THE
CRITERION
JOURNAL ON INNOVATION

VOL. 4 2019

Antitrust After Big Data

*John M. Yun**

The phrase, "big data," likely first entered the economic lexicon in 2000, when economist Francis Diebold used the term in his article, *"Big Data" Dynamic Factor Models for Macroeconomic Measurement and Forecasting*;[1] although Diebold acknowledges that "credit for the term Big Data must be shared."[2] Regardless of its origin, most definitions of big data are similar to the one provided by the Organisation for Co-operation and Development (OECD): "Big Data is commonly understood as the use of large scale computing power and technologically advanced software in order to collect, process and analyse data characterised by a large volume, velocity, variety and value."[3]

Since those early days, according to Matt Turck, interest in big data reached its apex between 2011 and 2014—at least in the world of venture capitalists (VC).[4] Since that time, VC interest has transitioned to the analytical tools needed to assess and unlock value from big data—namely, through

* Associate Professor of Law and Director of Economic Education, Global Antitrust Institute, Antonin Scalia Law School, George Mason University. Email: jyun9@gmu.edu. This article is published with the kind permission of the *Baltic Yearbook of International Law*. No outside funding was received or relied upon for this paper. I thank Ken Heyer, Seth Sacher, and Joshua Wright for tremendously helpful comments.

[1] Francis X. Diebold, *"Big Data' Dynamic Factor Models for Macroeconomic Measurement and Forecasting*, *in* ADVANCES IN ECONOMICS AND ECONOMETRICS: THEORY AND APPLICATIONS, EIGHTH WORLD CONGRESS 115 (Mathias Dewatripoint, Lars Peter Hansen & Stephen J. Turnovsky eds., Cambridge Univ. Press 2003).

[2] Francis X. Diebold, On the Origin(s) and Development of "Big Data": The Phenomenon, the Term, and the Discipline (Feb. 13, 2019) (unpublished manuscript) (manuscript at 2), https://www.sas.upenn.edu/~fdiebold/papers/paper112/Diebold_Big_Data.pdf.

[3] Big Data: Bringing Competition Policy to the Digital Era—Background Note by the Secretariat—29–30 November 2016, at 2, DAF/COMP/M(2016) 2/ANN4 final (Apr. 26, 2017).

[4] Matt Turck, *Is Big Data Still a Thing? (The 2016 Big Data Landscape)*, MATT TURCK BLOG (Feb. 1, 2016), http://mattturck.com/big-data-landscape ("[I]nterest in the concept of 'Big Data' reached fever pitch sometime between 2011 and 2014. This was the period when, at least in the press and on industry panels, Big Data was the new 'black', 'gold' or 'oil'. However, at least in my conversations with people in the industry, there's an increasing sense of having reached some kind of plateau. 2015 was probably the year when the cool kids in the data world (to the extent there is such a thing) moved on to obsessing over AI and its many related concepts and flavors: machine intelligence, deep learning, etc.").

machine learning and artificial intelligence. This shift prompted Turck to ask in 2016: "Is Big Data still a 'thing'?"[5]

While interest in big data had begun to wane among VCs, public interest in, and concern over, big data as a potential *competition issue* perhaps hit its apex in mid-2017 when *The Economist* contended that, in terms of its role in today's information economy, big data plays a role similar to that historically played by oil in powering the pre-information economy. Put succinctly, *The Economist* declared that big data was "the new oil."[6] Yet, well before that point, research into whether big data truly represents an antitrust problem had already began with a number of foundational questions, including: What exactly is big data? Do only big firms have it? Do you need it to enter and compete successfully in a digital economy? Is it a barrier to entry? Can it be a competition problem? And can it generate efficiencies, enhance consumer welfare, and perhaps even promote competition? Some early conclusions, which I detail in Part I, tended to suggest that competition agencies should use a strong presumption of harm when a firm with market power possesses big data.

Since those initial inquiries, there has been a flurry of research and commentary in a relatively short period of time—primarily, over a five-year period from 2013 to 2018. From this research, what has emerged are some key findings and insights that indicate we have moved into a new "era" in how we consider big data in the context of antitrust. This is not to suggest that big data is no longer a consideration in antitrust review; it is. But rather, there is growing recognition that big data in and of itself is not a competition problem. In this article, after discussing the primary antitrust concerns in Part I, I detail the three findings that I believe support this new era conclusion.

First, as detailed in Part II, we have a better understanding of the role that big data plays in the production and innovation process. Big data is one input, among many, that a firm uses to maximize profits, including through innovation. The success of a new product, new feature, or other types of innovation, whether from an entrant or an incumbent, relies on a mix of factors that the firm combines to help it take advantage of concurrent market opportunities—for example, shifts in consumer preferences or changes in rivals' behavior. Indeed, even with potential network effects and

[5] *Id.* Certainly, a comparison of the frequency that users searched for the terms "big data" and "machine learning" on Google Search supports Turck's conjecture, which indicates that searches for big data peaked around 2014, and searches for machine learning has outstripped that of big data after 2016. *See Google Trends*, Google, https://trends.google.com/trends/?geo=US.

[6] *The World's Most Valuable Resource Is No Longer Oil, But Data*, Economist, May 6, 2017, https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data. The idea of data as the new oil was likely coined by U.K. mathematician Clive Humby. *See* Michael Palmer, *Data Is the New Oil*, ANA Blogs (Nov. 3, 2006), https://ana.blogs.com/maestros/2006/11/data_is_the_new.html.

associated feedback effects from the use of big data,[7] the fundamental reality is that data lacks worth unless it is combined with ingenuity, skill, and market conditions to unlock its value. Nearly all research on big data has affirmed this central point.

Second, in competition analysis, there is often confusion regarding what is and what is not a barrier to entry—as evidenced by a lack of consensus regarding its definition—which I explore in detail in Part III. While big data frequently represents an important asset and resource that firms use, that fact, by itself, does not demonstrate that it constitutes a barrier to entry. As Darren Tucker and Hill Wellford state, "[t]he fact that some established online firms collect a large volume of data from their customers or other sources does not mean that new entrants must have the same quantity or type of data in order to enter and compete effectively. . . . [L]ack of asset equivalence should not be a sufficient basis to define a barrier to entry."[8] There is also the fact that similar data may be available to entrants due to the nonrivalrous nature of information, rather than most of it being owned exclusively by the incumbents using it. Ultimately, I find that labeling big data an entry barrier does not necessarily make it so. Rather, a proper analysis of the role played by big data in impeding, or perhaps even facilitating entry, is better performed by examining actual entry conditions on a case-by-case basis. Competition policy is ill-served by prejudging the outcome of the analysis by resorting to labels rather than by a careful examination of the facts.

Finally, as detailed in Part IV, competition authorities in the United States and Europe have now amassed considerable experience in assessing big data in actual cases and, thus far at least, it has yet to play a dominant role in any competition agency challenge—holding aside cases where data is the product itself.[9] What does this lack of agency challenge tell us? While it does

[7] The idea is that successful use of big data and analytics creates a positive feedback loop. For example, a greater volume of users can provide data to a search engine which, in turn, has a positive impact on search quality. A positive impact on search quality, in turn, can result in a greater volume of users. This can be labeled as "data-driven indirect network effects." *See* Cédric Argenton & Jens Prüfer, *Search Engine Competition with Network Externalities*, 8 J. Competition L. & Econ. 73 (2012); Jens Prüfer & Christoph Schottmüller, *Competing with Big Data* (Tilburg Law & Econ. Ctr. Discussion Paper No. 06/2017, 2017).

[8] Darren S. Tucker & Hill B. Wellford, *Big Mistakes Regarding Big Data*, 14 Antitrust Source, Dec. 2014, at 1, 7.

[9] *See* D. Daniel Sokol & Roisin Comerford, *Antitrust and Regulating Big Data*, 23 Geo. Mason L. Rev. 1129, 1130 ("In reality, there is no challenge at all, as the arguments for antitrust intervention when Big Data has come up as an issue have never carried the day for any merger or decided conduct case in any Department of Justice Antitrust Division ('DOJ'), Federal Trade Commission ('FTC') or Directorate-General for Competition ('DG Competition') case to date."); *see also* Margrethe Vestager, European Commissioner for Competition, Competition in a Big Data World (Jan. 17, 2016), https://ec.europa.eu/commission/commissioners/2014-2019/vestager/announcements/competition-big-data-world_en ("[W]e shouldn't take action just because a company holds a lot of data. After all, data doesn't automatically equal power. . . . The Commission has looked at this issue in two merger cases—Google's acquisition of DoubleClick, and Facebook's purchase of WhatsApp. In the particular circumstances of those cases, there was no serious cause for concern. . . . We continue to look carefully at this issue, but we haven't found a competition problem yet.").

not prove that big data is or is not a competition issue, it strongly supports that proposition. Moreover, an assessment of the agencies' decisions that involve big data reveals a lack of basis to suggest that competition agencies and the courts have been systematically neglecting this possible issue.[10]

In totality, these considerations suggest that either we are in a new era regarding big data in competition policy—or we should be. In this new era, big data and analytics are considered just as important to innovation; yet, there is a recognition that big data is not anything unique when it comes to competition. Thus, it should be treated no differently than assets such as highly skilled labor, specialized capital, and research and development (R&D) infrastructure. Thus, just as it is possible to formulate a theory of harm involving the foreclosure of a necessary input involving intellectual property and capital, it is also possible to formulate a theory of harm involving the foreclosure of vital data.[11] Yet, the same rule of reason paradigm applies, which is based on weighing alleged anticompetitive harms with potential procompetitive benefits using actual evidence. Therefore, the mere possession and use of big data, as well as the competitive advantages gained from big data and associated analytics, are neither necessary nor sufficient conditions for presuming harm, let alone for concluding it.

To slightly expand the point, there is a distinction between firms monopolizing data and firms collecting and utilizing data in their ordinary course of business. These are the not the same thing and should not be conflated with each other. Monopolizing data involves a firm engaging in anticompetitive means to exclusively control so much information that rivals are unable to effectively compete. Firms that collect a great deal of data and use the data to improve products are engaging, instead, in procompetitive conduct, suggesting that use or even ownership of lots of data need not necessarily lead to consumer harm.

I. The Rise of Big Data

Given that the use of data by suppliers to improve efficiency and gain a competitive advantage is certainly not a new concept,[12] when did the concern

[10] Of course, there are plenty of criticisms of the U.S. antitrust agencies' decisions to enforce and not to enforce. In support of this position, the most cited study is John Kwoka, Mergers, Merger Control, and Remedies: A Retrospective Analysis of U.S. Policy (MIT Press 2014). The study, however, has a number of weaknesses. *See* Michael Vita & David Osinski, *John Kwoka's* Mergers, Merger Control, and Remedies*: A Critical Review*, 82 Antitrust L.J. 361 (2018). We are not aware of any Kwoka-like critiques of EC decisions.

[11] However, Catherine Tucker argues convincingly that big data should not be considered an "essential facility." *See* Catherine Tucker, *Digital Data, Platforms and the Usual [Antitrust] Suspects: Network Effects, Switching Costs*, *Essential Facility*, 54 Rev. Indus. Org. 683, 684 (2019).

[12] *See* Competition Bureau Canada, Big Data and Innovation: Implications for Competition Policy in Canada 5 (2017), https://www.competitionbureau.gc.ca/eic/site/cb-bc.nsf/vwapj/CB-Report-BigData-Eng.pdf/$file/CB-Report-BigData-Eng.pdf ("[T]he use of data is hardly new. General

regarding big data as an *antitrust* problem begin? Certainly, the idea that possessing certain types of data can give firms a competitive advantage predates the digital era. Timothy Muris and Jonathan Nuechterlein convincingly argue that concerns about the use of data were present even during the heyday of the Great Atlantic and Pacific Tea Company (A&P).[13]

What has unquestionably changed is the rise of the digital economy and the staggering volume, velocity, and variety of data generated virtually instantaneously.[14] In 2013, IBM estimated that "2.5 quintillion bytes of data" are created daily and "90% of the data in the world today has been created in the last two years."[15] According to Domo, a software provider, "[b]y 2020, it's estimated that for every person on earth, 1.7 MB [megabytes] of data will be created every second."[16] In 2018, the International Data Corporation (IDC) estimated the global volume of data to be 33 zettabytes, which is equivalent to 33 trillion gigabytes—forecasting it to grow to 175 zettabytes in 2025.[17] Focusing more on the consumer end, global smartphone application (app) downloads exceeded 194 billion in 2018, which is 35 percent more than just two years before.[18]

Clearly, this colossal volume of data implies that greater resources are needed to collect, store, and manage the data. While this potentially increases the capital requirements to handle such large volumes of data, the ready availability of cloud computing services enables many firms to access

stores used to keep a record of information about customers located in their small town (including, for example, credit information). One of the first large-scale and systematic collections of data occurred in the mid-nineteenth century when the large American railroads began mandating regular, system-wide reports, which required the building of extensive comptroller departments, the hiring of full-time auditors, and the development of fundamental accounting concepts still in use today. At about the same time, mercantile agencies, such as precursors to the Dun & Bradstreet firm, began collecting and selling substantial amounts of credit reporting data."); Alfred D. Chandler, Jr., The Visible Hand: The Managerial Revolution in American Business 109 (Harvard Univ. Press 2002) (1977) ("[A] constant flow of information was essential to the efficient operation of these new large business domains. For the middle and top managers, control through statistics quickly became both a science and an art. This need for accurate information led to the devising of improved methods for collecting, collating, and analyzing a wide variety of data generated by the day-to-day operations of the enterprise. Of even more importance it brought a revolution in accounting.").

[13] Timothy J. Muris & Jonathan E. Nuechterlein, *Antitrust in the Internet Era: The Legacy of* United States v. A&P, 54 Rev. Indus. Org. 651, 657 (2018) ("A&P also succeeded because it did what many tech companies do today, albeit amid much controversy: use data to create greater consumer value.").

[14] *See* Big Data: Bringing Competition Policy to the Digital Era—Background Note by the Secretariat—29–30 November 2016, *supra* note 3, at 2.

[15] Ralph Jacobson, *2.5 Quintillion Bytes of Data Created Every Day. How Does CPG & Retail Manage It?*, IBM Consumer Prods. Indus. Blog (Apr. 24, 2013), https://www.ibm.com/blogs/insights-on-business/consumer-products/2-5-quintillion-bytes-of-data-created-every-day-how-does-cpg-retail-manage-it.

[16] *Data Never Sleeps 6.0*, Domo, https://www.domo.com/learn/data-never-sleeps-6.

[17] David Reinsel, John Gantz & John Rydning, The Digitization of the World: From Edge to Core 6 (2018), https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-data-age-whitepaper.pdf.

[18] App Annie, The State of Mobile 2019, at 9 (2019), https://www.appannie.com/en/insights/market-data/the-state-of-mobile-2019.

and utilize big data without having to make large investments in capital and infrastructure.[19]

Additionally, the sheer size and opportunities provided by big data has led, perhaps not surprisingly, to conjectures and concerns regarding its antitrust implications. The thinking has been that big data represents something new and potentially quite valuable, particularly with the rise and prominence of multi-sided platforms. And when something is potentially quite valuable, concerns naturally arise about whether it can be abused by large marketplace players. Further, though not as obviously antitrust-relevant, serious concerns have been raised about the implications of big data for consumer privacy. The antitrust concerns are summarized in the following argument from Nathan Newman, as applied to one particular user of big data:

> [C]ontrol of information has been skewed towards a few players with both the concentrated data processing power and supply of user data to dominate a particular sector. For example, as of 2012, Google processes sixty-five percent of U.S. search queries, and earns seventy-eight percent of U.S. search advertising revenue (and eighty-five percent of global search revenue). Thus, antitrust authorities around the world have with good reason targeted the company for its dominance.[20]

The implication is that the sheer volume of data that firms with high market shares regularly collect is *the reason* for their dominance and/or their maintenance of monopoly power. That logic is suspect. Even if data were the primary reason for a firm's dominance, this hardly implies that dominant firms are doing anything nefarious, or that there is a need for competition authorities to intervene. Ownership and/or use of big data by large firms, it has been alleged, is objectionable because "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."[21] Maybe it can, maybe it cannot. Theory alone does not lead to a single unambiguous conclusion, and available evidence does not do so either.

---

[19] *See* Ibrahim Abaker Targio Hashem, Ibrar Yaqoob, Nor Badrul Anuar, Salimah Mokhtar, Abdullah Gani & Samee Ullah Khan, *The Rise of "Big Data" on Cloud Computing: Review and Open Research Issues*, 47 Info. Sys. 98, 99 (2015) ("Cloud computing is one of the most significant shifts in modern ICT and service for enterprise applications and has become a powerful architecture to perform large-scale and complex computing. . . . Cloud computing can not only minimize the cost and restriction for automation and computerization by individuals and enterprises but can also provide reduced infrastructure maintenance cost, efficient management, and user access."); Amit Verma, *Big Data and Cloud Computing—A Perfect Combination*, Whizlabs Blog (July 21, 2018), https://www.whizlabs.com/blog/big-data-and-cloud-computing ("Cloud enables customers for big data processing without large-scale big data resources.").

[20] Nathan Newman, *Search, Antitrust, and the Economics of the Control of User Data*, 31 Yale J. on Reg. 401, 403 (2014).

[21] *Id.* at 404. Newman later refines his argument and states that "[a]t the core of Google's aggressive expansion of control over user data in recent years has been its rapid extension into a wide range of related product lines where it could collect ever more personal information about online users to sell to its advertisers." *Id.* at 426.

Allen Grunes and Maurice Stucke argue that "these features of big data [volume, velocity, variety, and value] have several implications for competition policy, including raising barriers to entry and foreclosing access to essential inputs."[22] Further, "companies, whose business model depends on securing a competitive advantage through big data, may also devise anticompetitive data-driven strategies. Such strategies may include preventing rivals from accessing the data (such as through exclusivity provisions with third-party providers) or foreclosing opportunities for rivals to procure similar data (such as making it harder for consumers to adopt other technologies or platforms)."[23] Further, it is a "myth" that "[t]he current antitrust tools fully address the big data issues."[24]

There is also a strand in the literature that suggests that big data leads to a loss in privacy, which could intersect with antitrust concerns.[25] Michael Katz notes that privacy could be considered an antitrust issue to the extent that it is part of the non-price dimension of competition.[26] He finds, however, "that viewing privacy as an element of product quality highlights the fact that it is not evident that competition promotes privacy or, indeed, that promoting greater levels of privacy is desirable."[27] Further, James Cooper, Deborah Feinstein, and Maureen Ohlhausen and Alexander Okuliar argue convincingly that privacy and antitrust considerations are best considered separately.[28]

In sum, with the rise of digital markets in the 2000s and the subsequent collection of large volumes of data, a number of researchers began asking whether or not big data is a competition issue. Boiled down to its essence, the concerns regarding big data—holding aside, for the moment, the privacy considerations—follow the standard litany of (1) creating and/or maintaining market power, (2) allowing for the foreclosure of rivals, and (3) serving as a barrier to entry that prohibits or severely hampers the entry of innovative firms. Where legitimate, these concerns, as suggested above and discussed

---

[22] Allen P. Grunes & Maurice E. Stucke, *No Mistake About It: The Important Role of Antitrust in the Era of Big Data*, 14 Antitrust Source, Apr. 2015, at 1, 3.

[23] *Id.*

[24] *Id.*

[25] *See, e.g.*, Dissenting Statement of Commissioner Pamela Jones Harbour at 4, Google/DoubleClick, FTC File No. 071-0170 (FTC Dec. 20, 2007) ("The transaction will combine not only the two firms' products and services, but also their vast troves of data about consumer behavior on the Internet. Thus, the transaction reflects an interplay between traditional competition and consumer protection issues. The Commission is uniquely situated to evaluate the implications of this kind of data merger, from a competition as well as a consumer protection perspective.").

[26] Michael L. Katz, *Multisided Platforms, Big Data, and a Little Antitrust Policy*, 54 Rev. Indus. Org. 695, 707 (2019).

[27] *Id.*

[28] James C. Cooper, *Privacy and Antitrust: Underpants Gnomes, the First Amendment, and Subjectivity*, 20 Geo. Mason L. Rev. 1129, 1133–34 (2013); Deborah Feinstein, *Big Data in a Competition Environment*, CPI Antitrust Chron., May 2015, at 1, 2 ("Despite calls to use the merger review process to improve privacy protections for consumers, the FTC continues to examine competition and consumer protection issues separately."); Maureen K. Ohlhausen & Alexander P. Okuliar, *Competition, Consumer Protection, and* The Right *[Approach]* to Privacy, 80 Antitrust L.J. 121 (2015).

below, apply to important inputs generally, not to big data specifically or especially.

Despite the strong concerns expressed by several commentators, including those quoted above, a number of researchers and commentators have examined the arguments and evidence and found no basis for concluding that big data is a unique competition issue that is different from how antitrust agencies analyze ownership and use of any other important inputs. Central to these arguments are the following points. First, big data is one among a host of potential factors that may determine market success. There is nothing particularly "unique" about big data in its ability to predict market success. Moreover, current antitrust tools actually *are* sufficient to address anticompetitive concerns that might arise from an alleged misuse of data. Second, while big data can be a highly valuable asset that firms expend resources to collect and analyze, it does not follow that its ownership or use reduces social welfare or constitutes an entry barrier that prevents innovative entrants from participating in markets. Finally, we now have over a decade of agency decisions where big data was considered, either explicitly or implicitly, and there has yet to be a challenge predicated primarily on abuse of big data. The following parts further detail each of these points and their relevance to the policy debate.

II. The Role of Big Data in a Firm's Production and Innovation Process

The paradigmatic antitrust concern involving big data is when it is perceived as an essential and exclusive factor in the production of a final good or service—typically as it relates to multi-sided platforms and/or other digital markets. Yet, all assessments of big data must, at some point, address how big data fits into the larger production function of firms—whether in a digital or widget market. While there are instances where data is the final product—for example, commercial databases[29]—this is not the primary concern in terms of competition policy as it relates to big data.[30]

Specifically, Turck states that "Big Data, fundamentally, is . . . plumbing. Certainly, Big Data powers many consumer or business user experiences, but at its core, it is enterprise technology: databases, analytics, etc[.]: stuff that runs in the back that no one but a few get to see."[31] Similarly, in a commentary

[29] *See, e.g.*, Complaint, Nielsen Holdings N.V. and Arbitron Inc., FTC File No. 131 0058, Dkt. No. C-4439 (FTC Sept. 30, 2013); FTC v. CCC Holdings Inc., 605 F. Supp. 2d 26, 38–39 (D.D.C. 2009).

[30] Joshua D. Wright & Elyse Dorsey, *Antitrust Analysis of Big Data*, 2 Competition L. & Pol'y Debate, Dec. 2016, at 35, 37 ("Consumer data, however, is predominantly used by firms as an input into a final product—but is not ever the product itself. In other words, firms are neither pricing consumer data nor otherwise competing to provide better consumer data.").

[31] *See supra* note 4; *see also* Palmer, *Data Is the New Oil*, *supra* note 6 ("Data is just like crude. It's valuable, but if unrefined it cannot really be used. It has to be changed into gas, plastic, chemicals, etc[.] to create

to the Canadian Competition Bureau's white paper on big data, Microsoft stated, "it also bears emphasizing that large sets of structured, unstructured, labeled or unlabeled data *may have no known value* in the development of ML [machine learning] or AI [artificial intelligence] applications and may in fact not have any significant value. Data is thus best viewed as an 'asset' that may or may not have value just like any other asset depending on its commercial utility."[32]

These statements highlight the central point that big data must be considered in the larger context of a firm's production function.[33] Further, it is not necessarily the relative size of the collected data that can confer a competitive advantage, but the quality of the data and/or the quality of the data analytics.[34] It boils down to the fundamental reality that the *mere possession* of big data does not accord a competitive advantage to a firm.[35] As in almost every dimension of competition, firms differ, not only in the attributes of their final products, but in how they organize themselves—*including the extent and efficiency of their use of big data via analytics*.[36]

a valuable entity that drives profitable activity; so must data be broken down, analyzed[,] for it to have value.").

[32] Letter from Greg Sivinski, Assistant Gen. Counsel, Microsoft Corp., to Competition Bureau Canada, *Microsoft's Comments Re Big Data and Innovation: Implications for Competition Policy in Canada* 2 (Nov. 17, 2017), https://www.competitionbureau.gc.ca/eic/site/cb-bc.nsf/vwapj/Microsoft_Comments_on_CB_Draft_Discussion_Paper.pdf/$file/Microsoft_Comments_on_CB_Draft_Discussion_Paper.pdf.

[33] *See* Marc Bourreau, Alexandre de Streel & Inge Graef, Big Data and Competition Policy: Market Power, Personalised Pricing and Advertising, Centre on Regulation in Europe 37 (2017), https://cerre.eu/sites/cerre/files/170216_CERRE_CompData_FinalReport.pdf ("The first principle is that data are one input, which is important but not unique, to develop successful applications and algorithms. Other inputs are also important such as skilled and creative labour force (in particular computer scientists and singers), capital and distribution channels. Above all, the skills and creativity of the labour force will make the success of the applications.").

[34] *See* Eliana Garcés, *Data Collection in Online Platform Businesses: A Perspective for Antitrust Assessment*, CPI Antitrust Chron., May 2018, at 1, 3 ("Access to data is not normally at the root of an online platform's success.").

[35] In fact, there are instances when big data gets stale. *See* Lockwood Lyon, *The End of Big Data*, Database J., May 16, 2016 ("As data ages it tends to become less relevant for the following reasons: Newly implemented operational applications will not have a data history; Older products are removed and replaced by new products; Older customers may no longer exist; As you apply maintenance to current operational systems, some analyses of 'old' behavior becomes irrelevant; Older data tends to be less accurate and sometimes is missing altogether; as operational systems are adjusted to fix these problems, inaccurate or missing historical data will skew analyses.").

[36] Global Antitrust Institute, Comment on the Canadian Competition Bureau's White Paper, "Big Data and Innovation: Implications for Competition Policy in Canada" 3 (Nov. 17, 2017), https://www.competitionbureau.gc.ca/eic/site/cb-bc.nsf/vwapj/GAI_Comment_on_Big_Data_FINAL_UPDATED.pdf/$file/GAI_Comment_on_Big_Data_FINAL_UPDATED.pdf ("The implication is that, when considering the role that big data plays in a given market, rivals might speciously suggest that a market leader is succeeding due to the leader's sheer volume of data, when it is not data which is scarce but the skill and talent needed to combine the data with other inputs to produce something of value."); *see also* Lyon, *The End of Big Data*, *supra* note 35 ("The next phase of big data is now here. Applications and analysts must now create value in two ways: by automating current analytical results to feed back into operational systems and by turning to analysis of unstructured and unmodeled data."). Consider the production of academic empirical economic studies using large datasets. For academics, what differentiates a successful publishing record from a poor one is not likely an inability to access data. While certain academics might have access to "better" data or more unique data—that process is endogenous and could be a result of superior effort on the part of the academic to obtain or construct a dataset. The point is that academic

What about the argument that, while big data might not be a sufficient condition to compete effectively in digital markets, it is perhaps a necessary condition? Again, this assertion misses the point that firms are differentiated, not only in their ability to unlock value from big data, but in terms of how reliant they are on big data to fuel innovation and quality products.[37] The classic economic production function is a paradigm where firms mix a variety of inputs, from labor, capital, technology, land, and, relevant for our purposes, data analytics to produce output with certain attributes including quality. How *exactly* firms mix their inputs to produce the final product will differ. For instance, one particular entrant could invest heavily in R&D and intellectual property to develop a differentiated product, while another is more heavily reliant on iterative machine learning as its user base grows. Similarly, an entrant could choose to invest and utilize big data—but perhaps not to the extent that an incumbent does.[38] Thus, improving quality through innovation can be achieved through a number of different paths. Consequently, asset equivalence is not a necessary condition for entry.[39]

For instance, Spotify overtook iTunes in terms of being the leading platform for online music—not because it had a larger amount of data, but because it mixed its ingredients in a different way than Apple.[40] Similarly, between

success comes down to the skill, effort, and ingenuity of the researcher—not whether or not they have access to bigger and better data.

[37] For example, there is evidence that, in the mid-2000s, Yahoo fell behind Google in its ability to monetize via search ads—not because Yahoo had less advertisers or data, but because Yahoo had inferior ad matching algorithms. *See* Danny Sullivan, *Yahoo's Poor Ad Targeting & Thoughts on Google-Yahoo*, Search Engine Land (Sept. 21, 2008), https://searchengineland.com/yahoos-poor-ad-targeting-thoughts-on-google-yahoo-14780.

[38] However, at least in one respect, entrants may have an advantage over incumbents. Specifically, startups and new entrants are not burdened by legacy systems and digital infrastructure from a prior age; consequently, they can be as nimble, if not more so, than entrenched incumbents. *See* Turck, *Is Big Data Still a Thing? (The 2016 Big Data Landscape)*, *supra* note 4 ("Other 'digital native' companies, including many of the budding unicorns, started facing similar needs as the large Internet companies, and had no legacy infrastructure either, so they became early adopters of those Big Data technologies."). Additionally, theoretical work shows that entrants can overtake incumbents if they are more efficient in extracting value per unit of data. *See* Maryam Farboodi, Roxana Mihet, Thomas Philippon & Laura Veldkamp, Big Data and Firm Dynamics (Jan. 14, 2019) (unpublished manuscript) (manuscript at 2), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3334064 ("We also learn that initial size is not the most important factor in the success of a firm. A small firm that uses data efficiently, meaning that it harvests more data per unit of production, may lose money initially while it builds up its data stock. But if the firm can finance this phase, it can quickly out-compete a larger, less data-efficient firm.").

[39] *See* Nils-Peter Schepp & Achim Wambach, *On Big Data and Its Relevance for Market Power Assessment*, 7 J. Eur. Competition L. & Prac. 120, 122 (2016) ("[P]otential competitors do not necessarily have to build a dataset equivalent to the size of the incumbent. . . . They rather need to find ways to accumulate highly relevant data to build a competitive [dataset], [but] not necessarily the same dataset."); *see also* Xavier Boutin & Georg Clemens, Big But Not Insurmountable? How the Definition of 'Big Data' Can Help in the Assessment of Entry 3 (2018), https://www.compasslexecon.com/wp-content/uploads/2018/04/Big-Data-Expert-Opinion.pdf ("The evidence from data-related markets shows that the possession of large data volumes alone rarely provides a competitive advantage.").

[40] *See, e.g.*, Ingrid Lunden, *In Europe, Spotify Royalties Overtake iTunes Earnings by 13%*, TechCrunch, Nov. 4, 2014, https://techcrunch.com/2014/11/04/in-europe-spotify-royalties-overtake-itunes-earnings-by-13; Mark Mulligan, *Mid-Year 2018 Streaming Market Shares*, Midia Blog (Sept. 13, 2018), http://www.midiaresearch.com/blog/mid-year-2018-streaming-market-shares. One might argue that

2011 and 2012, Chrome overtook both Firefox and Internet Explorer as the predominant web browser even though Chrome had a less than 5-percent share at the start of 2010.[41] Given that web browsers collect a tremendous amount of data on user behavior,[42] under the big data theory of entry barriers, iTunes, Firefox, and Internet Explorer should never have relinquished their market-leading positions.[43] Moving to social media, as of 2018, Snapchat is used more intensely per user than all other social media messaging apps in the United States, including Facebook, Instagram, Messenger, and Pinterest.[44] The point of these examples is not to demonstrate that big data is unimportant, but to illustrate that firms can achieve tremendous market success, including overtaking incumbents, using ingredients in different proportions based on their particular skills, sets of assets, and comparative advantages.

Of course, after entry, firms will likely adjust and reoptimize how they combine their various inputs to achieve their desired product features and levels of quality. For instance, a firm might have originally entered with lower costs or superior technology via a valuable patent or algorithm, but, over time, as it starts to gain users and data, it begins to incorporate the feedback and insights from the data into the design of its product and organization.[45] Similarly, an entrant might start with a basic product that fulfills a certain

Apple's iTunes is not exactly the same product as Spotify's streaming service—so it is not actually an apples-to-apples comparison. (Rather the proper comparison is with Apple Music.) The fact that Spotify (and Apple Music) are streaming services and not exactly the same as iTunes actually *proves* the larger point that markets evolve in ways that are not easily forecasted. (Apple's decision to discontinue iTunes in favor of its Apple Music product, among others, further proves the point.) Nonetheless, if we wanted to focus on an early-mover in streaming music services, one example would be Pandora, which currently sits seventh in global market share—despite the fact that it entered before all six firms ahead of it. *See* Mulligan, *Mid-Year 2018 Streaming Market Shares*, *supra*.

41 *See Desktop, Mobile & Tablet Browser Market Share Worldwide—September 2019*, StatCounter, http://gs.statcounter.com/browser-market-share/desktop-mobile-tablet/worldwide/#monthly-200901-201904. Of course, one could argue that Chrome's ownership by Google and its trove of data are responsible for its success. According to one article, however, Chrome's success was due to its sheer technical superiority. *See* Jonathan Tamary & Dror G. Feitelson, *The Rise of Chrome*, PeerJ Comput. Sci. (2015), https://peerj.com/articles/cs-28.pdf. Further, even if big data was primarily responsible for this superiority, and it is not clear that it is, this would be a welfare enhancing use of big data to facilitate entry and create a superior product.

42 *See, e.g.*, Dan Price, *10 Types of Data Your Browser Is Collecting About You Right Now*, MakeUseOf, Oct. 12, 2018, https://www.makeuseof.com/tag/data-browser-collects-about-you.

43 In Part III, I detail the use and abuse of the term "barriers to entry."

44 *See* App Annie, The State of Mobile 2019, *supra* note 18, at 48. Among users aged 12 to 17, Snapchat is the market leader with 16.8 million users, while Instagram and Facebook are second and third at 12.8 million and 11.5 million, respectively. *See* eMarketer Editors, *Facebook Is Tops with Everyone But Teens*, eMarketer (Aug. 28, 2018), https://www.emarketer.com/content/facebook-is-tops-with-everyone-but-teens.

45 An example would be Walmart. Although founded in 1962, today, Walmart analyzes a tremendous amount of data. Walmart processes 2.5 petabytes of data from 1 million customers every hour and uses big data to improve operational efficiency and increase online sales. *See How Big Data Analysis Helped Increase Walmart Sales Turnover*, Dezyre (May 23, 2015), https://www.dezyre.com/article/how-big-data-analysis-helped-increase-walmarts-sales-turnover/109; Walmart Staff, *5 Ways Walmart Uses Big Data to Help Customers*, Walmart Blog (Aug. 7, 2017), https://blog.walmart.com/innovation/20170807/5-ways-walmart-uses-big-data-to-help-customers.

niche in the market but then proceed to add new features, expand into adjacent markets, or change its overall business strategy.[46]

In sum, for the purposes of competition policy, it is critical to understand precisely *why* a product is successful. Even in the presence of big data, building a better mousetrap is the foundational paradigm of competition.[47] The path to a better mousetrap can differ across markets and within firms in a given market. For example, network effects—both direct and indirect—could be the primary reason a product is successful in gaining and retaining consumers, rather than big data in and of itself.[48] And network effects can cause a formerly leading product to lose its position quickly to a new entrant with better products.

Even if the use of big data is the primary reason for one firm's success, a relevant question is whether comparable, but not necessarily equivalent, data are costly for a rival firm to acquire. In the decade or so since the antitrust concerns over big data began in earnest, it is fair to say that the amount of data generated has further exploded and has been growing exponentially. Importantly, it is clear that there are often multiple sources of data available to the same consumer and provided by multiple entities. The average consumer has over 100 apps on her smartphone with about 30 to 40 apps

[46] The story of OpenTable is a case in point. David Evans and Richard Schmalensee detail how OpenTable originally attempted to sign up as many restaurants in as many cities as possible to participate on its restaurant reservation platform. After a few years of experience and uneven results, however, OpenTable decided to shift its focus to restaurants in four major cities: San Francisco, Chicago, New York, and Washington, D.C. The shift in strategy, based in part on observing user behavior, unlocked the cross-group effects it needed to bring a critical mass of users and restaurants. *See* David S. Evans & Richard Schmalensee, Matchmakers: The New Economics of Multisided Platforms 10–11 (Harvard Bus. Rev. Press 2016).

[47] *See* Competition Bureau Canada, Big Data and Innovation: Implications for Competition Policy in Canada, *supra* note 12, at 3 ("But competition policy in Canada does not, and should not, assume that 'big is bad.' Companies that achieve a leading market position—even a dominant one—by virtue of their own investment, ingenuity, and competitive performance should not be penalized for doing so. Imposing a penalty for excellence removes the incentives to pursue excellence.").

[48] It is worth noting that, while big data and network effects are often discussed jointly and perhaps interchangeably, they are distinct concepts. *See supra* note 7. Ultimately, however, it is an empirical question regarding the strength of potential data-driven indirect network effects and whether or not there are diminishing returns. For instance, Lesley Chiou and Catherine Tucker find that large amounts of historical data may not be particularly useful for the relevancy of search results. *See* Lesley Chiou & Catherine Tucker, *Search Engines and Data Retention: Implications for Privacy and Antitrust* (Nat'l Bureau of Econ. Research, Working Paper No. 23815, 2017). Further, Hal Varian argues that improving search quality is more akin to learning-by-doing, rather than network effects. Hal Varian, Use and Abuse of Network Effects (Sept. 17, 2017) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3215488. Varian challenges the presumption that collecting more data wondrously results in greater quality and this unlocks an unstoppable feedback loop. Rather, Varian states, "Though this is a convenient modeling shorthand, it can be somewhat misleading as it suggests that 'learning' is a passive activity that automatically happens as more output is produced. Nothing could be further from the truth. Learning by doing necessarily requires investment in data collection, analysis, and experimentation." *Id.* (manuscript at 4); *see also* Tucker, *Digital Data, Platforms and the Usual [Antitrust] Suspects: Network Effects, Switching Costs, Essential Facility*, *supra* note 11, at 684 ("I find little evidence that digital data augments market power due to either network effects or switching costs. Instead, digitization may have weakened these two economic forces, because it frees a user from a particular hardware system.").

used per month.[49] Further, according to Deloitte, "estimates of the number of online accounts held by each individual across all online services range from 25 to 90, potentially increasing to around 200 in years to come."[50] Additionally, "[i]n the majority of categories included in the survey, most users use multiple providers rather than a single one, highlighting the ease with which people may access different services online."[51] In fact, the most common services where multiple providers are used include social and sharing platforms—online shopping, email, instant messaging (IM), video calling, travel, credit cards, and news.[52] Further, Deloitte concludes that "the same or similar data is often created across several providers" including browsing history, transaction data, location data, searches, and content viewed.[53] Thus, it is increasingly difficult to make a credible argument that what some firms or potential entrants in digital markets are missing today is access to big data. In fact, the problem is not how to get data, but what to do with it.

Finally, when assessing the role that big data should play in competition policy, what must not be lost is that big data can lower a firm's costs through a more efficient production process and/or result in greater demand through higher quality or a greater matching to consumers' taste and preferences. This benefit is decidedly procompetitive and consumer-welfare-enhancing.[54] In some respects, big data's role is similar to that of R&D expenditures, learning-by-doing, and economies of scale. What can too easily be lost in the debate regarding the *potential* anticompetitive impact of large users of big data is the fact that big data and analytics are, to one degree or another, procompetitive assets that help drive innovation and otherwise benefit consumers.[55] While almost all serious scholarship acknowledges this funda-

[49] App Annie, The State of Mobile 2019, *supra* note 18, at 13.

[50] Deloitte, The Data Landscape: A Report for Facebook 12 (2017), https://www2.deloitte.com/uk/en/pages/technology-media-and-telecommunications/articles/the-data-landscape.html.

[51] *Id.* at 23.

[52] *Id.*

[53] *Id.* at 26.

[54] *See* Letter from Greg Sivinski to Competition Bureau Canada, *supra* note 32, at 2 ("Using proprietary data to seek competitive advantage is economically efficient behavior that drives innovation."); *see also, e.g.*, Organisation for Economic Co-operation and Development (OECD), Data-Driven Innovation: Big Data for Growth and Well-Being 23 (2015), https://read.oecd-ilibrary.org/science-and-technology/data-driven-innovation_9789264229358-en#page1 ("By collecting and analysing 'big data', a large share of which is provided by Internet users (consumers), Internet companies are able to automate their processes and to experiment with, and foster, new products and business models at much a faster rate than the rest of the industry. In particular, the advanced use of data and analytics enables Internet firms to scale their businesses at much lower costs than other ICT [information-communication technologies] firms, a phenomenon that goes much further than what Brynjolfsson et al. (2008) describe as *scaling without mass*.") (emphasis in original) (citing Erik Brynjolffson, Andrew McAfee, Michael Sorrell & Feng Zhu, *Scale Without Mass: Business Process Replication and Industry Dynamics* (Harv. Bus. Sch. Tech. & Operations Mgmt. Unit, Research Paper No. 07-016 Sept. 30, 2008)).

[55] Importantly, the impact of data is not limited to digital platforms. Take for instance, big data in the healthcare sector. Data is being used to forecast demand for inpatient admissions, epidemic outbreaks, and overall health outcomes. The OECD finds that data-driven innovation (DDI) is "a disruptive new source of growth that could transform all sectors in the economy. Even traditional sectors such as retail, manufacturing and agriculture are being disrupted through DDI." OECD, Data-Driven Innovation: Big Data

mental reality, at least nominally, the debate over big data often focuses entirely on the *potential* for harm from big data, rather than on the clear and documented evidence of the large benefits that it commonly provides. It would be as if computer operating systems were viewed with inherent suspicion simply because of the possibility that an owner of one might become dominant and then might engage in anticompetitive behavior.[56] In the new era of big data, which I argue that competition policy has now entered (or should be entering), big data's procompetitive effects need to be afforded their proper weight—not only by competition agencies and courts, but also by policymakers and scholars.

III. Big Data and Barriers to Entry

Related to the prior part's discussion is the question of whether or not big data is a barrier to entry. Most commentators who favor aggressive antitrust enforcement as it relates to big data would argue in the affirmative.[57] I contend that it is important to consider how economists define barriers to entry and how this concept relates to the analysis of competitive outcomes. More importantly, rather than classifying big data, or other inputs, as barriers to entry, it is more relevant to detail what role big data plays in a particular market—both historically and in the present—in order to better forecast the hurdles that potential entrants must overcome to be effective and viable competitors.

Scholarly discussions of barriers to entry inevitably start with Joseph Bain, who defined barriers to entry as structural factors that allow incumbents to persistently price above the competitive level because they are not disciplined by a threat of entry.[58] Within this framework, economies of scale

for Growth and Well-Being, *supra* note 54, at 27. Further, "[t]he analysis of 'big data', increasingly in real time, is driving knowledge and value creation across society; fostering new products, processes and markets; spurring entirely new business models; transforming most if not all sectors in OECD countries and partner economies; and thereby enhancing economic competitiveness and productivity growth." *Id.* at 20; *see also* Deloitte, The Data Landscape: A Report for Facebook, *supra* note 50, at 9 ("A recent study for the European Commission estimated that the economic impact of data amounted to €300 billion in 2016 across EU28 countries, potentially rising to €430 billion by 2020. This includes the direct impacts in terms of revenues from big data, analytics and other data-related services, plus indirect impacts across sectors of the economy and induced impacts through increased economic activity.").

56 When this arguably happened, antitrust was certainly able to successfully intervene.

57 *See, e.g.*, Joshua Gans, Enhancing Competition with Data and Identity Portability 8 (2018), https://www.hamiltonproject.org/assets/files/Gans_20180611.pdf ("A firm's monopolization of data could harm consumers if it confers an incumbency advantage—supported by barriers to entry—that reduces the incentives for competing platforms to enter a particular market."); Terrell McSweeny & Brian O'Dea, *Data, Innovation, and Potential Competition in Digital Markets—Looking Beyond Short-Term Price Effects in Merger Analysis*, CPI Antitrust Chron., Feb. 2018, at 1.

58 *See* Joseph Bain, Barriers to New Competition 3 (Harvard Univ. Press 1956) ("A barrier to entry is an advantage of established sellers in an industry over potential entrant sellers, which is reflected in the extent to which established sellers can persistently raise their prices above competitive levels without attracting new firms to enter the industry.").

that require large capital expenditures, product differentiation, and absolute cost advantages would be considered barriers to entry. Kip Viscusi, John Vernon, and Joseph Harrington, Jr. highlight a fundamental problem with Bain's definition, which is that "it is a tautology: A barrier to entry is said to exist if existing firms earn above-normal profit without inducing entry. In other words, Bain defines a barrier to entry in terms of its outcome."[59]

While Bain's definition offers a useful starting point, it tends to create a paradigm where economic concepts such as economies of scale and product differentiation are considered exogenous factors that then determine the competitiveness of an industry. This is a remnant of the prior structure-conduct-performance paradigm in industrial organization. A decade or so later, George Stigler built off of Bain's definition and stated that a "barrier to entry may be defined as 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."[60] Two examples of this type of barrier to entry are patents and grandfathered government regulations. Economies of scale, on the other hand, would not be considered a barrier to the extent that an entrant has access to the same cost function—even if it operates on a different part of the same cost curve. One benefit of Stigler's definition is the explicit recognition that, if there is an asymmetry between incumbents and entrants—in that an entrant *must incur* a cost that an incumbent does not—this has the potential to confer, but does not guarantee, supranormal profits to the incumbent in equilibrium.

Neither Bain's nor Stigler's definitions explicitly incorporate assessments of efficiency and economic welfare. C. C. von Weizsäcker considers this a deficiency, given that "economists normally implicitly assume that barriers to entry are a distortion of the competitive process. They inhibit the proper working of the principle of the 'invisible hand,' and thus imply inefficiencies."[61] In the face of this, Franklin Fisher and von Weizsäcker developed normative statements regarding barriers to entry.[62] Fisher found "a barrier to entry exists when entry would be socially beneficial but is somehow prevented."[63] von Weizsäcker stated, "a barrier to entry is a cost of producing which must be borne by a firm which seeks to enter an industry but is not borne by firms already in the industry and which implies a distortion in the allocation of resources from the social point of view."[64] This is identical to Stigler,

---

[59] *See* W. Kip Viscusi, John M. Vernon & Joseph E. Harrington, Jr., Economics of Regulation and Antitrust 168 (MIT Press 4th ed. 2005).

[60] *See* George J. Stigler, The Organization of Industry 67 (Univ. of Chicago Press 1983) (1968).

[61] C. C. von Weizsäcker, *A Welfare Analysis of Barriers to Entry*, 11 Bell J. Econ. 399, 399 (1980).

[62] Franklin M. Fisher, *Diagnosing Monopoly*, 19 Q. Rev. Econ. & Bus. 7 (1979); von Weizsäcker, *A Welfare Analysis of Barriers to Entry*, *supra* note 61.

[63] Fisher, *Diagnosing Monopoly*, *supra* note 62, at 23.

[64] von Weizsäcker, *A Welfare Analysis of Barriers to Entry*, *supra* note 61, at 400.

except for the last clause invoking welfare considerations. He uses economies of scale to illustrate his point. Namely, if capturing economies of scale can increase overall welfare *and* we associate entry barriers with inefficiencies, then, "in which sense can we speak of [exploiting scale economies to be] a barrier to entry?"[65] The virtue of von Weizsäcker's and Fisher's formulations is that, when policy conclusions are on the line, it is imperative that welfare considerations are unequivocally stated.

What both Fisher and von Weizsäcker do is peel back the layers in defining and, more importantly, in understanding the role that market features such as economies of scale and product differentiation play in determining the market equilibrium. Fisher illustrates this idea when he states, "the right issue is not whether there are barriers to entry into the production of a particular mousetrap, but whether there are barriers to entry into innovation in mousetraps."[66] In other words, we should not be narrowly focused on a particular product or approach a firm uses—for example, one that involves use of a certain volume or type of big data in particular ways—but on the larger question of whether other viable approaches to entry are hindered and, as stated earlier, whether this hindrance results in a loss of welfare.

In competition policy, the term "barriers to entry" is generally used in one of two ways. Dennis Carlton summarizes this idea when he stated, "[t]rying to use 'barriers to entry' to refer to both the factors that influence the time it takes to reach a new equilibrium and to whether there are excess long-run profits is confusing."[67] The former definition is consistent, for instance, with the analysis of entry in the U.S. Horizontal Merger Guidelines.[68] The latter definition is more in line with the economic literature on barriers described above. The danger in labeling a factor as a "barrier to entry" is a lack of clarity regarding which definition one is considering.

Entry into almost all highly valued markets involves some hurdles that must be overcome and perhaps high costs that need to be incurred in order to compete successfully. Common examples include regulatory compliance costs, developing intellectual property, expenditures on specialized equipment, and hiring skilled labor. Labeling all these factors as barriers to entry effectively renders the term meaningless. Hal Varian highlights this point

---

[65] *Id.* at 401.

[66] Fisher, *Diagnosing Monopoly*, *supra* note 62, at 57.

[67] Dennis W. Carlton, *Barriers to Entry*, 1 Issues Competition L. & Pol'y 601, 606 (2008); *see also* Viscusi, Vernon & Harrington, Jr., Economics of Regulation and Antitrust, *supra* note 59, at 168 ("There is perhaps no subject that has created more controversy among industrial organization economists than that of barriers to entry. At one extreme, some economists argue that the only real barriers are government related. . . . At the other end of the spectrum, some economists argue that almost any large expenditure necessary to start up a business is a barrier to entry.").

[68] U.S. Department of Justice & Federal Trade Commission, Horizontal Merger Guidelines § 9 (2010) [hereinafter Horizontal Merger Guidelines], https://www.ftc.gov/sites/default/files/attachments/merger-review/100819hmg.pdf ("The Agencies examine the timeliness, likelihood, and sufficiency of the entry efforts an entrant might practically employ.").

when he states, "in starting a new business, is the problem with data or with knowledge? For example, I would like to enter the automobile manufacturing industry. Unfortunately, I know nothing whatsoever about how to build an automobile. Should that be considered a barrier to entry?"[69] Either the term barriers to entry is explicitly stated and the welfare consequences evaluated, or, as Carlton recommends, "rather than focusing on whether an entry barrier exists according to some definition, analysts should explain how the industry will behave over the next several years . . . [which] will force them to pay attention to uncertainty and adjustment costs."[70]

For the reasons just discussed, it makes little sense to label big data as a barrier to entry and thereby treat it as an inevitable impediment to competition and consumer welfare.[71] First, big data is one of many potential factors that influence "the timeliness, likelihood, and sufficiency of entry efforts an entrant might practically employ."[72] Certainly, effective investing in big data, machine learning, and artificial intelligence can create competitive distance between rivals. Yet this distance is a byproduct of competition on the merits and, as numerous examples confirm, is not necessarily an impediment to entry by firms clever enough to innovate a better mousetrap.[73] Rather than labeling big data as a barrier to entry, the focus should be on assessing what big data helps a firm accomplish, whether an incumbent's use of it furthers innovation and consumer welfare, and whether the incumbent's use of it actually prevents rivals from accessing it themselves or competing effectively in other ways.[74]

---

[69] Varian, Use and Abuse of Network Effects, *supra* note 48 (manuscript at 9).

[70] Carlton, *Barriers to Entry*, *supra* note 67, at 615. Similarly, Harold Demsetz observed that conditions frequently considered barriers to entry, such as scale economies, capital requirements, and advertising expenditures, are not the fundamental source of barriers; the fundamental barriers are rather the cost of information and the uncertainty that an entrant has to overcome. *See* Harold Demsetz, *Barriers to Entry*, 72 Am. Econ. Rev. 47 (1982).

[71] Anja Lambrecht and Catherine Tucker arrive at this conclusion through a slightly different approach. *See* Anja Lambrecht & Catherine Tucker, *Can Big Data Protect a Firm from Competition?*, CPI Antitrust Chron., Jan. 2017, at 1, 8 ("For a wide range of examples from the digital economy we demonstrate that when firms have access to big data, at least one, and often more, of the four criteria which are required for a resource to constitute a sustainable competitive advantage are not met."); *see also* Sokol & Comerford, *Antitrust and Regulating Big Data*, *supra* note 9, at 1135–40.

[72] Horizontal Merger Guidelines, *supra* note 68.

[73] Competition Bureau Canada, Big Data and Innovation: Implications for Competition Policy in Canada, *supra* note 12, at 14 ("[D]eveloping valuable data through competition on the merits does not run afoul of [Canada's Competition] Act even if it results in significant market power. For example, a firm can create market power by developing a high-quality product or an efficient production process[.]").

[74] Further, significant expenditures on big data collection and analytics can result in high fixed costs; thus, merely observing that a firm is pricing above marginal cost and equating this with supracompetitive pricing is not properly considering the rate of return required to cover all the prior periods' expenditures on big data. *See* Michael L. Katz & Carl Shapiro, *Systems Competition and Network Effects*, 8 J. Econ. Persp., Spring 1994, at 93, 107 ("[M]erely observing a firm with a position of market dominance does not imply that the firm is earning super-normal profits: the firm's quasi-rents may merely reflect costs incurred earlier to obtain the position of market leadership.").

Finally, it is worth noting that if big data is a barrier to entry that lowers social welfare, then this idea is in some tension with related theories of harm involving incumbent firms, particularly digital platforms, that are allegedly acquiring smaller potential or current rivals in order to preempt future competition.[75] If big data is such a significant barrier to entry, one might ask, how would these new entrants be able to threaten the incumbent acquirer's strong position in the marketplace? This line of inquiry leads to a number of critical implications. First, in markets characterized by big data, we should not expect to see entry that would be considered likely, timely, and sufficient. Second, in markets characterized by big data, nascent competitors (holding aside for the moment that their mere existence violates the prior implication) cannot be expected to represent a serious competitive threat to incumbents. Consequently, following this logic, acquisitions involving so-called "nascent competitors" and incumbents employing big data are likely procompetitive. We can perhaps label this the "big data paradox." The reality is that theories of harm based on data barriers to entry confuse data with knowledge.[76] Further, it illustrates the fundamental fragility of defining big data as a barrier to entry in the context of competition policy.

IV. Big Data and Competition Matters

It has been at least a decade since the emergence of big data as a relevant competition issue. During that time, competition agencies in the United States and Europe have considered various aspects of big data in matters ranging from mergers to unilateral conduct by firms classified as "dominant." Yet, despite the flurry of policy research calling for antitrust condemnation

[75] Facebook's acquisition of Instagram is considered the canonical example of an entrenched incumbent preempting future competition. Yet, this example likely suffers from a nirvana counterfactual fallacy. *See* John M. Yun, Douglas H. Ginsburg, Joshua D. Wright & Tad Lipsky, *Comment of the Global Antitrust Institute, George Mason University School of Law, on the Australian Competition & Consumer Commission's Digital Platforms Inquiry, Preliminary Report* 15 (Geo. Mason L. & Econ. Research, Paper No. 19-04, 2019) ("Central to the narrative that strategic acquisitions have entrenched market power is Facebook's acquisition of Instagram in 2012, from which some commentators have inferred that competition authorities are missing potential competition cases. At the time of the acquisition, Instagram had zero revenues and a handful of employees. Since Facebook's acquisition, Instagram has grown from 30 million users to well over one billion. During the same period, Facebook grew from approximately 900 million users to over two billion users. This substantial expansion in users and output is hardly indicative of an anticompetitive outcome. Of course, one could argue that, but for the acquisition, Instagram would have been just as successful, if not more, and would have remained an independent competitor. While this type of 'nirvana' counterfactual is frequently asserted, without more it is not a sufficient basis upon which retrospectively to condemn an acquisition. To treat the success and associated output expansion of an acquired product as evidence of an anticompetitive acquisition severely twists the meaning of 'anticompetitive.'").

[76] *See* Varian, Use and Abuse of Network Effects, *supra* note 48 (manuscript at 9–10) ("Knowledge is a critical part of production. In economic models, knowledge is embedded in the production function but in the real world, knowledge is embedded in people. . . . Today's successful businesses had no data when they started, but nevertheless they were able to acquire sufficient expertise to gain a foothold, gather what data they could, and extract information and knowledge from that data to create competitive advantage over the incumbents.").

of big data, competition agencies have yet to bring a challenge based primarily on big data considerations—other than as a tangential reference as a barrier to entry. What does this lack of agency challenge tell us? While it does not prove that big data cannot possibly present a competition issue, it certainly supports the proposition that big data is unlikely often to present one. Moreover, an assessment of the agencies' decisions in matters that involve firms utilizing big data fails to support the conjecture that competition agencies and the courts are falsely and systematically neglecting the issue entirely. Rather, in a number of cases, the agencies have fully considered big data issues and ultimately determined that they were an insufficient and inadequate basis upon which to issue a formal complaint.

I examined a number of U.S. and European cases involving big data to better understand why it has not been the centerpiece of any antitrust challenge. The point is not to offer a comprehensive overview of all competition cases involving big data, but to identify a number of highly prominent and significant cases that can help inform the issue. Of course, there could be a big data challenge at some point in the future. Even so, it would only support the proposition that big data concerns can exist and be taken seriously, not that they are typical or frequent. Further, in contrast with the claims of some who have expressed great concern over big data and the ability of competition agencies to deal with it, Joshua Wright and Elyse Dorsey argue that, if there are legitimate concerns, they can be assessed with the standard tools that agencies employ.[77]

One of the world's largest commercial users of big data is Google. In 2011, the U.S. Federal Trade Commission (FTC) began to formally investigate whether Google was engaging in anticompetitive conduct with its online search engine. Specifically, the agency scrutinized Google's "universal search" feature, which provides distinct, specialized search results in various categories such as for shopping queries (for example, listing stores that sell a particular product) and local business queries (for example, providing a map of local florists and their addresses). These specialized search results were interspersed with the standard links to web sites—that is, "blue links"—to form a composite search results page. The allegation was that Google demoted links to competing web sites that offered similar specialized search information—for example, shopping and local business results. Further, the allegation was that Google unfairly promoted its own specialized search results at the top of the page, which deprived traffic to rival sites that offered similar content. There has been no shortage of commentary on the case, both before

[77] Wright & Dorsey, *Antitrust Analysis of Big Data*, *supra* note 30, at 35 ("[T]he existing antitrust framework is well equipped with sophisticated tools and methods for analyzing just the kinds of issues big data presents.").

and after the FTC's ultimate decision to close the investigation.[78] Yet despite the fact that Google analyzes a staggering volume of data, that fact alone—or even in combination with other concerns—proved to be an insufficient basis upon which to bring a complaint. In fact, the FTC concluded just the opposite, when it considered Google's implementation of universal search to be procompetitive.[79] The lesson from the investigation is that competition cases require actual evidence of anticompetitive conduct and resulting harm.

In 2013, the U.S. Department of Justice (DOJ) challenged Bazaarvoice's acquisition of PowerReview, where "Bazaarvoice provides the market-leading R&R [Rating & Reviews] platform to manufacturers and online retailers."[80] In its complaint, the DOJ cited a document indicating that Bazaarvoice identified its "ability to leverage the data" as "a key barrier [to] entry."[81] After the court ruled in favor of the DOJ, the DOJ stated, "[t]he entry barriers identified by the Court include networks effects from syndication, switching costs, moderation, analytics, and reputation." While it is certainly true that the DOJ and the court considered data and analytics (presumably from big data) as a barrier to entry, they are only mentioned twice in these key DOJ documents; additionally, analytics is listed among a series of other "barriers" including switching costs and reputation.[82] Rather, the deciding factors in bringing the case were the standard litany of potential harms from horizontal mergers, including the fact that "PowerReviews' 'scorched earth approach to pricing' applied significant pressure to Bazaarvoice in competitive deals."[83]

In 2007, both the United States and Europe scrutinized, but eventually cleared, Google's acquisition of DoubleClick. The European Commission (EC) considered a theory of harm involving foreclosure based on the combination of Google and DoubleClick's assets. The essence of the theory was

[78] *See, e.g.*, Geoffrey A. Manne & Joshua D. Wright, *Google and the Limits of Antitrust: The Case Against the Antitrust Case Against Google*, 34 Harv. J.L. & Pub. Pol'y 1, 3 (2011); Michael A. Salinger & Robert J. Levinson, *Economics and the FTC's Google Investigation*, 46 Rev. Indus. Org. 25 (2015); Richard J. Gilbert, *U.S. Federal Trade Commission Investigation of* Google Search *(2013)*, *in* The Antitrust Revolution 489 (John E. Kwoka, Jr. & Lawrence J. White eds., Oxford Univ. Press 7th ed. 2018); John M. Yun, *Understanding Google's Search Platform and the Implications for Antitrust Analyses*, 14 J. Competition L. & Econ. 311 (2018).

[79] *See* Statement of the Federal Trade Commission Regarding Google's Search Practices *In the Matter of Google Inc.* at 2, FTC File No. 111-0163 (FTC Jan. 3, 2013), https://www.ftc.gov/system/files/documents/public_statements/295971/130103googlesearchstmtofcomm.pdf ("[T]he . . . quantitative evidence the Commission examined are largely consistent with the conclusion that Google likely benefited consumers by prominently displaying its vertical content on its search results page.").

[80] Competitive Impact Statement at 3, United States v. Bazaarvoice, No. 13-cv-00133, 2014 WL 203966 (N.D. Cal. Jan. 8, 2014), 2014 WL 2142181.

[81] Complaint at 18, *Bazaarvoice*, No. 13-cv-00133, 2014 WL 203966, 2013 WL 127168 [hereinafter Complaint, *Bazaarvoice*].

[82] As discussed in Part III, the idea that switching costs and reputation are barriers to entry further supports the point made by von Weizsäcker, Fisher, and Demsetz that the use of the term "barriers to entry" is generally used in a very imprecise manner and often divorced from dynamic welfare considerations.

[83] Complaint, *Bazaarvoice*, *supra* note 81, at 13. Further, if Bazaarvoice's ownership of big data served as a strong barrier to entry, how did the acquired firm manage to become such a threat? Rather, Bazaarvoice was successful because it was highly skilled in using the data that it collected.

that combining Google's user data with DoubleClick's user data "would allow the merged entity to achieve a position that could not be replicated by its integrated competitors. . . . Google's competitors would be progressively marginalised which would ultimately allow Google to raise prices for its intermediation services."[84] The EC also recognized that "[s]uch information could potentially be used to better target ads to users."[85] The EC ultimately concluded, "[e]ven if Google's and DoubleClick's data collections were available as input for DoubleClick it would therefore be unlikely that its competitiveness would be enhanced in a way that would confer on the merged entity a competitive advantage that could not be matched by its competitors."[86] This led the EC to the conclusion that "the possible combination of data of Google and DoubleClick post-merger is very unlikely to bring more traffic to AdSense so as to squeeze out competitors and ultimately enable the merged entity to charge higher prices for its intermediation services."[87] Thus, it is clear that Europe explicitly considered and ultimately rejected a theory of harm where data played a central role—as the evidence did not support such concerns.

In the United States, in a 4-to-1 vote, the FTC cleared the Google-DoubleClick merger. In doing so, the Commission explicitly considered and rejected "the theory that the combination of their respective data sets of consumer information could be exploited in a way that threatens consumers' privacy."[88] Relatedly, the Commission "assessed the suggestion that the combination of Google's database of user information and the data respecting users and competitive intermediaries collected by DoubleClick on behalf of its customers would give Google an overwhelming advantage in the ad intermediation market."[89] Ultimately, the Commission concluded that "the transaction is not likely to create, enhance, or facilitate market power."[90] In other words, big data considerations were explicitly considered but the evidence did not support such an action. In fact, the agency explicitly stated:

> At bottom, the concerns raised by Google's competitors regarding the integration of these two data sets—should privacy concerns not prevent such integration—really amount to a fear that the transaction will lead to Google offering a superior product to its customers. Yet, the evidence indicates that neither the data available to Google, nor the data available to DoubleClick,

[84] Commission Decision Declaring a Concentration to Be Compatible with the Common Market and the Functioning of the EEA Agreement ¶ 359, Case No. COMP/M.4731—Google-DoubleClick, C (2008) 927 final (Mar. 11, 2003).

[85] *Id.* ¶ 360.

[86] *Id.* ¶ 364.

[87] *Id.* ¶ 366.

[88] Statement of Federal Trade Commission Concerning Google/DoubleClick at 2, FTC File No. 071-0170 (FTC Dec. 20, 2007).

[89] *Id.* at 12.

[90] *Id.* at 2.

> constitutes an essential input to a successful online advertising product. A number of Google's competitors have at their disposal valuable stores of data not available to Google.[91]

Similarly, in approving Facebook's acquisition of WhatsApp in 2014, the European Commission concluded that, "regardless of whether the merged entity will start using WhatsApp user data to improve targeted advertising on Facebook's social network, there will continue to be a large amount of Internet user data that are valuable for advertising purposes and that are not within Facebook's exclusive control."[92] Additionally, for Microsoft's acquisition of LinkedIn in 2016, the EC concluded that "the Transaction does not raise competition concerns resulting from the possible post-merger combination of the 'data' (essentially consisting of personal information, such as information about an individual's job, career history and professional connections, and/or her or his email or other contacts, search behaviour etc. about the users of their services) held by each of the Parties in relation to online advertising."[93]

Finally, in 2008, while the EC ultimately cleared TomTom's acquisition of Tele Atlas,[94] the Directorate-General for Competition concluded that entry into navigable digital map databases was "unlikely" due to the "vast volumes of data [that] have to be collected from various sources and field survey teams."[95] That conclusion was soon to be proved by the market to be incorrect. As Xavier Boutin and Georg Clemens highlight, "[i]n fact, using its competitive advantages relating to variety and velocity, Google soon started to offer navigation for free. Google Maps Navigation integrated a series of features . . . using the data that was collected from Google's search engine. Moreover, it had the significant advantage of relying on the updates sent by its users. The combination of a high volume and variety of data updated at a

---

[91] *Id.* at 12.

[92] Commission Decision Pursuant to Article 6(1)(b) of Council Regulation No. 139/2004 ¶ 189, Case M.7217—Facebook/WhatsApp, C (2014) 7239 final (Oct. 3, 2014).

[93] Commission Decision Pursuant to Article 6(1)(b) in Conjunction with Article 6(2) of Council Regulation No. 139/2004 and Article 57 of the Agreement on the European Economic Area ¶ 176, Case M.8124—Microsoft/LinkedIn, C (2016) 8404 final (Dec. 6, 2016); *see also id.* ¶ 180 ("In the case at hand, however, the [Microsoft-LinkedIn] Transaction does not give rise to this type of concerns in relation to online advertising. First, Microsoft and LinkedIn do not make available their data to third parties for advertising purposes, with very limited exceptions. Second, the combination of their respective datasets does not appear to result in raising the barriers to entry/expansion for other players in this space, as there will continue to be a large amount of internet user data that are valuable for advertising purposes and that are not within Microsoft's exclusive control. Third, the Parties are small market players and compete with each other only to a very limited extent in online advertising and its possible segments.").

[94] The EC found that the vertical transaction did not raise concerns of input foreclosure or a greater likelihood of anticompetitive coordinated conduct. *See* Commission Decision Declaring a Concentration to Be Compatible with the Common Market and the EEA Agreement, Case No. COMP/M.4854—TomTom/Tele Atlas, C (2008) 1859 (May 14, 2008).

[95] *Id.* ¶ 132.

high velocity caught incumbents such as TomTom off-guard."[96] The EC inaccurately forecasted the future market environment because it believed big data insulated TomTom from competition; yet, it was a different type of big data that helped fuel Google's entry. This episode of disruptive entry further suggests that caution should be exercised when attempting to determine whether or not entry will be precluded by incumbent providers' access to and use of a specific type of big data.

Conclusion

Calls for a strong presumption of harm, or, at the extreme, recommendations to make the possession and use of big data by firms with considerable market power akin to a *per se* antitrust violation, are simply not supported either by a consensus in the academic literature, by real world marketplace experiences, or by significant past investigations into firms possessing and using big data. It follows that there is no evidence to justify extraordinary scrutiny or strong presumptions of big data as an antitrust concern. Since initial concerns regarding big data were first expressed, experience has shown that big data is not a particularly scarce resource in the digital economy—but rather it is the ability to analyze it in a manner that unlocks value that can be in relatively short supply.[97] Most certainly, calls to mandate that data be shared, and that it be declared to be some sort of "essential facility," are wrought with further problems and would open a Pandora's box of unintended consequences—including those associated with setting appropriate terms and conditions under which sharing is to be performed—not to mention the likely conflict such policies may raise with intellectual property laws.

[96] Boutin & Clemens, Big But Not Insurmountable? How the Definition of 'Big Data' Can Help in the Assessment of Entry, *supra* note 39, at 2.

[97] *See* Turck, *Is Big Data Still a Thing? (The 2016 Big Data Landscape)*, *supra* note 4 ("As Big Data continues to mature, however, the term itself will probably disappear, or become so dated that nobody will use it anymore. It is the ironic fate of successful enabling technologies that they become widespread, then ubiquitous, and eventually invisible.").