TLI *Think!*

The Data Paradox in Competition Enforcement

José Tomás Llanos

Abstract: There is a data paradox in competition enforcement. Data is an infinitely scalable, widely available, non-rivalrous and non-exclusive asset. The European Commission's merger decisional practice suggests that these features make data a kind of asset unlikely to be conducive to competition issues. Yet, some observe that digital markets are increasingly dominated by incumbents in possession of large troves of data, raising alarm bells of market power and anticompetitive conduct. To craft sound competition policy, this paradox must be unravelled. To this effect, this article examines the test devised by the Commission to assess a 'data advantage', identifying several drawbacks from which that test suffers. It argues that excessive reliance on data's wide availability and ubiquity to dismiss competition concerns is problematic, since data is not fungible and may not be readily accessible to competitors. Additionally, other factors such as data's volume, variety, velocity, and spill-overs are decisive for the magnitude a data advantage may reach. Therefore, in spite of its inherent features, data may nevertheless entrench dominance in digital sectors. To solve the deficiencies of the Commission's test, this article proposes a comprehensive data advantage test that takes due consideration of the dynamics and structural conditions of data-driven markets.

Keywords: Competition law; Antitrust; Big data; Data-driven markets; Mergers; Data advantage

Institutional affiliation:

Dr. *José Tomás Llanos*
Associate Lecturer in Competition Law
King's College London
Research Fellow, British Institute of International
and Comparative Law

Email: jose.llanos@kcl.ac.uk

1

The Data Paradox in Competition Enforcement

José Tomás Llanos[†]

Abstract

There is a data paradox in competition enforcement. Data is an infinitely scalable, widely available, nonrivalrous and non-exclusive asset. The European Commission's merger decisional practice suggests that these features make data a kind of asset unlikely to be conducive to competition issues. Yet, some observe that digital markets are increasingly dominated by incumbents in possession of large troves of data, raising alarm bells of market power and anticompetitive conduct. To craft sound competition policy, this paradox must be unravelled. To this effect, this article examines the test devised by the Commission to assess a 'data advantage', identifying several drawbacks from which that test suffers. It argues that excessive reliance on data's wide availability and ubiquity to dismiss competition concerns is problematic, since data is not fungible and may not be readily accessible to competitors. Additionally, other factors such as data's volume, variety, velocity, and spill-overs are decisive for the magnitude a data advantage may reach. Therefore, in spite of its inherent features, data may nevertheless entrench dominance in digital sectors. To solve the deficiencies of the Commission's test, this article proposes a comprehensive data advantage test that takes due consideration of the dynamics and structural conditions of data-driven markets.

Introduction

Whilst the benefits of big data are widely acknowledged,[1] the question of whether big data leads to any competitive concerns is a highly disputed topic. In this connection, it is possible to identify two main opposite stances. Some argue that data is inconsequential from a competition standpoint, or at least that it does not lead to any competition concerns. For example, former Google CEO Eric Schmidt has publicly stated that '[Google's] experience is that you don't need data to compete online.'[2] There are powerful reasons to be doubtful as to the sincerity of this statement, since for years Google executives have been boasting about the scale of data Google enjoys.[3] However, it is indeed true that data has certain features (i.e. non-rivalrous and non-

---

[†] Associate Lecturer in Competition Law, King's College London; Research Fellow, British Institute of International and Comparative Law. Email: jose.llanos@kcl.ac.uk

[1] Big data can lead to the improvement of digital undertakings' products and services, the exploitation of new business opportunities and the provision of more personalised services, such as behavioural advertising. See Autorité de la Concurrence and Bundeskartellamt, 'Competition Law and Data' (2016) 9–11.

[2] Eric Schmidt, 'The New Gründergeist' (*Google Europe Blog*, 2014) <https://europe.googleblog.com/2014/10/the-new-grundergeist.html>.

[3] For example, in 2009 Eric Schmidt stated: 'Scale is key. We just have so much scale in terms of the data we can bring to bear.' Fair Search, 'Fact-Checking Google: Scale Is a Barrier to Entry in Search' (*Fair Search*, 2011) <http://fairsearch.org/fact-checking-google-scale-is-a-barrier-to-entry-in-search/>; See also statement by Google's chief scientist Peter Norvig: 'We don't have better algorithms than everyone else; we just have more data'. Matt Asay, 'Tim O'Reilly: "Whole Web" Is the OS of the Future' (*CNET*, 2010) <https://www.cnet.com/news/tim-oreilly-whole-web-is-the-os-of-the-future/>; and statement by Google Cloud Platform's director of product management Tom Kershaw: 'Never delete anything, always use data - it's what Google does'. Quentin Hardy, 'Google Offers Cheap Storage for Certain Kinds of Data' (*Bits Blog*, 2015) <https://bits.blogs.nytimes.com/2015/03/11/google-offers-cheap-storage-for-certain-kinds-of-data/> It is unsurprising, however, the fact that Google is now trying to downplay the significance of data for its business model, on account of the severe antitrust scrutiny to which it has been subject during this decade.

exclusive nature, wide availability and ubiquity[4]) that make it 'different from the industry structures typically seen as conducive to competition problems.'[5] As a matter of fact, scarce and rival assets within the exclusive control of the incumbent tend to be those that are likely to raise barriers to entry and dampen competition.[6]

At the other end of the spectrum, it is contended that data is capable of conferring a competitive advantage. Early in 2011, a MGI report observed: 'the impact of developing a superior capacity to take advantage of big data will confer enhanced competitive advantage over the long term and is therefore well worth the investment to create this capability. But the converse is also true. In a big data world, a competitor that fails to sufficiently develop its capabilities will be left behind.'[7] This 'data advantage', if large enough, may lead to industry concentration, strong barriers to entry and market power posing significant risk of anticompetitive conduct.[8] For example, there is a risk that a particularly large dataset may lead to a quality gap between the incumbent's and its competitors' services, as a result of which the market is bound to tip.[9] Additionally, possession of certain data may enable the incumbent to seamlessly expand into adjacent markets,[10] a possibility which market operators without access to such data would not have. Also, unilateral conduct by an incumbent may lead to 'data foreclosure', thereby denying competitors the scale necessary to compete.[11] Data advantages are also a concern in merger review, where it is feared that the merger of firms that hold valuable data may lead to a combined vast and comprehensive dataset capable of deterring entry and inhibiting competition.[12]

---

[4] See below Section I.

[5] Darren S Tucker and Hill B Wellford, 'Big Mistakes Regarding Big Data' (2014) Antitrust Source, American Bar Association 4 <http://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2549044>.

[6] For example, if a sole manufacturer controls the main (scarce and rival) raw material to produce an end product, the market for this product will be effectively foreclosed for potential entrants which cannot access the raw material (i.e. barriers to entry in that market are high, perhaps insurmountable.). As the OECD notes, 'control over access to scarce or non-duplicable resources such as land, natural resources and distribution channels' is capable of deterring entry. See OECD, 'Policy Roundtables: Barriers to Entry' (2005) DAF/COMP(2005)42 104 <https://www.oecd.org/daf/competition/abuse/36344429.pdf>.

[7] McKinsey Global Institute, 'Big Data: The next Frontier for Innovation, Competition and Productivity' (2011) 8 <http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation>

[8] OECD, 'Hearing on Big Data - Note by BIAC' (2016) 5.

[9] By using user data to adapt a product better to users' preferences, a platform operator increases the product's perceived quality in the future. Accordingly, higher initial demand reduces the marginal cost of innovation, as it makes it cheaper to produce one additional unit of product or service quality, as perceived by users. In this context, the market will eventually tip and one firm will dominate the market. See Jens Prufer and Christoph Schottmüller, 'Competing with Big Data' [2017] Tilburg Law School Research Paper No. 06/2017 1–2 <https://papers.ssrn.com/abstract=2918726>.

[10] See text accompanying footnotes 124-129.

[11] For example, in the 2012 investigation launched by the U.S. Federal Trade Commission ('FTC') into Google's potential anticompetitive practices in the search market, the FTC found that Google's exclusive agreements with website publishers reduced the ability of Google's competitors to achieve minimum efficient scale and thus to compete effectively with Google. As a matter of fact, website publishers represented to the FTC that Bing's search syndication offering was inferior at least partly because Microsoft's network of advertisers is smaller than Google's. With its considerably larger advertiser base Google is more likely to have high-quality and more relevant ads for any given query, which improves its monetisation rate relative to Microsoft to a significant extent. See FTC Staff, 'FTC Staff Report on Google - File No. 111-0163' 56–58 <http://graphics.wsj.com/google-ftc-report/>.

[12] In the context of the Google/DoubleClick merger review conducted by the FTC, then-Commissioner Harbour had a different view than the majority's as to the evolution of the online advertising market. In particular, she held: [S]earch information gathered by Google, combined with browsing information gathered by DoubleClick, will create a far richer source of data to enable highly targeted advertising […] marrying the two datasets raises

The conflicting views above are seemingly hard to reconcile. Whilst the latter view finds a significant degree of support in the literature,[13] the decisional practice of the European Commission (the 'Commission') seems to adhere to the former stance.[14] In this article, it is argued that both views are the two components of the 'data paradox in competition enforcement': data, a widely available, non-rivalrous, non-exclusive and ubiquitous asset is capable of raising barriers to entry and leading to high concentration levels in digital markets. To craft and implement sound competition policy in digital markets, this data paradox must be unravelled. To this effect, this article begins by explaining data's main features, identifying the key developments that turned data into a source of economic value (Section I). Then, the test the Commission has defined to assess a potential data advantage is explained (Section II), and subsequently a number of drawbacks from which that test suffers are assessed (Section III). It will be seen that excessive reliance without qualifications on the wide availability and ubiquity of data to dismiss competition concerns is problematic. Generally speaking, data may be widely available, non-rivalrous and non-exclusive, but in concrete cases data is not fungible, and may not be readily accessible to actual and potential competitors. Also, other factors such as *inter alia* data's volume, variety, velocity, and spill-overs are decisive for the magnitude a competitive advantage arising from data may reach. Therefore, in spite of its inherent features that at first glance allow to support the view that data is an immaterial or at least neutral asset for competition policy, in practice data may still raise barriers to entry and entrench dominance. To address the pitfalls of the Commission's test, this article proposes an alternative data advantage test that takes due account of the dynamics and structural conditions of data-driven markets (Section IV). Finally, some conclusions are presented, wrapping up the discussion.

I.   Data's main features and economic value

The transition from the industrial and post-industrial economic models of the nineteenth and twentieth centuries to the 'information economy'[15] signalled a move from economic value being

---

long-term competition questions that beg further inquiry: In a post-merger online advertising market driven by the value of behavioral targeting, will Google/DoubleClick face meaningful competition?; Will any other firm be able to amass a dataset of the same scope and size?; Will any other company be able to overcome network effects and offer an equally focused level of behavioral targeting?; If advertisers and publishers have to channel their online advertising through Google/DoubleClick in order to access the best dataset that supports targeted advertising, will any other firms have the ability or incentive to compete meaningfully in this market? See Pamela Jones Harbour, 'Dissenting Statement of Commissioner Pamela Jones Harbour - Google/DoubleClick, FTC File No. 071-0170', 7–8 <https://www.ftc.gov/sites/default/files/documents/public_statements/statement-matter-google/doubleclick/071220harbour_0.pdf>.

[13] See inter alia Ariel Ezrachi and Maurice Stucke, *Virtual Competition* (Harvard University Press 2016); Maurice Stucke and Allen Grunes, *Big Data and Competition Policy* (Oxford University Press 2016); Xavier Boutin and Georg Clemens, 'Defining "Big Data" in Antitrust' (2017) 1 Competition Policy International: Antitrust Chronicle 22; Nathan Newman, 'Data Justice: Taking on Big Data as an Economic Justice Issue' (2015); Nathan Newman, 'Search, Antitrust and the Economics of the Control of User Data' (2014) 41 Yale Journal on Regulation; Autorité de la Concurrence and Bundeskartellamt (n 1); Inge Graef, *Eu Competition Law, Data Protection and Online Platforms: Data as Essential Facility* (Kluwer Law International 2016).

[14] See below Section II.

[15] The concept of 'information economy' can be traced back to the notion of 'knowledge-based industry' coined by Fritz Machlup, who found that by 1959 knowledge-producing occupations had outnumbered other traditional (i.e. agricultural and industrial) occupations in the U.S, and to the work of Marc Porat, who distinguished between two distinct yet inexorably linked domains: one that is involved 'in the transformation of matter and energy from one form into another' (which included the agriculture and industrial sectors), and the other 'in transforming information from one pattern into another' (referred to as the 'information sector').

sited within physical goods, to economic value being sited within information. Whilst industrialised nations made money by manufacturing products such as ships, weapons, clothing, railway tracks and aircrafts, in the information economy, conversely, control over information encoded in bits[16] is the main source of wealth.[17]

Bits are the basic building blocks of digital information. Virtually all information that not so long ago was recorded in a physical carrier, such as words on a notebook, sounds and videos on tape and images on photographs, can be now encoded in binary notation (i.e. words on a word processor program file, sounds on an MP3 file and images on JPEG files). And as digital information is increasingly cheaper to store, distribute and encode, a large-scale migration from cue technologies to digital technologies, and with it a shift in the economic values of information, has taken place.

In his book *Information Technology Law: The Law and Society*, Andrew Murray explains that 'informational goods' (i.e. goods capable of being encoded in bits such as ideas, radio-communication broadcasts, images and sounds) share a common feature: they all are 'nonrivalrous'.[18] Rivalry in consumption is the property of a good whereby one person's use limits other people's use of said good. For example, an ice-cream cone is a rivalrous good, because if one person eats an ice-cream cone, another person cannot eat the same cone.[19] Conversely, nonrivalrous goods are those the consumption of which by one person does not limit another person's ability to consume the same good (i.e. they may be consumed by several consumers simultaneously), such as television broadcasts.[20] As a natural consequence of their nonrivalrous nature, informational goods are also non-exclusive or non-excludable. Excludability is the property of a good whereby a person can be prevented from using or accessing it.[21] An ice-cream cone is an excludable good, since a person can be prevented from eating it by simply denying him access to the ice cream cone, as opposed to a television broadcast, which can be enjoyed by anybody with a TV set. Additionally, informational goods are (almost) infinite. Digital technology and the Internet have made possible the effortless and costless production of exact copies of text, images, audio, video and other information of varied nature by anybody with a computer, a smartphone or another similar device. If a person wants, for example, the last Iron Maiden album but cannot afford it, he may be able to get it for free from another person, who would not surrender the original but would rather make a brand-new copy of that album. Bits never run out and therefore are infinitely scalable, the only limit being on how many bits it is

---

An economy can be considered an information economy when information-related work, that is, work related to the production, processing, storage and dissemination of information, surpasses the work related to other sectors. According to Porat's measurements of a number of sectors, this transition took place in the U.S. in 1967, when 53% of the workforce was involved in 'information work'. Generally speaking, the information economy represents the structural shift in the global economy away from a purely manufacturing or agriculturally based economy to one dominated by services with a disproportionate emphasis on digitized information. See Fritz Machlup, 'The Production and Distribution of Knowledge in the United States' (*Princeton University Press*, 1962) <https://press.princeton.edu/titles/1510.html>; Marc Porat, *The Information Economy: Definition and Measurement.* (US GPO 1977); IGI Global, 'What Is Information Economy?' <https://www.igi-global.com/dictionary/information-literacy-digital-divide/14385>.

[16] Margaret Rouse, 'What Is Bit (Binary Digit)? - Definition from WhatIs.Com' (*WhatIs.com*) <https://whatis.techtarget.com/definition/bit-binary-digit>.

[17] Andrew Murray, *Information Technology Law: The Law and Society* (Third Edition, Oxford University Press 2016) 4–13.

[18] ibid 12.

[19] NG Mankiw, *Principles of Economics, 5th Edition* (South-Western Cengage Learning 2011) 226.

[20] Murray (n 17) 12.

[21] Mankiw (n 19) 226.

possible to store (i.e. storage capacity).[22] As a consequence, data is widely available and ubiquitous. A study conducted by IDC notes that there were 4.4 trillion gigabytes of data produced globally in 2013, an amount which is forecast to double every two years, in such a way that by 2020 around 44 trillion gigabytes will have been generated.[23]

Murray observes that the inherent features of digital data pose a paradox. Traditional private property theory attempts to define a system of allocation of all things on earth. As individuals claim the material resources provided by our planet, the ensuing scarcity, partly derived from the rivalrous nature of such things, requires said system in order to preserve peace and order.[24] In this system, wealth can be amassed by occupying and owning things, that is, by excluding others from using and enjoying scarce, rivalrous physical resources. However, in the information economy, scarcity is no longer an issue: as seen above, bits never run out insofar as there is storage capacity left. In addition, the setting upon which traditional private property theory is predicated, where valuable goods are physical, tangible and rivalrous or where intangible goods (such as ideas protected by intellectual property rights) are fixed onto some type of physical carrier such as a book or a patent specification, no longer holds. As Barlow observes, the advent of digitisation has made it 'possible to replace all previous information storage forms […] with complex and highly liquid patterns of ones and zeros',[25] as a result of which valuable content (for example, a song) can be now separated from its tangible carrier (a CD), and therefore tangibility and rivalrousness are no longer requirements to be met for an asset to be valuable. Accordingly, a data paradox emerges: data is valuable. At the same time, it is (almost) infinitely scalable, nonrivalrous and intangible.[26]

This data paradox can be unravelled. Digital information has evident advantages as compared to analogue information. Take the example of medical records. They were traditionally held on a number of manual filing systems by physicians and hospitals, each of whom and which would normally keep a rather limited set of patient records. Accordingly, a patient attending for example three different hospitals for different treatments would have at least four sets of medical records (one at each hospital and one at the physician's surgery), for which reason no set would be all-encompassing or final. Also, searching for information in patients' files was highly time-consuming, as the data kept on the records was manually recorded and indexed, and given that indexing was a skilled job, and therefore expensive, only key data would be indexed.[27] With information technology, conversely, a single comprehensive record of a patient can be shared and accessed by all carers at the same time, and information therein can be effortlessly searched and found by reference to any keyword.[28] As digital information is easier and increasingly cheaper to generate, manipulate, transmit and store, it developed an intrinsic value that is not found in analogue information, due to its very nature.[29] Put in other words, it attracted a premium in the marketplace that encouraged data collectors to prefer the collection of digital information to analogue information, which led to a major increase in the volume of digital data

---

[22] Murray (n 17) 14.

[23] IDC, 'The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things' (2014).

[24] Without this system there would be constant disputes as to what belonged to one person and what to another. Sukhninder Panesar, 'Theories of Private Property in Modern Property Law' (2012) 15 The Denning Law Journal 113, 116.

[25] John Perry Barlow, 'The Economy of Ideas' [1994] *Wired* <https://www.wired.com/1994/03/economy-ideas/>.

[26] Murray (n 17) 14.

[27] ibid 42.

[28] ibid.

[29] Fred H Cate, *Privacy in the Information Age* (Brookings Institution Press 1997) 14–15.

available.[30] Increased availability of digital data, in turn, enabled the emergence of innovative business models based on the collection and processing of large volumes of data ('big data')[31] with the aid of advanced algorithmic technology ('big analytics').[32] The analysis of big data performed through big analytics allows for specific patterns to be found and new correlations to be made between several datasets coming from combined different sources, as a result of which new information can be deduced or inferred, trends and behaviour can be accurately predicted and the likelihood for certain events to occur can be assessed with astonishing precision.[33] The more data are available for processing, irrespective of their apparent significance or value, the higher are the chances to obtain unexpected and potentially valuable information.[34] This information, in turn, can be used for myriad commercial and other purposes.[35] Therefore, data is valuable, especially in large volumes, in spite of the fact that digital data's features are diametrically opposite to those which have historically explained the value that sits within traditional tangible goods.

The shift in the economic value of information is intrinsically linked to the emergence of data-driven markets. These markets may be defined as those where use by consumers of one provider's services generates user information (or user data) as a costless by-product, this information being subsequently analysed and used by the provider to adapt its services better to users' preferences, incorporate new features or develop new products and services.[36] Importantly, the most popular data-driven markets are 'multisided',[37] where economic operators are 'platforms' that mediate interactions between two or more separate groups of customers, thereby generating value. The markets for search engines, social networks and ecommerce platforms are all examples of data-driven, multisided markets. For example, search engines and social networks connect users on one side (the user or 'free' side) with advertisers on the other side (the advertiser or 'paid' side). Ecommerce platforms, in turn, connect buyers with sellers. Critically, the services provided and the 'connection' performed by the platform rely to a great extent on the processing of big data through big analytics. Thus, based on *inter alia* search query data, crawled and indexed websites and other information provided by users (such as IP address and language preferences), search engine algorithms are able to render 'more relevant' results to user queries, and at the same time, that data is processed to render search-based advertisements that are also relevant to a given user query. Equally, based on the data gathered from user-generated content and user interactions with the platform, social network algorithms can increase the relevance of social network engagement, suggested friends or suggested interests that are shown to users, and in turn, the same data collected from the user side is processed to identify consumers that may be interested in specific ads that are then shown to those consumers during

---

[30] Murray (n 17) 43.

[31] Big data can be defined as the '[t]he ability to collect and analyze a large volume of data which contains a variety of information in a timely manner'. This definition takes into consideration the three 'Vs' of big data: volume, variety and velocity. See Boutin and Clemens (n 13) 4.

[32] Advancements in data science have led to the ability to learn fast and deep from big data with the aid of algorithms that access and analyse vast amounts of information (i.e. Big Analytics). Said advancements include *inter alia* data-mining techniques such as association analysis, data segmentation and clustering, classification and regression analysis, anomaly detection and predictive modelling. Daniel L Rubinfeld and Michal S Gal, 'Access Barriers to Big Data' (2017) 59 Ariz. L. Rev. 339, 347.

[33] Primavera De Filippi, 'Big Data, Big Responsibilities' (2014) 3 Internet Policy Review 2.

[34] Directorate General for Internal Policies, 'Big Data and Smart Devices and Their Impact on Privacy' (2015) 11 <http://www.europarl.europa.eu/RegData/etudes/STUD/2015/536455/IPOL_STU(2015)536455_EN.pdf>.

[35] See next paragraph.

[36] Prufer and Schottmüller (n 9) 1–2.

[37] See generally David S Evans and Richard Schmalensee, *The Matchmakers: The New Economics of Multisided Platforms* (Harvard Business Review Press 2016).

their interaction with the platform. Similarly, when users shop and buy on an ecommerce platform they provide the platform operator with information about their habits, interests and preferences, information which is processed by the platform to tailor recommendations and future deals, thereby driving more sales. Therefore, user data plays a fundamental role in the business proposition of data-driven markets' players.

However, whereas the fact that big data has become both a source of economic value and a driver of technological progress is virtually undisputed, its impact on the competitive process remains a contentious issue. In particular, whilst the characteristics of data explained above effectively makes it an asset unlikely to be conducive to competition concerns, it is nevertheless true that some data-driven markets are increasingly becoming more concentrated, with leadership positions becoming entrenched instead of challenged. Ultimately, the data paradox identified by Murray takes a specific form in competition enforcement: data is a widely available, non-rivalrous, non-exclusive and ubiquitous asset. At the same time, it is capable of raising barriers to entry and leading to high concentration levels in digital markets. Is it possible to unravel this paradox?

II.   The Commission's Test to Assess a 'Data Advantage'

A good starting point to unravel the data paradox in competition enforcement is a careful examination of the Commission's view as to the competitive implications of concentrations of large datasets. This view can be deduced from the Commission's decisional practice in important merger cases involving big data. In these cases, in order to assess the extent to which, if any, data provides a competitive advantage and/or raises barriers to entry, the Commission assessed whether post-transaction there remained available an amount of data sufficient for competitors to match the competitive advantage arising from the transaction for the merging firms (the 'Commission's test'). The question that follows is whether this test is sufficient and sound enough to capture the competitive implications of a firm holding and/or gaining access to a valuable dataset.

This Section examines the merger cases in which the Commission defined the test referenced above. It is worth pointing out that in neither of these cases did the Commission find that the data advantage arising from the concentration led to anticompetitive risks that could justify blocking the transaction. The Commission's reasoning in these cases and its reliance on data's wide availability and ubiquity to dismiss competition concerns reveal its view that, on account of its features, data is unlikely to raise barriers to entry and dampen competition in data-driven markets.

A.   *Google/DoubleClick*[38]

In this case both Google and DoubleClick were active in the online advertising industry, albeit providing different services. Google was active as ad publisher with its own search engine properties, and as ad intermediary with its ad network AdSense. On the other hand, DoubleClick offered ad serving tools[39] to both publishers and advertisers.

As a result of their commercial activities, the merging firms collected and processed large amounts of data. Consequently, one of the theories of harm was that the merged entity could

---

[38] Case COMP/M4731, *Google/DoubleClick* (2008).
[39] ibid para 26.

attempt to combine Google's and DoubleClick's datasets, thereby achieving a position that could not be replicated by its competitors.[40]

The Commission, however, dismissed the possibility of a combination of the merging firms' datasets, noting that the merged entity would have neither the ability nor the incentive to do so.[41] Crucially, the Commission argued that even if the merged firm combined Google's and DoubleClick's datasets, it would be unlikely that said combination could confer upon the merged entity a competitive advantage that could not be matched by its competitors, given that the combination of data about searches with data about users' web surfing behaviour was already available to a number of Google's competitors.[42] The Commission elaborated further noting that both Microsoft and Yahoo! ran search engines and offered ad serving, and competitors could also purchase data or targeting services from third parties such as comScore.[43]

From the Commission's analysis transpires that the competitors' option to access data similar to that of the merged entity was the key factor in establishing whether a competitive advantage that could not be matched by them would arise from the transaction. Ultimately, since similar data could be available for competitors, anticompetitive effects arising from the combination of datasets were dismissed.

2.   *Telefónica UK/Vodafone UK/Everything Everywhere/JV*[44]

This transaction concerned a joint venture ('JV') set up by the UK mobile network operators ('MNOs') Telefónica UK, Vodafone UK and Everything Everywhere. The JV was aimed at providing mobile services to the parties to the JV and third party mobile operators, including mobile advertising services for advertisers and media agencies wishing to develop advertising campaigns targeted at customers of mobile operators and data analytics services. The JV intended to collect and analyse the data generated from its services in order to provide its customers with valuable insights into consumer behaviour,[45] including basic customer data such as age, residential status, profession and location and data collected on the basis of contracts with merchants.[46]

---

[40] ibid para 359.

[41] In particular, the Commission noted that the merged firm would not have the ability to combine the datasets, given that DoubleClick was contractually prohibited from using the data regarding which webpages a user visited to improve targeting of search ads on Google's sites or in the AdSense network, nor would it have the incentive to do so, as such combination would prompt DoubleClick's customers to switch to competitors (given that advertisers have no interest in other advertisers having access to their data). See ibid paras 361 and 363. Time would prove the Commission's predictions wrong. In 2016 Google amended its privacy policy, deleting the part that promised keeping DoubleClick's database of web-browsing records separate from the personal data Google collects from its many 'free' services. After the amendment, users' activity on websites and apps are associated with their personal information in order to improve Google's services and ad targeting. See Julia Angwin, 'Google Has Quietly Dropped Ban on Personally Identifiable Web Tracking' (*ProPublica*, 21 October 2016) <https://www.propublica.org/article/google-has-quietly-dropped-ban-on-personally-identifiable-web-tracking>.

[42] *Google/DoubleClick* (2008) (n 38) para 365.

[43] ibid para 365; The Federal Trade Commission reached similar conclusions. See *Google/DoubleClick* [2008] Federal Trade Commission FTC File No. 071-0170 12–13.

[44] Case COMP/M6314*, Telefónica UK/Vodafone UK/Everything Everywhere/JV* (2012).

[45] ibid para 531.

[46] ibid para 532.

The Commission assessed whether the JV would foreclose competing providers of data analytics or advertising services by combining personal information, location data, response data, social behaviour data and browsing data, thereby creating a unique database essential for targeted mobile advertising that no competing provider of said services would be able to replicate.[47]

Noting that the information available to the JV was also available to a large extent to both existing and new market players such as Google, Apple, Facebook, card issuers, reference agencies or retailers, which were already using the information to provide targeted advertising, the Commission dismissed the risk of foreclosure. Notably, the Commission held that customers normally tend to give their personal data to many market players, for which reason this type of data 'is generally understood to be a commodity'.[48]

The Commission's decision in this case reflects its view that concentrations of vast amounts of data in one undertaking are unlikely to give rise to anticompetitive risks when data is widely available, especially if the data at sake may in possession of different firms at the same time.

3.   *Facebook/WhatsApp*[49]

In the review of this acquisition the Commission found that Facebook was active in the market for social networking services and online advertising services, and that Facebook collected data regarding the users of its social network and analysed it in order to serve advertisements that are as much as possible 'targeted' at each particular user. [50] Moreover, the Commission found that WhatsApp was active in the market for consumer communications services, and that WhatsApp did not sell any form of advertising and did not collect or store data about its users that would be valuable for advertising purposes.[51]

One of the theories of harm analysed by the Commission was that the merged entity could start collecting data from WhatsApp users with the aim of improving the accuracy of the targeted ads served on Facebook's social networking platform to WhatsApp users that are also Facebook users.[52] However, it was suggested in the proceedings that the merged entity would not have the incentive to start collecting data from WhatsApp users, as this data collection could prompt some users to switch to other consumer communications apps that they could perceive as less intrusive.[53]

Moreover, the Commission analysed a potential concentration of Facebook's and WhatsApp's data only to the extent that it was likely to strengthen Facebook's position in the online

---

[47] ibid para 539.

[48] ibid para 543.

[49] Case COMP/M7217, *Facebook/WhatsApp* (2014).

[50] ibid para 70.

[51] ibid para 71.

[52] ibid para 180.

[53] ibid para 186. Time also proved this prediction wrong. In less than two years following the approval of the concentration WhatsApp's privacy policy was amended, after which the app started sharing user data with the 'Facebook family of companies', enabling the possibility of linking WhatsApp user phone numbers with Facebook user identities. See Tas Bindi, 'WhatsApp, Facebook to Face EU Data Protection Taskforce' (*ZDNet*, 27 October 2017) <https://www.zdnet.com/article/whatsapp-facebook-to-face-eu-data-protection-taskforce/> And whilst some consumers retaliated by migrating to competing messaging apps such as Telegram and Signal, the number of WhatsApp users has steadily increased since the privacy policy amendment. See Statista, 'WhatsApp: Number of Users 2013-2017' (*Statista*) <https://www.statista.com/statistics/260819/number-of-monthly-active-whatsapp-users/>.

advertising market or any sub-segments thereof,[54] and found that even if the merged entity started using WhatsApp user data to improve targeted advertising, competition concerns were unlikely to arise, as there would still remain large volumes of user data that were valuable for advertising purposes and that were not within Facebook's exclusive control.[55]

In the Commission's analysis above the key factor to determine whether anticompetitive effects are likely to arise from a transaction that entails the combination of large datasets is whether there would remain sufficient data available for competitors post-transaction.

4.   *Microsoft/LinkedIn*[56]

In the assessment of this concentration, the Commission found that both Microsoft and LinkedIn offered online advertising services.[57] Microsoft provided non-search advertising (mainly through its web properties)[58] and search advertising (through its search engine Bing),[59] whereas LinkedIn only offered non-search advertising (through its Professional Social Network ('PSN') LinkedIn).[60]

The Commission assessed whether the post-merger combination of the parties' data in relation to online advertising could raise barriers to entry/expansion for actual or potential competitors which 'could be required to collect a larger dataset in order to compete effectively with the merged entity than absent the merger',[61] and dismissed this concern on the basis that 'there [would] 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.'[62]

The Commission's assessment in this merger review confirms the notion that post-transaction availability of data is the key factor to dismiss a concern of heightened barriers to entry/expansion.

III.   Unravelling the Data Paradox in Competition Enforcement

In each of the cases above Commission concluded that the combination of data that would take place post-transaction was unlikely to raise competition concerns, as there remained large amounts of data available to competitors to match the data advantage resulting from said combination. At first glance, this conclusion seems correct. As the Autorité de la Concurrence and the Bundeskartellamt observe, '[i]f the data are widely available, no undertaking, even a dominant one, would benefit from a competitive advantage thanks to these data because competitors are able to get the same information and efficiency from data as the leading operators of the market.'[63]

---

[54] *Facebook/WhatsApp* (2014) (n 49) paras 164 and 187.
[55] ibid para 189.
[56] Case M8124, *Microsoft/LinkedIn* (2016).
[57] ibid para 152.
[58] ibid para 172.
[59] ibid paras 2 and 170.
[60] ibid para 173.
[61] ibid para 179.
[62] ibid para 180.
[63] Autorité de la Concurrence and Bundeskartellamt (n 1) 36.

In addition, data may be obtained from third parties, and is inherently non-rivalrous and non-exclusive. These features seem to confirm the soundness of the Commission's test. However, upon close inspection of data's features, a more nuanced picture emerges. Data may be generally widely available, non-rivalrous and non-exclusive, but in specific cases data is not fungible, and may not be readily accessible to actual and potential competitors. Moreover, other factors such as the volume and variety of data, the speed at which value can be extracted from it, learning-by-doing effects and spill-overs play a fundamental role in the extent to which data may confer a competitive advantage. Accordingly, in spite of its features, data may still raise barriers to entry and entrench dominance.

1.   Factors Suggesting the Soundness of the Commission's test

A.   Data can be obtained from third parties

Data intermediaries or data brokers are companies the primary business of which is the collection, aggregation, analysis and disclosure of personal information about consumers for purposes such as marketing products, verifying an individual's identity, or detecting fraud.[64] They routinely collect data from several sources, including governmental sources, publicly available sources such as social media, blogs and the Internet, and commercial sources.[65] In particular, some data brokers obtain information 'by crawling social media sites, such as Bebo and LinkedIn, where individuals have not set their privacy settings to restrict access to their information and the social media sites have given the data brokers access to such information'.[66] They can also obtain behavioural data derived from tracking cookies or pixels, or transaction-specific data about purchasers from retailers and catalogue companies. This data may include the types of purchases (for example, videogames, products associated with a medical condition or a sports activity), price and date of the purchase, and payment method used.[67]

According to the Autorité de la Concurrence and the Bundeskartellamt, firms may resort to the use of third parties' data as an alternative to direct data collection, as such intermediated access may be less costly: the fixed costs associated with data collection are spread over a larger number of customers, and a firm may buy from the intermediary only the data it needs (in terms of volume and variety) without incurring a large fixed cost.[68] The fact that data can be readily and cheaply purchased from data brokers suggests that the Commission's test is sound, given that the risk of data foreclosure following a merger between firms in possession of large datasets can be ameliorated by the competitors' possibility to buy similar data from third parties.

B.   Data is non-rivalrous and non-exclusive

As explained in Section I, data is non-rivalrous and non-exclusive, which means that the collection of certain data by a specific firm does not prevent other companies from gathering the same type of data through similar or other means and using such data for multiple purposes. Users regularly provide information such as gender, age, home address and name upon registering for online services (for instance, when creating a profile on a social network or registering for online games), for which reason the same data may be in possession of several firms at the same time. In addition, since the value of data derives from the insights it is possible

---

[64] FTC, 'Data Brokers: A Call for Transparency and Accountability' (2014) 3.
[65] ibid 11.
[66] ibid 13.
[67] ibid.
[68] Autorité de la Concurrence and Bundeskartellamt (n 1) 39.

to extract from it rather than from just amassing the data,[69] different firms may extract the same knowledge from different sources and types of data. For instance, Apple may know the music preferences of a specific user based on his last purchases on iTunes, whilst Google may hold the same information based on the music videos he has searched for on YouTube.

Moreover, online platforms and firms in general do not have exclusive access to any specific user, as there is neither exclusive contracts nor a pricing structure that may lock-in users to a specific platform or firm.[70] The lack of exclusivity over users is confirmed by the fact that users have the possibility to 'multi-home', and therefore may share their information with several online platforms and firms. In this connection, Lerner notes that 'there is extensive user multi-homing whether one looks at websites that provide similar services (*e.g.*, general search providers such as Bing, Yahoo!, Google, and Ask!) or differentiated websites that nevertheless may compete for users with regard to a particular activity (*e.g.*, Amazon and Google)'.[71]

As Lambrecht and Tucker observe, '[f]or there to be a sustainable competitive advantage, the firm's rivals must be unable realistically to duplicate the benefits of [the] strategy or input.'[72] Since data is non-rivalrous and non-exclusive, large online firms cannot foreclose rivals from accessing the same data they collect through their services and consequently from replicating the same benefits arising from that data collection. This observation seems to confirm the robustness of the Commission's test and the contention that big data in possession of large online firms is unlikely to be the cause of high concentration levels and significant antitrust risk in digital markets.

2. Close inspection of data's features and implications

A. Limited impact of data brokers on the wide availability of data. Also, data is not fungible, and may not be readily available to actual and potential competitors

As noted in a Report prepared for the White House, big data can be 'bought, bartered, traded and sold'.[73] However, the data collected by data brokers is very limited as compared with the volume and variety of data collected directly by large online platforms like Google or Facebook. For example, browsing data collected through tracking cookies or pixels are likely to be limited in scope and therefore imperfect substitutes with personal data voluntarily provided by users as a result of their interaction with a platform, since tracking cookies only render information on the websites that a given user has visited, but may not allow to gather detailed socio-demographic information, which may be particularly valuable.[74] As seen above, some data brokers obtain data from social media profiles with privacy settings disabled.[75] The data they are likely to obtain cannot possibly be compared to the data gathered by a social media platform operator, especially

---

[69] Geoffrey A Manne and Ben Sperry, 'The Problems and Perils of Bootstrapping Privacy and Data into an Antitrust Framework' (2015) 2 CPI Antitrust Chronicle 9.
[70] D Daniel Sokol and Roisin Comerford, 'Antitrust and Regulating Big Data' (2015) 23 Geo. Mason L. Rev. 1129, 1137–1138.
[71] Andres Lerner, 'The Role of "Big Data" in Online Platform Competition' (2014) SSRN paper 22 <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2482780>.
[72] Anja Lambrecht and Catherine E Tucker, 'Can Big Data Protect a Firm from Competition?' (2015) SSRN paper 4–5 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2705530>.
[73] Executive Office of the President, 'Big Data: Seizing Opportunities, Preserving Values' (2014) 50 <http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf>.
[74] Autorité de la Concurrence and Bundeskartellamt (n 1) 40.
[75] See above text accompanying footnote 66.

a dominant one. As will be seen below,[76] the greater the variety of data sources, the more valuable and accurate the predictions and insights that can be derived from data. Data brokers rely on restricted data sources, as opposed to vertically integrated platforms like Apple or Google, which have *inter alia* mobile OSs, App Stores and online map apps to extract personal data from large customer bases. In this connection, writing on the U.S. context, Newman observed: '[t]here has recently been a flurry of political interest in abusive practices by data brokers who buy and sell personal data, with major reports released by both the U.S. Senate and the Federal Trade Commission […] it is worth noting that the companies involved are relative minnows in the big data ecosystem compared to the major big data platforms […] Experian is one of the largest at $4.8 billion in sales per year, while Acxiom, a data broker often cited as having one of the largest datasets on consumers, has only about $1 billion per year in revenue. Even collectively, these data brokers are dwarfed by a company like Google with over $60 billion in annual revenue.'[77]

The fact that data brokers' data is limited in scope suggests that its impact on data availability is likely to be also limited. As a matter of fact, the CMA observes that 'the more sensitive the type of data, the less likely it is to be available from an intermediary.'[78]

Moreover, the non-rivalrous and non-exclusive character of data will not always justify a broad statement such as 'data is widely available'. It is true that consumers can surrender the same kind of data to different offline and online service providers, and that the same data can be used by multiple firms to draw inferences, create consumer profiles or for myriad other purposes. But the fact that two or more companies <u>could</u> have the same data does not necessarily entail that they <u>will</u> have it.[79] Apple and Google can both have profiles of the same consumer, but that hardly means that the comprehensiveness of such profiles is the same, or that both companies have made the same inferences. Moreover, although certain data is effectively non-exclusive, 'a lot of data that are of particular relevance for companies are in the exclusive control of the companies that collected it and [therefore they] decide about its use, often denying access to competitors'',[80] as for example, search data derived from queries entered in websites having exclusive search and search advertising syndication agreements with Google,[81] Facebook's friends lists,[82] or Yelp's and TripAdvisor's user reviews.[83]

Yet, these considerations have been overlooked in the Commission's assessments outlined above.[84] In its *Facebook/WhatsApp* merger decision the Commission held that the transaction would only raise competition concerns if the concentration of data within Facebook's control

---

[76] See below Section III.2.B.2.

[77] Newman, 'Data Justice: Taking on Big Data as an Economic Justice Issue' (n 13) 15.

[78] CMA, 'The Commercial Use of Consumer Data – Report on the CMA's Call for Information' (2015) 82.

[79] Robert Mahnke, 'Big Data as a Barrier to Entry' (2015) 5 Antitrust Chronicle 3 <http://econpapers.repec.org/RePEc:cpi:atchrn:5.2.2015:i=18431>.

[80] Nils-Peter Schepp and Achim Wambach, 'On Big Data and Its Relevance for Market Power Assessment' (2015) 7 Journal of European Competition Law & Practice 120, 2; In the same vein CMA (n 78) 87.

[81] European Commission 'Press Release - Statement of VP Almunia on the Google Antitrust Investigation Press Room Brussels, 21 May 2012' <http://europa.eu/rapid/press-release_SPEECH-12-372_en.htm> .

[82] It has been reported that when Twitter-owned video app Vine tried to use Facebook to let users find friends on the platform, Facebook denied Vine access to its users' friend lists. See Colin Lecher, 'Facebook Fends off New Anti-Monopoly Questions after UK Email Release' (*The Verge*, 5 December 2018) <https://www.theverge.com/2018/12/5/18127483/facebook-emails-uk-parliament-power-documents-antitrust>.

[83] FTC Staff (n 11) 34–36.

[84] See above Section II.

were to allow it to strengthen its position in advertising,[85] and dismissed that scenario because there were 'a significant number of market participants that collect user data alongside Facebook', including Google and other companies such as *inter alia* Apple, Amazon, eBay, Microsoft, AOL, Yahoo!, Twitter, IAC, LinkedIn, Adobe and Yelp,[86] and there would be a large amount of Internet user data valuable for advertising purposes not within Facebook's exclusive control.[87] For illustrative purposes, the Commission calculated the share of data collection across the Internet:[88]
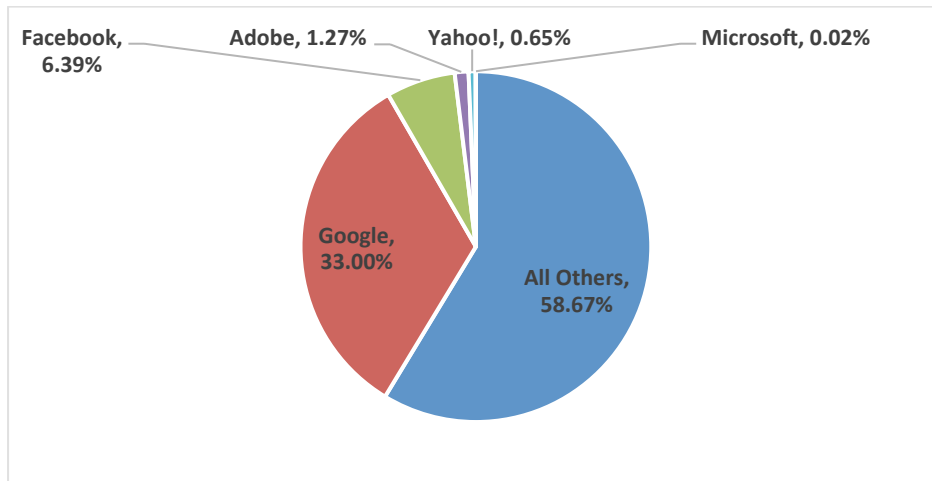
The problem with the Commission's reasoning and calculation of data collection share is that it confuses 'availability of data' with 'substitutability of data': the fact that data is widely available does not entail that different data is substitutable with each other. As Mahnke observes, '[t]here are data, and then there are data'.[89] It is highly uncertain whether the data Microsoft, Twitter, Adobe and Yelp collect are equally valuable to Facebook for its advertising business (or for other purposes) as the data it could collect from WhatsApp after the consummation of the merger. Put in other words, data is not fungible, as the relevance of some kind of data is likely to vary to a great extent with respect to different business models.[90] For example, the information social network platforms have on their users is likely to be as rich or even richer than that gathered by search engines from search queries, but unlike social network data, search data have the advantage of allowing to identify consumers who are in active search for a specific product or service,[91] for which reason they are unlikely to be substitutable.

Curiously enough, the Commission has engaged in assessments of substitutability of data in other cases. For example, in *TomTom/Tele Atlas* the Commission concluded that digital map databases for non-navigational purposes could not be deemed substitutes to navigable digital map databases, as these databases have much more detailed information (such as road type, road class, traffic flow information and turn restrictions) and are therefore more costly and time-consuming to produce.[92] Similarly, in the *Telefónica UK/Vodafone UK/Everything Everywhere* joint venture decision, the Commission found that there were possibly separate markets for online

---

[85] See above text accompanying footnote 54.
[86] *Facebook/WhatsApp* (2014) (n 49) para 188.
[87] See above text accompanying footnote 55.
[88] *Facebook/WhatsApp* (2014) (n 49) para 188.
[89] Mahnke (n 79) 3.
[90] Schepp and Wambach (n 80) 121.
[91] Autorité de la Concurrence and Bundeskartellamt (n 1) 44.
[92] Case COMP/M4854, *TomTom/Tele Atlas* (2008) para 22.

and mobile data analytics,[93] as data analytics for static online advertising cannot be substituted by data analytics for mobile advertising. Lack of substitutability arose from the fact that the two services collect 'different type of information and amount of consumer details'; for example 'the information collected via mobile data analytics is usually more personal, geo-located, and can be cross referenced with call behaviour, which cannot be offered by online data analytics to a comparable extent.'[94] Similarly, to dismiss a potential input foreclosure strategy, in *Microsoft/LinkedIn* the Commission held that LinkedIn's full data would constitute only one of the many types of data needed for machine learning in CRM software solutions,[95] and also, regard being had to the quality of the data collected by LinkedIn (such as for example users' professional details, connections, interests, posts and endorsements), that LinkedIn full data was likely to be relevant only for certain purposes (such as for the CRM B2B Sales and B2B Marketing sub-segments), but not for others.[96] It is submitted that assessments of data substitutability like the aforementioned should be the starting point of a sound data advantage test.

In addition, the Commission's assessment in *Facebook/WhatsApp* also confuses 'availability of data' with 'accessibility to data'. Other competitors may collect data alongside Facebook, and there might well have been a large amount of data up-for-grabs post-transaction, but that hardly means that Facebook's actual and potential competitors could have had ready access to the kind of data Facebook was able to access post-merger.

As seen above, some online platforms offer 'free' services in exchange for personal data, which is subsequently monetised on any of the sides of the markets they serve (mostly in the form of targeted advertising). Accordingly, any entrant that wishes to compete in an online platform market (for example, in a search engine, social network or App store market) must build a platform capable of providing similar services allowing for the extraction of similar data. Setting up a platform demands high investments in R&D. For example, FTC Staff found that search and search advertising platforms 'require enormous investments in the technology and infrastructure required to crawl and categorize the entire Internet', noting that Microsoft invested in 2010 more than USD 4.5 billion to develop its algorithms and building the physical capacity necessary to operate Bing.[97] In addition, the collection, storage, processing and analysis of user data involves substantial fixed costs and low or negligible marginal costs,[98] which means that established firms have cost advantages over entrants in this regard.

Furthermore, setting up the platform is not enough. To launch an effective challenge (and therefore to have access to the data necessary to compete), competitors must be able to attract a sufficiently large user base. To this effect, they must overcome direct and/or indirect network effects, which is a rather daunting task. Direct network effects arise when there is interaction between the users of a product, and having more users makes the product more useful and valuable for all users. A good example is given by social network platforms such as Facebook: the more users on the network, the higher its attractiveness is for its users, as they have a larger

---

[93] *Telefónica UK/Vodafone UK/Everything Everywhere/JV* (2012) (n 44) para 202.

[94] ibid para 200.

[95] Noting that the data needed for ML in CRM software solutions comes mainly from two data sources: in-house customer data uploaded in the CRM software (which relates to accounts, service tickets, interactions, leads, etcetera, the availability of which for each relevant provider of CRM software solutions would not be affected by the Transaction) and complementary third party data (which could be different depending on the use case and the relevant industry). *Microsoft/LinkedIn* (2016) (n 56) paras 259–261.

[96] ibid.

[97] FTC Staff (n 11) 76.

[98] CMA (n 78) 75.

universe of people with whom they can interact. As a matter of fact, every new member of Facebook brings in 200 friends on average.[99] In turn, indirect network effects arise when the increasing use of a product enhances its attractiveness to another economic group, which in turn renders indirect benefits for the original users of the product. Think about OSs: widespread adoption of an OS attracts application developers who will devise and make available new applications compatible with such OS, thereby increasing its value for its users. Increased value to the platform's different customer groups results in a positive feedback loop, whereby more customers on one side attract more customers on the other side. Additionally, as will be seen below,[100] data-driven economies of scale, scope and speed, learning-by-doing and spill-overs reinforce the positive feedback loop and enhance the quality (and hence the attractiveness) of the incumbent's products and services, thereby making a successful challenge against it all the more difficult. This difficulty may be compounded even further by switching costs and/or status quo bias[101] arising from an unmatchable distribution channel (such as for example, a search engine app preinstalled in a dominant mobile OS).

The limited impact of data intermediaries on data availability in combination with the lack of substitutability and inaccessibility of data in specific cases calls into question the soundness of the Commission's test, as these considerations greatly limit the countervailing effect of data's wide availability when concentrations of valuable datasets take place. Indeed, a post-merger combination of datasets composed of widely available, non-rivalrous and non-exclusive data such as, for example, search query data with browsing data or social media user profiles with messaging data, may be unproblematic without qualifications under the Commission's test; however, since actual and potential competitors may neither be able to replicate such data (as it is unsubstitutable with other data they hold such as locational or transactional data) nor have ready access to it (as compiling that data may be prohibitively costly or time-consuming), barriers to entry may nevertheless be raised, thereby entrenching the merged entity's market power.

B.   The role of data's volume/learning-by-doing, variety, velocity and spill-overs

The Commission's test is excessively focused on the volume of data available post-transaction.[102] This is problematic, because although the volume of data is certainly important for the provision of data-driven products and services, the mere possession of large volumes of data will not make a firm better off. Rather, a firm will enjoy a data advantage depending on what it can do with the data it holds, and this in turn depends not only on the volume of the data, but also on its variety, the speed at which it is processed, learning-by-doing effects and spill-overs[103] that flow from the extent to which the firm holding the data is horizontally and/or vertically integrated. Ultimately, data is an asset in flux, the competitive significance of which depends on the structural and technological ability of a firm to collect, process and extract value from it. These factors, which must be at the core of the analysis of any data advantage, make data an extraordinary asset well capable of leading to competition concerns in a manner not previously seen in 'traditional' scarce and exclusive assets.

---

[99] House of Lords, 'Online Platforms and the Digital Single Market' 24.
[100] See below Section III.2.B.
[101] See for example *Facebook/WhatsApp* (2014) (n 49) para 111.
[102] See for example ibid para 189 and *Microsoft/LinkedIn* (2016) (n 56) para 180, where the Commission refers to 'large amount of Internet user data'.
[103] Stucke and Grunes (n 13) 170.

1.   Volume/Learning-by-doing effects

The scale of data is important, although not in and of itself. Rather, benefits arise from the collection of large volumes of data insofar as more data allows a firm to train its algorithms to provide better quality services, thereby eliciting learning-by-doing effects. Take the example of search engines. Online search services are provided on the user side free of charge, for which reason search engines compete on the basis of quality and innovation.[104] Perhaps the most important dimension of search services' quality is the provision of fast 'relevant' search results to users. When confronted with a given search query, the search engine must 'guess' which links the user entering the query is more likely to click on.[105] In particular, when a user enters search terms, such terms are processed by the search engine's algorithms, which determine the probability that a specific webpage will be relevant to the search. Then, the user 'receives results that are rank-ordered based on the search engine's judgment of the likelihood that each result matches what the user was seeking in entering the search terms. This process necessarily depends on multiple variables and constant refinement.'[106]

The more users a search engine has, the more data the search engine has at its disposal to improve the relevance of its results, and therefore the more trials its algorithms will be able to conduct to guess consumer preferences. Accordingly, there is a mutually reinforcing relationship between data, trials and quality: more data enables more trials thereby leading to enhanced quality, and greater quality, in turn, attracts more users. Indeed, the OECD has noted that '[t]he accumulation of data can lead to significant improvements of data-driven services which in turns can attract more users, leading to even more data that can be collected […]For example, the more people use services such as Google Search, or recommendation engines such as that provided by Amazon, or navigation systems such as that provided by TomTom, the better the services as they become more accurate in delivering requested sites and products, and providing traffic information, and the more users it will attract.'[107] Put in other words, there are increasing returns to scale from data collection.

In *Microsoft/Yahoo! Search Business*,[108] the Commission analysed Microsoft's acquisition of Yahoo!'s business in online search and search advertising. The Commission observed that 'scale is an important element to be an effective competitor', and that the majority of respondents to the market investigation considered that Microsoft did not have enough traffic volume to be an attractive alternative to Google.[109] In addition, it found that 'the effects of scale [were] likely to allow the merged entity to run more tests and experiments on the algorithms in order to improve its relevance.'[110] The Commission ultimately approved the merger, as it predicted that the

---

[104] See generally Maurice E Stucke and Ariel Ezrachi, 'When Competition Fails to Optimize Quality: A Look at Search Engines' (2016) 18 Yale Journal of Law and Technology 70.

[105] For example, if a user enters the query 'Beatles', the search engine must guess whether the user is looking for the band, the Volkswagen car, or the insect.

[106] Eric Schmidt, The Power of Google: Serving Consumers or Threatening Competition? Hearing before the Subcommittee on Antitrust, Competition Policy and Consumer Rights, US Senate Judiciary Committee 2011.

[107] OECD, 'Data-Driven Innovation for Growth and Well-Being: Interim Synthesis Report' (2014) 29.

[108] Case COMP/M5257, *Microsoft/Yahoo! Search Business* (2010).

[109] ibid para 153.

[110] ibid para 223; The US Department of Justice arrived at similar conclusions. See DOJ, 'Statement of the Department of Justice Antitrust Division on Its Decision to Close Its Investigation of the Internet Search and Paid Search Advertising Agreement Between Microsoft Corporation and Yahoo! Inc.' (2010)

merged entity would enjoy greater scale of data and therefore would be able to improve its algorithms through trial and error, thereby exerting more competitive pressure on Google. The evolution of the search and search advertising markets, however, disproved the Commission's prediction, as Google has consolidated itself as the dominant player in those segments. Different explanations may be proffered for the manner in which those markets evolved. Firstly, Microsoft did not have the required business acumen to catch up with Google. On account of Microsoft's longstanding presence in high technology segments, however, this explanation seems too dismissive. Secondly, Microsoft could not overcome Google's head start in collecting data and training its algorithms. After all, Google's search algorithms had been fed with data and improved for at least 12 years by the time Microsoft's Bing was launched.[111] And thirdly, greater scale of data was simply not enough. For example, Google might have been able to combine search query data with data collected from its myriad 'free' services, thereby being able to enhance its search algorithms in a way Microsoft could not do.

Irrespective of the actual factors that determined the evolution of the search and search advertising markets, this discussion illustrates a fundamental problem in the Commission's test. By dismissing competition concerns based on the wide availability of data following a transaction, the fact that the merging firms might have had a head start in collecting data and eliciting learning-by-doing effects is not taken into account. Such a head start may determine that the products and services of the merging firms be already of a quality that competitors cannot match, and the combination of datasets enabled by the transaction is likely to broaden the quality gap between the merging firms' and their competitors' products and services even further, regardless of the 'widely available' data the latter group may access.

2.   Variety (economies of scope)

Depending on the usage of data, the scope of data may be as important as its scale.[112] The integration of data from different sources may significantly increase the value of the dataset.[113] For example, to deliver relevant results in response to queries a search engine has never seen before, data from different sources may be required. Relevance of results returned to a query consisting of a specific product model number, for instance, can depend on whether the search engine has 'crawled' web pages containing such exact model number, or whether such data is obtained from other sources such as product data feeds from manufacturers or retailers.[114]

If a platform offers a variety of services that collect data, economies of scope are likely to arise insofar as data linkage is possible. Linked data is a source of 'super-additive insights' and value that are greater than the sum of its isolated parts (data silos).[115] As Schepp and Wambach explain, '[t]he linkage of [data from different sources] can give companies more insights into user habits, enabling them to further improve their services and reinforce their market position. Generally speaking, the more data a company can combine, the better its chances to gain knowledge that can be used to strengthen its market position.'[116]

<https://www.justice.gov/opa/pr/statement-department-justice-antitrust-division-its-decision-close-its-investigation-internet>.

[111] Google was launched in August 1996, whilst Bing was launched in June 2009.

[112] Autorité de la Concurrence and Bundeskartellamt (n 1) 51.

[113] Rubinfeld and Gal (n 32) 347.

[114] See Lerner (n 71) footnote 123.

[115] OECD, 'Data-Driven Innovation for Growth and Well-Being: Interim Synthesis Report' (n 107) 29.

[116] Schepp and Wambach (n 80) 121.

Google is perhaps the most salient example of data-driven economies of scope. Not only does it collect user data from its search engine, but also from a plethora of data-driven products and services it offers at zero-prices, such as its mobile operating system (Android), web browser (Chrome), email service (Gmail), video streaming site (YouTube), mapping service (Google Maps), social networking service (Google Plus), website analytics tool (Google Analytics), cloud platform service (Google Apps), and many others,[117] in addition to the products offered by its ad-serving companies DoubleClick and AdMob. By the same token, Amazon guides consumers' shopping experience and 'guesses' what sort of products a consumer may like by applying data collected from its e-commerce platform, ad-serving tools and mobile and Kindle devices.[118] Diverse datasets allow large online platforms to create highly detailed user profiles that could not be created with each single service.[119]

In a report by the UK Information Commissioner's Office, variety was reported to be the most important characteristic of big data.[120] This is particularly the case in the world of Artificial Intelligence. Think of, for example, Google's 'Assistant'. The higher users' reliance on Google's plentiful services (such as Search, Maps, Android and Chrome), the greater the variety of personal data about particular users it will have, the better the digital assistant can segment results by user profiles, and the better the digital assistant can personalise results.[121]

Therefore, the assessment of a potential data advantage must also pay heed to the variety of the data to which the merging firms' competitors may have access. The acquisition of a firm in possession of a valuable dataset by a highly horizontally and/or vertically integrated data-driven undertaking active in, for example, the display advertising market, is likely to enhance its data advantage to the detriment of non-integrated providers of display advertising, in spite of the fact that post-transaction there remains 'a sufficient amount of data' valuable for display advertising that is not within the merged entity's exclusive control.

3.   Velocity (economies of speed)

As some types of data lose its value rather quickly,[122] data-driven firms have the necessity to keep gathering up-to-date information about consumers, be it their interests and preferences, purchasing behaviour, location or other relevant information. If, for example, users' interests suddenly change as a consequence of a recent event, data-driven firms need to react rapidly and adapt to the new scenario. In this connection, first access to data and the ability to process it more quickly than competitors may confer a competitive advantage. As Microsoft's consultant

---

[117] Omer Tene and Jules Polonetsky, 'Big Data for All: Privacy and User Control in the Age of Analytics' (2013) 11 Northwestern Journal of Technology and Intellectual Property 239, 250.

[118] Marcus Wohlsen, 'Amazon's Ad Business Makes Twitter's IPO Look Puny' (*Wired*, 2013) <https://www.wired.com/2013/09/amazon-tops-twitter-in-ads/>.

[119] OECD, 'Data-Driven Innovation for Growth and Well-Being: Interim Synthesis Report' (n 107) 29.

[120] ICO, 'Big Data and Data Protection' (2014) 7 <https://ico.org.uk/media/for-organisations/documents/1541/big-data-and-data-protection.pdf>.

[121] 'So the feedback loop adds a dimension: it is no longer the trial-and-error, learning-by-doing from earlier queries, but trial-and-error in predicting individual tastes and preferences from the variety of personal data the company collects across its platform (such as the person's email, geo-location data, social network, browser history) and Internet (from the cookies placed when the person visits a website). Now the digital assistant – in personalizing results – can target users with specific sponsored advertisements that they will more likely click (thereby generating more revenue for the platform operator).' Stucke and Grunes (n 13) 186–187.

[122] For example, current locational data is important for search queries such as "restaurants near me", but historic location data is clearly of less value in this regard.

Susan Athey explains: '[w]hen Michael Jackson died […] there was a huge spike in internet traffic, and the search engine companies wanted to be able to figure out in the first 30 seconds to stop sending people to general pages about the performer and start sending them instead to the latest news. By using the latest data — crowd-sourcing what you want — a search engine can be a quick learner. All search engines try to do that, but how well they do it is a function of how fast they get the data. So Google will do it faster than Bing, because more people come to Google first.'[123]

Moreover, the ability to collect and process data at unparalleled velocity may facilitate the leveraging of a data advantage from one data-driven market to another. In its *TomTom/TeleAtlas* merger decision, the Commission dismissed the possibility that Google could quickly enter the market for navigable map databases, noting that end-user feedback was unsuitable to produce a navigable digital map database. [124] The Commission distinguished between basic digital map datasets and digital map datasets for navigation purposes, noting that a supplier of basic digital map datasets wanting to switch production to navigable digital map datasets would have to commit 'substantial resources' to gather all additional information necessary to achieve navigational capability.[125] This required the deployment of field forces driving the roads to record road features to be included in the database. In addition, after completion of a database, it had to be permanently updated to incorporate all changes made to the road network.[126] The Commission found that even undertakings offerings Internet-based map applications were unlikely to enter the market, as it would be too costly and time-consuming for them to obtain and process the necessary data.[127] However, one year after the *TomTom/Tele Atlas* decision, Google announced the launch of navigational features in Google Maps for Android based on a navigable digital map database powered by end-user feedback.[128] Google's successful entry into the market for navigable map databases can be explained by its ability to combine, process and update a high volume of varied data at a velocity its competitors, including TomTom, could not match.

In particular, Google had the ability to easily integrate into its map database data collected from its search engine such as store locations, restaurant locations and private addresses, and despite the Commission's predictions, it also had the ability to rely on updates sent by Google Maps' users to easily integrate live traffic data and update maps more quickly than TomTom.[129]

Therefore, in addition to the volume (including learning-by-doing effects) and variety of data, a data-advantage test must also take into consideration the technological capacity of the merging firms and their competitors to quickly collect, analyse, update and extract value from large volumes of data. In this light, the conclusions at which the Commission arrived in *Telefónica UK/Vodafone UK/Everything Everywhere* seem correct.[130] The JV did not have access to a volume and variety of data comparable to that of established providers of data analytics and online

---

[123] Kathleen O'Toole, 'Susan Athey: How Big Data Changes Business Management' (*Stanford Graduate School of Business*, 2013) <https://www.gsb.stanford.edu/insights/susan-athey-how-big-data-changes-business-management>.
[124] *TomTom/Tele Atlas* (2008) (n 92) para 120.
[125] ibid para 25.
[126] ibid.
[127] ibid para 158.
[128] Google, 'Announcing Google Maps Navigation for Android 2.0' (October 2009) <http://googlemobile.blogspot.com/2009/10/announcing-google-maps-navigation-for.html>.
[129] Boutin and Clemens (n 13) 2.
[130] See above text accompanying footnote 48.

advertising services like Google, and since it was a newly created entity, it seems safe to assume that it did not have the technological ability to collect, process and update data at a speed unavailable for competitors. Conversely, the Commission's conclusions in *Google/DoubleClick* appear to be mistaken,[131] given that in addition to gaining greater volume and variety of data post-transaction, Google already had the technological ability to collect, process and extract value from data at a speed its competitors could not replicate,[132] and this ability was enhanced after the merger. Indeed, following the acquisition, on top of being able to use the DoubleClick cookie to collect browsing data and improve the quality of ad targeting on the AdSense network (thereby gaining data-driven economies of scale and scope),[133] Google was able to achieve important economies of speed. For example, in the behavioural online advertising industry, Google's AdWords and DoubleClick Bid Manager (which rely on all the data gathered from Google's services, including Google Search) can respond faster to DoubleClick AdX bid requests[134] (which rely on DoubleClick cookie data) as compared to bid requests from other ad exchanges, since Google's data processing technology ensures that there are no latency or timeout issues between these platforms.[135] This means that publishers on DoubleClick AdX always receive bids from AdWords and DoubleClick Bid Manager advertisers, which is highly attractive for them, as more bids increase the profitability of their advertising space.

4.   Spill-overs

As explained above, online platforms benefit from indirect, and sometimes direct, network effects. These effects are amplified by increasing returns to scale, learning-by-doing effects, economies of scope and economies of speed, thereby giving rise to spill-overs across the different data-driven segments served by the relevant platform.

Take the example of Facebook' social networking platform. On the user side, more users increase the value of the platform to other users, thereby attracting new users and consequently more traffic (i.e. direct network effects). A higher number of users and increased traffic translate into more data. The more data users provide, the more data the social network has to carry out experiments to render more 'relevant' social network interactions and generally make its platform more attractive to users. At the same time, increased volume, variety and velocity of data help to improve ad-targeting, thereby increasing advertising revenues, which in turn enables Facebook to acquire data-driven firms[136] and develop new services and features that enhance Facebook's data collection capabilities. More users and improved ad-targeting in turns attract more advertisers (i.e. indirect network effects), thereby increasing advertising revenues even further. Crucially, the

---

[131] See above text accompanying footnotes 42-43.

[132] See above text accompanying footnote 123.

[133] Australian Competition & Consumer Commission, 'Digital Platforms Inquiry - Preliminary Report' (2018) 49.

[134] When a person visits a website that shows behavioural advertising, a bid request is made on behalf of the website by either Supply Side Platforms (SSP) or Ad Exchanges. This bid request is a broadcast of data about the person visiting the website to multiple Demand Side Platforms (DSP) which decide, on behalf of advertisers and based on the profile of the person that the advertiser has instructed them to target, whether or not to place bids for the possibility to serve an ad to the person in question. See Johnny Ryan, 'Behavioural Advertising and Pesonal Data' (2018) 2–5.

[135] Autorité de la Concurrence, 'Opinion No. 18-A-03 of 6 March 2018 on Data Processing in the Online Advertising Sector' (2018) 60 <http://www.autoritedelaconcurrence.fr/doc/avis18a03_en_.pdf>.

[136] Such as *inter alia* Instagram, Onavo and WhatsApp.

data Facebook collects can be processed and reprocessed for subsequent use on any segment of its ecosystem.[137]

When a platform is highly vertically integrated, spill-overs are bound to be more pronounced. As the OECD observes, 'consumers that appreciate customized search results and ads by Google's search and webmail platform will spend more time on the platform, which allows Google to gather even more valuable data about consumer behavior, and to further improve services, for (new) consumers as well as advertisers (on both sides of the market). These self-reinforcing effects may increase with the number of applications provided on a platform, e.g. bundling email, messaging, video, music and telephony as increasing returns to scope kicks in and even more information becomes available thanks to data linkage.'[138] These spill-overs are likely to derive into a snowball effect whereby the big becomes bigger, barriers to entry are raised, and incumbents can maintain and strengthen their dominant position and leverage their market power onto adjacent and even distant markets.[139]

Therefore, a data advantage test must also take into account the extent to which a concentration of data enables and/or reinforces a competitive advantage in different data-driven market segments. The magnitude of such advantage will vary greatly depending on the type of data being combined[140] and the degree to which the merged entity may use the data in other data-driven properties. For example, Google's acquisition of Waze[141] surely enhanced its data advantage in the market for turn-by-turn navigation applications for smartphones and tables. By gaining access to Waze's user base, Google was able to improve its mapping data and update it more quickly, on account of Waze's crowd-sourcing technology that 'enables drivers to build and use live maps, real-time traffic updates and turn-by-turn navigation'.[142] However, the data advantage resulting from the transaction did not end there, as real-time location data is key for advertising purposes.[143] Indeed, when an individual uses Waze, Google is able to process his location data with his search query data and any other '[i]nformation from [Google's] ad partners, including entities that belong to the Google group of companies' to show 'personalised ads'.[144]

IV.   Proposal of a 'Data Advantage' Test

In view of the above, it is argued that the Commission's test to assess a data advantage is overly broad and therefore likely to underestimate the extent of the competitive advantage arising from

---

[137] 'The reuse of data generates huge returns to scale and scope which lead to positive feedback loops in favour of the business on one side of the market, which in turn reinforces success in the other side(s) of the market.' OECD, 'Data-Driven Innovation for Growth and Well-Being: Interim Synthesis Report' (n 107) 29.
[138] ibid.
[139] For example, Google's search query data enabled Google to identify the vertical segments which were posing a competitive threat (such as Shopping, Travel and News) and expand accordingly. In turn, Google's expansion into vertical search segments and mobile OS gave Google access to data that helped it cementing its dominance in its core search segment (as it was able to access data from more sources to improve relevance of search results and target ads on the basis of location and other data collected from mobile devices). In addition, the data Google is able to collect from Google Maps, Waze and Android location services has enabled Google to gain the upper hand in the race for the development of self-driving cars.
[140] This reflects the fact that data is not fungible, as different data are more relevant for the provision of some services than other data. See above text accompanying footnotes 89-91.
[141] Completed acquisition by *Motorola Mobility Holding (Google, Inc) of Waze Mobile Limited,* ME/6167/13.
[142] ibid para 3.
[143] Stucke and Grunes (n 13) 93.
[144] Waze, 'Waze Privacy Policy' <https://www.waze.com/legal/privacy#advertising-campaigns>.

a concentration of data. To solve the deficiencies of the Commission's test, the following data advantage test is proposed:

Whether accessing, processing and extracting value from the data required to provide the products and services of the merging firms, on the scale and scope, and at the speed available to them, taking into account learning-by-doing effects and other relevant structural conditions such as direct and indirect network effects and their degree of horizontal and/or vertical integration in data-driven sectors, is costly, resource-intensive and time-consuming for actual and potential competitors to replicate.

The application of this test is more likely to fully capture the competitive implications of a concentration of data in data-driven sectors. Take the example of the *Facebook/WhatsApp* merger decision. As seen above, the Commission noted that the transaction would only raise competition concerns if the concentration of data within Facebook's control were to allow it to strengthen its position in advertising,[145] and dismissed competition concerns on the basis that many firms collect user data alongside Facebook and post-transaction there would be a large amount of Internet user data not within Facebook's exclusive control.[146] The Commission, therefore, did not consider that the data those firms collect may not be substitutable with the data required to provide Facebook's and WhatsApp's services, and therefore the disciplining effect of the wide availability of data was overstated.[147]

In addition, had the Commission focused on the economies of scale, scope and speed benefiting the merging firms, it would have realised that the data required to compete against Facebook was not readily accessible for actual and potential competitors. Stand-alone providers of display advertising, for example, cannot match the volume[148] and variety[149] of Facebook's data, nor the velocity[150] at which Facebook can collect, process and extract value from it. Similarly, a focus on Facebook's degree of integration would have shed light on the fact that the data advantage resulting from the transaction was likely to have spill-overs across Facebook's segments. In particular, the combination of data post-transaction is almost certain to have entrenched

---

[145] *Facebook/WhatsApp* (2014) (n 49) para 187.

[146] See above text accompanying footnotes 85-88.

[147] Accordingly, a substitution analysis of data such as that conducted in *TomTom/TeleAtlas, Telefónica UK/Vodafone UK/Everything Everywhere* and *Microsoft/LinkedIn* should be the starting point of this assessment. See above text accompanying footnotes 92-96.

[148] Facebook has 2.32 billion monthly active users providing a rich stream of data as of 31 December 2018. See Form 10-K 'Annual Report pursuant to Section 13 or 15(d) of the Securities Exchange Act of 1934' for the fiscal year ending on 31 December 2018, filed by Facebook with the U.S. Securities and Exchange Commission, available at https://www.sec.gov/Archives/edgar/data/1326801/000132680119000009/fb-12312018x10k.htm

[149] Facebook collects data from all of the 'Facebook Products', which include Facebook (including the Facebook mobile app and in-app browser), Messenger, Instagram (including apps like Direct and Boomerang), Portal-branded devices, Moments, Bonfire, Facebook Mentions, Spark AR Studio, Audience Network, Facebook Business Tools, which are tools used by website owners and publishers, app developers, business partners (including advertisers) and their customers to support business services and exchange information with Facebook, such as social plugins (like the 'Like' or 'Share' button) and its SDKs and APIs, and other features, apps, technologies, software, products, and services. See Facebook, 'What Are the Facebook Products? | Facebook Help Centre' <https://www.facebook.com/help/1561485474074139?ref=dp>.

[150] Given its unparalleled audience, Facebook has first access to data about recent events, which enables it to update relevant content more quickly than competing social network platforms, thereby generating more traffic and therefore more data. For example, within the first twelve hours of news that David Bowie had died, thirty-five million people had one hundred million interactions about Bowie's passing on Facebook. See Colin Stutz, 'David Bowie's Death Leads to 100 Million Facebook Interactions in First 12 Hours' (*Billboard*, 2016) <http://www.billboard.com/articles/columns/rock/6836601/david-bowie-death-100-million-facebook-interactions-12-hours>.

Facebook's position in the market for social networking services. Indeed, since August 2016 the phone numbers of WhatsApp users began to be shared with Facebook, which enabled the latter to run analytics on user activity and make friends suggestions based on people with whom users talk on WhatsApp.[151] Lastly, a look at the learning-by-doing and direct and indirect network effects benefiting Facebook[152] would have confirmed that access to the data required to compete against it was costly and time-consuming.

It must be noted that a competitive advantage arising from a concentration of datasets is not necessarily anticompetitive. After all, a data-driven merger may elicit efficiencies to the benefit of consumers in the form of, for example, higher quality products and services or more innovation. However, it is important to establish with the greatest precision possible to what extent a concentration of data makes entry more difficult, so adequate measures can be adopted to keep data-driven markets open, whether in the form of commitments or, if warranted, an outright merger prohibition. Given the data-driven effects explained in Section III.2.B, unrestricted combinations of datasets in the hands of one or a small number of firms have the ability to lead to undesirable market developments which are highly difficult to reverse post-merger.[153]

Conclusions

The advent of the information economy, and with it a shift in the economic value of information, has posed a data paradox. Tangible, scarce, rivalrous and exclusive assets have traditionally been those capable of producing wealth. In the information economy, conversely, control over data, an intangible, ubiquitous, non-rivalrous and non-exclusive asset, is the main source of wealth. This data paradox can be explained by the advantages digital information has relative to analogue information, and by the manifold applications and opportunities that the processing of large amounts of data has afforded.

The data paradox takes a specific form in competition enforcement. Rivalrous assets within the exclusive control of the incumbent are normally those capable of raising barriers to entry. However, in a number of data-driven markets it has been seen that data, in spite of its features, may confer a competitive advantage and make entry more difficult. How can this be possible?

To crack the data paradox in competition enforcement, this article has assessed the test the Commission has devised in its merger decisional practice to determine whether a concentration of data may confer a competitive advantage and raise barriers to entry. The application of that test, which mainly examines whether post-transaction there would be a sufficient amount of available data to compete, suggests that combinations of data are unlikely to lead to competition problems.

---

[151] Michael Duran, 'How to Stop WhatsApp From Giving Facebook Your Phone Number' [2016] *Wired* <https://www.wired.com/2016/08/how-to-stop-whatsapp-from-sharing-your-phone-number-with-facebook/> In turn, more friends connections translate into more user engagement and therefore more data to train its social network algorithms and enhance further its ad targeting capabilities.

[152] Regrettably, the Commission only assessed the significance of network effects in the market for consumer communications services. See *Facebook/WhatsApp* (2014) (n 49) Section 5.1.3.5.

[153] For example, Facebook recently announced that it is planning the integration of Messenger and two of its major acquisitions, Instagram and WhatsApp. Cross-platform integration is likely to make it easier for Facebook to share data across them to improve its ad targeting capabilities. At the same time, it is also likely to make they main components of Facebook's conglomerate more difficult to break up and spin off, in case governments and/or regulators decide that is warranted. See Dave Lee, 'WhatsApp, Instagram and Messenger to "Merge"' (25 January 2019) <https://www.bbc.com/news/technology-47001460>.

However, this article proves that the Commission's test is overly broad, as it does not take into account important considerations regarding the nature and dynamics of data. Some types of data are not substitutable with other types of data, for which reason clarifications must be made as to the exact kind of data that would be available post-transaction. Moreover, even if data is available from many sources (as users may, for example, multihome and surrender their data to different market players), it may not be readily accessible on the scale, scope and at the speed needed to compete effectively against the incumbent. In addition, learning-by-doing, direct and indirect network effects have the ability to reinforce the strength of a data advantage. Therefore, in spite of data's features, a concentration of data is nevertheless capable of leading to a scenario where the big becomes bigger and barriers to entry are raised in data-driven markets.

To solve the flaws of the Commission's test, this article proposes an alternative 'data advantage' test that departs from the Commission's marked focus on the theoretical wide availability of data, allowing for more accurate, comprehensive and sound assessments of the competitive implications of concentrations of data in data-driven markets.