsoft's popularity, or, as asserted by a Microsoft witness, the company's efficiency.[88]

The point about quality may be true, and it may even be true that the extent of the purported barrier didn't correlate with Microsoft's popularity or efficiency. But it is *not* true that the applications barrier to entry was independent of Microsoft's efforts or investment: it was not merely a "characteristic of the operating system market," as if exogenous to any conduct undertaken by Microsoft in order to obtain its scale in the first place.

Rather, as noted, Microsoft invested heavily to create the network of developers in the first place. It entered a market with a unique barrier to entry of its own — the absence of any applications for its platform — and proceeded to expend considerable resources to facilitate their creation.

Moreover, having done so, Microsoft created a huge positive externality for new entrants: existing knowledge and organizations devoted to development, industry knowledge, reputation, awareness, incentive for schools to offer courses, etc. It could well be that new entrants in fact faced *lower* barriers with respect to app developers than did Microsoft when it entered.

This is crucial in considering the distinction between data pre- and post-entry. Much of the "analysis" of data as a barrier to entry casually speaks as if, because an incumbent has data, new entrants must also have data in order to compete. But the reality is that incumbents entered *without* data and produced it subsequent to entry — again, sometimes creating entirely new businesses and business

---

[86] *United States v. Microsoft Corp.*, 253 F.3d 34, 56 (D.C. Cir. 2001) ("When Microsoft entered the operating system market with MS-DOS and the first version of Windows, it did not confront a dominant rival operating system with as massive an installed base and as vast an existing array of applications as the Windows operating systems have since enjoyed.").

[87] *Id.*

[88] *Id.*

models around it. Facebook is an obvious example of this dynamic, but so are Uber and Google and many others.

Data in this respect is like reputation. Nearly all new entrants suffer reputational disadvantages. And yet new entry happens all the time. Likewise, the more successful the incumbent — the larger its network, the stronger its reputation, the better its product — the more difficult is new entry. And yet this *is* competition.

In the US, courts have consistently rejected the idea that reputation operates as a barrier to entry. The Ninth Circuit has noted:

> We agree with the unremarkable proposition that a competitor with a proven product and strong reputation is likely to enjoy success in the marketplace, but reject the notion that this is anticompetitive. It is the essence of competition.[89]

Or the Third Circuit, for example, noted:

> New entrants and customers in virtually any market emphasize the importance of a reputation for delivering a quality good or service.... [Plaintiff's] argument, without some limiting principle (that it fails to supply), implies that there are barriers to entry, significant in an antitrust sense, in all markets. We find this proposition implausible and... precluded by Supreme Court precedent.[90]

It is possible that, under some conditions, reputation or product differentiation can operate as a barrier to entry.[91] But there must be special circumstances for that to be true — most notably it has arisen in cases where incumbents have undertaken actions to prevent or preclude new entrants from developing their own brand reputation in order to compete.[92] But it can't be always and everywhere true, or else every market would be characterized by anticompetitive barriers.

The same holds true for data. Data is typically generated by companies *after* they enter markets, as a by-product (or intended consequence) of their operations, or else in some case it is *purchased* beforehand.[93] It cannot be the case that doing so in the abstract creates an entry barrier, or else every market would be marked by entry barriers and the risk of antitrust liability for incumbents —

---

[89] *Omega Environmental, Inc. v. Gilbarco, Inc.*, 127 F.3d 1157, 1164 (9th Cir. 1997) (*Citing American Professional Testing Service, Inc. v. Harcourt Brace Jovanovich Legal and Professional Publications, Inc.*, 108 F.3d 1147, 1154 (9th Cir.1997) ("[R]eputation alone does not constitute a sufficient entry barrier in this Circuit."); *United States v. Syufy Enterprises*, 903 F.2d 659, 669 (9th Cir.1990) ("We fail to see how the existence of good will achieved through effective service is an impediment to, rather than the natural result of, competition.").

[90] *Advo, Inc. v. Philadelphia Newspapers, Inc.*, 51 F.3d 1191, 1201-02 (3d Cir. 1995).

[91] *See Id.* at 1202.

[92] *See US Philips Corp v. Windmere Corp.* 861 F.2d 695 (Fed. Cir. 1988).

[93] Daniel L. Rubinfeld & Michal S. Gal, *Access Barriers to Big Data*, 59 ARIZ. L. REV. 339, 357 (2017) ("More commonly, data are collected as a (valuable) side-effect of other productive activities.").

including offline markets. By definition, data produced as a consequence of ongoing market operations is something only incumbents will have — and incumbents will always have. Defining the possession of data in this context as an entry barrier would be tantamount to inviting antitrust challenges on the basis of a company's mere existence (and even more so, success) in a market competitors wish to enter.

What seems to be required in order that data may be treated as a potential entry barrier is that the data at issue be some combination of essential, unique, exclusive, and rivalrous. If a suitable dataset can be created by new entrants or obtained elsewhere, or if other data can be used in its stead, or if alternatives other than data can be used (*e.g.*, synthetic data or artificial intelligence), then it is hard to see any relevant competitive significance from data, *regardless of the amount*.

A key aspect of the mistake here is a sort of availability heuristic: It is often assumed that the successful way something has been done, and is done today, is the *only* way to do it, or the only way new entrants can do it and be competitive.

But of course that's never actually true. Facebook uses a very different method and different data than does Google to match advertisers and users — and yet it entered the online advertising/matchmaking market and became enormously successful without adopting Google's model (and without obtaining Google's (or anyone else's) existing data). Uber entered the transportation network market with a business model that didn't require capital outlay on a large fleet of vehicles. Digital cameras made film irrelevant and didn't need to rely on suppliers of film to enter. Fax machines went through a series of improvements — until email and cloud services completely replaced them.

The examples are endless. But they are key to understanding the non-essentiality of data: For some entrants — those adopting incumbents' business models, minimizing their own innovations, or even piggy-backing on incumbents — it *seems* indispensable. And they may find a willing ear at some antitrust agencies. But innovation has never required implementation of the same business model as incumbents, and especially not access to the particular, proprietary inputs incumbents have created.

And, as noted above, new entrants may face even more welcoming environments *because of* incumbents. Consider how much Google contributed to the creation of the online advertising industry and consumer acceptance of advertising-financed websites, and web page and app developers' expectations that advertising would need to be accommodated. Whatever the data used to deliver it, there can be no doubt that a new provider of online advertising today faces an environment in which its product is known, and even invited. That wasn't always true in the past.

2. *Data as a simulacrum and information asymmetry*

As discussed above,[94] it can be hard for users to know the value of their data ex ante, but it can be hard for platforms or other intermediaries to know underlying information *at all*. On the one hand, platforms incur significant costs in order to obtain basic information that users know about themselves; on the other, they know the value of that data only if and when they apply high-quality (and expensive) processing to it.

Data is a simulacrum: Platforms are locked in an ever-evolving battle to identify, collect, process, interpret, and use data in order to figure out user preferences or to predict consumer behavior. There is no silver bullet amount and kind of data to accomplish this. Every data set represents some collection of pieces of information that are an effort to guess at the user's mind, as is every aggregate set of data about a large group of people (for which errors are more likely to cancel out, but for which the representational value of the data is less likely to be very accurate or useful because it encompasses significant noise relative to signal). Big data sets do, however, allow for pattern recognition (i.e., in order to plot out likely traffic issues, mapping apps don't really need to know where a given user is, *per se*, but only whether a large mass of drivers are likely to be in the same place at the same time…).

This is complicated by multi-homing and product differentiation, as well as by tools users use to hide their data. Asymmetry re value means that despite our concerns about big data, arguably users are *under*-producing data, not overproducing it, and/or they are spreading their data too thin.

Which is why it's also key to keep markets in mind: The story is different for advertisers than it is for users. But then, so are the ramifications of data. Advertisers want targeting, of course, but advertisers have enormous amounts of information on their own. They decide what keywords to bid on, for example, based at least in part on information they already have and would bring to any platform.

This information asymmetry point is important. It's commonly said or assumed that platforms have much more information than users, and can use it to their advantage. The same is said for incumbents versus new entrants. But is it really true? Whatever Google knows about a user, if a new entrant were to ask the right questions, or buy the right data, it could easily know more. Which is a key reason why Amazon is such a threat to Google: It knows what users shop for, what they buy, at what price, etc. That's of enormous value. Whatever Google knows about how often users search for terms like "Stigler entry barriers," it pales in importance compared to what Amazon knows about what books I buy, or what Facebook knows about who my friends are and how I interact with them. And tomorrow — who knows what will be most relevant? Even today, if big data were so good at predicting users' behavior, then tech firms would be very good at, for example, predicting what future products and R&D projects will be most profitable. They are not, of course.

---

[94] *See supra*, Section II.C.1.

It is also important to account for incumbent platforms as facilitators of new entry. Without generalizing, there are some obvious examples, like Amazon's Web Services, that reduce the cost to smaller entrants of obtaining scale in backbone technology, or Google's services making it easier for users to find new entrants that otherwise have to overcome the problem of anonymity.

In fact, to the extent that lack of information is a real entry barrier, the role of incumbent intermediaries in reducing search and other information costs (like providing reputation markets, etc.) can actually operate to *overcome* entry barriers. It is crucial in assessing the extent to which data might operate as a barrier to also assess the mechanisms it enables for reducing barriers, even for a company's direct competitors.

As suggested by the U.S. *Microsoft* court, however, the relevant question concerns not the "initial acquisition of monopoly power"; it concerns a company's "efforts to maintain this position through means other than competition on the merits."[95] It is, presumably, possible for a company to deploy, use, or limit access to data in order to impede competition at the platform level, rather to compete — but this *possibility* doesn't convert data into an entry barrier *per se*.[96]

B. The investment costs of innovation

However important incumbents' data may be, it is never as important as many make it out to be at the margin. Consumers want accurate video recommendations, for example, but they also want a variety of content, an attractive and functional user interface, high-quality streaming, etc. Even in something like online search, users care about interfaces, mobile-specific (including voice) input, attractive results pages, limited clicking, etc. These elements of design and of algorithmic processing are arguably decisively important, while the relative "quality" and amount of data may be significantly less important by comparison.

Can a new entrant make it without *some* of that data to begin with, though? An attractive interface is hardly going to drive users to a new search platform if it doesn't also offer search results that are roughly comparable. But a new entrant need not have the best search results in order to get off the ground. Even without the same data as the incumbent, the new rival can differentiate its product, offer other services designed to attract users to the platform and then obtain data (the old fashioned way), offer an alternative not dependent on data, find ways to make better use of more limited amounts or different kinds of data, or, finally, purchase the relevant data. Moreover, data are not

[95] *U.S. v. Microsoft*, 253 F.3d at 56.

[96] It should also be noted that examples of conduct that might amount to the erection of unjustified barriers to competition are few and far between, and may not even be identifiable in actual markets. See, for example, Rubinfeld & Gal, *supra* note 93, which attempts to canvass possible "behavioral" data barriers, but essentially identifies only a limitation imposed on a national census form as a constraint employed without business justification. *Id.* at 363.

monolithic. They vary along multiple dimensions, any of which can be more significant than the others. Even incumbents' business models were built using data with different characteristics:

> [T]he quality and value of data are affected not only by their volume, but also by their velocity, variety, and veracity. As a result, once one characteristic of big data exhibits high entry barriers, another characteristic might grow in importance in order to overcome the competitive advantages created by the first. For example, where past data are not easily available (therefore reducing the volume or temporal variety of data available), veracity or variety might gain importance in order to create a higher level of predictive certainty based on a smaller data panel.[97]

And recall that every incumbent had to face the same constraints itself.

There is a longstanding debate over whether an entry barrier is properly conceived of as "some source of disadvantage to potential entrants as compared with established firms" (the Joe Bain version[98]), or "a cost of producing (at some or every rate of output) which must be borne by a firm which seeks to enter an industry but is not borne by firms already in the industry" (the George Stigler version[99]).

The former rejects sunk costs as relevant and looks only at relative costs *ex post* — meaning every advantage enjoyed by incumbents can be an entry barrier, regardless of what it cost to obtain in the first place. The latter says that costs incurred similarly by both incumbents and entrants impose the same constraints on each (and are thus not appropriately conceived of as entry barriers), regardless that one has already incurred the costs and the other has yet to do so.

In part, the claimed justification for the *ex post* approach is the importance of entry to police incumbents in the short term. At the time when a new rival would enter the market to compete, according to this approach, it hardly matters that the costs it faces are the same as those that confronted the incumbent: if they would deter its entry, they constitute a competitively relevant barrier. On the other hand, the sunk cost approach recognizes that giving a regulatory leg up to new entrants that have yet to incur the costs that the incumbent has already incurred would reduce the incentive of the incumbent to incur those costs in the first place. This is socially costly, as it thus reduces the incumbent firm's incentive to invest, increase productivity, etc.

The problem with the *ex post* approach (at least as it is commonly applied) is that it does not account for opportunity costs. New entrants may face the cost of, say, *investing* in a new factory, but incumbents face the equivalent economic cost of not *selling* their existing factory. The two costs are

---

[97] *Id.* at 370.

[98] JOE S. BAIN, BARRIERS TO NEW COMPETITION: THEIR CHARACTER AND CONSEQUENCES IN MANUFACTURING INDUSTRIES (1956).

[99] GEORGE STIGLER, THE ORGANIZATION OF INDUSTRY 67 (1968).

equivalent, and thus the new entrant does not face a "disadvantage" simply because its cost is in the form of a required outlay and the incumbent's cost is in the form of foregone revenue.

More important perhaps, it is imperative to consider what the "real" sources of barriers to entry are, and whether they first provide important benefits that should not lightly be taken away or discounted. In many cases (and leaving aside government-created barriers), they come down to information costs. Why does reputation matter? Because it conveys information to consumers. Why does longevity in a market matter? Same reason. Scale economies are just a manifestation of the same thing in markets with declining marginal costs; they are indicia of established quality. A large installed base of users providing data is the same thing, as well.

That the higher capital cost facing a new entrant is the result of the success of the incumbent rather than a "disadvantage" protecting a sub-par incumbent doesn't make new entry any less costly, but it does suggest why being quick to use antitrust or other regulatory measures to overcome such barriers is a problem: It means less such capital will be created in the first place. We'll get less longevity, less investment to build reputation, and smaller scale, if those things are going to be used as a pretext to apply regulatory levers to favor new entrants. While the result of greasing the skids for new entry *might* be more competition, it's at least as likely that it would simply transfer more of the information costs back onto consumers, at a rate that more than offsets whatever gains there may be from having more firms — especially in a market that tends, because of its fundamental economics, toward a single firm or small number of firms.[100]

And it must be noted that making it easier for new entrants to replicate what incumbents are doing is likely to result in more replicas and less long-term innovation. Treating data as an essential facility and mandating its sharing with rivals in order to facilitate their ability to compete is akin to removing IP protection: It may lower the costs of existing products, but it also lowers the incentive to create new and different competing products. The perceived need to amass large amounts of data (if effectively employed, and if they do deter imitators) may raise the costs of entry, but it also creates an incentive for new entrants to innovate around the costs and to differentiate their products. As one antitrust authority has noted:

> Antitrust issues generally do not arise when firms collect more data and antitrust does not usually impose on firms an obligation to share data that they have collected and developed. To do so may very well chill innovation, which is the very behaviour that antitrust is designed to protect.[101]

---

[100] *See generally* Giuseppe Colangelo & Mariateresa Maggiolino, *Data Accumulation and the Privacy-Antitrust Interface: Insights from the Facebook case for the EU and the U.S.*, TTLF Working Paper No. 31 (2018), *available at* https://ssrn.com/abstract=3125490.

[101] Canadian Competition Bureau, Big data and Innovation: Implications for Competition Policy in Canada, Discussion Paper (2017), *available at* http://www.competitionbureau.gc.ca/eic/site/cb-bc.nsf/eng/04304.html.

If all we want is multiple exact copies of existing firms, with minimal further innovation, then viewing data as an entry barrier and treating it as a common good or essential facility may be fine. But if not, it makes no sense to do so.

IV. The problems with common regulatory interventions around data

The rapid proliferation of data-rich markets has also given rise to numerous other concerns. From scholars who believe that online markets require data portability mandates to be made more competitive, to others who argue that big data will lead to harmful price discrimination and algorithmic collusion, there is no shortage of calls for policymakers to intervene. And yet, these claims often fall at the first hurdle of regulatory intervention: in most instances, scholars have failed to make a solid case that data-reliant markets are not achieving efficient outcomes. The ills perceived by these scholars are often greatly inflated, as is their confidence in the ability of regulators to ameliorate matters.

A. The pitfalls of mandating data portability

The role of data portability (notably when it is mandated by regulatory or competition authorities) in promoting competition between online platforms has also been the source of much confusion among scholars. Data portability can broadly be understood as the right for platform users to remove or copy data relating to them from one platform and convey it to another.[102] The underlying intuition is that by mandating data portability policymakers may reduce switching costs, thus making it easier for new firms to enter a market by attracting its current consumers.

This section focuses on two important aspects of mandated data portability. First, data portability does not necessarily increase competition, and it may have a detrimental effect on innovation. Second, even if it did mechanically increase competition, mandated data portability may give rise to data security problems that could potentially outweigh competition-related benefits.

1. Mandated data portability does not necessarily increase competition

Enforcers may be tempted to mandate data portability because they believe it will decrease switching costs and thus prevent consumers from being locked-in to a single online platform. Lock-in occurs when various costs make it prohibitive for an "installed base" of consumers to switch to a rival's product.[103] These costs primarily stem from network effects, contractual provisions, and path dependence (notably learning costs).[104] Although lock-in is not a standalone theory of harm under

[102] This is roughly the definition that the EU adopted in the GDPR (art. 20).

[103] *See generally* Joseph Farrell & Paul Klemperer, *Coordination and Lock-in: Competition with Switching Costs and Network Effects*, 3 HANDBOOK OF INDUSTRIAL ORGANIZATION 1967 (Mark Armstrong & Robert H. Porter, eds 2007).

[104] *See, e.g.*, SHAPIRO & VARIAN, *supra* note 7, at 117.

US antitrust law, it can have a significant bearing on the outcome of cases. This was notably evidenced in the *Microsoft* antitrust proceedings.[105]

The inability to easily move data across platforms has sometimes been identified as a source of consumer lock-in. For instance, numerous commentators have argued that the inclusion of a right to data portability in the European Union's General Data Protection Regulation ("GDPR") could promote competition between online platforms.[106] Although there is a theoretical link between data portability and switching costs, it is unclear how much effect, if any, these data-related switching costs exert on consumers' actual behavior. Though consumers may find it less costly to switch platforms if they can carry their data with them (rather than reencode it in the new platform), they likely take dozens of other parameters into account and there is no clear sense that data-related switching costs have a significant bearing on these calculations.

For a start, these costs might be irrelevant depending on a rival's unique characteristics, particularly what it would do with existing data and how well it would process it. For example, critics of Facebook often refer to Instagram as a threat to Facebook, which Facebook neutralized by purchasing the company. Had this rivalry emerged, how important would it have been for users to be able to port data from Facebook? While some Facebook data — specifically relating to a user's network of friends, uploaded photos, and his or her interactions with those photos — would have been relevant to users moving to Instagram, other data — like interactions with news items, event check-ins, etc. — would not obviously be of much use. In addition, even access to old photos might be of limited value to users, most of whom use social media for ongoing engagement rather than as a repository of previous interactions.

At the same time, mandating data portability does not ensure that new entrants can or will offer consumers higher value uses of the ported data. Yet, given free access to data, new entrants might expend resources to try to do so, leading to (at least) two effects. First, new entrants might make better use of the same data by being able to devote more of their resources to data processing rather than collection. Second, new entrants that would have devoted more resources to improving users' experience with *new* data or under a different business model may devote some of those resources to utilizing the ported data, thus reducing the average quality of their products. While obviously the first effect could lead to useful competition while the second is unlikely to do so, in *both* cases, the overall effect is a likely reduction in innovative competition, as new entrants attempt to largely

[105] The cost of switching from Microsoft's products played an important role in defining the relevant market and in establishing Microsoft's market power. *See U.S. v. Microsoft*, 253 F.3d 34 (2001).

[106] *See* Wolfgang Kerber, *Digital markets, data, and privacy: competition law, consumer law and data protection*, 11 J.OF INTELL. PROP. L. & PRAC. 861 (2016). *See also* Aysem Diker Vanberg & Mehmet Bilal Ünver, *The right to data portability in the GDPR and EU competition law: odd couple or dynamic duo?*, 8 EUR. J.L OF L. & TECH. 2 (2017). *See also,* Orla Lynskey, *Aligning data protection rights with competition law remedies? The GDPR right to data portability*, 6 EUR. L.J. 793 (2017).

replicate incumbents. Whether the net effect is positive or not is highly fact dependent, but this cost — a reduction in product differentiation — must be considered and may be significant.

Also noteworthy, users often multi-home across platforms, which suggests two things. First, data portability is not a critical consideration if users are already willing to engage on multiple, differentiated networks without it, and, second, data portability mandates would affect only the subset of consumers who are currently unwilling to move across platforms.

But even if consumers' inability to move data across platforms *did* create switching costs and lock-in, this would have ambiguous welfare implications.[107] Although switching costs may lead to higher prices and fewer choices *once consumers are locked-in*, these effects are often counterbalanced by lower prices overall.[108] The lure of *ex post* profits may induce firms to compete aggressively in order to acquire valuable consumers.[109] Moreover, the ability to lock in consumers may play a crucial role in launching new products, especially in digital markets.[110]

For these reasons, antitrust authorities must approach potential consumer lock-in with caution and intervene only if there is clear evidence that lock-in will lead to higher prices over the whole "lock-in cycle," rather than higher prices at a given point in time. Moreover, authorities should question whether there is competition between platforms to acquire new customers, as this will generally constrain their *ex post* behavior.

Likewise, lock-in and switching costs are often the by-product of important product design choices. Agreeing upon a single, market-wide standard may delay the introduction of new products and make existing ones less reliable.[111] Moreover, learning costs will often reflect the rich set of features that a product offers its users rather than a naked attempt to lock them in. The upshot is that policy aimed at undermining switching costs can sometimes be highly counterproductive.

---

[107] *See, e.g.*, P. BELLEFLAMME & M. PEITZ, INDUSTRIAL ORGANIZATION: MARKETS AND STRATEGIES 167 (Cambridge University Press. 2010).

[108] *See* SHAPIRO & VARIAN, *supra* note 7, at 133. (The authors stress that lock-in must always be addressed by looking at the entire "lock-in cycle".).

[109] This notably occurs with exclusivity arrangements, which cause firms to compete aggressively "for the contract". *See* Benjamin Klein & Kevin M. Murphy, *Exclusive Dealing Intensifies Competition for Distribution*, 75 ANTITRUST L. J. 433 (2008).

[110] *See* SHAPIRO & VARIAN, *supra* note 7, at 142 ("*Companies unwilling or unable to offer concessions to gain locked-in consumers cannot prevail in a competitive battle*").

[111] *See, e.g.*, Michael L. Katz & Carl Shapiro, *Product Compatibility Choice in a Market with Technological Progress*, 38 OXFORD ECON. PAPERS,146, 147 (1986). ("*Typically, achieving technical compatibility will be costly*").

2. *Mandated data portability can decrease data security*

A final area of concern is that mandated data portability may have negative ramifications as far as data security and privacy are concerned.[112] The key requirement of data portability mandates is that individuals should be able to access data on one platform, in a format that they can easily read by other platforms.[113] This is not without issues.

By forcing firms to be in a position to provide users with personal data in an easily accessible format, mandated data portability creates a risk of identity theft and of personal data leaks.[114] For a start, mandating that firms should supply data without hindrance raises the risk that the relevant data could be requested by an imposter with the aim of identity theft.[115] This is because data portability mandates, notably the GDPR, often guarantee that subjects should be able to retrieve their personal data without hindrance, which gives firms the perverse incentive not to closely monitor the person to whom they are giving the requested data. To make matters worse, because firms must provide all data about a subject in a structured and readily accessible manner, data portability increases the risks associated with data leaks and hacking. Rather than acquiring encrypted, anonymous and/or piecemeal information about data subjects, these mandates guarantee that unscrupulous individuals may have access to a far more comprehensive set of information.

Finally, mandated data portability increases the risks that individuals will provide their information to less-than-reputable firms. At the very least, it is somewhat paradoxical that privacy regulations, such as the GDPR, that are partly premised on consumers' lack of information in dealing with online platforms, should also hand these same individuals the key to their own undoing, as they are now free to rapidly move all of their information to firms they may sometimes know very little about. In other words, if we think that consumers' lack of information and the risk that they will make poor decisions is significant (and there is every reason to be skeptical about such claims), it makes little sense to facilitate their ability to hand over all of their information to different firms.

B. Price discrimination is ambiguous, and often procompetitive

Much of the concern about big data seems to be that it will facilitate price discrimination — or, to use a less charged term, price differentiation. But differential pricing is frequently a good thing: tailoring prices more closely to an individual's willingness to pay means that more transactions will take place than otherwise would have, with more consumers having access to a product and more

---

[112] *See* Peter Swire & Yianni Lagos, *Why the Right to Data Portability Likely Reduces Consumer Welfare: Antitrust and Privacy Critique*, 72 MD. L. REV. 335 (2012).

[113] GDPR, art. 20.

[114] *See* Swire & Lagos, *supra* note 102, at 373-75.

[115] *See GDPR in Context: Data Portability*, Matheson, May. 24, 2017, *available at* https://www.matheson.com/images/uploads/documents/GDPR_in_Context_-_Data_Portability.pdf.

consumers' desires being satisfied.[116] And price differentiation generally means that the poor pay less, not more, for the same goods. That's precisely why it was so extraordinary when, in late 2012, a Wall Street Journal study found that Staples might be charging higher online prices to consumers who lived farther away from a Staples brick-and-mortar store — which would generally mean higher prices for poorer consumers (since this watershed revelation, many other online retailers have been shown to adopt similar practices):[117]

> What economists call price discrimination — when companies offer different prices to different people based on their perceived willingness to pay — is commonplace and can be beneficial.... But using geography as a pricing tool can also reinforce patterns that e-commerce had promised to erase: prices that are higher in areas with less [brick-and-mortar] competition, including rural or poor areas. It diminishes the Internet's role as an equalizer.[118]

In the Journal's examination of Staples' online pricing, the weighted average income among zip codes that mostly received discount prices was roughly $59,900, based on Internal Revenue Service data. Zip codes that saw generally high prices had a lower weighted average income, $48,700.52. For an economist, this highly anecdotal study would be merely the beginning of a larger inquiry. In particular, an economist would ask the obvious question: If customers are already shopping online, why would their proximity to a physical store be decisive? If they're willing to make purchases online, why would they not comparison shop to check prices at Staples' online competitors? In other words, could consumers not easily circumvent such price differentiation? If so, why would we expect price differentiation to be effective? And, if not, the problem isn't one of price discrimination per se, but of abuse of market power.

But there is a more complicated point that deserves greater discussion among specialists in the economics of information. The theoretical basis for concern about price differentiation appears to be that advertisers and merchants will use big data to take advantage of consumers through a greater information asymmetry, charging each consumer as much as he or she appears willing to bear. This is predicated on the idea that Internet users undervalue their data and overvalue the benefits of the content to which they receive access.[119]

---

[116] *See, e.g.*, Arthur C. Pigou, *Discriminating Monopoly*, in THE ECONOMICS OF WELFARE (1938). The conclusion is even more true in the presence of high fixed costs. *See, e.g.*, Hal Varian, *Differential Pricing and Efficiency*, 1 FIRST MONDAY (Aug. 5, 1996), *available at* http://firstmonday.org/ojs/index.php/fm/article/view/473/829.

[117] *See, e.g.*, Neil Howe, A Special Price Just for You, FORBES (Nov. 17, 2017), *available at* https://www.forbes.com/sites/neilhowe/2017/11/17/a-special-price-just-for-you/#3af79ae390b3.

[118] Jennifer Valentino-Devries, et al., *Websites Vary Prices, Deals Based on Users' Information*, WALL ST. J. (Dec. 24, 2012), *available at* http://online.wsj.com/articles/SB10001424127887323777204578189391813881534.

[119] *See, e.g.*, Nathan Newman, *The Costs of Lost Privacy: Consumer Harm and Rising Economic Inequality in the Age of Google*, 40 WM MITCHELL L. REV. 849 (2014).

This information asymmetry story has a superficial plausibility: consumers in real-world markets never have the perfect information that certain models presume. But businesses also fail to have perfect information, and this fact is glossed over by critics.[120] The market process is useful because it encourages participants to gain information efficiently as the costs and benefits of decisions are borne by each decisionmaker herself.[121] Further, whatever information asymmetry persists is not likely in the favor of even most informed business. No one knows more about a consumer's preferences than the consumer herself; all the tracking in the world will, at best, allow online advertising networks to play catch-up. It is surely overstated to claim that businesses in possession of information, even that enabled by big data, will be at an informational advantage compared to the consumers about whom they are supposedly informed.

Another plausible sounding argument about the harm of price differentiation is that because major advertising networks and retailers are able to collect a great deal of data about their users for analysis, businesses could segment consumer groups based on certain characteristics and offer different deals. The resulting price differentiation could lead to many consumers paying more than they would have in the absence of big data. Therefore, it has been argued, big data facilitates price differentiation, and that harms consumer welfare.

This argument misses a large part of the story, though. The flip side is that price differentiation could have benefits to those who receive lower prices from the scheme than they would have in the absence of big data. If this group is as big as or bigger than the group that pays higher prices, or if the price reductions are larger in magnitude than the price increases, then it is difficult to say with any certainty that the practice leads to a reduction in consumer welfare, even if this can be divorced from total welfare.[122]

Further, this analysis fails to consider the dynamic efficiencies of price differentiation, as suggested above. In a static model of third-degree price differentiation, some buyers receive lower prices (and purchase higher quantities), while other buyers receive higher prices (and purchase lower quantities). Thus, the net impact of price differentiation on output is ambiguous.[123] But, in a dynamic model, price differentiation may often be procompetitive because the profits provide incentives for entry and allow for additional investments in innovation and increasing product variety, expanding retail

---

[120] *See supra* Section II.1.

[121] *See* F.A. Hayek, *The Use of Knowledge in Society*, 35 AM. ECON. REV. 519 (1945).

[122] Ariel Ezrachi & Maurice Stucke assert that *increased* prices, with occasional downward deviations (rather than *lower* prices with occasional upward deviations) will be the norm. *See* ARIEL EZRACHI & MAURICE E. STUCKE, VIRTUAL COMPETITION 106-30 (2016). But their aerguments rest on behavioral assumptions regarding consumer responses to different pricing schemes that are weak, at best. *See* Todd J. Zywicki, Geoffrey A. Manne & Kristian Stout, *Behavioral Economics Goes to Court: The Fundamental Flaws in the Behavioral Law & Economics Arguments Against No-Surcharge Laws*, 82 MO. L. REV. 769 (2017).

[123] *See, e.g.*, Joshua D. Wright, *Missed Opportunities in Independent Ink*, CATO SUP. CT. REV., at 348 (2005-2006), available at http://object.cato.org/sites/cato.org/files/serials/files/supreme_court_review/2006/9/wright.pdf.

outlets, or research and development.[124] Price differentiation may allow for increased competition to all consumers, including previously unreached poorer consumers — a procompetitive outcome.[125] And contrary to the received wisdom,[126] economists have more recently noticed that price differentiation is present even in competitive markets.[127]

While critics of big data have focused on the possible negative effects of price differentiation to one subset of consumers, they usually ignore the positive effects of businesses being able to expand output by serving previously underserved consumers. It is unlikely that a business relying on metrics would want to serve only those who can pay more by charging them a lower price, while charging those who cannot afford it a larger one. If anything, price differentiation would likely promote egalitarian outcomes by allowing companies to offer lower prices to poorer segments of the population, which can be identified by data collection and analysis.

In an error cost framework, courts and antitrust regulators should refrain from declaring conduct anticompetitive unless the likelihood of procompetitive outcomes is extremely low.[128] It may be difficult for the enforcers to distinguish positive price differentiation from negative price differentiation. Here, it seems unlikely that the price differentiation "facilitated" by big data is anti-consumer welfare. Big data helps drive the Internet ecosystem, and many businesses that previously did not compete with each other do now compete in the same Internet markets in an effort to offer the best deal to consumers through targeted advertising. It seems just as likely, if not more so, that Big data is increasing consumer welfare by helping businesses find consumers interested in their products and serving up more relevant advertisements to those consumers — thus increasing the amount of positive sum transactions overall.

C. Algorithmic collusion fears are overblown

A final area where big data-related fearmongering has been ever-present is the question of algorithmic collusion. The basic intuition of these criticisms is that the use of big data might facilitate collusion in online markets, because rivals can immediately detect defections from a collusive outcome, thanks to automated algorithmic price surveillance. In turn, this deters co-conspirators from cheating on their cartel and is thus said to encourage collusion. As Ariel Ezrachi and Maurice Stucke assert:

> [T]wo technological advancements can amplify tacit collusion, creating a new level of stability and scope. The first advancement involves computers' ability to process high

[124] *Id.* at 530.

[125] *Id.*

[126] *See* William M. Landes & Richard A. Posner, MARKET POWER IN ANTITRUST CASES, 94 HARV. L. REV. 937, 977 (1981).

[127] See, e.g., 70 ANTITRUST L.J. 593 (2003) (symposium articles discussing competitive price discrimination)

[128] Frank H. Easterbrook, *The Limits of Antitrust*, 63 TEX. L. REV. 1 (1984). The error cost model is well-accepted in the antitrust law and economics literature. *See, e.g.,* RICHARD A. POSNER, ANTITRUST LAW, at ix (2nd ed. 2001).

> volumes of data in real time to achieve a God-like view of the marketplace. The second advancement concerns the increasing sophistication of algorithms as they engage in autonomous decision making and learning through experience — that is, the use of artificial intelligence. These two technological advances enable a wider, more detailed view of the market, a faster reaction time in response to competitive initiatives, and dynamic strategies achieved by "learning by doing." Thus they can expand tacit collusion beyond price, beyond oligopolistic markets, and beyond easy detection.[129]

But Ezrachi and Stucke's reading of the tacit collusion literature is myopic, to say the least. For a start, the authors' underlying premise is particularly dubious. Indeed, it is far from clear that the rise of big data will necessarily lead to increasingly transparent markets. As we highlighted in the previous section (see Section IV.B., *supra*), one of the basic competitive implications of big data is that companies are more likely to price discriminate between consumers. But price discrimination works only if a seller can prevent higher-price consumers from purchasing at lower-price consumer prices. This necessarily entails some degree of both complexity and obscurity in pricing — particularly as the same algorithmic technology that permits price discrimination by sellers also permits price discovery by users.

As a result, markets are just as likely to become *less* transparent for consumers, physical rivals, and algorithms. For instance, firms may want to offer discount codes to a select group of consumers. Sellers have every reason to keep these discounts secret, not least because less-price-elastic consumers would likely balk if they knew they were paying a premium compared to their peers. In order to be effective, this type of price-cutting would have to escape not only human notice, but also the notice of sophisticated price monitoring algorithms. To make matters worse, it has been shown that big data may sometimes lead firms to adopt price dispersion strategies in equilibrium: i.e., partly randomized prices.[130] By definition, a randomized pricing strategy masks the price that a company charges each type of consumer for given a product.

The upshot is that Ezrachi and Stucke's first premise ignores current trends in online retail and other services by assuming that online transparency has and will increase when, in fact, the opposite is just as likely.

Moreover, even if it were true that price transparency has increased with the advent of online commerce, big data, and algorithmic pricing, there is no reason to believe that this will necessarily lead to more collusive outcomes. A basic tenet of collusion economics is that "private" price

---

[129] Maurice E Stucke & Ariel Ezrachi, *How pricing bots could form cartels and make things more expensive*, HARV. BUS. REV. (Oct. 27, 2016), *available at* https://hbr.org/2016/10/how-pricing-bots-could-form-cartels-and-make-things-more-expensive. *See also* EZRACHI & STUCKE, *supra* note 122, at 56-70.

[130] *See, e.g.,* Eric K Clemons, Il-Horn Hann & Lorin M Hitt, *Price dispersion and differentiation in online travel: An empirical investigation*, 48 MANAGEMENT SCIENCE 534-549 (2002). *See also*, Rosa-Branca Esteves & Joana Resende, *Competitive targeted advertising with price discrimination*, 35 MARKETING SCIENCE 576-587(2016).

transparency tends to favor collusion, while "public" privacy tends to undermine it.[131] In other words, if a market becomes more transparent for sellers alone, then it mainly allows them to detect potential defections from a collusive outcome. Conversely, increasing market transparency for buyers and sellers alike generally has the opposite effect because buyers are in a better position to shop around. Firms that cheat a cartel are thus more likely to earn a one-shot payoff that outweighs the losses that might stem from rivals' retaliation. This possibility is particularly likely in online markets characterized by the attributes we see today: where sales are heavily grouped in discreet periods of the year[132] (notably Black Friday and the holiday season), intermediaries offering price comparisons and product reviews abound, and the costs to consumers of finding and switching between sellers is trivial.

A final point is that attempts to collude online will likely be fought off by online platform operators like Amazon. Imagine a group of retailers that establish a cartel involving sales on the Amazon Marketplace. Such an arrangement would cut against Amazon's interests because it would lead to double marginalization and thus lost sales and lost revenue relative to monopoly pricing (assuming Amazon even has monopoly power).[133] Yet it is unlikely that Amazon would be unable to police such conduct on its platform. Unlike online retailers, who can only directly observe their own sales (or otherwise purchase non-confidential, aggregated sales information from Amazon), Amazon can monitor every transaction that takes place on its platform. This puts it in a unique position to weed out potential collusion. Amazon could exclude colluding retailers from its platform, or simply vertically integrate (or even threaten to do so) into those segments where it has reason to believe that retailers have been colluding. The risk of either outcome would significantly undermine the likelihood of algorithmic collusion.

The upshot is that assertions of algorithmic collusion seem untethered from the underlying reality of online markets.

---

[131] *See* MASSIMO MOTTA, COMPETITION POLICY: THEORY AND PRACTICE 156 (2004).

[132] *See id.* at 145.

[133] *See, e.g.*, Lester G. Telser, *Why should manufacturers want fair trade?*, 3 J. L. & ECON. 86 (1960). ("The manufacturers' interests seem to be best served when distributors resell their products under such competitive conditions as may exist at the level of distribution and at the lowest prices resulting from that competition. If manufacturers set a floor to the resale price then they also set a ceiling to their sales and thus apparently support a policy that runs counter to their own self-interest.").