CAN BIG DATA PROTECT A FIRM FROM COMPETITION?

BY ANJA LAMBRECHT & CATHERINE E. TUCKER[1]

I. INTRODUCTION

The digitization of the offline and online economy alike means that firms are naturally collecting "big data," distinguished by its volume,[2] variety of formats spanning text, image and video and velocity, meaning that data is recorded in real time.[3]

There is much hype surrounding big data. Firms are constantly exhorted to set strategies in place to collect and analyze big data, and warned about the potential negative consequences of not doing so. For example, the Wall Street Journal recently suggested that companies sit on a treasure trove of customer data but for the most part do not know how to use it.[4]

However, despite the excitement surrounding big data, its long-term strategic, rather than operational, implications for firms are less clear. Some observers have concluded that big data may lead to a new type of competitive advantage.[5] But others have questioned whether this is indeed the case.[6] The question of whether big data can confer a sustainable competitive advantage to a firm has, to our knowledge, received surprisingly little systematic attention. However, understanding the potential strategic implications of big data is important for firms who want to comprehend whether ownership of big data can protect their business from current or future competition.

To evaluate the strategic role of big data as a source of sustainable competitive advantage or as a barrier to entry, we use a classic framework in strategic management sometimes referred to as the resource-based view of the firm. This literature is useful because it sharply distinguishes factors that enhance an entire industry from a "sustained competitive advantage" that benefits a single firm. For there to be a sustainable competitive advantage, the firm's rivals must be unable realistically to duplicate the benefits of this strategy or input. Specifically, for a firm resource to be a source of competitive advantage, the resource has to be inimitable, rare, valuable and non-substitutable.[7]

II. IS BIG DATA INIMITABLE?

For big data to be inimitable, no other firm should easily be able to replicate the advantage. There are two underlying economic reasons for why big data in many instances is unlikely to be inimitable. First, big data is non-rivalrous, meaning consumption of

---

1 Anja Lambrecht is an Assistant Professor at London Business School, London NW1 4SA, UK. Catherine Tucker is the Distinguished Professor of Management Science at MIT Sloan School of Management, MIT, Cambridge, MA 02139, USA. The authors thank the Computer and Communications Industry Association for generous funding of this research. All mistakes are our own.

2 Companies such as Amazon and Walmart already work with petabytes of data in a single data set.

3 Traditionally definitions of big data have focused on its functional characteristics such as volume, variety and velocity rather than the nature of consumer insights it provides. This means that big data spans anonymized user data, personally identifiable information, search query data, web browsing data or data on consumer sentiments or purchase intentions. We recognize that depending on the specific type of data under consideration, the precise implications with respect to how it is of value to the firm may differ. One aim of the framework we set out is to provide firms with a structure that can guide the analysis of whether their "big data" provides a sustainable competitive advantage.

4http://www.wsj.com/articles/the-untapped-value-of-customer-data-1444734633?mod=djem_jiewr_MK_domainid.

5 See for example McGuire, T., J. Manyika, and M. Chui (2012),"Why big data is the new competitive advantage," Ivey Business Journal 76 (4), 1-4.

6 See: https://hbr.org/2015/01/why-nordstroms-digital-strategy-works-and-yours-probably-doesnt. This article highlights that because digital technologies are visible and accessible to competitors, it is hard to generate a competitive advantage.

7 Barney, J. (1991),"Firm resources and sustained competitive advantage," Journal of Management 17 (1), 99-120.

the good does not decrease its availability to others. Second, big data has near-zero marginal cost of production and distribution even over long distances. These two basic characteristics, combined with the fact that customers constantly leave footprints on the internet, have led to a thriving industry where consumer big data is resold.

This type of commercially available big data typically has broad reach and coverage, allowing many firms whose business does not usually generate big data to gain insights similar to those available to firms that own big data on a large number of customers. There are many examples for very big commercially available data sets. Acxiom has "multi-sourced insight into approximately 700 million consumers worldwide" with over 1,600 pieces of separate data on each consumer and Datalogics asserts that its data "includes almost every U.S. household."[8] Comcast is planning to license TV viewing data collected through set top boxes and apps.[9] Other companies, such as the Oracle-owned Bluekai, sell cookie-based user information online to allow for targeting advertising based on a user's past activities or demographics. Bluekai states that it has data on "750 million unique users per month with an average of 10-15 attributes per user."[10] To protect both their customers and themselves, such companies ensure that their data collection is done in full compliance with data protection rules.

Given the different possible types of big data, an obvious question is whether this analysis extends to cases where the big data has what appears to be unique or individual insights. For example, recently the retail store Target hit the headlines because of its alleged ability to use its retail shopping data to predict a pregnancy even before close relatives knew about it.[11] However, even such highly specific and timely data-driven insights are easy to imitate for firms that do not own a national database of retail sales. For example, a marketing unit of the credit-scoring agency Experian sells frequently updated data on expecting parents, along with income and first-birth information.[12]

In addition, data that is available due to individual consumer-level tracking is complemented by the explosion of user-generated content where consumers themselves create a footprint of their behavior, likes, opinions and interests across the internet. Recent research in computer science has emphasized that by combining a myriad of external online profiles external firms can gain huge insights into any one customer. Firms can also use such content as a direct substitute for customer data. For example, Zillow.com was able to build a successful home-buying digital platform by relying on existing town assessment data.

In short, where a market for data exists it is unlikely that big data is inimitable.

III. IS BIG DATA RARE?

For Big Data to be a "rare" resource would mean that few other firms possess it. However, there are two reasons why this is unlikely to hold. First, large shifts in supply infrastructure have rendered the tools for gathering big data commonplace. Cloud-based resources such as Amazon, Microsoft and Rackspace make these tools not dependent on scale[13] and storage costs for data continue to fall, so that some speculate they may eventually approach zero.[14] This allows ever smaller firms to have access to powerful and inexpensive computing resources. Furthermore, free open source technologies such as Hadoop that allow users to analyze large datasets are widely available and accessible.

Second, as consumers' lives increasingly shift to the web, consumers leave traces of their needs and preferences ev-

8 See Acxiom Corp., 2013 10K Annual Report for the Period Ending March 31, 2013 and Staff of S. Comm. on Commerce, Sci., and Transp., Office of Oversight & Investigation, A Review of the Data Broker Industry: Collection, Use and Sale of Consumer Data for Marketing Purposes.

9 See: http://www.wsj.com/articles/comcast-seeks-to-harness-trove-of-tv-data-1445333401.

10 See: https://docs.oracle.com/cloud/latest/daasmarketing_gs/DSMKT/GUID-418EDA59-1BD9-40F6-9D57-DD7C266555FF.htm#DSMKT3616.

11 See: http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?_r=0. Note, however, there are some doubts over the origin of this story and whether Target actually did this: http://www.kdnuggets.com/2014/05/target-predict-teen-pregnancy-inside-story.html.

12 See: http://www.experian.com/small-business/prenatal-lists.jsp.

13 See: http://betanews.com/2014/06/27/comparing-the-top-three-cloud-storage-providers/.

14 See: http://www.enterprisestorageforum.com/storage-management/can-cloud-storage-costs-fall-to-zero-1.html.

erywhere. Firms who embrace these low-cost digital technologies have many opportunities to gather customer data. Telecom companies can collect data on calling behavior and browsing on their phones; Amazon, Macy's and Walmart collect detailed consumer-level purchase data, while platforms such as Bluekai collect a large range of detailed consumer browsing and purchasing information across multiple websites.[15]

Indeed, such "multi-homing," that is the use of multiple different digital services by consumers, means that similar pieces of information are often available to many different companies. Take, as an example, consumers who use multiple online social media such as Facebook, Twitter, LinkedIn or Instagram and share broadly similar information through each of them. Or, consider the access to information in the app ecosystem: many apps, and not only those related to location or weather, regularly ping location data – as many as hundreds of times a week – meaning that a user's location is always available to a wide range of firms. Of course, as we later discuss, these firms will still have to invest in ensuring that they have the technical skills to transform this data into valuable insights.

Seeing that big data is not inimitable or rare, we turn to the question of whether and when big data is valuable for firms.

IV. IS BIG DATA VALUABLE?

Much of the current managerial literature is focused on whether or not big data is indeed valuable for firms in that it enhances a firm's ability to have profitable relationships with customers. There are three open problems currently challenging analysts and researchers faced with ensuring that big data is valuable to organizations. We discuss these challenges in turn and conclude that by itself big data is not sufficient to create profit-enhancing opportunities.

The first challenge limiting the value of big data to firms is compatibility and integration. One of the key characteristics of big data is that it comes from a "variety" of sources. However, if this data is not naturally congruent or easy to integrate, the variety of sources can make it difficult for firms to indeed save cost or create value for customers. Such hindrances may prove particularly burdensome in industries such as healthcare, where prior research has shown that firms have strategic incentives to ensure that data is siloed and hard to integrate.

The second challenge to making big data valuable is its unstructured nature. Specialized advances are being made in mining text-based data, where context and technique can lead to insights similar to that of structured data, but other forms of data such as video data are still not easily analyzed. One example is that, despite state-of-the-art facial recognition software, authorities were unable to identify the two bombing suspects for the Boston Marathon from a multitude of video data, as the software struggled to cope with the full-frontal nature of the photo of their faces.[16]

Given the challenges of unstructured data, firms tend to find big data most valuable when it augments the speed and accuracy of existing data analysis practices. In oil and gas exploration, big data is used to enhance existing operations and data analysis surrounding seismic drilling. However, engineers have been using massively parallel processing capabilities of high-performance computing to perform analysis on large quantities of data for decades. In other words, though big data may be a new label for such practices, and the volume of data may have increased, such big data is valuable in oil and gas as an extension of existing practices and infrastructure. In general, for the large majority of firms, their ability to analyze the "variety" of types of big data does not yet match the ability to record its volume and velocity.

The third challenge, and in our opinion the most important factor that limits how valuable big data is to firms, is the difficulty of establishing causal relationships within large pools of overlapping observational data. Very large data sets usually contain a number of very similar or virtually identical observations that can lead to spurious correlations and as a result misguide man-

15 The European Commission spoke similarly in 2014 when concluding its investigation into Facebook's acquisition of WhatsApp. It concluded that "there are currently a significant number of market participants that collect user data alongside Facebook, including Google, Apple, Amazon, eBay, Microsoft, AOL, Yahoo, Twitter, IAC, LinkedIn, Adobe and Yelp and that, in addition, 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." See Case No COMP/M.7217 - *FACEBOOK/ WHATSAPP*.

16 See: http://www.wired.com/2013/05/boston-marathon-investigation/.

agers in their decision making. The Economist recently pointed out that "in a world of big data the correlations surface almost by themselves"[17] and a Sloan Management Review blog post emphasized that while many firms have access to big data, such data is not "objective,"[18] since the difficulty lies in distilling "true" actionable insights from the data. Similarly, typical machine learning algorithms used to analyze big data identify correlations that may not necessarily offer causal and therefore actionable managerial insights. Recent work suggests that machine learning algorithms should be used as a "guide to further investigation" in order that we might be able to "predict the effect of our actions."[19] In other words, the skill in making big data valuable is being able to move from mere observational correlations to correctly identifying, potentially outside of big data, what correlations should form the basis for strategic action.

To take a specific example, imagine a shoe retailer that advertises to consumers across the web who have previously visited their website. Raw data analysis would suggest that customers exposed to these ads are more likely to purchase shoes. However, these consumers, who have previously visited the website have already demonstrated their interest in the specific retailer even prior to viewing the ad, and so are more likely than the average consumer to purchase. Was the ad effective? It is hard to say. Indeed, big data here does not allow any causal inference about marketing communication effectiveness. To understand whether such ads are effective, the retailer needs to run a randomized test or experiment, where one subset of consumers are randomly not exposed to the ad. By comparing the purchase probabilities across consumers who were exposed to the ad and those who were not, the company can then determine whether exposing consumers to an ad made them more likely to buy. Value is delivered in such instances not primarily by the access to data, but by the ability to design and implement meaningful experiments.

Therefore the primary avenue by which a firm can understand whether a data relationship is merely correlational or might be predictive (because it is causal) is through experimentation. While it may be challenging for a manager to improve profitability using even one petabyte of observational data describing customer behavior, comparing the behavior of a customer who was exposed to a marketing activity to that of a customer who was by chance unexposed may lead a marketer to conclude whether the activity was profitable.
Implementing field experiments, drawing the right conclusion and taking appropriate action is not necessarily easy.[20] But successful companies have developed the ability to design, implement, evaluate and then act upon meaningful field experiments. It is this "test and learn"environment, coupled with the skill to take action on the insights, which can make big data valuable.[21]

However, because of diminishing returns to increasingly large data samples, such experimentation does not necessarily require big data. For example, Google reports that it typically uses random samples of 0.1 percent of available data to perform analyses.[22] Indeed, a recent article suggested that the size of big data can actually be detrimental as "the bigger the database, the easier it is to get support for any hypothesis you put forward."[23] In other words, because big data often offers overlapping insights, a firm can get similar insight from one-thousandth of the full dataset as from the entire dataset.

Experimentation is not the only method companies can use to infer valuable insights from big data. Another potential skill firms can develop is the ability to build better algorithms to deal with big data. One example for such algorithms is recommender systems. Recommender systems rely on algorithms trained on correlational data to recommend the most relevant products to a customer. Yet, again, it is not the size of the underlying data, but the ability to identify the critical pieces of information that best predict a customer's preferences. For example, it has been shown that to predict preferences for movies, ten movie ratings

17 The Economist,(2010). Data, data everywhere. The Economist Newspaper Limited.

18See: http://sloanreview.mit.edu/article/for-better-decision-making-look-at-facts-not-data/.

19 Domingos, P. (2012, October). A few useful things to know about machine learning. Commun. ACM 55 (10), 78-87

20 See: https://hbr.org/2015/11/run-field-experiments-to-make-sense-of-your-big-data?utm_campaign=HBR&utm_source=facebook&utm_medium=social.

21 Note that even when using insights from experiments, managers need to carefully consider the scope of any findings and how replicable they will be in different contexts.

22 Varian, H. R. (2014). Big data: New tricks for econometrics. The Journal of Economic Perspectives, 3-27.

23 See: https://www.london.edu/faculty-and-research/lbsr/diie-nov-drowning-in-numbers#.Vk-OZvmrRNO.

alone are more helpful than extensive metadata.[24] Indeed, often not the size of the data but the machine learning algorithm used determine the quality of the results.[25] While predictive power may increase with the size of the data available, in many instances the improvements in predictions show diminishing returns to scale as data sets increase in size.[26]

Our analysis demonstrates that, by itself, big data is unlikely to be valuable. It is only when combined with managerial, engineering and analytic skill in determining the experiment or algorithm to apply to such data that it proves valuable to firms.[27] This suggest for firms the primary challenges lie in determining a big data strategy[28], implementing the systems and tools to analyze the data[29] and adapting organizational capabilities.

Given that our previous analyses suggest that big data is neither rare nor inimitable, we conclude that the search for competitive advantage in the new digital economy should focus on attracting the kind of skilled workers who are able to transform big data into valuable tools.

V. IS BIG DATA NON-SUBSTITUTABLE?

For a resource such as big data to provide a sustainable competitive advantage, there has to be no other means of achieving success in the specific industry. Yet, in the digital world, perhaps more so than offline, there are many examples of firms that came from nowhere and, without any embedded data advantage, were still able to disrupt an industry and attract more customers because of a superior value proposition. In this section, we discuss five settings where alternative firm capabilities have proved to be compelling substitutes to big data and consequently where big data has not been a sufficient sustainable source of competitive advantage.

First, it is natural to focus on an industry where data has, even before the internet, offered operational advantages. The communications industry offers such a case study due to its long history of using extensive data to both improve operations and offer better value to customers. Many traditional communications firms such as AT&T and Verizon as well as newer online firms such as Skype and Facebook have large datasets covering messaging services. However, even though incumbents owned massive data bases, the messaging app WhatsApp became a serious competitor to established messaging and social network services by offering a product that satisfied social media users' latent needs – an easy-to-use interface and an extremely low-cost messaging solution. Even when acquired by Facebook for USD $22 billion, WhatsApp had only 55 employees, suggesting its success was not due to large-scale data analytics capacity.[30] A similar example is Snapchat, which succeeded in competing in this space without access to big data because of its insight that people wanted to share personal information more privately.

Another industry where big data could provide insights into consumer preferences and therefore give advantages to large

24 Pilaszy, I. and D. Tikk (2009). Recommending new movies: even a few ratings are more valuable than metadata. In Proceedings of the third ACM conference on Recommender systems, pp. 93-100. ACM.

25 See: http://www.slideshare.net/xamat/10-lessons-learned-from-building-machine-learning-systems, http://stackoverflow.com/questions/25665017/does-the-dataset-size-influence-a-machine-learning-algorithm.

26 Junqué de Fortuny, Enric, David Martens, and Foster Provost, "Predictive modeling with big data: is bigger really better?" Big Data 1.4 (2013): 215-226.

27 One potential way of evaluating whether this insights holds in a specific context is to examine the relative pricing of data relative to firm processing skills. In contexts where data is cheap relative to processing skills this is suggestive that indeed processing skills are more important than data itself in creating value for a firm.

28 See: http://www.cio.com/article/2395010/data-management/the-big-data-challenge--how-to-develop-a-winning-strategy.html.

29 See: http://sloanreview.mit.edu/article/overcoming-legacy-processes-to-achieve-big-data-success/.

30 See: http://www.forbes.com/sites/parmyolson/2014/10/06/facebook-closes-19-billion-whatsapp-deal/, http://www.businessinsider.com/why-facebook-buying-whatsapp-2014-2?IR=T, http://www.bloomberg.com/news/articles/2014-10-28/facebook-s-22-billion-whatsapp-deal-buys-10-million-in-sales.

digital firms when launching new products, is online gaming. Yet, King Digital Entertainment was not among the dominant digital gaming companies, nor supported by firms with access to big data such as Google and Facebook, when it launched the smartphone hit Candy Crush Saga. By 2014, 93 million people played Candy Crush Saga more than 1 billion times a day.[31] The fact that Candy Crush is playable in short sessions and does not require extensive time investment explains its appeal to the non-gaming population of time-strapped parents, or commuters, "from office juniors through to CEOs."[32] One challenge for new games is discovery to speed up adoption. But when players progress in Candy Crush, Candy Crush displays the progress of the player's Facebook friends, fostering competition in the player's social network to keep them engaging with the game.[33] This example illustrates that a superior value proposition to a new group of consumers can be more important than access to data, even in a sector where companies routinely have access to big data.

Second, it is natural to ask whether there is a substitute for insights from big data in sectors where there has historically been little use of data. It is possible that in such contexts, firms in adjacent sectors who do have big data have an executional advantage in terms of modernizing these sectors. However, the rise of the new "sharing economy"provides evidence that to build up entirely new digital industries in traditional sectors does not require access to big data. Uber and Lyft had no superior access to data compared to established taxi services, but they were better at putting together a product that met consumer needs for a convenient and reliable taxi service. AirBnB entered a highly competitive industry where large travel companies have access to large swathes of data and regularly run experiments to interpret their data in a meaningful way to constantly improve business practices. Yet, despite the lack of data, AirBnB quickly became a dominant player because of its superior value proposition. Google's purchase of ITA along with its flight data and data-processing capabilities did not give Google a significant presence in the flight search market. This contrasts with the growth of Kayak – a travel search engine – which grew from 2004 from a small startup with no user data to being acquired in 2012 by Priceline for USD $1.8 billion.[34] Indeed, recent spectators have argued that for the sharing economy the secret sauce is not data by itself, but instead the systems that such platforms build around ensuring there is "trust and reputation"among users of the platform.[35]

Third, industries where data is important for delivering a personalized experience, and where this personalized system of recommendations is particularly important for customer experience, may be another natural setting where big data might have few substitutes. One obvious example of such an industry is online dating, where the difficulty of predicting human relationships likely puts a premium on the availability of large data sets. However, Tinder entered the online dating market in September 2012 with no access to existing data and quickly became a dominant player with 1.6 billion Tinder profiles, making more than 26 million matches per day (as of April 2015). More than 8 billion matches have been made since Tinder launched.[36]

Tinder succeeded not because of big data but because it offers a better solution for its market. Critically, this included a simple user interface that does not require users to fill out long surveys and personal questions but instead allows quick sign-in with Facebook. It also allows for "liking" (but no rejections) using a simple "swipe right." Another feature that makes Tinder attractive to users is the "double opt-in," that is, both users must agree before they can message each other. These points illustrate that Tinder was very good in understanding how people would like to use dating services and in mirroring offline interactions where normally two people would only strike up a conversation in a bar when there were signs of interest on both sides. This is especially important as on other dating sites women often receive many messages, making them feel overwhelmed, while men receive few messages, making them feel disheartened.[37]

---

31 See: http://www.theguardian.com/technology/2014/mar/26/candy-crush-saga-king-why-popular, https://thinkgaming.com/app-sales-data/2/candy-crush-saga/. While Candy Crush Saga is free to download and play, it makes its money from in-app purchases of extra moves, lives and power-ups, with estimated daily revenues of over USD $700,000, as of November 23, 2015.

32 See: http://www.theguardian.com/technology/2014/mar/26/candy-crush-saga-king-why-popular.

33 See: http://blog.upsight.com/blog/breaking-down-candy-crushs-formula-for-success.

34 See: http://thenextweb.com/insider/2012/11/08/priceline-com-acquiring-travel-company-kayak-for-1-8b-in-cash-and-stocks/.

35 See: http://sloanreview.mit.edu/article/data-at-the-heart-of-the-sharing-economy/?utm_source=facebook&utm_medium=social&utm_campaign=sm-direct.

36 See: https://en.wikipedia.org/wiki/Tinder_(app).

37 See: https://pando.com/2013/08/26/laid-to-paid-how-tinder-set-fire-to-online-dating/.

By allowing women and men to decide who could contact them, Tinder gave them more control over their dating experience. Additionally, the double opt-in reduces non-responses and so avoids feelings of rejections. This stands in contrast to other online dating sites where men or women often send many messages that are not responded to, ultimately demotivating them to continue using the service. Last, the easy swipe to the next profile makes the service more like a game and so more enjoyable to use. Notably, to build up its user base, Tinder did not advertise or use mass emails based on big data bases but hosted "exclusive"parties on college campuses with admittance based on having downloaded the app.[38] By signing up hundreds of available singles in dense geographic areas, Tinder could benefit from more traditional forms of word of mouth communications.

Fourth, another natural place to look for non-substitutability is industries with switching costs and network effects. Switching costs are the costs (both perceived and real) incurred by customers when they switch brands or suppliers. Network effects occur when the usefulness of a product, service or platform increases as more people use it. Historically, switching costs and network effects have been highlighted by economists as potential sources of incumbent competitive advantage, especially in digital environments. Therefore it is natural to ask whether big data in combination with switching costs and network effects might lead to a setting where potential rivals struggle to compete or find sufficient substitutes to compete with. Social network sites exhibit both potential network effects, because consumers value being able to communicate with their friends, and switching costs, as customers invest time and money in curating their online profiles.

However, the history of social networking sites suggests that big data has not protected larger firms in this industry. Rather, this industry has experienced a succession of large firms, even though at each point in time the incumbent had access to big data whereas the new entrant was, in terms of data availability, at a disadvantage. For example, Myspace replaced Friendster and was then replaced by Facebook as the leading social network site. What ultimately made Facebook successful was the ability to build a product that was more focused on customer needs for their social media interactions. This included giving customers more control over their social media interactions (for example Facebook allowed users more control relative to the public nature of MySpace about what content observers could see about a user), and increasing the usability of the site (for example, MySpace was seen by many as too cluttered, Facebook offered a much cleaner design).[39]

Fifth, one potential way that big data could be non-substitutable is if it is necessary for attracting capital investment. However, it is notable that venture capital does not view big data as "non-substitutable," in that it continues to fund startups to compete in spaces where other firms are demonstrably in possession of "big data." For example, despite "Amazon Fresh" and "Google Express"having access through their parent companies to big data about potential customers, there is vibrant funding of new startups that are trying to compete in the local delivery space who do not have this data advantage. For example, Instacart has received USD $275M in funding,[40] Jet has received $220M in funding,[41] and Postmates has received $138M in venture capital funding.[42]

Overall, big data is not a non-substitutable requirement for offering online services, though ownership of big data is often the natural consequence of being successful in offering such online services. Instead, in a similar manner to the offline world, what determines success online is a superior ability to understand and meet customer needs. The unstable history of digital business offers little evidence that the mere possession of big data is a sufficient protection for an incumbent against a superior product offering.

38 See: https://www.quora.com/How-did-Tinder-grow-so-quickly.

39 Decisions on the size, quality and placement of ads on MySpace were less influenced by needs of the users and more by the imperative to monetize the site, leading to an even more ad-cluttered site. For a comprehensive account of what happened to MySpace, see: http://www.bloomberg.com/bw/magazine/content/11_27/b4235053917570.htm#p3.

40 See: https://www.crunchbase.com/organization/instacart\#/entity.

41 See: https://www.crunchbase.com/organization/jet\#/entity.

42 See: https://www.crunchbase.com/organization/postmates\#/entity.

VI. IMPLICATIONS

Can big data confer a sustainable competitive advantage for firms, which can help them persistently deflect current and future competition? To analyze whether big data can act as a barrier to entry in this manner, we use the classic resource-based view of strategic management, which emphasizes that to qualify as a sustainable competitive advantage a resource needs to meet four criteria. It has to be inimitable, rare, valuable and non-substitutable. 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.

Our aim is not to suggest that firms cannot derive benefits from owning and evaluating big data. Instead, we highlight that the simple act of amassing big data by itself does not confer a long-term competitive advantage. We conclude that to build up a competitive advantage related to big data firms need to develop two new competencies.

First, firms need to attract employees who have the ability to develop and train algorithms or to design and/or to set up and run meaningful experiments since it is insights from such efforts that may be able to turn big data into a meaningful competitive advantage. Instead firms needs to develop complementary organizational skills.

Second, firms need to use big data to look forward and understand evolving customer needs rather than simply use past historic big data to make incremental improvements to their current product offering or service. The unstable history of digital business offers little evidence that the mere possession of big data is a sufficient protection for an incumbent against a superior product offering. To build a sustainable competitive advantage, the focus of a digital strategy should therefore be on how to use digital technologies to provide value to customers in ways that were previously impossible.

In addition to our managerial implications this paper also contributes to a policy literature. This literature is concerned with the question whether big data can constitute a barrier to entry which is in a sense the flipside of the question we focus on – whether big data constitutes a competitive advantage. In contrast to this largely legal literature, which grapples with how to frame big data in the context of traditional antitrust analysis, we use a long-established strategic framework to evaluate whether big data indeed merits consideration as a source of sustainable competitive advantage.


CPI
COMPETITION POLICY
INTERNATIONAL