Can Big Data Protect a Firm from Competition?

Anja Lambrecht and Catherine E. Tucker*

December 18, 2015

*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.

Contents

Executive Summary

There is plenty of hype around big data, but does it simply offer operational advantages, or can it provide firms with sustainable competitive advantage? To answer this question, we look at big data using a classic framework called the 'resource-based view of the firm,' which states that, for big data to provide competitive advantage, it has to be inimitable, rare, valuable, and non-substitutable.

Our analysis suggests that big data is not inimitable or rare, that substitutes exist, and that by itself big data is unlikely to be valuable. There are many alternative sources of data available to firms, reflecting the extent to which customers leave multiple digital footprints on the internet. In order to extract value from big data, firms need to have the right managerial toolkit. The history of the digital economy offers many examples, like Airbnb, Uber and Tinder, where a simple insight into customer needs allowed entry into markets where incumbents already had access to big data.

Therefore, to build sustainable competitive advantage in the new data-rich environment, rather than simply amassing big data, firms need to focus on developing both the tools and organizational competence to allow them to use big data to provide value to consumers in previously impossible ways.

1 Introduction

The digitization of the offline and online economy alike means that firms are naturally collecting 'big data', distinguished by its *volume*[1], *variety* of formats spanning text, image and video, and *velocity*, meaning that data is recorded in real time.[2]

There is plenty of hype around big data. Firms are constantly exhorted to set strategies in place to collect and analyze big data (Bughin et al., 2010; Biesdorf et al., 2013), 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.[3] Recent articles such as McGuire et al. (2012) and McAfee et al. (2012) have made cases for why big data offers a short-term operational advantage, both in terms of cost and performance, for firms who find ways of using it successfully.

However, big data's long-term strategic, rather than operational, implications for firms are less clear. Academic opinion differs on whether it will lead to a new type of competitive advantage (McGuire et al., 2012) or not.[4] The question of whether big data can indeed confer sustainable competitive advantage is critical for firms but has, to our knowledge, received surprisingly little systematic attention.

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' (Wernerfelt, 1984; Barney, 1991; Peteraf, 1993; Barney, 2001). This literature is useful because it sharply distinguishes factors that enhance an entire industry from a 'sustainable 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

[1]Companies such as Amazon and Walmart already work with petabytes of data in a single data set (McAfee et al., 2012).

[2]This functional definition of bid data does not specify the depth of consumer insight it can provide. Big data spans anonymized user data, personally identifiable information, search query data, web browsing data or data on consumer sentiments or purchase intentions. Depending on the specific type of data under consideration, how valuable it is to the firm may differ.

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

[4]`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.

this strategy or input. Specifically, Barney (1991) suggests that for a firm resource to be a source of competitive advantage, the resource has to be inimitable, rare, valuable, and non-substitutable. In a similar spirit to Markman et al. (2004)'s analysis of patents, we examine along each of these dimensions whether big data is a source of sustainable competitive advantage to firms.

2 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 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 (Shapiro and Varian, 1999). These two basic characteristics, combined with the fact that customers constantly leave footprints on the internet, have lead to a thriving industry where consumer big data is resold.

This type of commercially available big data allows new entrants to gain insights similar to those available to firms that own big data on a large number of customers. There are many examples of large 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; Datalogix asserts that its data 'includes almost every U.S. household.'[5] Comcast is planning to license TV viewing data collected through set-top boxes and apps.[6] 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.'[7] To protect both their customers and themselves, such companies undertake to ensure that their data collection complies fully with data protection rules.

[5] 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.

[6] http://www.wsj.com/articles/comcast-seeks-to-harness-trove-of-tv-data-1445333401

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

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.[8] 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.[9]

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 (Narayanan and Shmatikov, 2008; Calandrino et al., 2011). Firms can also use such content as a direct substitute for customer data. For example, Edelman (2015) discusses that Zillow.com was able to build a successful home-buying digital platform by relying on existing town assessment data.[10]

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

3 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 (Greenstein et al., 2013). Cloud-based resources such as Amazon, Microsoft, and Rackspace make these tools not dependent on scale[11] and storage

[8] http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?_r=0 There are some doubts over the origin of this story and whether Target actually did this - see for example http://www.kdnuggets.com/2014/05/target-predict-teen-pregnancy-inside-story.html

[9] http://www.experian.com/small-business/prenatal-lists.jsp

[10] He highlights the interaction between publicly available information and user generated content, saying, 'Zillow's initial information was good enough to attract consumer interest, at which point property owners happily contributed corrections, photos, and other information. Indeed, real estate agents were soon willing to pay to show their advertisements in and around Zillow's property listings.'

[11] http://betanews.com/2014/06/27/comparing-the-top-three-cloud-storage-providers/

costs for data continue to fall, so that some speculate they may eventually approach zero.[12] This allows ever smaller firms to have access to powerful and inexpensive computing resources. 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 everywhere. 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 website.[13]

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 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 (Almuhimedi et al., 2015). Of course, as we later discuss, these firms 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.

[12] http://www.enterprisestorageforum.com/storage-management/can-cloud-storage-costs-fall-to-zero-1.html

[13] 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 (Tucker and Wellford, 2014) as well as "Case No COMP/M.7217 - FACEBOOK/ WHATSAPP", http://ec.europa.eu/competition/mergers/cases/decisions/m7217_20141003_20310_3962132_EN.pdf

4 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 (Chen et al., 2012). Cuzzocrea et al. (2011) point to 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 (Miller and Tucker, 2014).

The second challenge to making big data valuable is its unstructured nature. As discussed by Feldman and Sanger (2007), 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.[14]

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, as emphasized by Feblowitz et al. (2013), ‘Geologists, geophysicists, and reservoir engineers have been using massively parallel processing capabilities of high-performance computing (HPC) to perform analysis on petabytes (PB) of data to inform exploration since the late 1990s.’ In other words, though big data may be a new label for such practices, and the volume of data may have

[14] `http://www.wired.com/2013/05/boston-marathon-investigation/`

increased, big data is valuable in oil and gas as an extension of existing practices and infrastructure. Most firms' ability to analyze the 'variety' of types of big data does not yet match their 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 mislead managers in their decision making. The Economist recently pointed out that 'in a world of big data the correlations surface almost by themselves' (Economist, 2010) and a Sloan Management Review blog post emphasized that while many firms have access to big data, such data is not 'objective',[15] 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. Domingos (2012) 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.' 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.

One well-known example of big data is Google Trends, which uses Google's records of aggregate search queries. However, it is also an example of a case where the fact that the data is merely correlational limits its usefulness. Butler (2008) argued that this data could be used to project the spread of flu. However, later researchers found that because the data was backward-looking rather than forward-looking, using search data only marginally improved performance relative to a 'simple autoregressive model' (Goel et al., 2010).[16]

To take a more specific example, imagine a shoe retailer that advertises to consumers who have previously visited their website on other websites. Raw data analysis would suggest that customers exposed to these ads are more likely to purchase shoes. However, these consumers, who have

[15] `http://sloanreview.mit.edu/article/for-better-decision-making-look-at-facts-not-data/`

[16] There were also other critiques of the usefulness of the initial predictive model: (Cook et al., 2011; Lazer et al., 2014)

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 (Lambrecht and Tucker, 2013). Was the ad effective? It is hard to say.[17] 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 experiment, where one subset of consumers is randomly not exposed to the ad.[18] 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, experiments are the main way firms can understand whether a data relationship is merely correlational or might be predictive (because it is causal). Implementing field experiments, drawing the right conclusion and taking appropriate action is not necessarily easy (Lambrecht and Tucker, 2015)[19]. However, 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, that can make big data valuable.[20]

Thanks to 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% of available data to perform analyses (Varian, 2014). 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'.[21] 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

[17] This is emphasized by work such as Lewis et al. (2011) who show that this kind of activity bias, that is the mere fact of being present on a website signalling something about the consumer, makes the use of non-experimental data in assessing advertising effectiveness almost impossible.

[18] Across many industries, field experiments have widely been used to evaluate advertising effectiveness (Lambrecht and Tucker, 2013; Draganska et al., 2014; Lambrecht et al., 2015; Lewis and Rao, 2015; Blake et al., 2015; ?)

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

[20] 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 (Ioannidis, 2005).

[21] `https://www.london.edu/faculty-and-research/lbsr/diie-nov-drowning-in-numbers#.Vk-OZvmrRNO`

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 alone are more helpful than extensive metadata (Pilászy and Tikk, 2009). Indeed, often not the size of the data but the machine learning algorithm used determines the quality of the results.[22] While predictive power may increase with the size of the data available, in many instances the improvements in predictions show diminishing returns to scale (Junqué de Fortuny et al., 2013).

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.[23] This suggests that for firms, the primary challenges lie in determining a big data strategy,[24] implementing the systems and tools to analyze the data[25] and adapting organizational capabilities (McAfee et al., 2012; Bughin et al., 2010).

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.

5 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

[22] 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

[23] One potential way of evaluating whether this insights holds in a specific context is to examine the pricing of data relative to firm processing skills. Data being very cheap relative to processing skills suggests that processing skills are more important than data in creating value for a firm.

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

[25] http://sloanreview.mit.edu/article/overcoming-legacy-processes-to-achieve-big-data-success/

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 sufficiently 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 has long used large data sets to both improve operations (e.g. data network flow) and offer better value to customers (e.g. through pricing plans that meet customer needs, see for example Lambrecht and Skiera (2006); Ascarza et al. (2012)). Many traditional communications firms such as AT&T and Verizon as well as newer online firms such as Skype and Facebook have large data sets covering messaging services. Despite this, 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 $22 billion, WhatsApp had only 55 employees, suggesting its success was not due to large-scale data analytics capacity.[26] 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 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.[27] The fact that Candy Crush is playable in short sessions and does not require extensive

[26] 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

[27] 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 $700,000, as of November 23, 2015

time investment explains its appeal to the non-gaming population of time-strapped parents, or commuters, 'from office juniors through to CEOs'.[28] 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 start up with no user data to being acquired in 2012 by Priceline for $1.8 billion.[29] 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.[30]

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

[28] http://www.theguardian.com/technology/2014/mar/26/candy-crush-saga-king-why-popular

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

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

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.[31] 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 long surveys and allowed users to express interest using a simple game-like 'swipe right' and a 'double opt-in' for matches, where both users must agree before they can message each other. To build up its user base, Tinder did not advertise or use mass emails based on big data bases, but instead hosted 'exclusive' parties on college campuses with admittance based on having downloaded the app.[32]

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 (Farrell and Klemperer, 2007). 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 (Tucker and Marthews, 2011). 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

[31] https://en.wikipedia.org/wiki/Tinder_(app)
[32] https://www.quora.com/How-did-Tinder-grow-so-quickly

customer needs. This included giving customers more control over their social media interactions. For example, Facebook allowed users more control over what content observers could see about a user, relative to the public nature of MySpace. MySpace was seen by many as too cluttered, and Facebook offered a much cleaner design.[33]

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 $275M in funding[34], Jet has received $220M in funding,[35] and Postmates has received $138M in venture capital funding.[36]

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 sufficient protection for an incumbent against a superior product offering.

6 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 of 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

[33] 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`

[34] `https://www.crunchbase.com/organization/instacart\#/entity`

[35] `https://www.crunchbase.com/organization/jet\#/entity`

[36] `https://www.crunchbase.com/organization/postmates\#/entity`

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. This builds on earlier work such as Porter and Millar (1985) who argued that information technology can confer a competitive advantage but that the simple presence of data is not sufficient for competitive success. 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 of 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 (Stucke et al., 2015; Grunes and Stucke, 2015; Tucker and Wellford, 2014), we use a long-established strategic framework to evaluate whether big data indeed merits consideration as a source of sustainable competitive advantage.

References

Almuhimedi, H., F. Schaub, N. Sadeh, I. Adjerid, A. Acquisti, J. Gluck, L. F. Cranor, and Y. Agarwal (2015). Your location has been shared 5,398 times!: A field study on mobile app privacy nudging. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 787–796. ACM.

Ascarza, E., A. Lambrecht, and N. Vilcassim (2012). When talk is free: The effect of tariff structure on usage under two-and three-part tariffs. *Journal of Marketing Research 49*(6), 882–899.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of management 17*(1), 99–120.

Barney, J. B. (2001). Resource-based theories of competitive advantage: A ten-year retrospective on the resource-based view. *Journal of management 27*(6), 643–650.

Biesdorf, S., D. Court, and P. Willmott (2013). Big data: What's your plan. *McKinsey Quarterly 2*, 40–51.

Blake, T., C. Nosko, and S. Tadelis (2015). Customer heterogeneity and paid search effectiveness: A large scale field experiment. *Econometrica forthcoming*.

Bughin, J., M. Chui, and J. Manyika (2010). Clouds, big data, and smart assets: Ten tech-enabled business trends to watch. *McKinsey Quarterly 56*(1), 75–86.

Butler, D. (2008). Web data predict flu. *Nature News 456*(7220), 287–288.

Calandrino, J., A. Kilzer, A. Narayanan, E. Felten, and V. Shmatikov (2011, May). You might also like: Privacy risks of collaborative filtering. In *Security and Privacy (SP), 2011 IEEE Symposium on*, pp. 231–246.

Chen, H., R. H. Chiang, and V. C. Storey (2012). Business intelligence and analytics: From big data to big impact. *MIS quarterly 36*(4), 1165–1188.

Cook, S., C. Conrad, A. L. Fowlkes, and M. H. Mohebbi (2011). Assessing google flu trends performance in the united states during the 2009 influenza virus a H1N1 pandemic. *PloS one 6*(8), e23610.

Cuzzocrea, A., I.-Y. Song, and K. C. Davis (2011). Analytics over large-scale multidimensional data: The big data revolution! In *Proceedings of the ACM 14th International Workshop on Data Warehousing and OLAP*, DOLAP '11, New York, NY, USA, pp. 101–104. ACM.

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

Draganska, M., W. R. Hartmann, and G. Stanglein (2014). Internet versus television advertising: A brand-building comparison. *Journal of Marketing Research 51*(5), 578–590.

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

Edelman, B. (2015, April). How to launch your digital platform. *Harvard Business Review*.

Farrell, J. and P. Klemperer (2007). Coordination and lock-in: Competition with switching costs and network effects. *Handbook of industrial organization 3*, 1967–2072.

Feblowitz, J. et al. (2013). Analytics in oil and gas: The big deal about big data. In *SPE Digital Energy Conference*. Society of Petroleum Engineers.

Feldman, R. and J. Sanger (2007). *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge University Press.

Goel, S., J. M. Hofman, S. Lahaie, D. M. Pennock, and D. J. Watts (2010). Predicting consumer behavior with web search. *Proceedings of the National academy of sciences 107*(41), 17486–17490.

Greenstein, S. M., A. Goldfarb, and C. Tucker (2013). *The Economics of Digitization*. Edward Elgar.

Grunes, A. P. and M. E. Stucke (2015). No mistake about it: The important role of antitrust in the era of big data. *Antitrust Source (Apr. 2015)*.

Ioannidis, J. P. (2005). Why most published research findings are false. *Chance 18*(4), 40–47.

Junqué de Fortuny, E., D. Martens, and F. Provost (2013). Predictive modeling with big data: Is bigger really better? *Big Data 1*(4), 215–226.

Lambrecht, A. and B. Skiera (2006). Paying too much and being happy about it: Existence, causes, and consequences of tariff-choice biases. *Journal of Marketing Research 43*(2), 212–223.

Lambrecht, A. and C. Tucker (2013). When does retargeting work? Information specificity in online advertising. *Journal of Marketing Research 50*(5), 561–576.

Lambrecht, A. and C. Tucker (2015). Field experiments in marketing. *Working paper (September 10, 2015)*.

Lambrecht, A., C. Tucker, and C. Wiertz (2015). Advertising to early trend propagators: Evidence from twitter. *Working paper (March 18, 2015)*.

Lazer, D., R. Kennedy, G. King, and A. Vespignani (2014). The parable of Google Flu: Traps in big data analysis. *Science 343*(14 March).

Lewis, R. A., J. Rao, and D. Reiley (2011). Here, there, everywhere: Correlated online behaviors can lead to overestimates of the effects of advertising. In *International World Wide Web Conference*.

Lewis, R. A. and J. M. Rao (2015). The unfavorable economics of measuring the returns to advertising*. *The Quarterly Journal of Economics*.

Markman, G. D., M. I. Espina, and P. H. Phan (2004). Patents as surrogates for inimitable and non-substitutable resources. *Journal of Management 30*(4), 529–544.

McAfee, A., E. Brynjolfsson, T. H. Davenport, D. Patil, and D. Barton (2012). Big data. *The management revolution. Harvard Bus Rev 90*(10), 61–67.

McGuire, T., J. Manyika, and M. Chui (2012). Why big data is the new competitive advantage. *Ivey Business Journal 76*(4), 1–4.

Miller, A. R. and C. Tucker (2014). Health information exchange, system size and information silos. *Journal of Health Economics 33*, 28 – 42.

Narayanan, A. and V. Shmatikov (2008). Robust de-anonymization of large sparse datasets. In *Security and Privacy, 2008. SP 2008. IEEE Symposium on*, pp. 111–125. IEEE.

Peteraf, M. A. (1993). The cornerstones of competitive advantage: A resource-based view. *Strategic management journal 14*(3), 179–191.

Pilászy, 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.

Porter, M. E. and V. E. Millar (1985). How information gives you competitive advantage.

Shapiro, C. and H. R. Varian (1999). *Information rules: a strategic guide to the network economy.* Harvard Business Press.

Stucke, M., A. Grunes, et al. (2015). Debunking the myths over big data and antitrust. *Antitrust Chronicle 5.*

Tucker, C. and A. Marthews (2011). Social networks, advertising, and antitrust. *Geo. Mason L. Rev. 19*, 1211.

Tucker, D. S. and H. B. Wellford (2014). Big mistakes regarding big data. *Antitrust Source, American Bar Association.*

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

Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic management journal 5*(2), 171–180.