RESEARCH OUTPUTS / RÉSULTATS DE RECHERCHE

Big data and competition policy

Bourreau, Marc; De Streel, Alexandre; Graef, Inge

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Link to publication

Citation for pulished version (HARVARD):
Bourreau, M, De Streel, A & Graef, I 2017, *Big data and competition policy: market power, personalised pricing and advertising : project report*. CERRE, Bruxelles.

Big Data and Competition Policy:

Market power, personalised pricing and advertising

Project Report

Marc Bourreau (CERRE and Telecom ParisTech)

Alexandre de Streel (CERRE and University of Namur)

Inge Graef (Tilburg University and KU Leuven)

With contribution from Tommaso Valletti (Imperial College)

16 February 2017

Table of contents

About CERRE

Providing high quality studies and dissemination activities, the Centre on Regulation in Europe (CERRE) promotes robust and consistent regulation in Europe's network industries. CERRE's members are regulatory authorities and operators in those industries as well as universities.

CERRE's added value is based on:

- its original, multidisciplinary and cross-sector approach;
- the widely acknowledged academic credentials and policy experience of its team and associated staff members;
- its scientific independence and impartiality;
- the direct relevance and timeliness of its contributions to the policy and regulatory development process applicable to network industries and the markets for their services.

CERRE's activities include contributions to the development of norms, standards and policy recommendations related to the regulation of service providers, to the specification of market rules and to improvements in the management of infrastructure in a changing political, economic, technological and social environment. CERRE's work also aims at clarifying the respective roles of market operators, governments and regulatory authorities, as well as at strengthening the expertise of the latter, since in many Member States, regulators are part of a relatively recent profession.

About the authors[1]

Marc Bourreau is a Joint Academic Director of CERRE, Professor of Economics at Telecom ParisTech, and director of the Innovation & Regulation Chair at Telecom ParisTech. He is also affiliated with the interdisciplinary institute for innovation (i3) for his research. Marc graduated in engineering from Telecom ParisTech in 1992. He received his doctorate in economics from University of Paris 2 Panthéon-Assas in 1999, and a "Habilitation à Diriger des Recherches" from University of Paris 1 Panthéon-Sorbonne in 2003. From 1997 to 2000, he worked as a regulatory economist at France Telecom/Orange. He became assistant professor at Telecom ParisTech in 2000. Marc has published widely in leading economics journals. He is Co Editor-in-Chief of Information Economics & Policy, and a member of the editorial boards of the Review of Network Economics, Telecommunications Policy and the DigiWorld Economic Journal (formerly Communications & Strategies). He is also a member of the scientific committee of the Florence School of Regulation at the European University Institute in Florence (Italy), an associate researcher of the Laboratory of Industrial Economics (LEI), and an associate researcher of Cepremap. His main research interests are in industrial organisation, regulation, telecommunications, and digital economics.

Alexandre de Streel is a Joint Academic Director of CERRE, Professor of European law at the Universities of Namur and Louvain in Belgium and the Director of the Research Centre for Information, Law and Society (CRIDS), focusing his research on Regulation and Competition Law in the network industries. He is also a member of the Scientific Committee of the Florence School of Regulation (FSR) at the European University Institute in Florence, and Research Fellow at the European Institute of Public Administration (EIPA) in Maastricht. Alexandre regularly advises international organisations (such as the European Commission, European Parliament, OECD, EBRD) and national regulatory authorities on regulatory and competition issues in network industries. He is also an Assessor (member of the decisional body) at the Belgian Competition Authority.

Inge Graef is a postdoctoral researcher at the Tilburg Institute for Law, Technology, and Society (TILT) and the Tilburg Law and Economics Center (TILEC), as well as guest lecturer at KU Leuven. The focus of her research is on competition enforcement in the digital economy. Inge is particularly interested in the interface between competition law and other fields of EU law such as data protection, intellectual property and electronic communications law. Prior to joining Tilburg University, Inge was affiliated to the Centre for IT & IP Law (CiTiP) of KU Leuven where she prepared a doctoral dissertation about the interaction between competition law and data on online platforms. Her PhD thesis was published as a book in 2016 by Kluwer Law International under the title 'EU Competition Law, Data Protection and Online Platforms: Data as

---

Essential Facility'. Inge has also been involved in a number of policy studies for the European Commission and the Belgian government on topics such as net neutrality, electronic communications law and the regulation of business-to-business relationships in the online platform environment.

Executive Summary

Building on the CERRE Report of January 2016 on the economics and the regulation of personal data written by Pierre Larouche, Martin Peitz and Nadya Purtova, this CERRE Report studies three specific issues around the application of competition policy to big data. The first issue relates to market power assessment and analyses the power given by data control in the big data value chain. The second and third issues relate to abuse of dominance assessment and analyse the use of data to personalise prices and to target advertising. To deal with those three issues, we recommend an analytical and governance framework for competition agencies.

Data and Market power

To assess the importance of data in determining market power in a big data value chain, we recommend an analytical framework based on three principles and two questions.

The first principle is that **data are one input, which is important but not unique, to developing successful applications and algorithms**. Other inputs are also important, such as a skilled and creative labour force (in particular computer scientists and engineers), capital and distribution channels. Above all, the skills and creativity of the labour force will be key to the success of the applications.

The second principle is that the big data value chain (data collection, storage and analysis) exhibits **many direct and indirect network effects that need to be captured by the competition authorities**. That requires an understanding and an analysis of each part of the value chain. However, it also requires an understanding of the interaction and possible feedback loops between its different parts, rather than analysing one part of the value chain in isolation. Over 10 years ago, in a more specific context, Julian Wright already warned competition authorities not to apply a one-sided logic to two-sided markets.

The third principle is that **each big data application and algorithm is different and should be analysed on a case-by-case basis**. It would be inappropriate to propose detailed recommendations at a general level beyond a broad framework for analysis because each case is different.

With those principles in mind, a competition authority trying to assess market power in the big data value chain should answer **two main questions**:

The first question relates to the **value of the data** under examination for the applications and the algorithms under consideration. Answering that question depends on determining:

- the extent of the economies of scale in the data, in particular what is the marginal benefit of having more data under examination to improve the quality of the application under consideration;

- the extent of the economies of scope in the data, in particular how important it is to combine different types of data to improve the quality of the application under consideration;
- the time depreciation value of the data, in particular the relationship between the age of the data and its relevance to developing or improving the application under examination.

The second question relates to the **availability of the data** under examination for the applications and the algorithms under consideration. Answering that question depends on determining:

- the possibility and the costs for an application developer to collect data directly from users or machines,
- the possibility and the costs for the application developer to buy the data from a data broker and in a data market place.

Such **data availability is to a large extent influenced by the legal framework** regarding data collection and use. As this framework is different in the EU than in other parts of the world (and within the EU, different in some Member States than in others), as well as being different for firms offering some services than for competitors offering other services, those legal differentiations should be factored into the competition analysis.

Use of data and personalised prices

The welfare effects of personalised pricing are a priori ambiguous. Price discrimination is not necessarily detrimental to welfare or consumer surplus and can increase welfare and/or consumer surplus in comparison to uniform pricing. From an economic viewpoint, there is therefore **no rationale for banning personalised pricing** *per se*.

One concern is that price discrimination could be used as a monopolisation device. For example, an incumbent firm may pre-empt entry in a given market or consumer segment by setting very low prices in this market or segment. Incumbent firms could also offer loyalty discounts to prevent entry of competitors. This type of concern could be aggravated if possibilities of price discrimination hinge on detailed consumer data, and incumbent firms have exclusive access to this consumer data. Such potential exclusionary practices call for an intervention from competition authorities, but not for a ban of personalised pricing *per se*.

Our main policy recommendation would therefore be that **personalised pricing strategies, if they exist, should be transparent** to ensure consumers' trust in online markets, which would positively affect all players. This may require **clarification on the implementation of the general principles of EU consumer rules to personalised pricing** with more developed guidelines than the ones adopted by the European Commission in 2016.

This also requires an **effective application of consumer protection** in all the Member States. For instance, online prices may need to be monitored by consumer protection agencies upon

complaints. One possibility would be to audit pricing algorithms, but it seems far too complex and costly, and it may have a negative effect on firms' incentives to innovate through sophisticated pricing algorithms. In addition, a requirement to disclose algorithms goes to the heart of a business' operations and interferes with valuable trade secrets. Therefore, this approach seems neither viable nor desirable. Another solution would be to use automated tests, for example via virtual mystery shoppers. We do not necessarily recommend a continuous monitoring program, but rather that **on a regular basis, the pricing strategies of online players be monitored**.

Use of data and target advertising

It is hard to make a general statement on whether targeted advertising is a good or a bad from a welfare point of view. There is a complex trade-off between the potential good matches that effective targeted ads can generate (which benefit both consumers and sellers), the pass-through of advertising revenues to final consumers in terms of lower prices for online services, and the nuisance for consumers from intrusive ads.

Competition authorities seem to be concerned that some players in the online advertising market may have enough market power, thanks in particular to a better access to data, to distort competition. Large players are also vertically integrated at various stages of the online advertising ecosystem. Though vertical integration can, in some cases, raise competition concerns, it can also increase efficiency. Overall, **standard competition policy suffices to deal with potential competitive problems in online advertising**.

Online advertising seems to be an extremely dynamic and innovative market. The organisation of this market has changed drastically over the last few years, with the emergence of new categories of players. Competition for the best algorithms, the best advertising technologies, also seems fierce. This is to the benefit of advertisers – when ads become more effective – and consumers – when they become less of a nuisance. One main recommendation would therefore be to make sure that no player blocks the innovation dynamics that are at play. Given the dynamic nature of the market, our view is that **anything beyond monitoring is not warranted as long as no new evidence appears that suggests that there is market power by some players and that this is abused**.

A governance framework

Competition issues in big data can be complex and evolve quickly. They are shaped by technological progress and influenced by legislative changes, in particular the rules for consumer protection, data protection and intellectual property. Moreover, as we saw for personalised pricing and target advertising, good market functioning requires transparency which may be guaranteed by clarification and effective implementation of the consumer protection rules.

Those assessments call for the following reforms in the governance framework:

- **C**ompetition agencies should improve the understanding of the functioning of the three main parts of the big data value chain: collection, storage and processing. In order to do so, they may set up structured dialogues with the big data industry and stakeholders, organise hearings, exchange information and ideas with foreign competition agencies, or even launch a formal sector enquiry provided it does not create an unreasonable burden for the sector.

- The **expertise of competition agencies in data and computer science should increase** either in-house or with external experts.

- **Antitrust agencies should cooperate closely with agencies in charge of consumer protection, data protection and intellectual property** to better understand the common problems they face, to better understand how consumer protection rules, data protection rules and IP law will influence the functioning of big data markets and, when intervention is needed, to secure that all those authorities adopt consistent decisions. However, institutional cooperation does not mean legal fusion. The role of each agency and legal instrument should be differentiated, as they are complements and not substitutes.

1. The big data value chain

According to De Mauro et al (2016), big data *'is the information asset characterized by such a high volume, velocity and variety to require specific technology and analytical methods for its transformation into value.'* Thus, the difference between data and big data is the famous four V's made possible by technological progress: the *volume* of data processed, the *variety* of data aggregated, the *velocity* at which data is collected and used and the *value* of the information found in the data.

As explained in OECD (2015, ch.2 and 2016) as well in Mayer-Schonberger and Cukier (2013), big data are collected, exchanged, stored and value is extracted in a complex eco-system made of many related markets which are often multi-sided:

1. Data are collected directly from users and from machines in many different ways or can be bought from data brokers;
2. Data are stored on internal servers or on external cloud computing services;
3. Data are analysed with software analytics and the valuable information can be used to improve and personalise products' characteristics and prices as well as their marketing, to improve process and organisation or for many other purposes such as controlling epidemics or managing emergencies.

1.1 Data collection

A firm may collect data directly, usually having a direct contact with the person or the object from which the data is collected, or indirectly, usually by buying the data from data brokers.

1.1.1 Direct collection

Firms may collect personal and non-personal data about users, as well as machines, in many different online and offline ways. For the particular case of the online collection of personal information, such ways include:

- First, some information is *publicly observed* through device, operating system, IP address, etc.

- Second, some information is *voluntarily provided* by the consumers either with knowledge when registering to a website, such as name, date of birth, email or postal address for delivery, etc. or often without knowledge when logging into a website (login-based data) such as products the consumer is looking for, purchases, etc.

- Third, some information can be collected by *tracking the consumer* online which can be achieved in different ways: (i) *tracking cookies*, which are a specific type of cookie that is distributed, shared, and read across two or more unrelated websites for the

purpose of gathering information or presenting customised data to a consumer;[2] (ii) *browser and device fingerprinting*, which is a method of tracking web browsers by the configuration and settings information they make visible to websites; (iii) *history sniffing*, which is the practice of tracking which sites a user has or has not visited (by hacking its browser history list); (iv) *cross-device tracking* offers the ability to interact with the same consumer across her desktop, laptop, tablet, wearable, and smartphone, using both online and offline information; (v) through the use of applications by the user, this information is accessible for the Operating System owner as well as for the developer of the application.

1.1.2 Data exchange and data intermediaries

Firms may also procure data from third parties, even though this remains a marginal practice in Europe today. On the basis of a business models survey done by Deloitte, the European Commission observes that "*in the vast majority of cases (78% of the companies surveyed) data is generated and analysed in-house by the company or by a sub-contractor. Vertical integration remains the principal strategy in the sectors surveyed. Data stays within an organisation and is not traded with third parties. This is particularly the case in sectors with a high presence of large, technologically advanced companies, such as banks and telecom providers or automotive and machinery producers*."[3] According to this study, data trading only accounted for 4% of the companies surveyed.

In this case, data can be obtained from data brokers which have been defined by the US FTC (2014:3) as '*companies whose primary business is collecting personal information about consumers from a variety of sources and aggregating, analysing, and sharing that information, or information derived from it, for purposes such as marketing products, verifying an individual's identity, or detecting fraud*'. The consultancy IDC (2016) has a broader concept of data marketplace defined as "*a third party, cloud-based software platform providing Internet access to a disparate set of external data sources for use in IT systems by business, government or non-profit organizations. The marketplace operator will manage payment mechanisms to reimburse each dataset owner/provider for data use, as necessary. Optionally, the marketplace provider may provide access to analysis tools that can operate on the data.*" IDC notes that in their simplest form, data marketplaces are online stores where firms can buy and sell data.

---

[2] For instance, when a consumer visits a website that places an ad from a third-party vendor, the third-party vendor can place a cookie on the consumer's computer. If the consumer visits other websites that display ads from the same vendor, the vendor knows that the same consumer visited the two websites

[3] Commission Staff Working Document of 10 January 2017 on the free flow of data and emerging issues of the European data economy, SWD(2017) 2, p. 15.

The secondary market for data is not yet well understood by public authorities. In the US, a Committee of the Senate in 2013 and the Federal Trade Commission in 2014[4] conducted inquiries to better understand those markets. The FTC concluded that:

- the US data broker industry is complex with multiple companies exchanging lots of data between themselves;
- data brokers have a vast amount of data on almost every US household and commercial transaction;[5]
- data brokers combine offline and online data from multiple different sources and it is very difficult for a consumer to know and determine how a data broker obtained her data;
- Data brokers analyse the data and make inference about the consumers, placing them into categories that may be sensitive.[6]

1.2 Data storage and cloud computing

The storage of massive quantities of data requires large data centres consisting of big clusters of computers connected by fast local area networks. Those data centres are expensive to build and characterised by economies of scale.

However, the development of cloud computing[7] allows small firms to rent instead of owning the data centres, thereby converting their fixed costs into variable costs (Greenstein et al., 2013). For the cloud computing market to function properly, the costs of switching between providers need not to be too high, which raises the issues of interoperability and portability in the cloud. Moreover, the competition among cloud providers may be limited by data localisation

---

[4] This report examines the following 9 data brokers: Acxiom, Corelogic, Datalogix, eBureau, ID Analytics, Intelius, PeekYou, Rapleaf, and Recorded Future.

[5] According to the FTC Report, one data broker's database has information on 1.4 billion consumer transactions and over 700 billion aggregated data elements; another data broker's database covers one trillion dollars in consumer transactions; and yet another data broker adds three billion new records each month to its databases. Most importantly, data brokers hold a vast array of information on individual consumers. For example, one of the nine data brokers has 3000 data segments for nearly every U.S. consumer. Based on companies' reports, Lambrecht and Tucker (2015:5) note that: Acxiom has multi-sourced insight into approximately 700 million consumers worldwide with over 1,600 pieces of separate data on each consumer; Datalogix asserts that its data includes almost every U.S. household; Bluekai states that it has data on 750 million unique users per month with an average of 10-15 attributes per user.

[6] Potentially sensitive categories include those that primarily focus on ethnicity and income levels, such as "Urban Scramble" and "Mobile Mixers," both of which include a high concentration of Latinos and African Americans with low incomes. Other potentially sensitive categories highlight a consumer's age such as "Rural Everlasting," which includes single men and women over the age of 66 with "low educational attainment and low net worths," while "Married Sophisticates" includes thirty-something couples in the "upper-middle class with no children." Yet other potentially sensitive categories highlight certain health-related topics or conditions, such as "Expectant Parent," "Diabetes Interest," and "Cholesterol Focus."

[7] Cloud computing service is defined by EU law as a digital service that enables access to a scalable and elastic pool of shareable computing resources: Article 4(19) of the Directive 2016/1148 on Network Information Security.

restrictions which can be important for certain types of privately owned data, in particular for health, financial and gaming/gambling data as well as for publicly owned data.[8]

1.3 Data analytics and use

The third step in the big data value chain is the analysis of those data to extract relevant information, mainly with correlation patterns. This is done by applications and algorithms which are increasingly learning by themselves. The development and the improvement of those algorithms are based on many inputs such as data, skilled and creative labour force (in particular computer scientists and engineers) or capital. Thus, data are important but probably not the most important input as mentioned by Lerner (2014).

Analytical applications and algorithms can be developed in-house and, for some, may require important investment in getting the best skills and volume of data. They may also be obtained from a third party. In this case, as for cloud computing, the fixed development costs can be converted into variable costs.

The information found in the data can have multiple uses: they can improve products for all thanks to a better understanding of consumers' needs. Those improved products can be data-rich (mainly intangible) such as a map or data-less rich (more tangible) such as a drive-less car;[9] they can better personalise products' prices or marketing strategies (see sections 4 and 5 for more details); they can also improve process, marketing and organisation, thereby increasing productive and dynamic efficiencies;

Thus, information inferred from big data can increase consumer welfare and GDP growth. Although the efficiency gains are hard to measure, OECD (2015) suggests that data-driven innovation firms benefit, on average, from a 5% to 10% faster productivity growth than similar companies that do not use data-driven innovation. Buchholtz et al. (2014) estimate that big data applications would allow the EU economy to grow by an additional 1.9% by 2020. McAfee and Brynjolfsson (2012) estimate that companies in the top third of their industry in the use of data-driven decision making were, on average, 5% more productive and 6% more profitable than their competitors. [10]

---

[8] Commission Staff Working Document of 10 January 2017 on the free flow of data and emerging issues of the European data economy, SWD(2017) 2, pp. 5-10.

[9] Prüfer and Schottmüller (2016) have built a model where firms can leverage their position from data-rich markets (for instance maps) to data-less rich markets (for instance drive-less cars) with products improvements based on data.

[10] In those studies, it is not clear whether there is a causal relationship between data-driven innovation and productivity increase or whether it is a mere correlation.

2. Main legal constraints on data collection and use

2.1 Legal typology of data

Currently, EU data protection legislation is in a period of transition from the Data Protection Directive that was adopted in 1995 to the General Data Protection Regulation, which will apply as of 25 May 2018.[11] One of the objectives of these legislative instruments is to protect the fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data. To achieve those objectives, these instruments set up a typology of data that reflects the extent to which the fundamental rights of individuals should be safeguarded in the context of the processing of particular types of personal data in line with the associated data protection risk.[12]

Table 1: Typology of data with the associated data protection risk

As such, the typology is unrelated to the underlying economic value of the data. Certain types of non-personal data might have more economic value than personal data while the processing of only the latter is subject to the requirements of EU data protection legislation. Many

---

[11] Directive 95/46 of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (*Data Protection Directive*) [1995] OJ L 281/31; and Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (*General Data Protection Regulation*) [2016] OJ L 119/1.

[12] In its January 2017 proposal for a Regulation on Privacy and Electronic Communications, the European Commission introduced new terminology: (i) *electronic communications data* consists of electronic communications content and electronic communications metadata; (ii) *electronic communications content* means 'the content exchanged by means of electronic communications services, such as text, voice, videos, images, and sound'; (iii) *electronic communications metadata* means 'data processed in an electronic communications network for the purposes of transmitting, distributing or exchanging electronic communications content; including data used to trace and identify the source and destination of a communication data on the location of the device generated in the context of providing electronic communications services, and the date, time, duration and the type of communication'.

applications of big data analytics enable companies to produce more efficiently, while no collection or use of personal data is involved. For example, supermarket chain Tesco developed a technology relying on the real-time analysis of refrigeration data from in-store sensors. This technology enabled Tesco to cut refrigeration energy costs by up to 20%.[13] While the collected data thus had significant economic value, it did not concern information relating to individuals.

In addition, the typology does not equate with market definition as performed under competition law. A relevant upstream, intermediate or downstream market for data may include personal as well as non-personal data depending on the particular circumstances. The definition of a relevant market for data is complex and requires the competition authority to establish what types of data are substitutable for one another (see Graef, 2016).

Thus, the purpose of the typology discussed here is merely to identify the relevant legal constraints applicable to the collection and use of particular types of data.

2.1.1 Personal data

Definition of personal data

The key notion of personal data is defined as '*any information relating to an identified or identifiable natural person (data subject)*'. An identifiable natural person is, in turn, defined as '*one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person*'.[14] Thus, the notion of personal data includes information directly identifying natural persons such as a list of customer names and contact details but also, for instance, IP addresses because they can be linked to an individual.[15]

In order to determine whether a natural person is identifiable, attention should be paid to the means reasonably likely to be used either by the controller or by any other person to identify the said person. Further guidance on how to interpret this concept of identifiability of natural persons was given by the Article 29 Working Party[16] and has now been codified in the General

---

[13] B. Goodwin, 'Tesco uses big data to cut cooling costs by up to €20m', *ComputerWeekly.com*, 22 May 2013, available at http://www.computerweekly.com/news/2240184482/Tesco-uses-big-data-to-cut-cooling-costs-by-up-to-20m.

[14] These are the definitions contained in Article 4(1) of the General Data Protection Regulation. For the previous wording, see Article 2(a) of the Data Protection Directive.

[15] Opinion 4/2007 of the Article 29 Working Party on the concept of personal data, WP 136, 20 June 2007. This is now acknowledged in recital 30 of the General Data Protection Regulation. In addition, the Court of Justice made clear in its *Breyer* judgment that so-called dynamic IP addresses which change each time there is a new connection to the internet also constitute personal data in the case of a website operator who has the legal means which enable it to identify the data subject with additional data which the internet service provider has about that person (Case C-582/14, *Patrick Breyer v Bundesrepublik Deutschland*, ECLI:EU:C:2016:779).

[16] The Article 29 Working Party was set up under Article 29 of the Data Protection Directive and is composed of a representative from the national data protection authority of each EU Member State, a representative of the European Data Protection Supervisor (the independent supervisory authority that is responsible for ensuring that all EU institutions and bodies respect people's right to personal data protection and privacy when processing their personal data) and a representative of the European Commission.

Data Protection Regulation which states that '*account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments*'.[17] The Article 29 Working Party made clear that the state of the art in technology at the time of the processing and new identification possibilities during the period for which the data will be processed should be taken into account.[18]

Subcategories of personal data

Based on the different levels of protection offered under EU data protection law, a distinction can be made between several types of personal data.

Stricter rules apply to the processing of **special categories of personal data, referred to as sensitive data,** which includes information revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation. Processing of this type of information is prohibited unless one of the lawful grounds of processing listed in EU rules applies.[19]

Pseudonymised data is personal data that is processed in such a way that the data can no longer be attributed to a specific data subject without the use of additional information.[20] An example of pseudonymisation is key-coded data used for statistical or scientific purposes. Key-coded data is information relating to a data subject who is earmarked by way of a code that is not derived from information identifying the individual. The key making the correspondence between the code and the common identifiers of the individual (such as name, address, date of birth) is kept separately. This way, taking into account all the means reasonably likely to be used, it is possible only to trace the information back to the data subject by referencing the key.[21] Since data that has undergone pseudonymisation may still identify a natural person, it is considered as personal data to which the data protection rules fully apply.[22]

---

[17] Recital 26 of the General Data Protection Regulation.

[18] In particular, the Article 29 Working Party made a distinction between two scenarios: (i) if the data are intended to be stored for one month only, '*identification may not be anticipated to be possible during the "lifetime" of the information, and they should not be considered as personal data*'; (ii) if the data are intended to be kept for 10 years, '*the controller should consider the possibility of identification that may occur also in the ninth year of their lifetime, and which may make them personal data at that moment*': Article 29 Working Party, 'Opinion 4/2007 on the concept of personal data', WP 136, 20 June 2007, p. 15.

[19] Article 8 of the current Data Protection Directive or Article 9 of the future General Data Protection Regulation

[20] Article 4(5) of the General Data Protection Regulation contains the following definition of pseudonymisation: '*the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person*'.

[21] See the discussion in Article 29 Working Party, 'Opinion 4/2007 on the concept of personal data', WP 136, 20 June 2007, p. 18-20.

[22] Recital 26 of the General Data Protection Regulation.

2.1.2   Non-personal data

Non-personal data is information that is not protected under data protection law. A distinction can be made between anonymous and anonymised data.

Anonymous data is information which does not relate to an identified or identifiable natural person taking account of all the means likely reasonably to be used either by the controller or by any other person to identify that individual. An example of anonymous data is machine data that relates to and is created by the activity of computers, mobile phones and other systems or devices for instance in the context of the Internet of Things. However, data generated by machines without human intervention can also constitute personal data where information is collected about the behaviour of an identifiable natural person, such as call detail records generated by telephony systems and search logs generated by online search engines.

Anonymised data is personal data that is rendered anonymous in such a manner that the data subject is no longer identifiable.[23] In other words, anonymised data is information that used to refer to an identifiable person, but where that identification is no longer possible. Anonymisation is a technical measure that may be combined with legal, organisational or contractual obligations. Anonymised data is excluded from the scope of application of the EU data protection rules if the anonymisation is irreversible.[24]

2.2   Rules applicable to personal and non-personal data[25]

2.2.1   Consumer protection rules

EU consumer law may provide protection to consumers in relation to how (personal and non-personal) data is used by businesses. The main horizontal rules are: the Directive 2005/29 concerning *unfair commercial practices* and the accompanying Commission Guidance which have been recently updated to take into account the development of online platforms;[26] the directive 93/13 on *unfair commercial terms* in consumer contracts;[27] and the directive 2011/83

---

[23] Recital 26 of the General Data Protection Regulation. Also Opinion 4/2007 of the Article 29 Working Party of 20 June 2007 on the concept of personal data, WP 136, p. 21. See also the Code of practice of November 2012 of the UK Information Commissioner's Office on Anonymisation drawing a distinction between *anonymisation* techniques used to produce aggregated information and those processes such as *pseudonymisation* – that produce anonymised data but on an individual-level basis. The latter can present a greater privacy risk but not necessarily an insurmountable one.

[24] Article 29 Working Party, 'Opinion 05/2014 on Anonymisation Techniques', WP 216, 10 April 2014.

[25] See also Osborne Clark (2016).

[26] Directive 2005/29 of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market, OJ [2005] L 149/22 and Commission Staff Working Document of 25 May 2016 on Guidance on the implementation/application of the Directive 2005/29 on Unfair commercial practices, SWD(2016) 163.

[27] Council Directive 93/13 of 5 April 1993 on unfair terms in consumer contracts, OJ [1193] L 95/29 as amended.

on *consumer rights*.[28] Those directives are under review[29] and additional rules have been proposed by the Commission, in particular for the supply of digital content.[30]

The scope of application

The EU consumer protection rules applies to B2C relationships[31] between a *trader*, acting for the purposes relating to his trade, business, craft or profession and a *consumer* acting outside his trade, business, craft or profession.

Those rules apply for goods or services which are exchanged against remuneration. However, in an online environment where more and more services (or goods) are paid with data and not with money, a crucial issue is what constitutes a remuneration and a payment. EU law is evolving in that point. In 2014, the Court of Justice of the EU, in case involving a Greek news Internet site, recognised that remuneration may come from the recipient but also from income generated by advertisements posted on a website.[32] In 2016, the Commission proposed the draft Digital Content Directive where consumer protection rules apply to digital content provided in exchange for a price or counter-performance in the form of personal or any other data.[33] This is because a differentiation depending on the nature of the counter-performance would discriminate between business models and would provide an unjustified incentive for businesses to move towards offering digital content against data.[34] Currently, the application of the Consumer Rights directive to goods or services paid with data is being discussed in the context of the REFIT evaluation.

Rules for protection

Regarding data use, the most important consumer protection instrument is the Unfair Commercial Practice Directive. It sets three levels of unfair practices that are prohibited:

- First, practices which (i) are against professional diligence and (ii) materially distorts or is likely to materially distort the economic behaviour of an average consumer;[35]

---

[28] Directive 2011/83 of the European Parliament and of the Council of 25 October 2011 on consumer rights, OJ [2011] L 304/64. See also DG Justice Guidance document of June 2014 on the Consumer Rights Directive.
[29] See: http://ec.europa.eu/consumers/consumer_rights/review/index_en.htm
[30] Proposal for a Directive of the European Parliament and of the Council on certain aspects concerning contracts for the supply of digital content (proposal for a Digital Content Directive), 9 December 2015, COM(2015) 634 final.
[31] Note that some Member States apply also those legislations to B2B relations.
[32] Case 291/13, *Papasavas*. ECLI:EU:C:2014:2209, point 30. Recital 18 of the directive 2000/31 also provides that: "*information society services are not solely restricted to services giving rise to on-line contracting but also, in so far as they represent an economic activity, extend to services which are not remunerated by those who receive them, such as those offering on-line information or commercial communications, or those providing tools allowing for search, access and retrieval of data*".
[33] Article 3(1) of the proposal for a Directive of the European Parliament and of the Council on certain aspects concerning contracts for the supply of digital content (proposal for a Digital Content Directive), 9 December 2015, COM(2015) 634 final.
[34] Recital 13 of the proposal for a Digital Content Directive.
[35] Professional diligence is defined as '*the standard of special skill and care which a trader may reasonably be expected to exercise towards consumers, commensurate with honest market practice and/or the general principle of*

- Second, practices which are (i) misleading by action or omission or aggressive using harassment, coercion or undue influence and (ii) cause or are likely to cause a transactional decision that an average consumer would not have taken otherwise;

- Third, specific misleading or aggressive practices that are listed in the Directive, independently of their concrete effects on the average consumer.

The 2016 Guidance of the Commission gives some clarification on the application of those prohibitions to the online sector, in particular to dynamic pricing, price discrimination and personalised pricing.[36] The main requirement is adequate information for consumers and transparency about how the prices are determined (for more details, see section 4.3 below on price discrimination).

The Unfair Contract Terms Directive also limits some data use. A standard contract term is considered unfair if, contrary to the requirement of good faith, it causes a significant imbalance in the parties' rights and obligations arising under the contract, to the detriment of the consumer.

The proposal for a Digital Content Directive[37] requires a supplier to provide a consumer who terminates a contract for the supply of digital content '*with technical means to retrieve all content provided by the consumer and any other data produced or generated through the consumer's use of the digital content to the extent that data has been retained by the supplier*'.[38] This provision overlaps to a certain extent with the right to data portability introduced in the General Data Protection Regulation. The latter, however, only covers personal data provided by the data subject while the Digital Content Directive would enable a consumer to retrieve any other data generated by the use of the digital content to the extent it has been retained by the supplier. This broad wording raises questions as to the practical feasibility of the data retrieval requirement as well as the relationship with the right to data portability in the General Data Protection Regulation.

---

good faith in the trader's field of activity': Article 2(h) of the Unfair Commercial Practices Directive. Materially distort the economic behaviour of consumers' refers to the use of '*a commercial practice to appreciably impair the consumer's ability to make an informed decision, thereby causing the consumer to take a transactional decision that he would not have taken otherwise*': Article 2(e) of the Unfair Commercial Practices Directive.

[36] Commission Guidance SWD(2016) 163, p. 143-147.

[37] At Article 13(2c).

[38] Article 16(4)(b) of the proposal for a Directive of the European Parliament and of the Council on certain aspects concerning contracts for the supply of digital content (proposal for a Digital Content Directive), 9 December 2015, COM(2015) 634 final, provides for a similar obligation for suppliers with regard to long term contracts for the supply of digital content.

2.2.2 The protection of intellectual property and trade secrets

Database protection

The Database Directive[39] provide two types of protection for the database, defined as a "*collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means*":

- First, *copyright protection* for the intellectual creation involved in the selection and arrangement of materials; that is those databases that meet the requirement for creativity of the work;
- Second, *sui generis protection* for an investment in the obtaining, verification or presentation of the contents of the databases.

However, the Directive does not protect the content of the database[40] or the software used to create the database. It is the scheme of the database that is protected.

Trade secret protection

The recently adopted Trade Secrets Directive[41] provides a protection against unlawful acquisition, use and disclosure of (i) a secret, (ii) which has commercial value and (iii) has been subject to a reasonable measure of protection. According to the Commission, such protection may apply to datasets when they have been protected, but not to individual raw data generated by machines, mostly because of a lack of commercial value.[42]

2.2.3 Competition rules

Competition law applies to big data business models as any other business models and, under some conditions, imposes an obligation to share data even when they are protected by an intellectual property right. The main cases on data sharing are *Magill*,[43] *IMS Health*,[44] *Microsoft*[45] and *Huawei*.[46] According to the Commission, the Court of Justice has set four conditions that need to be fulfilled to lead to an obligation to licence the use of commercially-held information:

---

[39] Directive 96/9 of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases.

[40] Case C-604/10, *Football Dataco and Others v. Yahoo! UK and Others*, ECLI:EU:C:2012:115, para 30; Case C-203/02, *The British Horseracing Board and Others v William Hill Organization,* ECLI:EU:C:2004:695; Case C-444/02, *Fixtures Marketing v Organismos prognostikon agonon podosfairou AE (OPAP)*, ECLI:EU:C:2004:697.

[41] Directive 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure, OJ [2016] L 157/1.

[42] Commission Staff Working Document of 10 January 2017 on the free flow of data and emerging issues of the European data economy, SWD(2017) 2, p. 20.

[43] Case C-241/91P and C-242/91P, *Radio Telefis Eireann (RTE) et Independent Television Publications (ITP) v Commission*, ECLI:EU:C:1995:98.

[44] Case C-418/01, *IMS Health v.NDC Health*, ECLI:EU:C:2004:257.

[45] Case T-201/04, *Microsoft v. Commission*, ECLI:EU:T:2007:289.

[46] Case C-170/13, *Huawei Technologies v. ZTE*, ECLI:EU:C:2015:477.

(i) the data is indispensable for the downstream product, (ii) there would not be any effective competition between the upstream and downstream product, (iii) refusal prevents the emergence of the second product, (iv) there is no objective reason for the refusal.[47] They are commonly referred to as the conditions for controlling an essential facility and are only exceptionally met in practice.

2.2.4   Specific rules imposing access to data

Next to those general rules, some specific rules regulate the access to data in certain contexts. The most important rule concerns data owned by the public authorities that should, in principle, be in open access.[48]

In addition, some rules impose access to data in some specific sectors of the economy. In its Staff working document on the data economy,[49] the European Commission mentioned access to in-vehicle data for after-sale services such as maintenance and repair[50] or the access to payment information for the Fintech.[51]

2.3   Rules applicable to personal data and privacy rules[52]

2.3.1   General data protection law

EU data protection rules lay down a number of obligations with which controllers and processors have to comply. A *controller* is the person or organisation who determines the purposes and means of the processing of personal data.[53] The *processor* is the natural or legal person who processes personal data on behalf of the controller.[54]

---

[47] Commission Staff Working Document of 10 January 2017 on the free flow of data and emerging issues of the European data economy, SWD(2017) 2, p. 22. See also Guidance of 3 December 2008 on the Commission's Enforcement Priorities in Applying Article 82 EC Treaty to Abusive Exclusionary Conduct by Dominant Undertakings O.J. [2009] C 45/7, paras 75-90.

[48] Directive 2003/98 of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information as amended by Directive 2013/37.

[49] Commission Staff Working Document of 10 January 2017 on the free flow of data and emerging issues of the European data economy, SWD(2017) 2, p. 21.

[50] Regulation 715/2007 as amended.

[51] Directive 2015/2366.

[52] As made clear by the Article 29 Working Party, other sets of rules such as tort law, criminal law or antidiscrimination law may provide additional protection to individuals in cases where various legitimate interests are at stake: Article 29 Working Party, 'Opinion 4/2007 on the concept of personal data', WP 136, 20 June 2007, p. 24.

[53] Article 2(d) of the Data Protection Directive and Article 4(7) of the General Data Protection Regulation.

[54] Article 2(e) of the Data Protection Directive and Article 4(8) of the General Data Protection Regulation.

Material and geographical scope of application

EU rules are applicable to the processing of personal data by automated means as well as to manual processing, if the data are contained or are intended to be contained in a filing system.[55] The term 'processing' is given a broad definition and consists of '*any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means*'.[56] As a result, almost every activity relating to personal data qualifies as a form of processing. Examples of processing activities include the collection, structuring, storage, use, dissemination and erasure of personal data.

The current Data Protection Directive applies to processing of personal data carried out in the context of a controller's EU establishment or, in case the controller is not established in EU territory, through the use of equipment situated in the EU.[57] The future General Data Protection Regulation will also be applicable if neither the controller, nor its equipment is situated in the EU. The processing of personal data of data subjects who are in the EU is caught by the General Data Protection Regulation as soon as the processing activities are related to the offering of goods or services to such data subjects, irrespective of whether a payment of the data subject is required, or to the monitoring of their behaviour as far as their behaviour takes place within the European Union.[58] As a result, businesses active in the EU will be subject to EU data protection rules even though they are based outside the Union.

Data quality requirements

EU rules contain the so-called data quality requirements under which controllers have to meet the following obligations:[59]

- *Lawful, fair and transparent processing* which requires the controller to have a legitimate ground to process personal data such as consent of the data subject, performance of a contract, legal obligation or legitimate interests of the controller;[60]
- *Purpose limitation* which requires that personal data are collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes;

---

[55] Article 3(1) of the Data Protection Directive and Article 2(1) of the General Data Protection Regulation. A filing system is defined as '*any structured set of personal data which are accessible according to specific criteria, whether centralised, decentralised or dispersed on a functional or geographical basis*' (Article 2(c) of the Data Protection Directive and Article 4(6) of the General Data Protection Regulation).

[56] Article 2(b) of the Data Protection Directive and Article 4(2) of the General Data Protection Regulation.

[57] Article 4(1a) and (c) of the Data Protection Directive.

[58] Article 3(2a) and (b) of the General Data Protection Regulation.

[59] Article 6(1) of the current Data Protection Directive and Article 5(1) of the future General Data Protection Regulation.

[60] See the legitimate grounds listed in Article 7 of the Data Protection Directive and Article 6(1) of the General Data Protection Regulation.

- *Data minimisation* which requires that personal data have to be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed;
- *Accuracy* which requires that personal data have to be accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay;
- *Storage limitation* which requires that personal data have to be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed;
- *Integrity and confidentiality* (as added in the General Data Protection Regulation): which required that personal data have to be processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures.

Rights to erasure and data portability

There already exists a right of access which entitles data subjects to obtain from the controller confirmation as to whether or not personal data are being processed, and, where that is the case, access to this data.[61] The General Data Protection Regulation extends this by introducing a right to erasure and a right to data portability.

The **right to erasure**, also referred to as the right to be forgotten, entitles a data subject to obtain from the controller the erasure of personal data in a number of situations such as where the personal data are no longer necessary in relation to the purposes for which they were processed, where the data subject withdraws consent and no other legal ground for the processing remains, or where the data subject objects to the processing.[62]

The **right to data portability** gives a data subject the right to receive her personal data that she has provided to a controller in a structured, commonly used and machine-readable format and to transmit this data to another controller.[63] Where technically feasible, the data subject also has the right to have the data transmitted directly from one controller to another. For instance,

---

[61] See Article 12 of the current Data Protection Directive and Article 15 of the future General Data Protection Regulation.

[62] Article 17(1) of the General Data Protection Regulation. In *Google Spain*, the Court of Justice has already established a related 'right to be delisted' with regard to the processing of personal data in the context of search engines. In its 2014 judgment, the Court held that a search engine provider is responsible for the processing that it carries out of personal information which appears on web pages published by third parties. If, following a search made on the basis of a person's name, the list of results displays a link to a web page which contains information on the person in question, that person may approach the search engine provider directly and request, under certain conditions, the removal of that link from the list of results: Case C-131/12, *Google Spain SL, Google Inc. v. Agencia Española de Protección de Datos, Mario Costeja González*, ECLI:EU:C:2014:317.

[63] Article 20 of the General Data Protection Regulation and Guidelines of the Article 29 Data Protection Working Party of 13 December 2016 on the right to data portability, WP 242.

Facebook and Google offer users the possibility to export their data by obtaining a copy of their accounts.[64]

Interestingly, data portability may also result from the application of competition law but with a different scope of application. If the prohibition or the refusal to port data, which may be personal or not personal, by a dominant company amounts to an abuse, a competition authority may impose data portability. Indeed, previous Competition Commissioner Almunia made clear in a 2012 speech that the right to data portability goes to the heart of competition policy.[65] In this case, the portability will apply to all data but only for dominant firms while under the General Data Protection Regulation, the portability applies only to personal data provided by the data subject but for all the firms[66] as illustrated in the Table 2 below.

Table 2: Scope of data portability obligation

	Dominant firms	Non dominant firms
Personal data		
Non-personal data	C	

Rights relating to profiling

Other requirements apply where necessary, depending on the circumstances and purposes of the processing. For instance, EU data protection rules provide for a right of the data subject not to be subject to a decision based solely on automated processing which produces legal effects concerning, or significantly affecting, her.[67] Under the General Data Protection Regulation, automated individual decision-making includes profiling which is defined as *any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements*.[68]

In addition, EU rules give the data subject the right to object to the processing of personal data for direct marketing purposes which, as the General Data Protection Regulation makes clear, includes profiling to the extent that it is related to such direct marketing.[69]

---

[64] Facebook offers the 'Download Your Info' feature (https://www.facebook.com/help/131112897028467) and Google has the 'Google Takeout' tool (https://takeout.google.com/settings/takeout).

[65] Former Competition Commissioner Almunia, 'Competition and personal data protection', Privacy Platform event: Competition and Privacy in Markets of Data Brussels, 26 November 2012, SPEECH/12/860.

[66] Swire and Lagos (2013) criticise the GDPR because the new portability right applies to all firms and not only to those holding a dominant position, hence even small start-ups have to bear the costs of developing expensive software codes to enable the porting of data.

[67] Article 15 of the current Data Protection Directive and Article 22 of future the General Data Protection Regulation.

[68] Article 4(4) of the General Data Protection Regulation.

[69] Article 14(b) of the Data Protection Directive and Article 21(2) of the General Data Protection Regulation.

2.3.2   Sector-specific data protection rules: ePrivacy Directive

The ePrivacy Directive[70] complements the current Data Protection Directive regarding the processing of personal data in the telecommunications sector and lays down a number of obligations on the providers of publicly available electronic communications services:

- They should take appropriate technical and organisational measures to safeguard *security* of their services;[71]
- They should ensure the *confidentiality* of communications, albeit allowing an almost unconditional exception to safeguard national security, defence, public security, and the prevention, investigation, detection and prosecution of criminal offences.[72] The Commission made clear in its January 2017 proposal for a Regulation on Privacy and Electronic Communications that the '*transmission of machine-to-machine communications involves the conveyance of signals over a network and, hence, usually constitutes an electronic communications service*'. On that basis, it is clarified that the principle of confidentiality enshrined in the Regulation should also apply to the transmission of machine-to-machine communications.[73]
- They should require *consent of users when cookies* or other forms of accessing and storing information in a user's terminal equipment are employed, which led to the implementation of the so-called cookie-banners ;[74]
- Moreover, *unsolicited communications* are only allowed for the purposes of direct marketing in case prior consent of the respective users (opt-in) has been obtained;[75]
- They can only process *traffic data* to the extent and for the duration necessary for the purpose of marketing electronic communications services or for the provision of value added services if the subscriber or user to whom the data relate has given consent.[76] Traffic data relating to subscribers and users must be erased or made anonymous when it is no longer needed for the purpose of the transmission of a communication;[77]

---

[70] Directive 2002/58 of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (*ePrivacy Directive*) [2002] OJ L 201/37 as amended by Directive 2009/136.

[71] Article 4 of the ePrivacy directive. These obligations formed the basis of Article 32 of the General Data Protection Regulation.

[72] Article 5(1) of the ePrivacy directive

[73] Recital 12 of the proposal for a Regulation on Privacy and Electronic Communications.

[74] Article 5(3) of the ePrivacy directive

[75] Article 6(3) of the ePrivacy directive. Note that If the provider has received the electronic contact details of a customer in the context of the sale of a product or service, it is allowed to use these details for direct marketing of its own similar products or services.

[76] Article 6(3) of the ePrivacy directive. Traffic data means '*any data processed for the purpose of the conveyance of a communication on an electronic communications network or for the billing thereof*'. Value added service as '*any service which requires the processing of traffic data or location data other than traffic data beyond what is necessary for the transmission of a communication or the billing thereof*'.

[77] It is worth noting that traffic data is also regulated in Article 5(1) in the context of the confidentiality of communications.

- They can only process *location data* (other than traffic data) when made anonymous or with the consent of the users.[78]

The rules relating to cookies and unsolicited communications are generally interpreted to apply to any entity making use of cookies or sending messages in the context of publicly available electronic communications services in public communications networks.[79]

This is different for the provisions relating to the protection of traffic and location data which are currently only applicable to traditional telecom operators. These asymmetries raise level playing field issues as other providers that do not offer electronic communications services, such as WhatsApp, Facebook or Skype, also process these types of data and offer, from the perspective of users, similar services over the internet.[80] The fact that telecom operators are subject to stricter obligations than these so-called Over-The-Top (OTT) players also has competitive effects considering that the limitations imposed by the ePrivacy Directive affect the extent to which telecom operators may process and use personal data in their business activities.[81] To address this issue, the CERRE report authored by Larouche, Peitz and Purtova on Consumer Privacy in network industries found that a future proof regulation requires a common approach to all industries and that sector-specific privacy regulations are inadequate in a dynamic environment and should be withdrawn.

In its current review of the ePrivacy Directive, the European Commission is looking into these issues as well as whether there is a need to retain the specific provisions of the ePrivacy Directive in light of the changes brought about by the horizontally applicable General Data Protection Regulation. The proposal for a Regulation on Privacy and Electronic Communications[82] indeed extends the scope of application of its obligations to other players (beyond telecom operators) also providing communication services such as providers of interpersonal communications services (which may or may not be number-based, such as for example, Voice over IP, messaging services and web-based e-mail services).

2.3.3 Privacy protection

Where the general or sector-specific data protection rules do not apply, an activity may still constitute an interference with the right to privacy as protected by the European Convention on

---

[78] Article 6(3) of the ePrivacy directive. Location data is defined as '*any data processed in an electronic communications network, indicating the geographic position of the terminal equipment of a user of a publicly available electronic communications service*'.

[79] For a detailed analysis, see Time Lex and Spark (2015, p. 51-63 for cookies and 89-93 for unsolicited communications). As such, the scope of Articles 5(3) and 13 is not limited to providers of electronic communications services[79] and also applies to providers of information society services.

[80] Ecorys and TNO (2016); WIK-Consult and TNO (2015).

[81] As has already been observed by Peitz, Schweitzer and Valletti (2014: p. 7 and 53).

[82] Proposal for a Regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (proposal for a Regulation on Privacy and Electronic Communications), 10 January 2017, COM(2017) 10 final.

Human Rights of the Council of Europe and/or the Charter on Fundamental Rights of the European Union. [83]

The scope of the right to privacy and the right to data protection overlaps but does not coincide. The former protects the private sphere of individuals irrespective of whether any processing of personal data is involved. For instance, voyeurism may be considered as an interference of someone's right to privacy while the right to data protection is not applicable as long as no personal data is processed. On the other hand, the right to data protection includes all personal data and not only information that is invasive of one's private life.

An interference with the right to privacy has to be justified in order to be lawful under the European Convention on Human Rights. A justification is examined in three steps:

(1) The interference must be in accordance with the law (i.e. have a sufficient legal basis);

(2) The interference must pursue a legitimate aim: national security, public safety or the economic wellbeing of the country, the prevention of disorder or crime, the protection of health or morals, or the protection of the rights and freedoms of others);

(3) The interference must be necessary in a democratic society (i.e. be proportional).

2.4 Conclusion

Thus, EU law contains many rules on the collection and the use of data which are shaping the functioning of big data markets. Those rules are more extensive for personal than for non-personal data.

Some of those rules make the collection and the use of raw data or processed data more difficult, thereby raising the static entry barriers to the data economy. This is, for instance, the case for most of the personal data protection rules but also for some rules relating to IP and trade secret protection. Other rules can make the collection or the use of data easier, thereby facilitating entry in the data economy. This is, for instance, the case of the data portability rule which facilitates switching between services providers or the specific rules imposing data sharing for public or privately owned data.

---

[83] Resp. Art. 8 ECHR and Article 7 Charter.

3. Data and Market Power

As data is having a key role in the big data value chain, this section analyses a specific but important question: whether and under which conditions the control of raw data can be a source of market power in a big data ecosystem.

3.1 The broader context

The big data eco-system is complex and involves many firms active on related markets which are often multi-sided. Therefore when assessing market power, competition authorities should keep a broad view taking into account the main characteristics of the eco-system and the relationship between markets (Autorité de la concurrence and Bundeskartellambt, 2016: 26-30; UK House of Lords, 2016: Chapter 4). Those main characteristics are:

- the presence of direct and indirect *network effects* which may lead to snowball effects where the markets tip in favour of a small number of players;[84] Network effects often change the type of competition, which is often *for* the market (Schumpeterian) rather than *in* the market;
- the steep *experience curve* of the self-learning algorithms which may substantially increase the first-mover advantage and also lead to snowball effects and market tipping;
- the *relationships between the different markets, often of multi-sided nature*, which may, in some circumstances, ease the leverage of a dominant position from one market to another;
- the extensive *multi-homing* of customers which may be affiliated to several online platforms at the same time;[85]
- The rate of *innovation* which is quick and often unpredictable and disruptive, leading to possible rapid displacement of powerful but lazy firms.

Therefore, when running an antitrust case in a big data industry, competition agencies should take into account the general characteristics of the big data eco-system which may, in some circumstances, amplify some of the effects of data control. In particular, they should take into account the direct and indirect network effects on the demand side and the multi-sidedness of the markets and leveraging possibilities on the supply side. The agencies should also adopt a dynamic view of market evolution without trying to predict, or worse shape, the future technology and market evolutions.

---

[84] Belleflamme and Peitz (2015), Shapiro and Varian (1999).

[85] Multi-homing has been recognised by the General Court of the EU as a mitigating factor for finding dominance when it upheld the Commission approval of the acquisition of Skype by Microsoft: Case T-79/12, *Cisco and Messagnet v. Commission*, ECLI:EU:T:2013:635, para 79 et sq

3.2 Dataset and market power

To determine whether the control of a dataset can be a source of market power in a big data value chain, the joint report by the Autorité de la concurrence and the Bundeskartellambt (2016:35) mentions two relevant factors to analyse: (i) the scarcity of data (or ease of replicability) and (ii) whether the scale/scope of data collection matters to competitive performance.

More generally, it is important to assess the effects of the possible entry barriers in the main steps of this value chain. This assessment should be done on a case-by-case basis and depends very much on the type of data and the type of use, in particular its algorithmic treatment, of such data.[86] This section focuses on data collection and data analysis where the risks of entry barriers, because of potential non-replicability, are *a priori* higher than for data storage.

3.2.1 Data collection

Costs of collecting data

One of the inherent characteristics of data, as many intangible goods, is the non-rivalry which means that the same data can be collected and used many times without losing value.[87] This fundamental characteristic decreases, *ceteris paribus*, the cost of collection as data collection by one firm does not impede other firms to do the same.

However, technical, legal or contractual restrictions may weaken or even remove the inherent non-rivalry of data to make them exclusive. [88]

- *Technical barriers* can make the collection of data more difficult or even impossible, for instance with encryption techniques.
- *Legal barriers*: some of the legal rules reviewed in section 2 above increase the costs of collection. This is in particular the case for personal data whose means of collection are limited by the general data protection rules and even more by the telecom specific data protection rules. As we have seen, some of those rules are more stringent in the EU than in other parts of world and within the EU, in some Member States more than in others. In addition, telecom-specific rules are more stringent than the general rules, thereby increasing the costs of collection for firms offering traditional telecom services compared to the others.

---

[86] Rubinfeld and Gal (2016). For instance, the 2014 DoJ's action against the merger of *Bazaarvoice* and its leading rival *Power-Reviews* established that data can serve as an entry barrier in the market for rating and review platforms: DOJ, Antitrust Division, Competitive Impact Statement of 8 May 2014, 13-cv-00133 WHO, http://www.justice.gov/atr/case-document/file/488826/download

[87] This is why the often used analogy between data and oil is misleading.

[88] See Graef (2016).

- *Contractual barriers*: Firms may also impose contractual restrictions, such as exclusivity clauses, on the transfer of data.[89]

In some cases, data are *collected for their own sake* and the collecting firm is ready to invest solely to gather data. This was the traditional business model of polling institutes and market research firms such as GFK or TNS or financial information companies such as Bloomberg or Reuters. This is also the new business model of the (often) Internet firms offering products which are supposedly free because they are not paid with money but with data (whose value has increased with the development of big data).[90] Interestingly, this new business model exhibits in general more network and experience effects than the traditional model and may lead to large platforms offering supposedly free services and collecting many (often personal) data (see also Autorité de la concurrence and Bundeskartellambt (2016:38).

In other cases, data are *collected as by-product* of the selling of another goods or services and the collecting firm does not invest specifically to gather the data. The standard example is customers' lists that firms are building over time as they sell their products. This data collection as by-product increases with proliferation of connected devices and the diminishing costs of information storage.

The cost of collection is obviously higher when data are collected for their own sake than as by-product. To be sure, those are two extreme cases and reality may lie in-between. Indeed, there are some cases where a firm may want to invest in improving its products just to get better data as by-products.

Antitrust assessment of data availability and replicability

On the basis of the availability of data and their collection costs, the competition agencies have determined in several cases whether dataset were replicable.

In some merger cases involving data-rich Internet firms, the Commission concluded that datasets of the merging parties were replicable, hence the combination of those data would not significantly impede competition.

- In *Google/DoubleClick*,[91] the Commission considered that the combination of the information on search behaviour from Google and web-browsing behaviour from DoubleClick was already available to a number of Google's competitors, hence would not give the merged entity a competitive advantage.
- In *Facebook/WhatsApp*,[92] the Commission considered that a large amount of Internet user data that are valuable for online advertising are not within the exclusive control

---

[89] For an analysis of some of those contractual clauses, see Osborne Clark (2016).

[90] An interesting and somewhat linked issue which is not dealt in this report is how competition analysis should deal with supposedly free markets where data is a means of remuneration, hence analysis purely based on monetary prices may be insufficient or even misleading.

[91] Commission Decision of 11 March 2008, Case M.4731 *Google/ DoubleClick*, para. 364-366.

[92] Commission Decision of 3 October 2014, Case M.7217 *Facebook/WhatsApp*, para. 167-189.

of Facebook. Therefore, even if Facebook would use WhatsApp as a potential source for user data to improve Facebook's target advertising, this would not significantly impede competition on the online advertising market.[93]

- In *Microsoft/LinkedIn* the Commission considered that no competition concerns on the market for online advertising arose from the concentration of the parties' user data that can be used for advertising purposes because a large amount of such user data would continue to be available after the transaction.[94]

However, in some abuse of dominance cases, some national competition authorities decided that customer lists gathered by firms enjoying a legal monopoly may not be reproducible by competitors[95] and cannot be used to launch other services which are under competition.

- In September 2014, the Autorité de la concurrence adopted an interim decision in which it found GDF Suez capable of taking advantage of its dominant position in the market for natural gas by using customer files it had inherited from its former monopoly status to launch offers at market prices outside the scope of its public service obligation. As regards the reproducibility of the database, the French agency considered that it was not reasonably possible for the competitors to reproduce the advantage held by GDF Suez or to rely on other databases from which information could be retrieved that was effective for prospecting new customers in the market for the supply of gas.[96]
- In September 2015, the Belgian competition authority imposed a fine on the National Lottery for having abused its dominance in the Belgian market for public lotteries in which it has a legal monopoly. When entering the competitive market for sports betting in 2013, the National Lottery used the contact details of individuals contained in a database that it had established in the context of its legal monopoly in order to send a one-off promotional email for the launch of its new sports betting product Scooore![97] Given its nature and size, the Belgian competition authority concluded

---

[93] Note that at the time of the merger in August 2014, Facebook indicated to the Commission that it was unable to establish a reliable automated matching between Facebook and WhatsApp user accounts. In August 2016, WhatsApp announced the possibility of linking WhatsApp user phone with Facebook user identities and it appears that technical possibilities of automatic matching existed already in 2014 contrary to Facebook allegation. On that basis, the Commission has sent a statement of objections against Facebook for having provided incorrect or misleading information: see Commission Press release of 20 December 2016, IP/16/4473.

[94] Commission Decision of 6 December 2016, Case M. 8124 *Microsoft/LinkedIn*.

[95] For a more elaborate analysis of indispensability of data, see Graef (2016: 270-273).

[96] Autorité de la concurrence, Décision 14-MC-02 du 9 septembre 2014 relative à une demande de mesures conservatoires présentée par la société Direct Energie dans les secteurs du gaz et de l'électricité, par. 147-154.

[97] Belgian Competition Authority, Beslissing BMA-2015-P/K-27-AUD van 22 september 2015, Zaken nr. MEDE-P/K-13/0012 en CONC-P/K-13/0013, *Stanleybet Belgium NV/Stanley International Betting Ltd en Sagevas S.A./World Football Association S.P.R.L./Samenwerkende Nevenmaatschappij Belgische PMU S.C.R.L. t. Nationale Loterij NV*, par. 44-48.

that the contact details could not have been reproduced by competitors in the market at reasonable financial conditions and within a reasonable period of time.[98]

3.2.2 Data analysis

The relationship between, on the one hand, the volume, the variety and the velocity of the data and, on the other hand, the quality of the analytical tools (often a computer algorithm) and the value of the inferred information is complex and changing with rapid progress in artificial intelligence. It depends on the type of data and analysis performed.

Volume of data: the economies of scale

The marginal return of having more data depends very much on the type of data and the type of analysis which is done. Economies of scale in data use may be low when data are used for inference purpose, but higher for other usages.

For search services, the economies of scale are lower for head queries which are frequently entered by users than for tail queries[99] which are rarer. As noted by the Monopolkommission (2015, para 202): *"While the added value of a frequently searched term can thus be very low, seldom-made search queries may make a major contribution towards improving search results. Such infrequent search queries are likely to particularly include those search queries concerning for instance current events with regard to which there is as yet no information on users' conduct, and search queries consisting of several terms, "long-tail queries".*

However, the level of those scale economies is not clear. Some authors like Lerner (2014:37), Lambrecht and Trucker (2015:10) or Sokol and Comerford (2016) submit that the scale economies are low even for tail queries and that there is a diminishing marginal return of data both for head and tail queries.[100] Others like Mc Afee (2015) find that more data matters for tail queries. In addition, in the context of the *Microsoft/Yahoo! Search Business* merger decision, Microsoft argued that with larger scale a search engine can run tests on how to improve the algorithm and that it is possible to experiment more and faster as traffic volume increases because experimental traffic will take up a smaller proportion of overall traffic.[101]

The extent of the economies of scale is thus an empirical question which should be tested in each case on the basis of the type of data and application at hand. In particular, the necessity of having more data to improve the quality of the application and the algorithm should be carefully analysed.

---

[98] *Ibidem*, par. 69-70.
[99] Lerner (2014) defines tail queries as including misspelled queries, addresses, specific product descriptions or model numbers, and detailed queries composed of multiple terms.
[100] See also Dou, Song and Wen (2007).
[101] Commission Decision of 18 February 2010, Case M. 5727 *Microsoft/Yahoo! Search Business*, par. 162 and 223.

Variety of data: the economies of scope

Another characteristic of the big data resides in the capacity and the importance of combining different types of data, which points to the economies of scope. As noted by the US Executive Office of the President (2014), the combination of data from different sources "may uncover new meanings. In particular, data fusion can result in the identification of individual people, the creation of profiles of an individual, and the tracking of an individual's activities."[102] Similarly, the European Commission states, in its *Google/DoubleClick* merger decision, that "competition based on the quality of collected data thus is not only decided by virtue of the sheer size of the respective databases, but also determined by the different types of data the competitors have access to and the question which type eventually will prove to be the most useful for internet advertising purposes."[103]

This is also an empirical question which should be tested in each case on the basis of the type of data and application at hand. In particular, the necessity of having more variety of data to improve the quality of the application and the algorithm should be carefully analysed.

Depreciation value of the data and velocity of the analysis

Many types of data are transient in value and only relevant over a short period of time, hence their depreciation rate is very high.[104] As noted by Autorité de la concurrence and Bundeskartellambt (2016:49), "historical data, while useful for analysing trends in advertising markets, may have comparatively little value for instant decision making such as the choice of which ad to display in real-time bidding. Moreover historical data may be of relatively low value for some actors like search engines in view of the high rate of new search queries: as reported by Google, 15 % of every day people's searches are new, implying that algorithms continuously need new data to be effective in providing the most relevant ranking of results to those new queries." Thus, the control over these types of data may not in itself give rise to a sustainable competitive advantage.[105] However, those data may be used to improve existing applications or algorithms or to develop new applications or algorithms and those improvements or creations will have more permanent value. In other words, the transient value of data may be capitalised and transformed into more permanent value of applications' improvements or developments.

Other data have more permanent value, such as names, gender, date of birth, address and their depreciation rate is much slower. Therefore, the control of those data gives a more permanent benefit than the control of transient data.

---

[102] Also the UK Information Commissioner's Office (2014, para 25) observes that variety is the most important characteristic of big data.

[103] Commission Decision of 11 March 2008, Case M.4731 *Google/ DoubleClick*, para. 273.

[104] Schepp and Wambach (2016); Sokol and Comerford (2016); UK Competition & Markets Authority (2015, para 3.6).

[105] As noted by Competition Commissioner Vestager: '*It might not be easy to build a strong market position using data that quickly goes out of date. So we need to look at the type of data, to see if it stays valuable*': Competition Commissioner Vestager, 'Competition in a big data world', DLD 16 Munich, Speech 17 January 2016.

Artificial intelligence

Finally, data analytics tools, like any other tools or workers, improve with experience. However, with the development of artificial intelligence and self-learning algorithm, the experience curve may become much steeper as computers can learn some tasks much faster than humans.[106] This may, in some circumstances, increase the first-mover advantage and make the entry of late-comer competitors more difficult. Their entry strategies may then be limited to acquiring the algorithms or investing in different algorithms trying to provide similar services to end-users, albeit in a different manner.

3.2.3   Relationships between data collection and analysis

As the different parts of the big data value chain are very much related, they may be some feedback loops between data collection and data analysis which decrease the cost of the former (Graef, 2016; OECD, 2016).

A first feedback loop, which constitutes a network effect, is linked to the *number of users* as depicted in Figure 1 below and runs as follows (Autorité de la concurrence and Bundeskartellambt, 2016:13):

(1)   more users means more data;

(2)   which in turn, means better quality of the service in a general way (on the basis of general indicators such as language, location etc.) as well as in a personalised way (on the basis of the profile that has been built of a specific user);

(3)   which in turn, attracts even more users to the service.

Figure 1: The user feedback loop
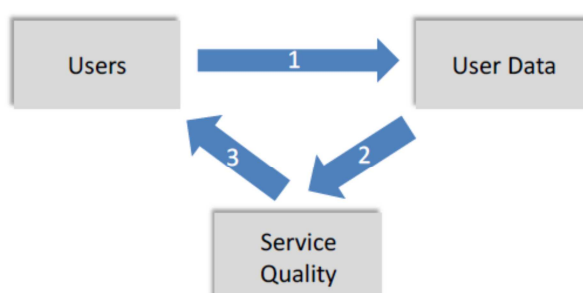
Source: Lerner, 2014, p. 19

If this feedback loop takes place, the cost of data collection is higher for a new and small platform than for a larger one (Pasquale, 2013; Stucke and Ezrachi, 2016). However, the

---

[106] OCDE (2016 :11) notes another reason why the experience curve can be steeper in big data: the lack of physical bounds to the quantity and variety of data that can be collected in a digital world and the unlimited knowledge that can be obtained by running data mining algorithms on a variety of datasets, or using data-fusion.

existence of the feedback loop depends on the relationship between the data and the service quality which in turn depends on the type of data and the type of application at hand. Balto and Lane (2016), Lerner (2014), Lambrecht and Trucker (2015) or Sokol and Comerford (2016) submit that in most cases, the service quality depends only marginally – if at all – of the user data, hence the feedback loop is rarely existent.[107] Moreover, even when the feedback loop exists, if the data collection cost is very small, the effects of the feedback loop will be very small as well.

A second feedback loop is linked to the *monetisation possibilities* as depicted in Figure 2 below and runs as follows:

(1) more users means more data;

(2) which in turn, means better targeting possibilities for advertisement when additional data are necessary to improve the targeting algorithms;

(3) which in turn, raises the likelihood that users click on the ads that are displayed to them, hence increase monetisation under the commonly used pay-per-click model;

(4) and which attracts more advertisers because of the higher probability that a user buys the advertised product;

(5) which in turn raises again the revenues of the provider;

(6) those increased possibilities of monetisation increase in turn the possibility of investment to improve the service and attract more users;

(7) which also contributes to the increase of advertisers.

Figure 2: The monetisation feedback loop
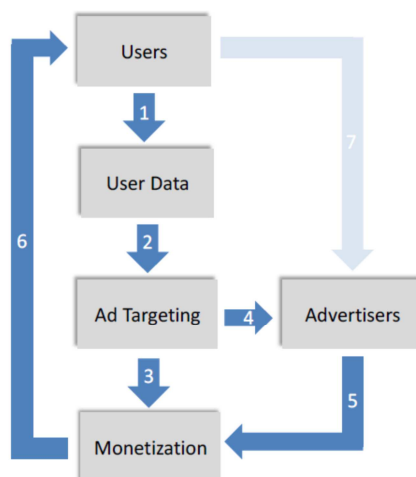
Source: Lerner, 2014, p. 40

If this second feedback loop takes place, a new and small platform with less users and less data will have more difficulty to target its ads and attract advertisers than large platforms, hence less money to improve its products. Here again, the existence and the intensity of this feedback loop

---

[107] According to Lambrecht and Trucker (2015:11), data being very cheap relative to processing skills suggests that processing skills are more important than data in creating value for a firm.

needs to be tested on a case-by-case basis and depends on (i) the relationship between the quantity of data and the improvement of ad targeting algorithm, (ii) the relationship between the quality of ad targeting and the attraction of advertisers, and (iii) how the platform invests advertisement revenues and finances service improvement. Lerner (2014:42) and Sokol and Comerford (2016) submit that empirical evidence does not show that a higher number of users necessarily leads to a better monetisation. Moreover, even when the feedback loop exists, if the data collection cost is very small, the effects of the feedback loop will be very small as well.

Finally, it is important to note that the development of artificial intelligence may in some circumstances re-inforce those feedback loops as it may strengthen the relationship between user data and service quality in the first loop and the relationship between user data and ad targeting in the second loop.[108] Again, this effect should be assessed on a case-by-case basis and very much depends on the type of data and on the type of self-learning algorithms at hand.

3.3   Recommendations for competition agencies

To assess the importance of data in determining market power in a big data value chain, we recommend an analytical framework based on three principles and two questions.

The first principle is that data are one input, which is important but not unique, to develop successful applications and algorithms. Other inputs are also important such as skilled and creative labour force (in particular computer scientists and engineers), capital and distribution channels. Above all, the skills and creativity of the labour force will make the success of the applications.

The second principle is that big data value chains (data collection, storage and analysis) exhibit many direct and indirect network effects that need to be captured by the competition authorities. That requires an understanding and an analysis of each part of the value chain but also the interaction and possible feedback loops between its different parts instead of analysing one part of the value chain in isolation. Already more than 10 years ago, in a more specific context, Julian Wright warned the competition authorities not to apply a one-sided logic to two-sided markets.

The third principle is that each big data application and algorithm is different and should be analysed on a case-by-case basis. Therefore, it would be inappropriate to propose detailed recommendations at a general level beyond a broad framework for analysis.

With those principles in mind, a competition authority trying to assess market power in the big data value chain should answer two main questions:

---

[108] As explained in Domingos (2015:12), "*the power of machine learning is perhaps best explained by a low tech analogy. For example in farming, we plant the seeds, make sure they have enough water and nutrients, and reap the grown crops. When it comes to artificial intelligence, learning algorithms are the seeds, data is the soil, and the learned programs are the grown plants. This means that the more data is available to a learning algorithm, the more it can learn. In short: No data? Nothing to learn. Big data? Lots to learn.*"

The first question relates to the value of the data under examination for the applications and the algorithms under examination. That question requires determining:

- the extent of the economies of scale in the data, in particular what is the marginal benefit of having more data under examination to improve the quality of the application under examination;
- the extent of the economies of scope in the data, in particular how important it is to combine different types of data to improve the quality of the application under examination;
- the time depreciation value of the data, in particular the relationship between the age of the data and its relevance to develop or improve the application under examination.

The second question relates to the availability of the data under examination for the applications and the algorithms under examination. This question requires determining:

- the possibility and the costs for an application developer to collect data directly from the users, machines, sensors, etc.;
- the possibility and the costs for the application developer to buy the data from data broker and in a data market place.

Such data availability is to a large extent influenced by the legal framework regarding data collection and use. As this framework is different in the EU than in other parts of the world (and within the EU, in some Member States than in others), as well as for firms offering some services (such as traditional telecommunications services) than competitors offering other services (such as the communication Over-the-top services), those legal differentiations should be factored in the competition analysis.

4.   Use of Data and Personalised Prices

In the two previous sections, we have defined big data and described the technologies that firms can use to produce them. In this section and the following section, we discuss the various uses that firms can make of big data. We explore two specific types of uses: we first look at the possibility for firms to set personalised prices in the absence of an advertising revenue channel (this section) and then address targeted advertising (the next section).[109]

4.1 Definition and types of price discrimination[110]

4.1.1   Definition and types of price discrimination

A firm price discriminates when it charges two consumers (or the same consumer) different prices for two units of the same product or similar products, and the price difference does not reflect cost differences. A firm price discriminates to extract as much as possible what the consumers are willing to pay for its products or services.

Price discrimination (PD, thereafter) is feasible under some conditions: (i) the firm should have some market power (price discrimination is not feasible under perfect competition); and (ii) there is no or limited possibilities of arbitrage or resale (otherwise, consumers who benefit from low prices would have an incentive to resell the goods at higher prices and compete with the high-priced versions).

The economic literature distinguishes between three types of PD (Pigou, 1920), as explained in Table 3 below.[111]

---
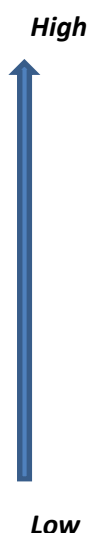
[109] Though targeted advertising constitutes a form of price discrimination, as we will see, the industrial organisation of the advertising market is very specific. Furthermore, consumers' responses to personalised prices and targeted ads are different. Therefore, we treat separately personalised pricing and targeted advertising.

[110] Using data for price discrimination is also briefly dealt with in the joint report of the French and German competition authorities, p. 21. The CERRE report by Larouche, Peitz and Purtova (2016) on consumer privacy in network industries also discusses briefly the effects on consumer welfare of price discrimination.

[111] A more detailed analysis of price discrimination can be found in any industrial organisation textbook. See, for example, Tirole (1988), chapter 3, or Belleflamme and Peitz (2015), chapters 8 to 10.

Table 3: Types of discrimination and associated information requirement

Types of discrimination	Shapiro and Varian (1999) terminology	Definition
First-degree PD	*Personalised pricing*	Firms observe all relevant heterogeneity and capture the entire consumer surplus.
Third-degree PD	*Group pricing*	Firms observe some but not all the relevant heterogeneity and construct groups. Using this direct signal about demand, they are able to price discriminate between groups, but not within groups.
Second-degree PD	*Versioning*	When the relevant heterogeneity cannot be observed, firms can still offer menus of options or packages, and consumers self-select an option/package based on their preferences.

High

Low

In this order, as we move from first-degree to third-degree PD and from third-degree to second-degree PD, there are less requirements in terms of information about consumers' preferences. In a standard monopoly context, the monopoly firm's profit is also decreasing: the more information about its consumers the monopoly can collect, the larger share of their surplus it can extract. The results are far less clear under competition though, as we will see below.

4.1.2   Data and price discrimination

Data facilitates price discrimination. Firms can use data to infer consumers' willingness-to-pay (WTP). The more information a firm can collect about its existing or potential customers, the more accurate its estimate of consumers' WTP. To the extent that the firm has some market power, it can then set individual prices based on this estimation.

At the extreme, the firm is able to set individual prices and fully extract consumers' WTP (first-degree PD). Perhaps more realistically, it is able to engage in group pricing (third-degree PD), with small targeted groups (e.g., fishing enthusiasts, etc.). The literature refers to both first-degree and third-degree PD with small targeted groups as *personalised pricing* or *price targeting*. We adopt this terminology.

Personalised prices have not been observed in practice according to the European Commission[112] as well as to the reports of OFT (2013) for the UK and to the CNIL-DGCCRF for France (2014). For example, the CNIL-DGCCRF report (2014) found no evidence of personalised prices based on IP addresses in France in e-commerce websites. On the same topic, Vissers et al.

---

[112] Commission Guidance SWD(2016) 163, p. 147.

(2014) ran a three-week experiment with 66 user profiles connecting 25 airlines twice a day, and found no evidence of price targeting, though prices were observed to be very volatile.[113]

However, this is a controversial area. An (in)famous case occurred in 2000 when a customer complained that, after erasing the cookies from his computer's browser, he obtained a lower price for a particular DVD on Amazon.com. Consumers were very upset, and Amazon's CEO, Jeff Bezos, promised that the company "never will test prices based on customer demographics".[114]

The fear of negative consumer reaction may explain why targeted pricing is hardly observed. However, there are subtler – and perhaps, more acceptable, from a consumer viewpoint – ways for a company to achieve the same outcome.

First, firms can offer the same uniform prices to all consumers, but with personalised discounts. Since discounts are less easily compared, negative reaction from consumers seems less likely. Since consumers end up paying different, personalised, net prices, this pricing strategy is equivalent to setting personalised prices.

Second, a firm can engage in "search discrimination" or "steering", which consists in showing different products to customers from different groups, based on the available information about consumers. For example, the Wall Street Journal (2012) reported that the travel agency OrbitzWorldwide was showing more expensive hotel offers to Mac users than to PC users. A similar practice has been employed by Staples.com: the same newspaper article revealed that this site displayed different prices once the potential buyers' locations had been identified. Other firms using customers' browsing history and geolocation to vary the offers and products were identified by the newspaper: Discover Financial Services, Rosetta Stone, Home Depot and Office Depot.[115]

In sum, with the advent of big data, we should expect more personalised prices, though firms may have to employ indirect methods (such as personalised discounts or search discrimination) to avoid upsetting their consumers.

4.2 Market outcomes and consumer welfare

In this subsection, we first discuss the impact of big data on firms' ability to price discriminate, and then the impact of price discrimination on firms' profits under monopoly and competition.

---

[113] Other types of practices may explain the high variability of online prices. In particular, it may be the case that firms use the possibility to change their prices online frequently to explore the demand curve (and estimate price elasticities).

[114] See https://en.wikipedia.org/wiki/Amazon.com_controversies#Differential_pricing.

[115] See also Mikians et al. (2012, 2013) who systematically collected data on various retailers and provide some empirical evidence of search discrimination.

4.2.1 Personalised prices and market outcomes

Monopoly

To begin with, consider a monopoly firm. If the firm can collect precise data on its consumers (e.g. demographics, online behaviour, etc.), it may increase its profits by offering personalised prices. For example, Shiller (2014) estimates the increase in profit for Netflix if Netflix would introduce personalised prices. According to the author, this would lead to an increase of profit for the company of between 0.8% (if the company used data on consumer demographics) and 12.2% (if it used the browsing history of its consumers).

There is one caveat though highlighted by the literature, when the monopolistic seller has repeated interactions with its consumers and cannot commit to future prices. The economic literature (e.g. Stokey, 1979) shows indeed that intertemporal price discrimination is not optimal for the monopolist. Acquisti and Varian (2005) revisit this result in a model where a monopolist has access to a tracking technology (e.g. by putting cookies on consumers' device) and consumers can use an anonymising technology (e.g. by erasing their cookies). Acquisti and Varian show that using past information about consumers benefits the monopolist either if a large share of consumers is myopic (i.e. they ignore the fact that paying a high price today makes it more likely that they will be offered a high price tomorrow) and/or tracking is also used to provide consumers with personalised (higher-quality) services.

Firms may set personalised prices based on consumers' search history rather than on their purchase history. Armstrong and Zhou (2010) study a model where firms can use information on consumers' search history to make their pricing decisions. They show that firms have an incentive to offer lower prices for first visits compared to second visits; this practice raises all prices and lowers consumer surplus and total welfare.

Competition

While a monopoly firm will in many environments benefit from personalised pricing, it is less clear if the firm faces competition.

First, consider a situation where all competing firms have access to the same information about consumers' tastes and preferences. What is the effect of a switch from uniform to personalised prices? The literature suggests that if only one firm introduces personalised prices, this firm can increase its profit. By contrast, if all firms switch to personalised prices, the intensity of competition can either increase or decrease (see, e.g. Corts, 1998). Therefore, from the industry point of view, the availability of big data containing consumer information can be either beneficial or harmful.

So far, we postulated that firms in the market have access to the same information about consumers. However, this is not necessarily the case, for example, when firms obtain data about consumers from independent data brokers. Montes et al. (2016) study a setting where two competitors obtain data from a data broker. They show that in equilibrium the data broker sells

its data to only one firm, and therefore only one of the competing firms can set personalised prices. The firm that has access to the data makes more profit than in the situation without information about consumers, and the firm without access to data makes a lower profit. In this case, a move from uniform to personalised prices decreases total welfare, though consumers benefit from this move.

Firms can also acquire information about consumers through previous market interactions. The literature on behaviour-based price discrimination (see Esteves (2009) for a survey) shows that in this case firms compete aggressively in earlier periods to learn about consumers, and then set personalised prices in later periods.

An additional insight from the literature is that competition may lead firms to differentiate with respect to how they use personal information from their consumers. Casadesus-Masanell and Hervas-Drane (2015) study a model where consumers provide voluntarily personal information to firms to obtain higher-quality personalised products, but firms can disclose and sell some of this information to outside firms, which decreases consumers' utility. The authors show that the market equilibrium involves vertical differentiation with respect to disclosure of consumers' information: one firm positions itself as a high-quality/low-disclosure provider, and the other firm as a low-quality/high-disclosure provider. Furthermore, more intense competition implies more disclosure by the firms. In other words, this paper shows that stronger market power does not necessarily imply that firms disclose more information about their users.

To sum up, while a monopolist can increase its profit by setting personalised prices, competing firms do not necessarily gain from personalised pricing. In addition, under competition asymmetric outcomes can arise where only some firms are able to engage in personalised pricing. Therefore, the availability of consumer data can increase market concentration, some firms benefiting more from big data than their competitors.

4.2.2   Effects of on personalised prices on consumer welfare

We discuss below the potential effects of personalised pricing on consumer welfare.

Appropriation vs. market expansion

When a firm sets personalised instead of uniform prices, a trade-off arises: some consumers with high WTP can be worse off (appropriation effect), while some consumers with low WTP can be better off (market expansion effect).

The appropriation effect states that the firm charges higher personalised prices to consumers with high WTP compared to uniform prices (note that this is true under monopoly, but not necessarily true in a competitive environment); those consumers are then worse off with personalised prices.

The market expansion effect means that the firm charges lower personalised prices to consumers with low WTP, and some consumers with low WTP who could not afford the good previously under uniform pricing can purchase it with the low personalised prices.

Ignoring other effects, personalised pricing increases consumer surplus if the demand expansion effect outweighs the appropriation effect. Typically (although not a 'general' result), if a monopolist firm charges different prices in different markets or market segments and these markets are "similar" enough in terms of consumer demand, personalised pricing is not good for welfare, although there will be winners and losers. Instead, if markets are different, personalised prices are good as they open up new markets (under uniform pricing, the smallest markets would shut down and everyone would be worse off).

The economics literature shows more formally that in a monopolistic context, personalised prices can increase or decrease consumer surplus and social welfare depending on demand conditions – see, e.g., Aguirre, Cowan and Vickers (2010), Bergemann et al. (2015), and Cowan (2016). For example, Bergemann et al. (2015) show that with a switch from uniform pricing to price discrimination, social welfare and consumer surplus can both increase, both decrease, or social welfare increases while consumer surplus decreases.

In a context of imperfect competition, as we already explained, personalised pricing can lead to lower (or higher) prices for all consumers, compared to uniform pricing.

In sum, the collection of detailed data on consumers' preferences allows firms, in theory at least, to set personalised prices, but it is not necessarily harmful to consumers.

Complex prices can soften competition

Personalised pricing, if it takes the form of sophisticated or complex prices, can increase consumers' search costs and thereby soften competition. The literature talks about "obfuscation strategies" as firms' actions to raise consumers' search costs (e.g., see Ellison and Ellison, 2009): a large range of sophisticated prices can make it hard for consumers to compare prices, which may increase their search costs and therefore soften competition. Whether complex prices lead to higher prices depends also on consumers' ability to compare prices across different channels or sources, and there is some empirical evidence that consumers indeed engage in price comparisons.[116]

One might also argue that the opposite may well hold true, that is, complex prices may intensify competition and lead to lower prices. For example, as we already outlined, with behaviour-based price discrimination, where prices are based on consumers' previous purchase decisions, firms compete more aggressively than with uniform pricing.

---

[116] For example, see the study on supermarkets by Seiders and Costley (1994).

Quality degradation

Personalised pricing may give firms incentives to degrade the quality of their products (Deneckere and McAffee, 1996). The idea is that when offering different qualities at different prices, firms want consumers with high WTP to pick the high-quality expensive offers. One possible strategy to force them to pay for high quality is to degrade the quality of the low-quality offer. This type of strategy is particularly feasible with digital products, since the costs of quality degradation are very low.[117] In other instances, personalised prices may lead to higher top quality.

Exclusionary strategies

An incumbent firm may use personalised pricing to offer low targeted prices (or discounts) to its customers who could be tempted to switch to a competitor or new entrant. It may constitute an abuse of dominant position.

Trust in online markets

Personalised pricing can reduce consumers' trust in online markets, in particular if it is realised in a non-transparent way. For example, the Annenberg Centre published a study in 2005 showing that two thirds of Internet users believed that it was illegal for online retailers to offer personalised prices.[118] An online player that would set personalised prices in a non-transparent way to exploit consumers' confidence could create negative externalities on other online players. Exploited consumers could indeed view online markets as unsecure, and restrain from buying online.

4.3   Legal limits to personalised pricing

Data protection rules

Firms that collect personal data in order to engage in personalised pricing qualify as controllers within the meaning of EU data protection rules. Therefore, they have to comply with the EU data protection rules related to profiling.[119] In practice, this means that for a controller to engage in personalised pricing, which is a form of profiling, it needs to have the explicit consent of the data subject involved.

Consumer protection rules

In principle, the consumer protection rules leave traders free to set prices as long as they inform consumers about the prices, or how they are calculated. However, the combination of

---

[117] See, for example, the discussion on versioning strategies in Shapiro and Varian (1999).
[118] See Turow et al. (2005).
[119] Article 15 of the Data Protection Directive and Article 22 of the General Data Protection Regulation.

personalised pricing with unfair commercial practices is prohibited. In its 2016 Guidance,[120] the European Commission gives the example of the use of information gathered through profiling to exert undue influence such as an airline or a railways company falsely claiming that only a few tickets are left after finding out that a consumer is running out of time to buy the ticket. That can be considered as a misleading commercial practice which is prohibited.[121]

Anti-discrimination law

Apart from national anti-discrimination law prohibiting differentiation based on factors such as gender, racial or ethnic origin, religion and age, there is no general prohibition of price discrimination under civil law rules.

When it comes to EU legislation, the Services Directive[122] prohibits discrimination based on the service recipient's nationality or residence. On this basis, the Commission started a probe against Disneyland Paris in Summer 2015 acting on the basis of allegations that cheap deals were only made available for residents of France or Belgium in violation of the Services Directive. In April 2016, Disneyland Paris changed its policy and brought its online booking procedures and payment methods for tickets in line with the principle of non-discrimination.[123]

The Commission also proposed in May 2016 a Regulation on geo-blocking. [124] Geo-blocking occurs where traders operating in one Member State block or limit access to their websites or apps of customers from other Member States wishing to enter into cross-border commercial transaction. In addition, discrimination occurs through other actions by traders involving the application of different general conditions of access to their goods and services with respect to such customers from other Member States, both online and offline. In particular, the proposed Regulation aims to prevent discrimination based on the nationality, place of residence or place of establishment of customers beyond the Services Directive which is argued not to have sufficiently addressed discrimination of customers.[125]

---

[120] Guidance of the European Commission services of 25 May 2016 on the Implementation/Application of Directive 2005/29/EC on unfair commercial practices, SWD(2016) 163, p. 146.

[121] In line with that, one of the commercial practices which are in all circumstances to be considered unfair constitutes a false statement '*that a product will only be available for a very limited time, or that it will only be available on particular terms for a very limited time, in order to elicit an immediate decision and deprive consumers of sufficient opportunity or time to make an informed choice*': Point 7 of Annex I of the Unfair Commercial Practice Directive.

[122] Article 20 of Directive 2006/123 of the European Parliament and of the Council of 12 December 2006 on services in the internal market (Services Directive) [2006] OJ L 376/36.

[123] See https://ec.europa.eu/germany/news/disneyland-paris-kommission-begr%C3%BC%C3%9Ft-%C3%A4nderung-der-preispolitik_de.

[124] Proposal of the Commission of 25 May 2016 for a Regulation of the European Parliament and of the Council on addressing geo-blocking and other forms of discrimination based on customers' nationality, place of residence or place of establishment within the internal market, COM(2016) 289.

[125] Recital 3 of the Proposal for a Regulation of the European Parliament and of the Council on addressing geo-blocking and other forms of discrimination based on customers' nationality, place of residence or place of establishment within the internal market and amending Regulation (EC) No 2006/2004 and Directive 2009/22/EC, 25.5.2016, COM(2016) 289 final.

Geo-blocking and competition law

In cases where geo-blocking involves vertical restrictions of competition, there is also scope for competition enforcement under Article 101 TFEU as evidenced by the proceedings of the Commission against Disney, NBC Universal, Sony, Twentieth Century Fox, Warner Bros, Paramount and Sky UK in the area of cross-border provision of pay TV services.[126]

4.4 Recommendations for competition agencies

4.4.1 Possible consumer exploitation

The welfare effects of personalised pricing are *a priori* ambiguous. As we have shown, the economic literature emphasises that price discrimination is not necessarily detrimental to welfare or consumer surplus, and that it can increase welfare and/or consumer surplus in comparison to uniform pricing. From an economic viewpoint, there is therefore no rationale for banning personalised pricing *per se* (as there is no rationale for banning price discrimination).

Also from a legal perspective, the Court of Justice instructed the competition agencies to be concerned about the consumer welfare of all consumers taken together and not each consumer individually. In the Case C-238/05 *Asnef*, the Court of Justice decided that:[127]

> "Under Article (101(3) TFEU), it is the beneficial nature of the effect on all consumers in the relevant markets that must be taken into consideration, not the effect on each member of that category of consumers."

Furthermore, as we have seen, large players may be concerned about their reputation, and hence hesitate over engaging in personalised pricing in fear of negative consumer reaction (as the Amazon example has shown). One could argue, though, that small players have less reputation concerns, and have therefore larger incentives to set personalised prices. However, even if these personalised prices would have negative welfare effects, the magnitude of such effects would most likely be very low with small players with little market power.[128]

4.4.2 Possible competitor exclusion

Another type of concern is that price discrimination could be used as a monopolisation device. For example, an incumbent firm may pre-empt entry in a given market or consumer segment by

---

[126] See Press release, 'Antitrust: Commission sends Statement of Objections on cross-border provision of pay-TV services available in UK and Ireland' , 23 July 2015, available at http://europa.eu/rapid/press-release_IP-15-5432_en.htm and Press release, 'Antitrust: Commission accepts commitments by Paramount on cross-border pay-TV services', 26 July 2016, available at http://europa.eu/rapid/press-release_IP-16-2645_en.htm.

[127] Case C-238/05, *Asnef-Equifax, Servicios de Información sobre Solvencia y Crédito, SL and Administración del Estado v. Asociación de Usuarios de Servicios Bancarios (Ausbanc)*, ECLI:EU:C:2006:734.
[128] Of course, if personalisation under competition leads to lower prices, there is no concern that personalisation harms welfare.

setting very low prices in this market or segment.[129] Incumbent firms could also offer loyalty discounts to prevent entry of competitors. This type of concern could be aggravated if possibilities of price discrimination hinge on detailed consumer data, and incumbent firms have exclusive access to this consumer data. Of course, such potential exclusionary practices would call for an intervention from competition authorities, but they do not call for a ban of personalised pricing *per se*.

4.4.3 Recommendations

Our main policy recommendation would therefore be that personalised pricing strategies, if they exist, should be transparent to ensure consumers' trust in online markets, which would affect positively all players (also UK House of Lords, 2016, para 291).

This may require clarification on the implementation of the general principles of EU consumer rules to personalised pricing with more developed guidelines than the ones adopted by the European Commission in 2016.

This also requires an effective application of consumer protection in all the Member States. For instance, online prices may need to be monitored by consumer protection agencies upon complaints. One possibility would be to audit pricing algorithms, but it seems far too complex and costly, and it may have a negative effect on firms' incentives to innovate through sophisticated pricing algorithms. In addition, a requirement to disclose algorithms goes to the heart of a business' operations and interferes with valuable trade secrets. Therefore, this approach seems neither viable nor desirable. Another solution would be to use automated tests, for example via virtual mystery shoppers. We do not recommend necessarily a continuous monitoring program but rather, that on a regular basis, the pricing strategies of online players be monitored.

---

[129] However, this form of limit pricing via price discrimination may not be credible.

5. Use of Data and Targeted advertising

For many online firms, advertising is a major source of revenue, if not the only one. Since it is costly to impress consumers with advertising, firms make efforts to increase the effectiveness of online advertising. In particular, firms can use data from consumer online behaviour (big data) to offer more targeted (hence, more effective) ads.[130]

5.1 Online advertising ecosystem

5.1.1 The online advertising ecosystem

Online advertising represents now a large share in total advertising expenditures. For example, in 2015 in Europe, according to IAB Europe, investments in online advertising were higher than investments in TV advertising (with total investments of €36.2bn for the former, against €33.3bn for the latter).[131]

Two main forms of online advertising coexist: search advertising and non-search advertising.[132]

Search advertising enables advertisers to target consumers based on their search queries. In 2015 in Europe, it represented 47% of the online advertising market according to IAB Europe. This form of advertising is naturally targeted, since it is primarily based on consumers' search queries.

The main form of non-search advertising is *display advertising*, which includes banner ads, plain text ads, audio and video ads, etc. In 2015 in Europe it represented 38% of the online advertising market according to IAB Europe, and it is growing fast. Display advertising includes both non-targeted and targeted advertising, and the latter is the sub-category growing the faster.[133] *Classified advertising* is the other firm of non-search advertising, and in 2015 in Europe it represented 15% of the online advertising market.

Targeted display advertising is becoming more and more important in the online advertising market. Since it involves a relatively complex ecosystem, in particular in comparison to search advertising, in the rest of this section we focus on display targeted advertising.

---

[130] This is not the only strategy followed by advertisers to increase ad effectiveness. For example, advertisers try also to improve ad *viewability*, to increase consumers' exposition to ads.

[131] See http://www.iabeurope.eu/research-thought-leadership/press-release-european-online-advertising-surpasses-tv-to-record-annual-spend-of-e36-2bn/.
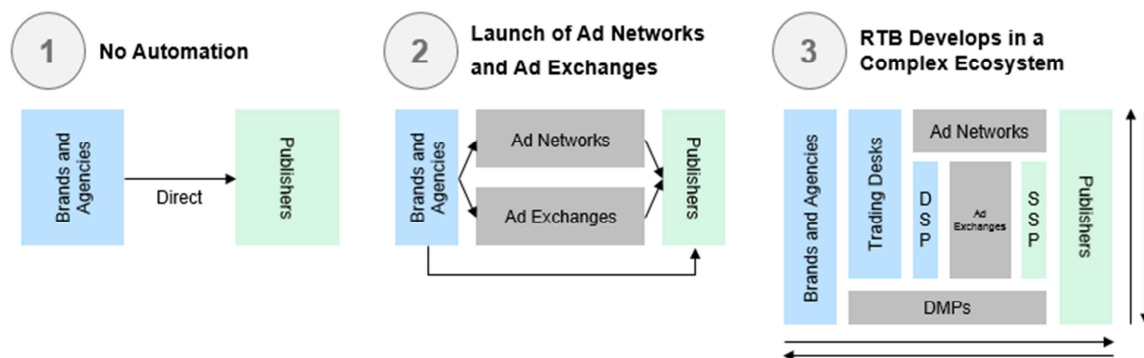
[132] See also Peitz and Reisinger (2016), who provide detailed information on advertising in online media.

[133] In 2010, according to the OFT report "Online Targeting of Advertising and Prices", targeted advertising represented 10 to 15% of display advertising. In 2015, in France, "programmatic" advertising (i.e., targeted advertising) represented 40% of the investments in display advertising. See http://www.sri-france.org/etudes-et-chiffre-cles/observatoire-de-le-pub-sri/15eme-observatoire-de-pub-sri/.

To understand the market power issues that may arise in the online advertising markets, it is useful to present more precisely the different intermediaries that can be involved in display advertising.

Display advertising can take different forms, as illustrated in Figure 3 below.

Figure 3: Different forms of target advertising

Source: Bulger Partners, http://www.bulgerpartners.com/five-ways-digital-advertising-is-rapidly-transforming/.

Case 1: No automation: In its simplest form, a publisher sells directly advertising space to advertisers. This type of advertising can be targeted (first-party behavioural targeting): the publisher uses the information (registration-based information or information collected via tracking cookies) it has about its consumers to place its own targeted ads or those of third parties.

Case 2: Ad networks and ad exchanges: The publisher can also contract with an ad network or an ad exchange to sell its advertising space:

- ad networks are intermediaries between websites that offer advertising space and advertisers looking for an audience. Examples include DoubleClick and AdMob (owned by Google), Atlas (owned by Facebook), AdColony, etc.;
- an ad exchange is an automated platform that auctions advertising inventory from multiple ad networks. Examples include DoubleClick (owned by Google), Microsoft Ad Exchange, AppNexus, etc.

Ad networks collect information about consumers' online behaviour, and can help advertisers target their ads for a specific audience (which is called third-party behavioural targeting). The ad network compiles information from the company and from other third-party members of the ad network to profile users.

Online publishers sell advertising space via ad networks. Small publishers often sell all their advertising space via ad networks, while large publishers may also sell some of their advertising space directly.

Case 3: Real-time bidding (RTB). Over the last few years, RTB systems have emerged. In RTB systems, ad exchanges act as intermediaries between supply and demand of advertising space. Ad exchanges connect on one side to supply side platforms (SSP), which help publishers sell their advertising space, and on the other side to demand side platforms (DSP), which give access to buyers. Data management platforms (DMPs) sell data about users to the other players.[134]

RTB works as follows (e.g., see Yuan et al., 2014). When a user visits the website of a publisher, the data management platform (DMP) associated with the publisher identifies the user's profile (interests, characteristics) through their cookies. Once the user connects to the server (URL) of the publisher, the latter sends the user profile to its supply-side platform (SSP), which in turn sends the information to an ad exchange. The ad exchange provides the user profile to demand-side platforms (DSP). A two-stage auction follows. First, based on information provided by the DMP on the user's profile, each DSP starts an auction with its advertisers. Second, the winners at each DSP enter a second-round auction at the ad exchange. The highest bidder of this second auction wins and their ads are then displayed when the website loads. The whole process takes less than 200 milliseconds, and is therefore hardly noticeable to the user.

5.1.2   Targeted advertising

Targeted advertising (or behavioural targeting) takes place when firms place ads that target a specific audience based on their estimated personal characteristics/interests. Those characteristics/interests are evaluated using information from consumers' online behaviour – the sites they visit, what they search for, etc.

Targeted ads have two benefits for advertisers relative to non-targeted (regular) ads: (i) ads can be displayed only to the consumers potentially interested in the product or service, thereby reducing wasteful advertising and (ii) the content of the ad can be personalised.

Because it is based on a consumer's interests, a targeted ad leads to a high probability of a good match and hence tends to increase the advertising effectiveness. It is expected to increase the click-through rate and the conversion rate of ads, though this is not always the case, as we shall see below. Because it improves advertising effectiveness, targeted advertising is also expected to lower the costs of advertising for large and small firms who want to place ads.

To the extent that targeted advertising is an improvement over existing (regular) forms of online advertising, an important question is whether all online players can implement this new advertising technology. If it were not the case, then targeted advertising could be expected to increase concentration and market power. The question is in particular whether all players have access to the consumer data necessary to target advertising. In the *Facebook/WhatsApp* merger case, the European Commission concluded that the market for online advertising was wider than social media, and therefore there was no antitrust concern since lots of internet user data

---

[134] Examples of SSP are Rubicon Project, Admeld (owned by Google) and Pubmatic. Examples of DSP are Criteo, AppNexus and MediaMath. Examples of DMP are Lotame and BlueKai.

would still be available and not within Facebook's exclusive control. The question here is whether data and knowledge would become uniquely available to the merged entity: different data sources may simply not replicate it – they may be available but not give rise to the same knowledge. The point is that data about one consumer collected by one company may not be a substitute for data about the same consumer collected by another: they may concern different aspects of the consumer and not have the same value on the advertising market.[135]

5.2 Market outcomes and consumer welfare

5.2.1 Targeted advertising and market outcomes

Targeted advertising is widespread nowadays, as evidenced by Carrascosa et al. (2015) for example. These authors have developed a methodology to detect targeted advertising and run a large number of experiments using 72 interest-based personas (i.e. virtual identities). They found that (i) 88% of the personas were displayed targeted ads based on their interests, and (ii) advertisers used sensitive individual characteristics related to health, politics, or sexual orientation.

Advertisers experience low click-through rates for generic online ads. They expect higher click-through rates for targeted ads, thanks to the better match between this type of ad and the consumers' interests. There is some empirical evidence that targeted ads are more effective than generic ads as expected, but details matter, as we will see below.

Most of the empirical evidence concerns a specific form of targeting, called 'retargeting'. It works as follows: (i) a consumer visits a company's website, showing interest for the brand; (ii) later on, while visiting another website, the consumer is shown an ad from the first brand. We talk of 'dynamic retargeting' when the ad concerns a specific product of the brand that the consumer may have looked at, and of 'generic retargeting' when the ad is only generic (e.g., it shows the brand's logo). Dynamic retargeting corresponds therefore to a higher level of targeting than generic retargeting.

However, Lambrecht and Tucker (2013) find that dynamic retargeting is more effective than generic retargeting only if consumers have well-defined preferences (e.g., they have visited various review sites to collect information), and that it is less effective otherwise. A possible interpretation of this finding is that in the early stage of their search process consumers have only a loose idea of what they are looking for. In such a context, a generic ad may be more effective. Once they have gathered enough information, consumers have a clearer idea of the

---

[135] Competition authorities have become interested in online advertising markets. For example, the French competition authority launched a study on online advertising in May 2016 (See http://www.autoritedelaconcurrence.fr/user/standard.php?id_rub=630&id_article=2780). The French AA wants to study the substitutability between different forms of targeted advertising and in particular between targeted advertising on social networks and other forms of targeted advertising.

good they are interested in; in this case, dynamic retargeting is more effective in pushing consumers to finalise their purchase.

Theoretical work has also shown that targeted advertising can soften competition between advertisers-sellers (Chen et al., 2001; Gal-Or and Gal-Or, 2005; Iyer et al., 2005), and therefore increase the profits of advertisers. The intuition is that targeted advertising allows sellers to differentiate from their rivals, thereby reducing the intensity of competition. Therefore, there might be not only direct benefits of targeted advertising for sellers, in terms of higher ad effectiveness, but also indirect benefits in terms of softened competition.

Because it is more effective and it may improve differentiation possibilities, targeted advertising is expected to benefit advertisers (though in response advertising prices may go up), and favour entry of small niche advertiser-sellers, which would be excluded with regular ads.

As for media platforms, targeted advertising may have differential effects. First, targeted advertising improves prospects of general outlets at the expense of tailored outlets. The idea is that tailored outlets enable advertisers to target an audience when no targeting technology is available. With the development of targeted advertising, tailored outlets lose this competitive advantage. Second, with targeted advertising the markets for a media outlet can become so thin that the advertising price may fall (Levin and Milgrom, 2010). Indeed, if the media outlet auctions out its advertising space, it needs to attract enough bidders. If the ad space is highly targeted, it might be unable to do so, leading to a low price for the ad. There might be a trade-off, therefore, between ad effectiveness and ad exposure.

5.2.2   Effects of targeted advertising on consumer welfare

Consumer benefits

With targeted advertising, consumers are exposed to more relevant ads that better match their interests, which facilitates access to products and services that correspond to their tastes. Because consumers are more and better informed, they should benefit from targeted advertising.[136]

Furthermore, more effective targeted advertising reduces advertising costs for advertisers/sellers, which stimulates the demand for online advertising. Online sellers/publishers' revenues are then expected to increase. These higher advertising revenues can be partially passed through to consumers in terms of lower subscription fees or higher quality of service.

---

[136] Lerner (2014:16) cites a 2013 poll showing that 40.5 percent of respondents indicated that they prefer targeted to non-targeted advertisements, while less than 5 percent of respondents had an unfavourable opinion of behaviourally-targeted advertisements: "Consumers Say They Prefer Targeted to Random Online Ads," Marketing Charts, citing research by Zogby Analytics, April 19, 2013, available at http://www.marketingcharts.com/wp/online/consumers-say-they-prefertargeted-to-random-online-ads-28825/.

Potential consumer harm

There is some evidence that consumers can react negatively to targeted advertising. One possible explanation is that consumers may foresee that firms could also use their personal information to set personalised high prices. De Cornière and de Nijs (2016) formalise this idea. In their theoretical investigation, an online platform sells advertising slots through auctions. The platform decides whether to allow advertisers to access consumer information or not. The main result is that disclosure of consumer information improves the match between advertisers and consumers, but also increases product prices.

It seems that consumers also have some direct disutility or distaste from intrusive (targeted) ads. The advertising literature talks about consumer "reactance" (see, e.g., Turow et al., 2009; Tucker, 2012). Some policy reports then highlight that these privacy concerns may more generally also undermine consumers' trust in online markets.

However, there is also evidence that consumers view regular ads as a nuisance. Whether the overall "ad pollution" is higher or smaller with targeted ads than with regular ads is an open question. On the one hand, consumers may view targeted ads as "creepy".[137] On the other hand, because they are more effective, consumers may be exposed to fewer and more relevant ads with targeted advertising than with regular advertising.

5.3 Legal limits to target advertising

Firms that collect personal data in order to engage in targeted advertising qualify as controllers within the meaning of the EU data protection rules. Therefore, they have to comply with the right of data subjects to object to the processing of personal data for direct marketing purposes.[138] This implies that a controller must stop sending targeting ads, constituting a form of direct marketing, if the individual receiving it objects to his personal data being processed for that purpose.

When sending targeted ads over electronic communications networks, firms must also comply with the ePrivacy Directive which stipulates that unsolicited communications are only allowed for the purpose of direct marketing in case prior consent of the user at issue is obtained.[139] As such, firms may only engage in targeting advertising over electronic communications networks after having received consent of users to do so.

---

[137] Note however that ad networks have incentives to make their targeted ads look less "creepy" to avoid negative consumer reaction.

[138] Article 14(b) of the Data Protection Directive and Article 21(2) of the General Data Protection Regulation.
[139] Article 13 of the ePrivacy Directive.

5.4 Recommendations for competition agencies

It is hard to make a general statement on whether targeted advertising is a good or a bad from a welfare point of view. There is a complex trade-off between the potential good matches that effective targeted ads can generate (which benefit both consumers and sellers), the pass-through of advertising revenues to final consumers in terms of lower prices for online services, and the nuisance for consumers from intrusive ads.

Competition authorities seem to be concerned that some players in the online advertising market may have enough market power, thanks in particular to a better access to data, to distort competition. Large players are also vertically integrated at various stages of the online advertising ecosystem. Though vertical integration can raise in some cases competition concerns, it can also increase efficiency. All in all, the standard competition policy suffices to deal with potential competitive problems in online advertising.

Online advertising seems to be an extremely dynamic and innovative market. The organisation of this market has changed drastically over the last few years, with the emergence of new categories of players. Competition for the best algorithms, the best advertising technologies, seems also fierce. This is to the benefit of advertisers – when ads become more effective – and consumers – when they become less of a nuisance. One main recommendation would therefore be to make sure that no player blocks the innovation dynamics that are at play. Given the dynamic nature of the market, our view is that anything beyond monitoring is not warranted as long as no new evidence appears that suggests that there is market power by some players and that this is abused.

References

Official Reports

Autorité de la concurrence and Bundeskartellamt (2016), *Competition law and data*.

CNIL-DGCCRF (2014), *IP Tracking : Conclusions de l'enquête conjointe menée par la CNIL et la DGCCRF*.

EDPS (2016), *Coherent enforcement of fundamental rights in the age of big data*, Opinion 8/2016.

Monopolkommission (2015), *Competition policy: The challenge of digital markets*, Special Report 68.

OECD (2013), *Exploring the economics of personal data: a survey of methodologies for measuring monetary value*, DSTI/ICCP/REG(2011)2.

OECD (2015), *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publications.

OECD (2016), *Big Data: Bringing competition policy to the digital era*, DAF/COMP(2016)14.

UK Office of Fair Trading (2010), *Online targeting of advertising and prices: A market study,* OFT 1231.

UK Office of Fair Trading (2013), *The economics of online personalised pricing*, OFT 1488.

UK Office of Fair Trading (2013), *Personalised pricing – increasing transparency to improve trust*, OFT 1489.

UK Information Commissioner's Office (2014), *Big data and data protection*.

UK Competition & Markets Authority (2015), *The commercial use of consumer data*, Report on the CMA's call for information, CMA 38.

UK House of Lords (2016), *Online Platforms and the Digital Single Market*, Report of the Select Committee on European Union.

US Executive Office of the President (2014), *Big Data and Privacy: a Technological Perspective*.

US Executive Office of the President (2015), *Big Data and Differential Pricing*.

US Federal Trade Commission (2014), *Data Brokers: A Call for Transparency and Accountability*.

US Senate Committee on Commerce, Science and Transportation (2013), *A Review of the Data Broker Industry: Collection, Use and Sale of Consumer Data for Marketing Purposes*, Staff Report.

World Economic Forum (2011), *Personal data: the emergence of a new asset class*.

Academic books and papers

Aguirre, I., Cowan, S. and Vickers, J. (2010), "Monopoly Price Discrimination and Demand Curvature", *American Economic Review*, 100(4), 1601-1615.

Argenton C. and Jens Prüfer J. (2012), "Search engine competition with network externalities", *Journal of Competition Law & Economics* 8(1), 73-105.

Armstrong, M. and Zhou, J. (2010), "Conditioning prices on search behaviour", Mimeo.

Balto D.A. and Lane M.C. (2016), "Monopolizing water in a tsunami: Finding sensible antitrust rules for Big Data", *Competition Policy International*.

Belleflamme P. and M. Peitz (2015), *Industrial Organisation: Markets and Strategies*, 2nd ed., Cambridge University Press.

Bergemann, D., Brooks, B. and Morris, S. (2015), "The Limits of Price Discrimination", *American Economic Review*, 105(3), 921-957.

Bleier, A. and Eisenbeiss, M. (2015), "Personalized Online Advertising Effectiveness: The Interplay of What, When, and Where", *Marketing Science*, 34(5), 669-688.

Buchholtz, S., M. Bukowski and A. Sniegocki (2014), *Big and Open Data in Europe - A Growth Engine or a Missed Opportunity?*, Warsaw Institute for Economic Studies.

Carrascosa, J.M., Mikians, J., Cuevas, R., Erramilli, V., and Laoutaris, N. (2015), "I Always Feel Like Somebody's Watching Me. Measuring Online Behavioural Advertising", CoNEXT'15 December 01-04, Heidelberg, Germany.

Casadesus-Masanell, R. and Hervas-Drane, R. (2015), "Competing with Privacy", *Management Science*, 61(1), 229-246.

Chen, Y., Narasimhan, C. and Zhang, J., (2001), "Individual Marketing With Imperfect Targetability," *Marketing Science*, 20, 23-41.

Corts, K. (1998), "Third-Degree Price Discrimination in Oligopoly: All-Out Competition and Strategic Commitment," *RAND Journal of Economics*, 29(2), 306-323.

Cowan, S. (2016), "Welfare-increasing third-degree price discrimination", *RAND Journal of Economics*, 47(2), 326-340.

de Cornière, A. and de Nijs, R. (2016), "Online advertising and privacy", *RAND Journal of Economics*, 47(1), 48-72.

De Mauro, A., M. Greco and M. Grimaldi (2016), "A Formal Definition of Big Data Based on its Essential Features", *Library Review* 65(3), 122-135.

Deneckere, R. and McAffee, R.P. (1996), "Damaged Goods", Jou*rnal of Economics & Management Strategy*, 5(2), 149-174.

Domingos P. (2015), *The Master Algorithm, The Machine Learning Revolution*, Basic Civitas Books.

Dou Z., Song R. and JR Wen (2007), A Large-scale Evaluation and Analysis of Personalizes Search Strategies, Paper presented at the IW3C Conference.

Ecorys and TNO (2016), *Future trends and business models in communication services*, Study for the European Commission.

Ellison, G. and Ellison, S. (2009), "Search, Obfuscation, and Price Elasticities on the Internet", *Econometrica*, 77(2), 427-452.

Ezrachi A. and Stucke M.E. (2016), *Virtual Competition: The Promise and Perils of the Algorithm-Driven Economy*, Harvard University Press.

Esteves, R., (2009), "A Survey on the Economics of Behaviour-Based Price Discrimination," NIPE Working Paper 5/2009.

Gal-Or, E. and Gal-Or, M. (2005), "Customized Advertising via a Common Media Distributor," *Marketing Science*, 24(2), 241-253.

Graef I. (2016), *EU Competition Law, Data Protection and Online Platforms: Data as Essential Facility*, Kluwer Law International.

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

IDC (2016), *Europe's Data Marketplaces – Current Status and Future Perspectives*, Report for the European Commission

Lambrecht, A. and Tucker, C. (2013), "When Does Retargeting Work? Information Specificity in Online Advertising", *Journal of Marketing Research*, 50(5), 561-576.

Lambrecht, A., Goldfarb, A., Bonatti, A., Ghose, A., Goldstein, D. G., Lewis, R., Rao, A., Sahni, N., and Yao, S. (2014), "How do firms make money selling digital goods online?" *Marketing Letters*, 25(3), 331-341.

Lambrecht A. and Tucker C. (2015), "Can Big Data Protect a Firm from Competition?", available on SSRN.

Laney, D. (2001), "3D Data Management: Controlling Data Volume, Velocity and Variety", Meta Group (Gartner Blog post), posted on 6 February 2001, available at http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Larouche P., M. Peitz and N. Purtova (2016), *Consumer privacy in network industries*, CERRE Policy Report.

Levin, J. and Milgrom, P. (2010), "Online Advertising: Heterogeneity and Conflation in Market Design", *American Economic Review: Papers & Proceedings*, 100, 603-607.

Lerner A. (2014), "The Role of 'Big Data' in Online Platform Competition", available on SSRN.

Mayer-Schönberger V. and Cukier K. (2013), *Big Data: A Revolution that will transform how we live, work and think*, Eamon Dolan/Mariner Books.

McAfee A. and E.Brynjolfsson (2012), *"Big Data: The Management Revolution", Harvard Business Review*, October, p. 60.

Mikians, J., Gyarmati, L., Erramilli, V., Laoutaris, N. (2012), "Detecting price and search discrimination on the internet", Hotnets'12, October 29-30, Seattle, WA, USA, 79-84.

Mikians, J., Gyarmati, L., Erramilli, V., Laoutaris, N. (2013), "Crowd-assisted Search for Price Discrimination in E-Commerce: First results", Mimeo.

Montes, R., Sand-Zantman, W., and Valletti, T. (2016), "The Value of Personal Information in Markets with Endogenous Privacy", Mimeo.

Newman N. (2014), "Search, Antitrust and the Economics of the Control of User Data", *Yale Journal of Regulation* 30(3).

Osborne Clark (2016), *Legal study on ownership and access to data*, Study for the European Commission.

Pasquale F.A. (2013), "Privacy, Antitrust and Power", *Georges Mason Law Review* 20(4), 1009-1024.

Peitz M., H. Schweitzer and T. Valletti (2014), *Market Definition, Market Power and Regulatory Interaction in Electronic Communications Markets*, CERRE Study.

Pigou, C. (1920), *The Economics of Welfare*, Macmillan.

Prüfer J. and Schottmüller C. (2016), "Competing with Big Data", CEPS seminar on Competition policy in the digital economy.

Schepp N. P. and A. Wambach (2016), "On Big Data and Its Relevance for Market Power Assessment", *Journal of European Competition Law and Practice* 7, 120.

Schumpeter, J.A. (1942) *Capitalism, Socialism and Democracy*, Harper and Brothers.

Shapiro C. and H. Varian (1999), *Information Rules – A Strategic Guide in the Information Society*, Harvard Business School Press.

Shaw, I. and Vulkan, N. (2012), "Competitive Personalized Pricing: An Experimental Investigation", Mimeo.

Shelanski H. (2013), "Information, Innovation, and Competition Policy for the Internet", *University of Pennsylvania Law Review* 1663.

Shiller, B. (2014), "First-Degree Price Discrimination Using Big Data", Mimeo.

Sokol D. and Comerford R. (2016), "Antitrust and Regulating Big Data", *Georges Mason Law Review*, 1129-1161.

Stokey, N. (1979), "Intertemporal price discrimination", *Quarterly Journal of Economics*, 93(3), 355-371.

Stucke M. and Grunes J. (2016), *Big Data and Competition Policy*, Oxford University Press.

Swire P. and Y. Lagos (2013), "Why the Right to Data Portability Likely Reduces Consumer Welfare: Antitrust and Privacy Critique", *Maryland Law Review* 72(2).

Time Lex and Spark (2015), *ePrivacy Directive: assessment of transposition, effectiveness and compatibility with the proposed Data Protection Regulation*, Study Prepared for the European Commission.

Tucker, C. (2012), "The economics of advertising and privacy", *International Journal of Industrial Organization*, 30(3), 326-329.

Tucker D. and Wellford H. (2014), "Big Mistakes Regarding Big Data", *Antitrust Source*.

Turow, J., King, J., Hoofnatle, C. J., Bleakley, A., and Hennessy, M. (2009), "Americans Reject Tailored Advertising and Three Activities That Enable It", Mimeo.

Varian H. (2013), "Beyond Big Data", available at http://people.ischool.

Vissers, T., Nikiforakis, N., Bielova, N., Joosen, W. (2014), "Crying Wolf? On the Price Discrimination of Online Airline Tickets", Mimeo.

Wall Street Journal (2012a), "On Orbitz, Mac Users Steered to Pricier Hotels", 23/08/2012.

Wall Street Journal (2012b), "Websites Vary Prices, Deals Based on Users' Information", 24/12/2012.

WIK-Consult and TNO (2015), *Over-the-Top (OTT) players: Market dynamics and policy challenges*, Study for the European Parliament.

Yuan, Y., Wang, F., Li, J. and Qin, R. (2014), "A survey on real time bidding advertising", *Proceedings of 2014 IEEE International Conference on Service Operations and Logistics, and Informatics*, Qingdao, 2014, pp. 418-423.