POLITECNICO DI TORINO

Corso di Laurea Magistrale
In INGEGNERIA GESTIONALE

Tesi di Laurea Magistrale

Big Data and competition policy:

Description of Big Data markets and critical analysis of recent cases

Relatore:
Prof. Carlo Cambini

Candidato:
Giovanna Dal Negro

Ottobre 2018

Summary

1. Introduction and scope of work

Data are, today, an important element in every sector and function of the global economy, and, like other essential factors of production such as hard assets and human capital, much of modern activity simply could not take place without them.

The use of Big Data is the basis for competition and growth for individual firms, enhancing productivity and creating significant value for the world economy by reducing waste and increasing the quality of products and services.

Firms want consumer data because it helps them to understand better what their customers want and how they respond to their goods and services. Firms can create more value for their customers by responding directly to consumer feedback to improve the services and products they offer as well as to make their operations more efficient (CMA, 2015).

Through constant innovation, firms are taking advantage of new technologies and new sources of data to find a competitive edge. This is creating a dynamic market environment where businesses and markets are evolving rapidly (CMA, 2015).

This work has the purpose to describe the principal elements that characterize the Big Data markets. The chapter number two reports the definition of Big Data and the classification proposed by the European Authority. The third chapter describes the economics features of data as an economic good and analyse the barrier to entry in data's markets. The fourth presents network effects and in particular, the Mitomo model related to data network effects that shows, in the case of user-generated data, that each user receives a benefit from the service depending on the number of the users. Thus, interdependencies among users will create a mass effect. This will result in the advantage of a service attracting a large number of users. If the services provided by competitive suppliers are homogeneous, such as online information retrieval system, the precursor can take this advantage.

The chapter number five, firstly, defines the two side markets and shows the model of Roche and Tirole. They found that in a two-sided market, the demand is influenced by the externalities and the elasticity.

In the last part of the chapter, is possible to describe the competition in two side markets, and the Argenton and Prüfer model. The model predicts market tipping: the market share of the dominant firm is expected to increase more and more, whereas the market shares of the other firms are expected to decrease. Second, the model predicts substantial profit growth for the market leader and decreasing profits for the following firms. Third, as a consequence of the second point, the model predicts market exit of one follower.

The sixth reports the different way to use data: price discrimination, targeted advertising, services personalisation and recommendations personalisation.

In the last part of the thesis, I choose to presents the Facebook/WhatsApp case about the merger of the two firms and the recent events about Google and Amazon. This analysis show how European Commission evaluate merge between two firms that possess large amount of data and how firms want use data to achieve a dominant position.

2. Big Data

"*Big Data is high-volume, high velocity and high variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making*".[1]

The difference between Data and Big Data is the four V's made possible by technological progress. The first V is the *volume* that referred to the huge amount of data where the data scale become increasingly high. The second one is the *variety*: it indicated the different type of data that can be produced in a structured and unstructured way. The third V is the *velocity* at which data is collected and used and the last one is the *value* of the information found in the data, this means not only create value for firms but also increase the welfare of the society.

As mentioned by Davenport (2014), Big Data can be classified into two main categories:

1. **Machine-generated**: this type of data is created by a machine without human intervention. It refers to audio, music, image, speech and video data, sensor data, Intelligent Lighting Control sensors used, for example, to identify the locations and the conditions of goods on a supply chain.
2. **Human-generated:** it refers to data that humans, in interaction with computers, supply. This type of data contain tweets, social media post, web contents, and clickstream data.

2.1. European regulation

EU data protection legislation[2] has the objective to protect the fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data. To do that it classifies data in a way that reflects the extent to which the fundamental rights of individual should be safeguarded in the context of the

[1] Gartner IT glossary, *Big data*

[2] EU data protection legislation is in a period of transition from Data Protection Directive that was adopted in 1995 to the General Data Protection

processing of particular types of personal data in line with the associated protection risk (Bourreau et al., 2017).

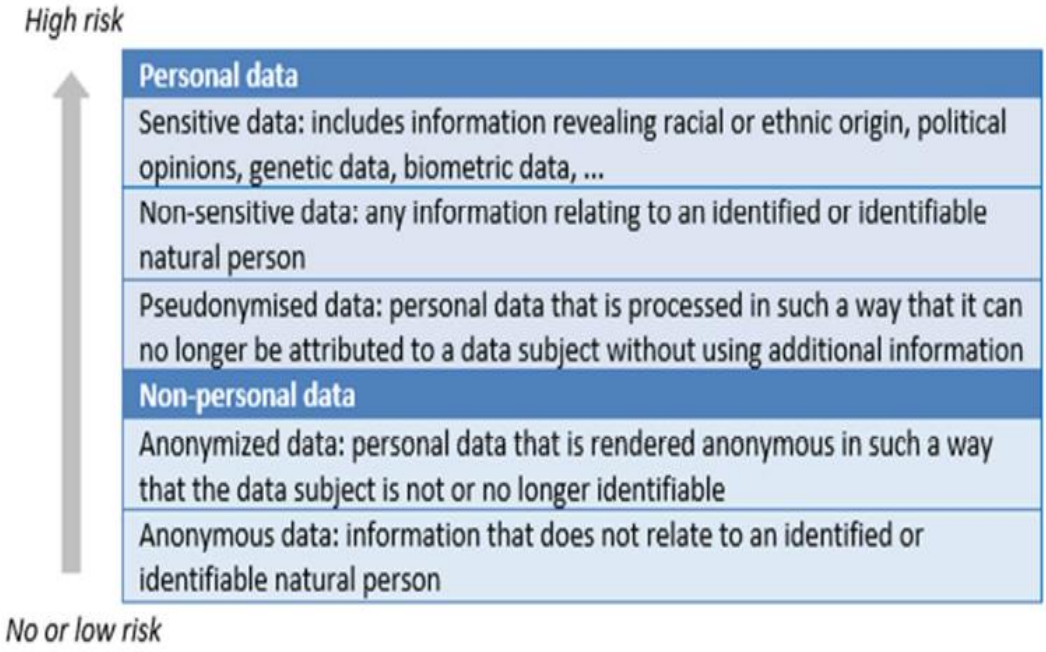

Fig. 2: Typology of data with the associated data protection risk[3]

The classification proposed by European Union is:

1. **Personal data:** "Any information relating to an identified or identifiable natural person (data subject)". An identifiable natural person is 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."[4]

   There two categories of personal data: sensitive data, which includes information, for example, that reveal racial or ethics origin, political opinions, religious or philosophical.

[3] Source: "Big Data and Competition Policy: Market power, personalised pricing and advertising", Bourreau et al., 2017, CERRE.
[4] Article 4(1) of the General Data Protection Regulation

Pseudonymised data is personal data that is processed in such a way that data can no longer be attributed to a specific data subject without the use of additional information.

2. **Non-personal data:** does not contain personally identifiable characteristics and cannot alone be used to identify individuals.

   Anonymous data is an information that is collected or used without any personal identifiers and where identification is unlikely to take place. For example, market research information collected from consumers that simply asks about what shops they have visited without collecting information about who they are.

   Pseudonymous data is an information collected and used at the level of individuals, which may contain personal information such as age or gender but where personal identifiers are no present.

"Aggregate meta data" is a data created by combining personal, anonymous data for multiple individuals as a group.

EU law contains many rules on the collection and use of data, which are shaping the functioning of big data markets. Those rules are more extensive for personal than for non-personal data. The typology is unrelated to the underlying economic value of data. Certain types of non-personal data might have more economic value than personal data while the processing of only the personal data is subject to the requirements of EU data protection legislation (Bourreau et al., 2017).

EU apply stricter rules to the processing of sensitive data: processing this type of information is prohibited unless one of the lawful grounds of processing listed in EU rules applies.

Some 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 rules

relating to IP and trade secret protection[5]. 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[6]. The right to data portability gives 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. The controller is the person or organisation who determines the purposes and means of the processing of personal data (Bourreau et al., 2017).

2.2. Value Chain

The big data value chain is a simple industrial process that encompasses different steps, among which the production of information. The information inferred from big data suggests the new goods that could better meet consumers' preferences. The same set of data can reveal different pieces of information and the same piece of information can be derived from diverse data sets (Colangelo et al., 2017).

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[7]:

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 in 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 personalize products' characteristics and prices as well as their marketing, to improve process and organization or for many other purposes (Bourreau et al., 2017).

[5] Trade Secrets Directive provides a protection against unlawful acquisition, use and disclosure of a secret, which has commercial value and has been subject to a reasonable measure of protection.
[6] Competition law applies imposes an obligation to share data even when they are protected by an intellectual property right.
[7] Multi-side market: market characterized by indirect network effects, supplier's marginal cost of innovation decreases in amount of user information, which increase with demand ($c_{x_i}, D_{x_i} < 0$)

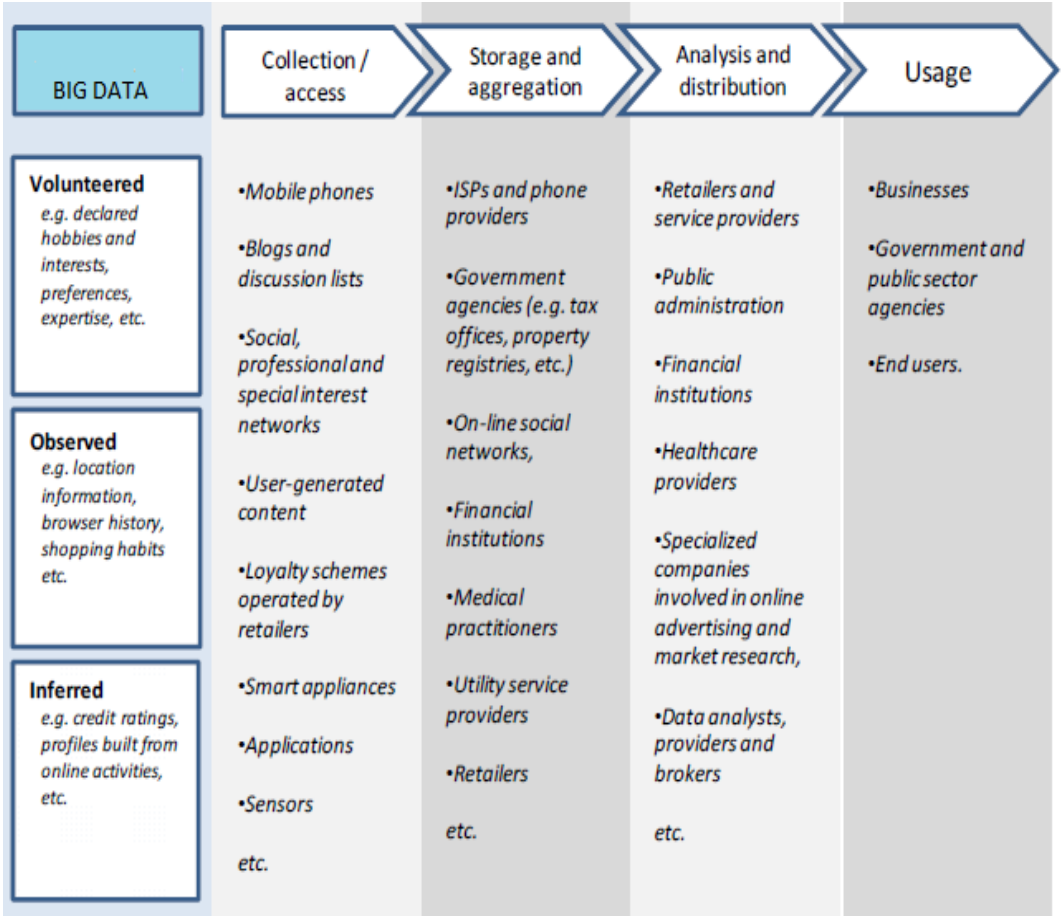

Fig. 1: Big Data value chain[8]

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

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

1. Some information is *publicly observed* through device, operating system, IP address, etc. (e.g. internet browsing preferences, location data when using cellular mobile phones);
2. Some information is *voluntarily provided* by the consumers when they explicitly share information about themselves or about third parties (e.g.

[8] Source: OECD

when someone creates a social network profile, enters credit card information for online purchases, provides his/her personal information as a condition of registration to a given on-line service, or posts information about a friend, colleague, family members, etc.);

3. Some information can be collected by *tracking the consumer* online (e.g. through tracking cookies[9], the use of application by the user, history sniffing[10], etc.) (Bourreau et al., 2017).

As already said, firms also obtain data from third parties, for example from data brokers, that are 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. Data marketplace are online stores where firms can buy and sell data (Bourreau et al., 2017).

Once data has been collected, it can be stored. The individual data elements are organized and stored in datasets that can be used for further processing and analysis. The storage of massive quantities of data requires large data centers consisting of big clusters of computer connected by fast local area networks. Those data are expensive to build and are characterized by economies of scale.

However, the development of cloud computing[11] allow small firms to rent instead of owning the data centers, thereby converting their fixed costs into variable costs. For the cloud computing market to function properly, the costs of switching between providers need 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 localization restriction that can be important for certain

[9] Tracking cookie: cookie that is distributed, shared, and read across two or more unrelated websites for the purpose of gathering information or presenting customized data to a consumer.
[10] History sniffing: practice of tracking which sites a user has or has not visited by hacking its browser history list.
[11] Cloud computing service: is defined as a digital service that enables access to a scalable and elastic pool of shareable computing resources.

type of privately owned data, in particular for health, financial data (Bourreau et al., 2017).

The third step in the value chain is the analysis of the data to extract relevant information. Applications and algorithms that are increasingly learning by themselves do this. The development and the improvement of those algorithms are based on many inputs such as data, skilled and creative labor force. Thus, data are important, but not the most important input (Bourreau et al., 2017).

Indeed, data does not typically provide value on a standalone basis. Mere possession of data alone therefore, even in large volume, does not secure competitive success, that can only be achieved through engineering talent, quality of service, speed of innovation, and attention to consumer needs. As such, the firm with the most data does not necessarily win (Sokol D. and Comerford R., 2016).

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 be also obtained from third party. In this case, as for cloud computing, the fixed development costs can be converted into variable costs (Bourreau et al., 2017).

The distribution of the processed information takes many forms. For major players, a good deal of the data collected and processed is used internally in CRM[12] software, through loyalty programs and transactions records to tailor product and service offering to the costumers. Other trade and transfer the data they collect and some purchase processed data as a service from data analysts and brokers and/or on data exchanges (OECD, 2013).

Once data have been collected, stored and analyzed, they are often made available to end users in markets representing the end of the value chain. The end user usually purchase profiles of individuals (or firms) in order to supplement their

[12] CRM: Costumer relationship management

own business activities (OECD, 2013). End user can be used data, for example, to reduce production costs, organize resources, to invent new services or products and to improve performance indicators.

It is important to say that both private and public sectors can obtain some advantages from data. They can improve the efficiency in the use of resources, create new services and improve the existing one.

3. Big Data and market power

Big Data differ from other goods and services in the economy in several dimensions.

The main economic characteristics of data are:

1. **Non-scarcity**: the scarcity is an important characteristic for the economic goods: the more a good is scarce the higher equilibrium price on the market is
2. **Non-rivalry**: when more than one person may use the same consumer data at the same time it is possible to speak of non-rival goods. This characteristic allows to partially consider data similar a public good[13]. However, restrictions can be placed on access to data through, for example, contractual condition. This implies that efficient markets may involve sharing data beyond those involved in the initial transaction to minimize the costs in multiple firms collecting, storing and processing the same data multiple times. It also means that failing to share, sell or license consumer data may, in addition to the potential to generate competition concerns, be a further source of inefficiency in data markets, leading to increased costs for consumers and firms (CMA, 2015).
3. **Non-excludability**: a good is non-excludable when it is impossible to expel third parties from their use (Agcom, 2018).
4. **Complementarity**: the capacity to aggregate heterogeneous information sources is fundamental to extract value from data (Agcom, 2018).
5. **Substitutability**: there is some degree of substitutability between data. Digital data on which firms work to extract information and develop new goods and processes often admit substitutes, and that firms interested in digital data can either generate them or get them from the market sometimes for free. In a marketing campaign, for example, data on

[13]Public good: a pure public good is a good that can be consumed simultaneously by everyone and anyone can be excluded from the use.

consumption choice of consumers can be partially replaced by data on actual consumption of services and goods (Agcom, 2018).

6. **Cost structure**: the collection, storage processing and analysis of consumer data is likely to involve relatively substantial fixed cost and low or negligible marginal costs. Firms need to acquire infrastructures, technologies, competences and specific analytical techniques to infer information from data. In markets with this structure, economies of scale[14] and scope[15] are common. This means that larger firms are likely to have cost advantages over smaller firms in collecting, storing and processing more and different type of data. These advantages can act as barriers to entry and expansion in markets, particularly where they are significant and where data is a key input into the products and services being developed. This suggests that some markets where data is important may be more likely than others to experience higher levels of concentration and so potentially lower levels of competition (CMA, 2015).
7. **Diversity in value**: there is significant diversity in the types of consumer data collected and used. Many types of data (such as the particular products a consumer has been searching for) are transient in value and are only relevant over a short time, hence their depreciation rate is very high. 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. Furthermore, historical data may be of relatively low value for some actors like search engines in view of the high rate of new search queries. For examples, Google reported that the 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

[14] Economy of scale: the cost advantages that firms obtain from size, output, or scale of their options. The cost of production typically decreases with increasing scale as fixed costs are spread over increased volume output.

[15] Economy of scope: the cost advantage that arises from firms undertaking a range of activities, where the average cost across the range of activities falls as volume of output increases.

over these types of data may not itself give rise to a sustainable competitive advantage. 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 value that is more permanent. In other words, the transient value of data may be capitalised and transformed into more permanent value through applications' improvement or developments. Other data, such as name, gender, data of birth, address have a more permanent value and their depreciation rate is much lower. Therefore, the control of those data gives more permanent benefits than the control of transient data (Bourreau et al., 2017).

Big Data can lead to markets that are more competitive thanks to making information more readily available, lower search costs and increasing propensity to switch, and even the creation of new channel for entry and expansion.

However, Big Data presents some challenges to competitive markets. Some of them are:

1. **Data network effects**: Big Data can drive network effects, raising barriers to entry and enabling the large incumbent firms to consolidate their position in the market. Data network effects (Matt Tucker, 2016) are the phenomenon whereby the product become smarter, and thus more attractive to consumers. The more consumers use a given product, the more data they contribute and the smarter the product become; this in turn attract more consumers, who contribute with new data, further increasing the products' performance, and so the cycle continues. These data network effects can combine with more traditional network effects, increasingly concentrating data and thus market power in the hands of fewer firms (Europe Economics, 2017).
2. **Platforms and market foreclosure**: Big Data can also decrease competitive pressures, by denying potential competitor firms access to this data: either through an outright refusal to supply, or a constructive refusal to supply,

for example by selling data to competitors significantly above the competitive rate. This is of particular concern where: few substitutes of data are available and the data are significant value to the production process, such absence of the data results in an inferior product offering. The data network effects, which sees Big Data increasingly concentrated in a small number of platforms may exacerbate this effect (Europe Economics, 2017).

3. **Collusion**: by increasing the speed at which price changes are observed, and thus the ability to detect and punish deviation, Big Data may also facilitate greater collusive practises. This is of particular concern as collusion in a Big Data context is likely to take more tacit forms. Ezrachi and Stucke (2015)[16] studied how algorithmic models built around Big Data can promote greater collusion: firstly, by omitting human biases from the strategy, algorithmic models should create the stability necessary for the tacit collusion[17]. Secondly, by knowing that rival firms operating similar models can capture and respond to competitive price changes very quickly, this diminishes the incentive to undertake such strategies (Europe Economics, 2017).
4. **Mergers**: in the case of data-driven mergers, firms want to take advantage of complementarities in the data they collect (Europe Economics, 2017).
5. **Behavioural discrimination**: the growth of Big Data has led to a move away from traditional price discrimination models to models of behavioural discrimination. Big Data is allowing firms to segment consumers into smaller groups to better identify their reservation price and ultimately extract greater surplus. Such behavioural discrimination by incumbent firms make it more difficult for potential entrants to compete, given that they lack sufficient scale or breadth of data (Europe Economics, 2017).

[16] Ezrachi A. and M. E. Stucke (2015), "Artificial Intelligence & Collusion: When computers Inhibit Competition". Oxford Legal Studies Research Paper No. 18/2015

[17] Tacit collusion – where firms make informal agreements or collude without actually speaking to their rivals. This maybe to avoid detection by government regulators.

3.1. Barriers to entry

Barriers to entry and expansion may affect competition in the long term, restricting the growth of smaller firms and the entry of new firms. These potential barriers fall within the following categories:

1. **Structural barriers**: arise from basic industry conditions, such as the structure and costs of production including the potential for economies of scale and scope, the technology used or other similar factors needed to become established in a market (CMA, 2015).

   One of the properties of collecting, storing, and analysing data is that the costs are typically mostly fixed, with low marginal costs of increased consumption, collection, storage and analysis. Therefore, once the necessary infrastructure and technical competences to generate and manage data have been acquired producing and distributing digital data is cheap. This gives rise to the presence of economies of scale and scope in these activities. Economies of scale and scope are usually mentioned as barriers to entry and expansion.

2. **Strategic barriers**: arise when the incumbent firms intentionally create or enhance the advantages they have over new or smaller rivals from their established position (CMA, 2015).

   One important potential barrier relevant to consumer data is first-mover advantage, which in data markets may be related to issues around trust, reputation and brand recognition. Some online firms, despite they are relatively new, have been active for a number of years and a significant number of consumer data are used regularly so have developed strong brands an reputations.

   Pricing strategies of incumbent firms could create barriers to entry and expansion when these costs are large for small or new firms to replicate. Such strategies may include discriminatory pricing, where a firm uses consumer data to charge different price to different group of customers. Small and new firms would not have a substantial fixed base of existing

customers, and so may be unable to compete as successfully to target customers through proposing them lower prices.

Where consumer data is important for the production of good or service, incumbent firms may decide to enable other firms, including data intermediaries and rivals, access to the data and benefit from the fees charged for the access. Alternatively, firms may choose to restrict access to the consumer data they collect and use internally, with the potential for this to give it better products or services and a competitive advantage over rivals.

3. **Absolute barriers**: these include legal barriers and technical advantages including preferential access to intellectual property (CMA, 2015).

   There are relatively few absolute barriers in data markets; firms that collect, purchase, store or use consumer data need to comply with data protection requirements.

   These regulations may discourage small and new firms from entering the market, reducing the competitive constraint and limiting the potential for innovative products and services.

4. Network Effects

A *network effect*, or network externality, is the effect that one user of a good or service has on the value of that product to other people. It can be positive, if any additional user increases the value of the product for other users, or negative, if it decreases such value. Network externalities exist when consumer utility in a certain market depends on the total number of consumers of the same good or service. There are a lot of example: the telephone system (the value of a phone increases if everyone has a phone), Facebook and many marketplaces. Data network effects produce many of the same benefits but they are more subtle. "*Data network effects occur when your product, generally powered by machine learning, becomes smarter as it gets more data from your users*" (Turck Matt, 2016). Therefore, this means the more users use your product, the more data they contribute. The more data they contribute, the smarter your product becomes (which can mean anything from core performance improvements to predictions, recommendations, personalization, etc.); the smarter your product is, the better it serves your users and the more likely they are to come back often and contribute more data – and so on and so forth. Over time, firms business becomes deeply and increasingly entrenched, as nobody can serve users as well.

Standard network effects and data network effects can be presented in the same company. For example, Uber benefits from standard network effect, indeed, it becomes more valuable for everyone as more drivers and more customers join the service. It benefits also from data network effect, more data enables Uber to constantly improve its routing algorithms to get customers a car as quickly as possible, to ensure its drivers get as many jobs as they can handle, making everyone happy and more likely to be long term members of the network.

Data network effects require at least some level of automated production of the learning. It is important to underlines that most businesses learn in some way from data, but that is typically done through analytics, with human analysts doing

a lot of the work, and a separate process to build insights into the product or service. The more automation you build into the loop, the more likely the data network effects rise (Turck Matt, 2016).

Google is a typical example where present data network effect are: the more people search, the more data they provide, enabling Google to constantly improve its core performance, as well as personalize the user experience.

Data can increase product value in different ways. If data is central element to the way the product benefits users, then the data of that product has the potential to be very powerful. If data is only marginal to the product, the Data will not matter much. When Netflix recommends a show to users, the algorithm is basing that recommendation on user viewing data. However, Netflix's discovery function is marginal; its real value comes from the inventory of television shows, movies and documentaries. Therefore, Netflix only has a marginal data network effect.

It is important to say that the relationship between product usage and the amount of useful new data collected can be asymmetrical. Yelp, a recommendation software in USA, has data network effect because a greater number of reviews for a greater number of restaurants makes the product more valuable. However, its network effect is weakened by the fact that only a small percentage of users produce the data; most people read from the Yelp database but they do not write to it.

Yelp[18] is also a good example of weak data network effects. They are asymptotic. The fifth review adds a lot more value than the thirtieth. Past a certain low level, more reviews on a restaurant do not increase the value to the user.

Instead of a good example of a service with a strong data network effects is Waze[19], a GPS software. Everyone consuming data on Waze also contribute useful data, and because the data is consumed in real time, the dataset needs to be continuously updated. Thereby the larger the network, the more accurate that

[18] https://www.nfx.com/post/network-effects-manual

[19] https://www.nfx.com/post/network-effects-manual

data will be at any instant for any given road. More data continues to produce value almost indefinitely, so there are no asymptotic data network effects.

4.1. Formal approach to Data Network Effects

One of the most typical data businesses is that a company collects data from its users and create a data platform. Users in return get useful information from the data platform. They can access both information about a service both a product but also additional information such as user evaluations. The abundance and reliability of information rely on how many users are contributing to the data platform. This type of service is expected to have demand externalities in that the benefit to a user depends on the total number of users. As the number of users increases, benefit that each user can obtain from the service also increase (Mitomo Hitoshi, 2017).

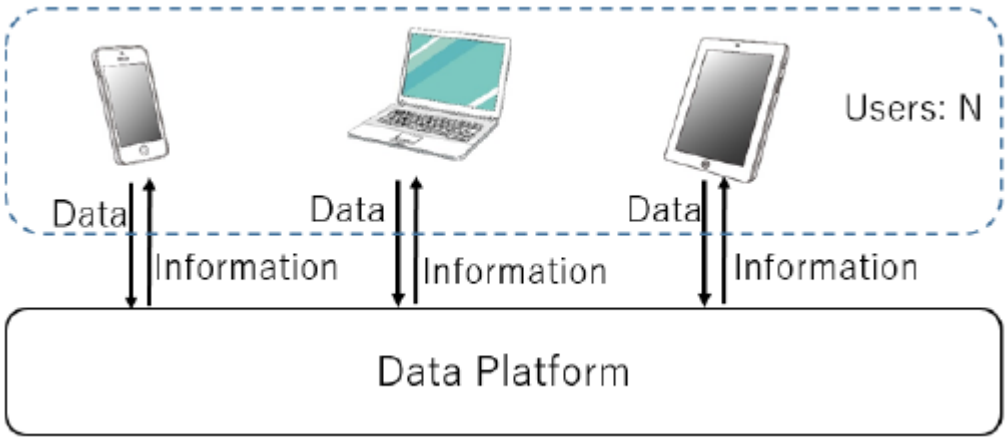

Fig. 3: Data service based on data collected from users[20]

Suppose there is a data platform service that is composed of a set of users. The number of the total users is denoted by y where y is a subset of the total potential users N. Each users acquires information from the platform. In return, he provides different data including personal information to the service. We assume that a

[20] Source: "Data Network Effects: Implications for data Business", Mitomo H. (2017).

benefit to a user does not depend directly on the information provided but simply on the number of the users. It is because both the quality and the quantity of the service depends on how many people use the service. Mitomo (2017) assume that each individual user is indexed by a unique index i, and without loss of generality that it distributes uniformly between 0 and N, such that $i \in [0, N]$.

In addition, he assume that users are distributes in order of the size of their potential demand. Let the index of the user who has the minimum potential demand be N and that of the user who have the maximum be 0.

The potential demand for the service can be defined as $v_p = D(p, i, y)$ where p is the unit price for the service. This demand function explicitly defines the existence of dependencies of each individual user's demand on the number of users. Recently, most platform services are provided for free or users do not pay for the service directly. Alternatively, some services are provided with a flat rate. A two-part tariff can deal with both usage sensitive and non-sensitive price settings.

The total charge C that the user i should pay for the service $C(v_p)$ is represented by a combination of usage and the flat fees:

$$C(v_p) \equiv C\big(D(p, i, y)\big) = p * D(p, i, y) + F$$

Now it is possible to assume that the demand for the service is finite even in the case where the service is provided free of charge:

$$D(0, i, y) = V(i, y)$$

Where $V(i, y)$ denotes the potential demand of user i for the service for the user set y. The gross benefit for user i from consuming the service depends on the unit price and the potential demand defined as:

$$B\big(p, V(i, y)\big) = \int_0^{v_p} D^{-1} dv \text{ where } v_p = D(p, i, y)$$

Therefore, the net benefit from this service is

$$NB(p,i,y) \equiv NB\big(p,V(i,y)\big) = B\big(p,v(i,y)\big) - C\big(v_p\big) = S(p,i,y) - F$$

Since the gross benefit can be represented by the area under the relevant demand curve, the net benefit is formulated as the consumer surplus $S(p,i,y)$ net of the fixed charge F.

For the user set to be feasible, the net benefit for the smallest user $i = y$ should be non-negative:

$$NB(p,i,y) \geq 0 \text{ for } i = y$$

According to Mitomo, stable and unstable equilibria can be defined in terms of the user set. For an equilibrium point $y = y^*$, if $\frac{dNB(p,y,y)}{dy}$ is negative, the user set [0, y^*] is a stable equilibrium, and if positive, it is an unstable equilibrium. An unstable equilibrium defines a "critical mass"[21], a concept in the diffusion theory.

In the case of user-generated data, each user receives the benefit from the service depending on the number of the users. Thus, interdependencies among users will create a mass effect. This will result in the advantage of a service attracting a large number of users. If the services provided by competitive suppliers are homogeneous, such as online information retrieval system, the precursor can take this advantage.

The figure 4 illustrates a case of a single modal net benefit function $NB(p,y,y)$ has a single modal parabolic curve at each unit price level. The fixed price is shown as F, which is a cutting plane parallel to the bottom plane. At the price p^*, the curve has two point of intersection with F. the lower intersection, y_0^*, is defined as a critical mass and the upper one, y_1^*, as an ultimate expansion level of the user set. The supplier once can attain a user set exceeding y_0^*, it will expand autonomously to y_1^* (Mitomo Hitoshi, 2017).

[21] Critical mass: is the minimum penetration level needed to a technology to remain in the market

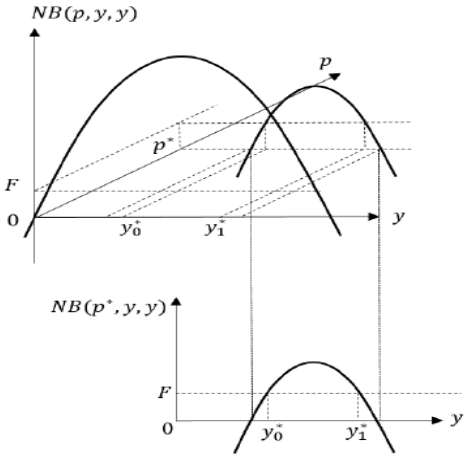

Fig. 4: The net benefit and the existence of the equilibria[22]

This suggests the existence of the first-mover advantage due to the existence of data network effects. If a precursor can overcome difficulties associated with the start-up stage of business and reach the critical mass level, the business can acquire a dominant position.

The revenue from the business is not depicted in the figure 4, but the combination of a unit price and a fixed charge can cover a variety of tariff settings and the supplier can selected an appropriate setting as a strategic tool for attracting users. In an early stage of service delivery, the service is less attractive because the users do not know how it is useful. In order to facilitate the subscription, the supplier can apply a low introductory price, a very low price or even zero price (Mitomo Hitoshi, 2017).

4.1.1. Competition in the presence of Data network effects

It is possible to see that the incumbent has an advantage in the provision of the service over potential entrants and can occupy a dominant position. Now it is possible to suppose, according with Mitomo model, that exists an entrant that seeks to provide service identical to incumbent service. The entrant, from marketing point of view, will employ a strategy of product differentiation to avoid fierce competition with the incumbent. If the service is homogeneous, a successful

[22] Source: "Data Network Effects: Implications for data Business", Mitomo H. (2017).

entry will be a cream-skimming[23] entry. Therefore, the entrant would focus on large-scale users (Mitomo Hitoshi, 2017).

From the assumption, the consumer surplus or the net benefit is monotone decreasing with respect to user index i.

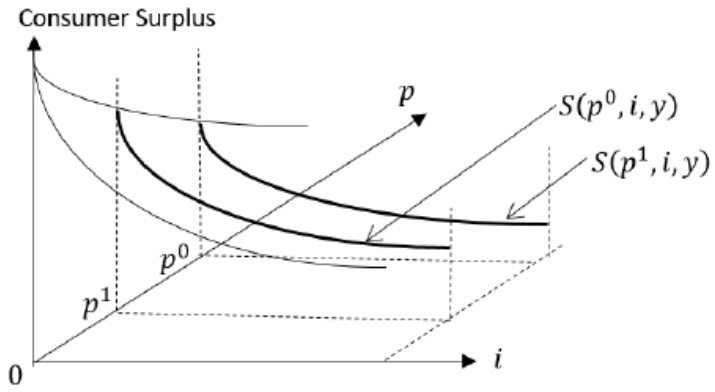

Fig. 5: Consumer surplus function[24]

At $i = y$, it should be equal zero since the net benefit for the smallest user must be equal zero at an equilibrium. The success or failure of the new entry totally depends on the shape of the net benefit function. The figure 5 shows the consumer surplus function, which is defined as the gross benefit net to the total unit usage charge. Assume p^0 and $p^1$are the prices for the services provided by the incumbent and the entrant. If the services from two suppliers are substantially homogeneous, the entrant cannot set the price higher than the incumbent one. So, $p^0 \geq p^1$. Since

$$\frac{\delta S(p,i,y)}{\delta p} = -D(p,i,y)$$

And the demand is monotone decreasing with respect to i, we obtain, for $p^0 \geq p^1$, that

$$\frac{\delta S(p,i,y)}{\delta p}\Bigg|_{p=p^0} \geq \frac{\delta S(p,i,y)}{\delta p}\Bigg|_{p=p^1}$$

[23] Cream-skimming: is a term used to refer to the perceived business practice of a company providing a product or a service to only the high-value or low-cost customers of that product or service, while disregarding clients that are less profitable for the company.

[24] Source: "Data Network Effects: Implications for data Business", Mitomo H. (2017).

This means that the consumer surplus curve for the incumbent is less steep than the curve for the entrant. On the base of the fixed charged, F^0 and F^1, an intersection can be found, as illustrated in figure 6.

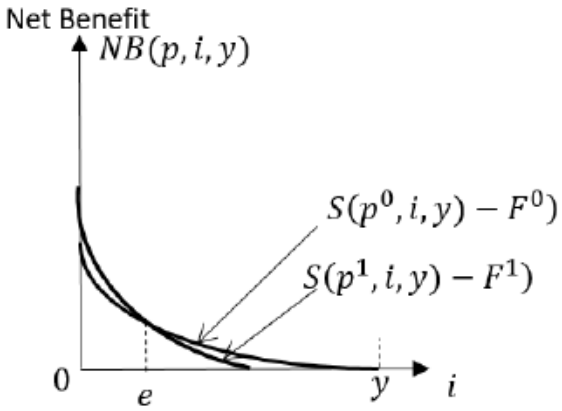

Fig. 6: Cream-skimming entry[25]

This implies that the entrant can obtain the users $0 \leq i \leq e$ and the incumbent the users $e \leq i \leq y$. However, there is a case where the benefit from the incumbent's service exceeds the entrant's service for all users.

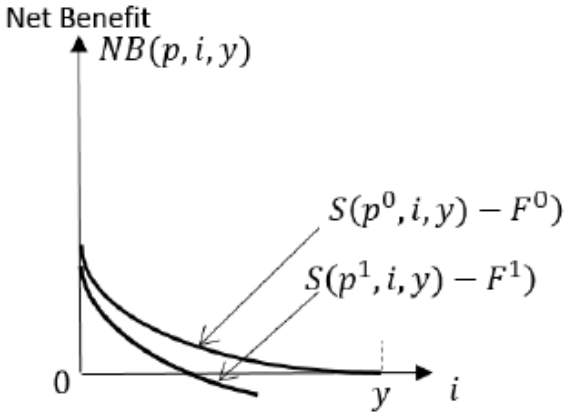

Fig. 7: Failure of entry[26]

The success or failure of the entrant depends on the shape of the benefit function and the tariff setting. If the incumbent's service is provided free or at very low

[25] Source: "Data Network Effects: Implications for data Business", Mitomo H. (2017).
[26] Source: "Data Network Effects: Implications for data Business", Mitomo H. (2017).

price by utilizing other revenue sources such as advertisement, it would be difficult for the entrant to get share in the market (Mitomo Hitoshi, 2017).

4.1.2. Policy and business implications

Even if it is already said that there are some possibility of competition in the market, it seems difficult to enhance competition in dominant platform business even where a potential entrant seeks to start a competitive service as far as there is no substantial product differentiation (Mitomo Hitoshi, 2017).

In the presence of positive externalities, the equilibrium diffusion level tends to be lower than the socially optimal level. The existence of data network effects can apply to this case. Let us suppose there exists a potential user who is willing to use the service. He will perceive the benefit from using the system with the total number of users $y + 1$. His perceived benefit is given by $NB(y + 1)$. The whole users are also benefited from his participation. The increase in the social benefit is:

$$(y + 1) * NB(y + 1) - y * NB(y) = y[NB(y + 1) - NB(y)] + NB(y + 1)$$

In addition to own benefit, $NB(y + 1)$, the new user will create an additional benefit to all other users, $y[NB(y + 1) - NB(y)]$. He will not perceive this additional benefit created by his participation. Thus, the private benefit is lower than the social benefit created by him. The equilibrium point where the marginal private benefit is equal to the marginal private cost is lower than the social optimal point where the social marginal benefit is equal to the marginal social cost. So, if it is left to the market mechanism, a lower diffusion level will be attained. The existence of the gap will justify a policy support to fill it and attain the socially optimal level of diffusion (Mitomo Hitoshi, 2017).

When data do not depend on users and are collected from alternative data sources, such as physical sensors or many unspecified users the model will be more complicated. In this case, the competitive advantage of the incumbent relies on the interdependencies among multi-sided markets.

5. Two-sided market

One characteristic of data markets is the speed with which products and services offered in these markets, as well as the firms in them, can change. This is driven by the growth in the collection and the availability of consumer data, together with increased processing power allowing for new and more sophisticated uses of consumer data to arise. There are different categories of markets that exploit consumer data, one of these are two-side platforms.

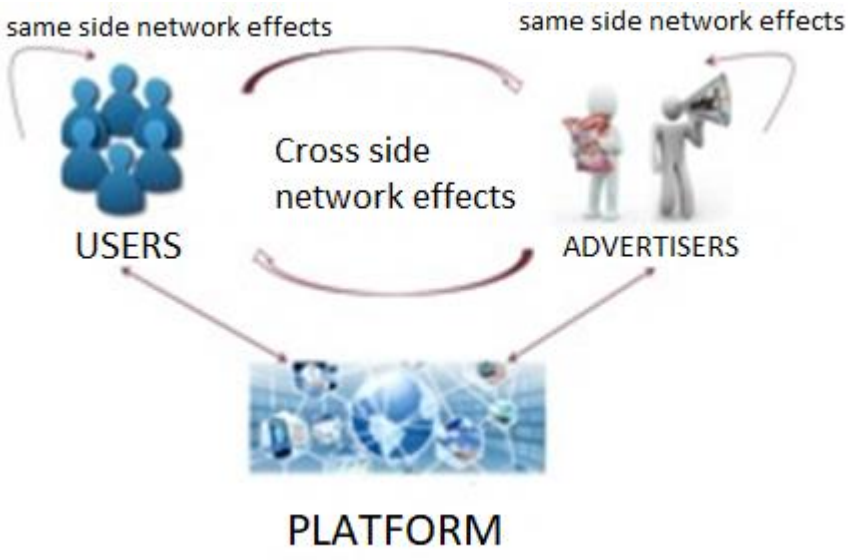

Fig. 8: Two-sided market model

Two-sided markets can be defined as markets in which special services are sold, allowing the interaction of two parties on a platform, managed by a third entity (Roson Roberto, 2004).

In other words, two-sided markets are markets in which one or several platforms enable interactions between end-users and try to get the two sides *"on board"* by appropriately charging each side. That is, platforms court each side while attempting to make, or at least not lose, money overall.

Two-sided markets are markets with a special type of network externality, *cross-side externalities or indirect network effects*. This externality does not depend on the number of agents in the same class (e.g., consumers of the same product), but on the number of different but "compatible", agents on an opposite market side (Roson Roberto, 2004).

Two-sided market	Subsidy side	Profit side	Examples
PC operating systems	Application developers	PC users	Microsoft Windows, Apple Macintosh
Online recruitment	Job seekers	Employers	Monster, Career Builder
Yellow Pages	Consumers	Advertisers	
Web search	Searchers	Advertisers	Google, Yahoo
Video games	Players	Game developers	Sony PlayStation, Microsoft Xbox, Nintendo Wii
Shopping malls	Shoppers	Retailers	
Streaming media	Consumers	Servers	Real Player, Windows Media Player
Web browsers	Users	Web servers	Microsoft Internet Explorer, Mozilla Firefox
Portals, newspapers and TV networks	Visitors, readers and viewers	Advertisers	
Text processing	Readers	Writers	Adobe Acrobat, Microsoft Word
Credit and differed debit cards	Cardholders	Merchants	Visa, MasterCard, American Express

Fig. 9: Different types of two-sided markets[27]

In buying a credit card, the number of merchants accepting the card for payment will be considered, in addition to its usage cost. Between the two elements, cost and diffusion level, there exists a clear trade-off. Indeed, a network may get revenue from both market sides and consequently can charge more some agents (e.g., merchants), and charge less some others (e.g., credit card owners). On the

[27] Source: Maffè Carnevale C. A. and Ruffoni G. (2009), "Two-sided markets: models and business cases".

other hand, the consumer utility does not depend on the price only: nobody would be interested in getting a cheap credit card that no merchant would accept. Given that the degree of diffusion on the other market side depends on the price that is applied there, the indirect utility[28] for an agent in a two-sided market depends on both prices (Roson Roberto, 2004).

Another example takes into account PC and video game network.

PC and video game networks look similar (Maffè Carnevale C. A. and Ruffoni G., 2009), with end users on one side wishing to link to software or games on the other side who buy a platform consisting of an operating system (OS) bundled with hardware – a PC or a game console. In addition, the two businesses show similarly positive cross-side network effects: end users privilege platforms that offer a wide variety of complements and developers privilege platforms with more end users because this improves the odds that they will recover the fixed, upfront costs of creating complements. In video games end users are subsidized and game developers are on the network's money side, while, in the PC industry end users are the money side, paying well above cost for the platform's essential element, its OS, and application developers are the subsidy side, as they pay no royalties and receive free software development kits from the OS vendors. This difference is because video game consoles users, typically teenagers, are both far more price sensitive and quality conscious than typical PC users. PCs are often purchased for work and are otherwise more likely viewed as household necessities than game consoles are, so price sensitivity is lower. Gamers' need for quality seems to be stronger, as does game developers' need for large numbers of consumers. PCs, on the contrary, accumulate many applications, with a huge range of price and quality levels (Maffè Carnevale C. A. and Ruffoni G., 2009).

[28] Indirect utility function gives the consumer's maximal attainable utility when faced with a vector P of goods prices and an amount of income Y. it reflects both consumer's preferences and market conditions.

5.1. Model two-sided markets

Now I consider the Rochet and Tirole (2003) model. This model is well suited to the credit card context although the analysis applies more widely. Economic value is created by "interactions" or "transactions" between pairs of end users, buyers (superscript *B*) and sellers (superscript *S*). Buyers are heterogeneous in that their gross surpluses b^B associated with a transaction differ. Similarly, sellers' gross surplus b^S from a transaction differ. Such transactions are mediated by a platform. The platform's marginal cost of a transaction is denoted by $c \geq 0$. As an illustration, consider the case of payment cards. The buyer wants to purchase a bundle of goods or services from the merchant at a certain price p. In our vocabulary, a "transaction" takes place if and only if the buyer pays by card instead of using another payment instrument (say, cash). Benefits b^B and b^S correspond to differences in utility of buyers and sellers when they pay by card rather than cash. Under the No Surcharge Rule (very often imposed by payment card networks), the merchant is not able to charge different retail prices for card and cash payments. Therefore the distributions of $b^B$and $b^S$are independent of the prices chosen by platforms and merchants, and can be taken as exogenous. In the absence of fixed usage costs and fixed fees, the buyers' (sellers') demand depends only on the price p^B(respectively,p^S) charged by the monopoly platform. There are network externalities in that the surplus of a buyer with gross per transaction surplus b^B, $(b^B - p^B)* N^S$, depends on the number of sellers N^S, but the buyers' "quasi-demand function"[29]:

$$N^B = \Pr(b^B \geq p^B) = D^B(p^B)$$

Is independent of the number of sellers. Similarly, let

$$N^S = \Pr(b^S \geq p^S) = D^S(p^S)$$

Denote the sellers' quasi-demand for platform services.

[29] The word "quasi-demand function" is used to reflect the fact that, in a two-sided market, actual demand depends on the decisions of both types of users (buyers and sellers in our terminology). In our specification, this demand is simply the product of the quasi-demands of buyers and sellers.

Assuming for simplicity the independence between b^B and b^S, the proportion (or volume) of transactions is equal to the product $D^B(p^B) * D^S(p^S)$.

Now it is possible to consider the private monopoly case. A private monopoly chooses prices to maximize total profit:

$$\pi = (p^B + p^S - c) * \ D^B(p^B) * D^S(p^S)$$

The authors derive the profit for p^B and p^S and the derivative were set equal zero:

$$\frac{\partial \pi}{\partial p^B} = \ D^S * \frac{\partial \ D^B}{\partial p^B} * (p^B + p^S - c) + \ D^B * D^S$$

$$\frac{\partial \pi}{\partial p^S} = \ D^B * \frac{\partial \ D^S}{\partial p^S} * (p^B + p^S - c) + \ D^B * D^S$$

They know that D^S and D^B are constant, so they simplify the equation:

$$(p^B + p^S - c) = -\frac{\partial p^B}{\partial \ D^B} \ D^B$$

$$(p^B + p^S - c) = -\frac{\partial p^S}{\partial \ D^S} \ D^S$$

They multiply the first equation for p^B and the second one for p^S and they find:

$$(p^B + p^S - c) = \frac{p^B}{\varepsilon^B} = \frac{p^S}{\varepsilon^S}$$

$$\varepsilon^B = \frac{\partial \ D^B p^B}{\partial p^B \ D^B}$$

$$\varepsilon^S = \frac{\partial \ D^S p^S}{\partial p^S \ D^S}$$

So Rochet e Tirole (2003) conclude that a monopoly platform's total price, $p = p^B + p^S$ is given by the standard Lerner formula for elasticity equal to the sum of the two elasticities, $\varepsilon = \varepsilon^B + \varepsilon^S$:

$$\frac{p - c}{p} = \frac{1}{\varepsilon}$$

Therefore, prices on market sides are imposed on the base of elasticity: in the side where elasticity is more rigid, price is more elevated.

Prices applied to the two market sides are both directly proportional to the price elasticity of the corresponding demand:

$$\frac{p^B}{\varepsilon^B} = \frac{p^S}{\varepsilon^S}$$

An increase in the elasticity of one sub-market reduces the aggregate price but increases the specific relative prices in that sub-market.

$$\frac{(p^B - (c - p^S))}{p^B} = \frac{1}{\varepsilon^B}$$

$$\frac{(p^S - (c - p^B))}{p^S} = \frac{1}{\varepsilon^S}$$

The relevant cost concept for two-sided markets is the *opportunity cost* ($(c - p^S)$ or $(c - p^B)$): marginal output in one side gives rise to a marginal cost, but also to the selling of an extra unit on the partner market. Also, notice that any factor, which is conducive to a higher price on one side, simultaneously reduces the marginal opportunity cost, and therefore the price, on the other sub-market. If $p^B > c$ one side compensates the marginal cost of the other and the price in the second side will be negative: when one side generates many externalities and the demand is elastic, this side will have low price. Typically, prices on the two sides move to opposite directions (Roson Roberto, 2004).

In conclusion, in a two-sided market, the demand is influenced by the externalities and the elasticity.

Is important to underline that the elasticities of demand are affected by platform competition and in particular by multi homing. Multi-homing is a term used to define those situations in which some agents, in one or both sides of a two-sided market, adopt more than one platform, so that interactions may occur through a series of alternative channels.

The elasticity of buyers' demand for a given platform increases, due to their ability to switch usage to a competing platform. On the other hand, the elasticity of sellers' demand is corrected by an index calls by Rochet and Tirole the "single-

homing index", it measures the loyalty of consumers of platform i, more precisely, the proportion of them who stop trading when platform i ceases to be available. So is possible to say that buyers' multi-homing allows platforms to address sellers that is to induce them to opt out of the competing platforms.

The smaller the single-homing index of buyers, the higher the incentive for platforms to address sellers.

Furthermore, Rochet and Tirole, considering a two sided market where in one side agents are able to multi-home, assert that single-homing side receives a large share of the joint surplus, while the multi-homing one receives a small share.

Multi-homing stems from the users' desire to reap the benefits of network externalities in an environment of non-interconnected platforms. For example, videogame developers may port their game to several game platforms. More generally, software developers may multi-home to competing but incompatible software platforms. In addition, because different payment card systems are not interconnected (a Visa cardholder cannot use her card at a merchant who accepts American Express or MasterCard but not Visa), merchants often accept and consumers often hold multiple cards. More generally, multi-homing by at least one side of the market is necessary for gains from trade when platforms are incompatible or not interconnected.

5.2. Two-sided markets and competition

Two-sided markets, as with other markets with network effects, often have high level of concentration, as customers gravitate toward companies that have already a large numbers of customers.

Two-sided data markets may therefore feature large firms holding a position of market power. In some cases, this may arise where on firm is an early innovator and builds up a strong market position on the basis of a first mover advantage. In other cases, there may be intense competition between a number of firms initially, as they tend to become established as preferred network (CMA, 2015).

When a new platform enters in the market, customers on either side of the two-sided market can choose whether to switch to the new platform ("single-homing"), or adopt both platforms ("multi-homing").

Understanding market participants' multi-homing tendencies on each side of the market has important strategic implications. Theoretical work has shown that when agents on one side of the market choose to multi-home, agents on the other side of the market will prefer to single-home as they can reach the same group on the other side through one platform. At the same time, when agents on one side of the market are multi-homing, the competition between the platforms on that side of the market will be not aggressive. When agents are single-homing, however, platforms will need to compete aggressively to attract them. For example, if all consumers choose to visit both Groupon and LivingSocial, merchants need to work with only one platform to reach all potential consumers. Groupon and LivingSocial do not have to compete for consumers but need to compete more aggressively to attract merchants than consumers. Furthermore, their strategies would be different if instead merchants were more likely to multi-home than consumers (Park et al., 2017).

Barriers to entry and expansion may be more significant in two sided-markets, especially, in those markets where consumers only use one platform. This is because of the difficulty, in such markets, for small and new firms to gain sufficient numbers of customers on both sides of two-sided platform, due to both customers and sellers being attracted to the largest network. Where consumers do not single-home, it is less likely that a two-sided market would have an impact on barriers to entry and expansion (CMA, 2015).

Competition and innovation in some data-rich two-sided markets can involve new developments that offer more functionality or a better service than that previously available. In this way, an innovative new provider can seek to overcome the difficulties of market entry and attract enough users and advertisers to make their platform operate effectively. For example, the development of Facebook gaining the previous customers of a similar service Myspace, and the entry of

Google's search engine compared to the established rivals at the time (CMA, 2015).

It is important to underline that, because platform are characterized by indirect network effects, strategic changes on one side of market in response to entry will require adjustment of strategies on the other side of the market. Given the complexity of the strategic responses, studies (Park et al., 2017) demonstrated that organizational learning from prior experience in responding to similar entrants to be particularly helpful to platform.

The literature on learning suggests that past experiences can generate learning that enhances growth, competitiveness, and survival. The literature on learning-by-doing provides evidence that firms benefit from replication of an experience by moving down a learning curve.

Learning-by-doing could be important for platforms in two-sided market settings because of the complexity involved in strategy. Experts expect that a platform that experiences a disruptive shock from another technology to learn how to respond to this shock, and be more readily able to respond to a similar technological shock in the future. Is important to say that firms need to consider how their customers' behaviour, whether they multi-home or single-home, may influence the trajectory of competition in a market (Park et al., 2017).

5.2.1. Search engines

Search engines are platforms that enable to collect data. Search engine respond to queries by providing relevant and valuable information about the topics that their users are looking for. They also are intermediaries, which match consumers with providers as services or sellers of products. Search engines obtain revenues through the collection and processing data from users and selling advertising slots to companies. By analysing the users' data, they can improve the quality of the search engine algorithm and provide more relevant organic search results, but they can also design personalised advertising strategies for companies' products and services. Advertising slots are attributed on a competitive basis to the

companies that are willing to pay the highest amount to get some prominence in the platform when the user is searching for relevant terms. Slots are auctioned and the winners have to pay a fee to the search engine whenever a potential consumer clicks on the ad. Search engines base their revenue almost exclusively on selling advertising slots.

The Financial Times (2010) says: “Google may have highly successful in search but competition is only a click away”. This point of view draws on the economic theory of contestable markets, which holds that an incumbent with market power cannot exploit consumers as long as he has to fear that competitors would just step in and offer their service at lower prices.

If the market for search engines is highly contestable, Google high market share, in Europe the market share is 91%, can then be taken as an indication that it is just the best search technology currently available.

Instead, if the market is non contestable, then its structure, rather than the conduct of any participants, is likely to be problematic and to call for direct regulatory intervention rather than antitrust action (Argenton C. and Prüfer J., 2012).

A peculiar type of indirect network externalities characterizes the production of search quality. Such indirect externalities (Argenton C. and Prüfer J., 2012) arise on the market for search engines because users will not consider, when deciding whether to run another query, that the results of their query and subsequent clicking behaviour on suggested links are stored by the search engine. Currently, this information, also known as query logs or search logs, is not public. Only the search engine that is used to run the query can aggregate it with the information gained from other users who entered a similar search keyword. Thereby, it can improve its guess as to what future users, on average or with certain revealed characteristics such as geographical location or language, are looking for when they enter a certain keyword. This translates into higher search quality perceived by users.

This mechanism is at the core of the model that Argenton and Prüfer (2012) propose to analyse competition in the search engine market. This implies for a competitive market, taking everything else equal, that a firm that had a modest lead in market share at some point in time can increase that advantage more and more, while the other firms' market shares decrease more and more. That is, the market `tips'.

The search engines quality perceived by users, may be proxied by the expected time a user needs to obtain a satisfactory result to his search query. This time depends on the inputs into search:

1. The sophistication of the search algorithm
2. The computer power of the server farms searched by the algorithm
3. The amount of potentially relevant data that the algorithm can search through

Since the quality of search engines increased with data collection and analysis, search platform with a larger stock of information than their competitors can have a competitive advantage. They can perform better and attract more users and consequently more advertisers. This suggests that the amount of data processed is linked with the market power in the search engine market. Incumbent with a large stock of information and experience in data analytics can protect their market penetration against new entrants and firms that are far behind in data accumulation (Argenton C. and Prüfer J., 2012).

The model

Search engines do not charges end-users for running queries. Nevertheless, some are highly profitable. Most of their revenues, as we have already said, come from selling advertisements related to search queries and displayed as “commercial results” or “sponsor links” next to the so-called “organic links”, which are the results to a search queries generated by the search engine’s algorithm. Because in literature it is a controversial question whether sponsored links bring direct utility to consumers, the Argenton and Prüfer model assumes that consumers neither derive positive or negative effects from them. The model assumes that the advertising revenue of a search engine is proportional to its market share.

Today's search engine market features one dominant firm that competes with two notable competitors. In Europe, in 2018, Google has a market share of 91.72%, Bing 3.65% and Yandex RU 2.38%. Therefore, it is possible to start by modeling a triopoly market with firms 1, 2, and 3.

On the demand side, in the model there is a unit mass of consumers, each of which has demand for one query. As has been business practice since the birth of the industry, nominal prices for using a search engine are zero. It is possible to assume that the market is fully covered[30].

The authors consider a class of models that allow consumers to choose which search engine they want to use, on the basis of those engines' characteristics, in particular their perceived quality level, as well as consumers' own preferences.

Discrete choice models of product differentiation allow consumers' utility to be randomly shocked in a way that makes sure that all engines capture at least a fraction of demand. In such models, consumers do not know in advance which product they will consume and it is possible to compute clear measures of economic welfare, such as expected consumer surplus. To simplify the presentation, the authors take a shortcut and work with a model that fits the stylized facts outlined above (as well as the properties of the functional form for market shares that would be derived from standard discrete-choice models).

Argenton and Prüfer (2012) model competition as a contest among search engines with simultaneous bids x_i, where x_i is firm i's search quality; $i \in \{1,2,3\}$. That is, each search engine simultaneously and independently chooses its quality level. The market share of firm i is then given by $D_i = \frac{x_i}{\sum_{j=1}^{n} x_j}$, where $j \in \{1, \dots, n\}$ and $n = 3$ firms are active in the market. Production of quality x_i comes at a cost, $C(x_i) = \frac{x_i}{N_i}$, where N_i is the "installed base" of firm i, i.e. the amount of previous search queries run on i. Without loss of generality, it is possible to

[30] market fully covered: market where the quality is high enough for every consumer to use a search engine.

assume $N_1 \geq N_2 \geq N_3 \equiv 1$. Moreover, each firm bears a fixed cost F for its operations.

This formulation interprets x_i as the quality of search engine i perceived by consumers. The cost to create a certain level of perceived quality depends on the resources spent on improving the search engine's algorithm and on the amount of private data accessible, which makes it cheaper to produce any quality level. It is as if search engines learned to produce quality more cheaply by using their stock of past queries.

Following these assumptions, search engine i solves the following program:

$$max_{x_i}\ \pi = \frac{x_i}{x_1 + x_2 + x_3} * p - \frac{x_i}{N_i} - F_i$$

Where p is the exogenously given advertising revenue associated to one consumer[31].

The model is a static game with complete information. If we solve the previous equation for all three firms, we obtain the following system of reaction functions:

$$x_1 = \sqrt{p * N_1 * (x_2 + x_3)} - x_2 - x_3$$

$$x_2 = \sqrt{p * N_2 * (x_1 + x_3)} - x_1 - x_3$$

$$x_3 = \sqrt{p * (x_1 + x_2)} - x_1 - x_2$$

Solving this system for optimal quality levels gives the following unique Nash equilibrium with positive market shares:

$$x_1^* = \sqrt{\frac{p^2 * {N_1}^2 * {N_2}^2}{(N_2 + N_1 * (1 + N_2))^2}} - \frac{4 * p * N_1 * {N_2}^2}{(N_2 + N_1 * (1 + N_2))^2}$$

$$x_2^* = \frac{2 * p * N_1 * N_2 * (N_1 * (N_2 - 1) + N_2)}{(N_2 + N_1 * (1 + N_2))^2}$$

$$x_3^* = \frac{2 * p * N_1 * N_2 * (-N_1 * (N_2 - 1) + N_2)}{(N_2 + N_1 * (1 + N_2))^2}$$

Nash equilibrium quality levels lead to the following market shares:

[31] In a first approximation, the marginal cost of running an additional query can be taken to be zero. In any case, variable costs can always be subsumed into the p variable, which would then stand for net revenue per user. This formulation implicitly assumes that quality affects the fixed cost of production rather than the variable cost.
This is likely: variable costs mostly come from the huge energy requirements needed to run server farms, whereas quality is directly related to the work of engineers and software developers.

$$D_1 = 1 - \frac{2 * N_2}{N_2 + N_1 * (1 + N_2)}$$

$$D_2 = 1 - \frac{2 * N_1}{N_2 + N_1 * (1 + N_2)}$$

$$D_3 = 1 - \frac{N_1 * (1 - N_2) + N_2}{N_2 + N_1 * (1 + N_2)}$$

The figure below shows equilibrium quality levels and market shares as a function of N_1, the amount of private data to be searched by firm 1.

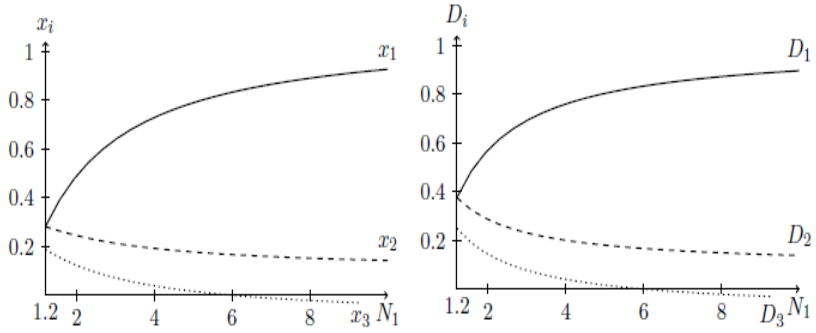

Fig. 10: Equilibrium quantity level and market share[32]

Although the authors model a simple one-shot game, N_i can be interpreted in a dynamic way: the market share of firm i in some period t influences the relative amount of private data that firm i has access to in period $t + 1$**.** Consequently, if firm 1 has access to more data than firm 2 and 3 at one point of time, the equilibrium predicts that this advantage increases over time.

This implies that the market is tipping, that is firm 1 is producing ever higher perceived quality and gains ever higher market share, whereas firms 2 and 3 decrease their respective quality levels and market shares over time. However, the survival of the weakest search engine, firm 3, is called into question as soon as the data access advantage of firm 1 or 2, expressed by N_1 and N_2 respectively, is sufficiently large. It is indeed costly to produce quality and advertising revenues, and thus market share, must be high enough to maintain positive profits.

[32] Source: "Search engine competition with network externalities", Argenton C., Prüfer J. (2012).

Therefore, the model predicts that firm 3 sooner or later exits the market, and the market structure turns from a triopoly into a duopoly.

Now it is possible to analyse the welfare effect of network externalities in the triopoly model.

Equilibrium profits are:

$$\pi_1 = \frac{(N_2 - N_1 * (1 + N_2))^2}{(N_2 + N_1 * (1 + N_2))^2} * p - F$$

$$\pi_2 = \frac{(N_2 + N_1 * (N_2 - 1))^2}{(N_2 + N_1 * (1 + N_2))^2} * p - F$$

$$\pi_3 = \frac{(N_2 + N_1 * (1 - N_2))^2}{(N_2 + N_1 * (1 + N_2))^2} * p - F$$

Producer surplus is the sum of all firms' profits:

$$PS = \pi_1 + \pi_2 + \pi_3$$

Expected consumer surplus is found by averaging equilibrium quality levels weighted with the market shares of active firms:

$$CS = D_1 * x_1^* + D_2 * x_2^* + D_3 * x_3^*$$

Summing up producer surplus and consumer surplus, we get total surplus:

$$W = \frac{(N_2 + N_1(3 * N_2 + 1)) * (3 * N_2^2 - 2 * N_1 N_2(1 + N_2) + N_1^2 * (3 + N_2(3N_2 - 2))}{(N_2 + N_1 * (1 + N_2))^2} * p - 3F$$

The figure above represents a graphical representation of firms' profits and total welfare as a function of N_1.

This figure shows that firm 1's profit increases in N_1, which implies that, due to the argumentation outlined above, its profit also increases over time. Instead, firm 2's and firm 3's profits decrease over time. Therefore, in the model it is very profitable to be the market leader and to maximize the advantage over competitors in accessing (private) data about past queries. Notably, consumer surplus and welfare are also increasing in N_1 and in N_2. This effect stems from the fact that network externalities decrease the cost of producing quality: more and more consumers enjoy the increasing quality of the market leader.

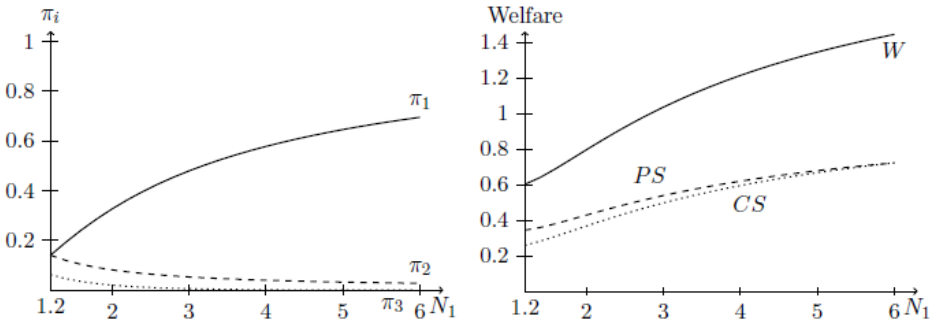

Fig. 11: Equilibrium profit and welfare[33]

Now, it is possible to consider that only 2 firms remain in the market:

$$max_{x_i}\ \pi_i = \frac{x_i}{x_1 + x_2} * p - \frac{x_i}{N_i} - F$$

For simplicity and without loss of generality, we set $N_2 \equiv 1$. Nash equilibrium in quality levels with positive market share is:

$$x_1^* = \frac{N_1^2}{(1 + N_1)^2}$$

$$x_2^* = \frac{N_1}{(1 + N_1)^2}$$

Which leads to the following market shares:

$$D_1^* = \frac{N_1}{(1 + N_1)}$$

$$D_2^* = \frac{1}{(1 + N_1)}$$

Also in the case of duopoly the market is tipping

$$\lim_{N_1 \to \infty} D_1 = 1 \lim_{N_1 \to \infty} D_1 = 0$$

In line with the ever decreasing market share of firm 2, for every $F > 0$, the profit of firm 2 turns negative if N_1 is sufficiently high. It follows that the model predicts that a duopoly, too, is not a stable market structure in the long run if the market leader retains its advantage regarding access to private data. In effect, the industry

[33] Source: "Search engine competition with network externalities", Argenton C., Prüfer J. (2012).

has the character of a natural monopoly. Therefore, it is possible to study equilibrium and welfare effects in the monopoly case.

The authors consider a contestable monopoly. The monopolist faces a potential entrant but may be able to prevent entry by making use of an appropriate “limit strategy".

First, the incumbent, firm 1, sets quality x_1. Second, the potential entrant, firm 2, decides whether to enter the market for a cost $K \geq 0$, or not. Third, if firm 2 entered the market, it sets quality x_2. Finally, consumers choose a search engine, as before. Given this structure, firm 1 can predict how firm 2 would behave, what quality it would produce, if it entered the market at the second stage. Therefore, firm 1 can calculate what entry would mean for its own profits and decide whether it wants to deter or accommodate it.

In this case, we have to solve a three stage sequential game by backward induction. At stage 3, given that firm 2 entered the market, it sets its quality, x_2, according to its duopoly reaction function:

$$x_2 = \sqrt{p * x_1} - x_1$$

Note that this only holds for $p \geq x_1$.

If $x_1 > p$, firm 2's best response is to set $x_2 = 0$. At stage 2, firm 2 enters the market if, and only if its expected profit from entry is positive. If we substitute x_2 into firm 2's duopoly objective function, assuming $N_2 = 1$, this requires:

$$\pi_2(x_1, F, K, p) = p - 2 * \sqrt{p * x_1} + x_1 - F - K > 0$$

If we solve the previous expression with the equality sign for zero shows that $\pi_2(x_1, F, K, p) \leq 0 \; \forall \; x_1 \in I$, where the interval I is defined as $I: [K + F - 2\sqrt{(K+F) * p} + p, K + F + 2 * \sqrt{(K+F) * p} + p]$. The upper bound of this interval, however, is larger than p, which hurts the condition above. It follows that the only condition that firm 1 has to meet such that even firm 2's best response

quality would leave it with a loss is $x_1 \geq K + F - 2\sqrt{(K+F) * p} + p$. Given that firm 1 has no intrinsic value in providing a higher quality than necessary, it follows that, if it wants to keep firm 2 out of the market, firm 1 sets equal x_1 to the lower bound of I. In other words, firm 1's "limit quality" is:

$$x_1^{lim} = K + F - 2\sqrt{(K+F) * p} + p$$

At stage 1, when determining its quality level, firm 1 effectively also chooses its competition because its quality determines whether firm 2 enters the market, or not. If it produces the limit quality, it keeps firm 2 out of the market.

If instead firm 1 sets x_1 such that firm 2 could possibly enter the market profitably, firm 2 will set its quality according to $x_2 = \sqrt{p * x_1} - x_1$. Substituting this in firm 1's unconstrained duopoly profit function shows that firm 1 solves:

$$max_{x_1} \ \pi_1 = \sqrt{p * x_1} - \frac{x_1}{N_1} - F$$

Which is solved by firm 1's Stackelberg quality:

$$x_1^{Stackelberg} = \frac{N_1^2 * p}{4}$$

If firm 2 enters, it will react to $x_1^{Stackelberg}$ by setting x_2 according $x_2 = \sqrt{p * x_1} - x_1$, to which gives:

$$x_2^{Stackelberg} = \frac{p}{4}(2 * N_1 - N_1^2)$$

Resulting market shares are:

$$D_1^{Stackelberg} = \frac{N_1}{2} \quad D_2^{Stackelberg} = 1 - \frac{N_1}{2}$$

Which implies that firm 2 gives up and sets $x_2 = 0$ as soon as $N_1 \geq 2$.

The key trade-off that arises for firm 1 is whether it wants to set the high quality x_1^{lim} and enjoy monopoly, or set low quality, $x_1^{Stackelberg}$, which saves on cost but does not necessarily foreclose the market.

The authors find that, if the advantage in private data access of firm 1 is sufficiently large, then firm 2 cannot enter the market and gain a positive market share even if firm 1 sets $x_1^{Stackelberg}$. Hence, in this case it is not even necessary for firm 1 to foreclose the market by making use of a limit quality strategy. In contrast, if firm 1's advantage in private data access is modest (for $N_1 < 2$), we find that firm 2's cost of market entry, K, becomes crucial. If K is rather low, firm 1 sets x_1^{lim}. In this case, the entrant knows that it would also have to set a rather high quality to compete with the incumbent. This high quality would be so costly for firm 2, however, that it could not profitably operate. It follows that firm 2 abstains from entering the market if the incumbent set the limit quality, x_1^{lim}. Finally, if K is high, firm 1's best choice is to set $x_1^{Stackelberg}$. This implies that firm 2 could indeed enter the market and gain a positive market share. However, it could not recoup the market entry cost with the modest operational profits this would bring. Understanding this, firm 2 would not enter the market in this case, either.

Reality check of the model

The model generates several implications. First, it predicts market tipping: the market share of the dominant firm is expected to increase more and more, whereas the market shares of the other firms are expected to decrease. Second, the model predicts substantial profit growth for the market leader and decreasing profits for the following firms. Third, as a consequence of the second point the model predicts market exit of one follower (Argenton C. and Prüfer J., 2012).

These predictions are well reflected in the history of the search engine market since 2003, when Google became the market leader. Google in 2003, in the US, has a market share of 35%, so it takes over the role of firm 1, Yahoo has the 32%, and Bing the 26%, they stand for firm 2 and firm 3.

Although starting from roughly the same market share as Yahoo in 2003, Google managed to increase its US market share up to nearly 70% by the end of 2010. In line with that success, Yahoo's and Microsoft's search engines reduce their market shares from a combined 50% in 2003 to 25% in 2010. The model's second prediction also corresponds to the empirically observed development of profits: while Google's profit rose by 36% to 2.51 billion dollars in the second quarter of 2011, Bing seems to lose money. Finally, the model's third prediction is reflected by the de facto market exit of Yahoo, which sold its search and advertising business to Microsoft in early 2010. Argenton and Prüfer (2012) conclude that, although the model is simple and stylized, it can reproduce different key developments that are well in line with the latest developments in the search engine industry.

Therefore, absent any other major changes, network externalities in the search engine market can be expected to drive out competitors of the dominant firm and to lead to a stable monopoly. Given the models, it is an open question how Google managed to become market leader and to outcompete both Yahoo and Microsoft by 2003. Likely explanations are that Google's search algorithm was of drastically higher quality than its competitors' at this stage or that query data mining was not yet in its mature phase. Probably, Yahoo and other search engines active before Google's market expansion were not able to exploit the network externalities present in the industry decisively. Since Google has taken over the pole position, however, we do not know whether it has increased its lead in market shares because its algorithm quality has become even better than its competitors' in relative terms or whether its success has mainly happened due to network externalities, in the absence of drastic innovation. Model suggests that the latter effect is sufficient to explain the market development throughout the last years. In order to reap the full benefits of competition in this highly innovative market, it is possible to intervene and reinstall merit-based competition (Argenton C. and Prüfer J., 2012).

6. Use of Data

Firms can use data in different ways, for example, they can use data to set personalised prices or address targeted advertising, to personalised services and create personalised recommendation.

6.1. Price discrimination

A firm price discriminates when it charges different consumers or different classes of customers different prices for the same product or similar products, and the price difference does not reflect cost difference. A firm price discriminates to extract as much as possible what the consumers are willing to pay for its products or service (Bourreau et al., 2017). Familiar forms include loyalty discount, volume or multi-buy discount, and the offering of status based discount for student, old-age pensioners and the unemployed.

Price discrimination is possible under some conditions: the firms should have some market power[34], and there is no or limited possibility of arbitrage or resale[35].

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.

Fig. 12: Types of discrimination and associated information requirements[36]

[34] Price discrimination is not possible under perfect competition.
[35] No arbitrage condition: the inability of any competitor, whether a low-price buyer also acting as a seller, or anyone else, to undercut discriminatory prices.
[36] Source: Source: "Big Data and Competition Policy: Market power, personalised pricing and advertising", Bourreau et al., 2017.

The economic literature defines three different types of price discrimination:

1. **First-degree price discrimination**: it occurs when a seller charges a different price to every buyer. First-degree discrimination has the highest requirements in term of information about consumers' preferences.
2. **Second-degree price discrimination or quantity discount**: it occurs when the pre-unit price falls with the amount purchased.
3. **Third-degree discrimination**: it occurs when sellers charge different prices to different consumers' group.

6.1.1. Price discrimination and Big Data

The advent of Big Data provides firms with more opportunities to obtain information on consumer preferences and their willingness to pay. This information may be public or private, and could be based both on the observed characteristics of the individual (e.g. their location, age, gender, employment status), or on their observed behaviour, (e.g. whether they have previously made a purchase), their responsiveness to previous price offers and their search histories. So, the technological capacities of Big Data substantially enhance the ability of digital retailers to engage in more precise, targeted and dynamic forms of price discrimination that were not previously possible. It allows for surveillance and tracking of online behaviour via the collection of individuals' 'digital breadcrumbs'[37], allowing firms to create detailed profiles of the tastes, habits and purchasing preferences of consumers at a highly personal and granular level. Secondly, Big Data allows personalised on-line strategic experimentation to acquire information concerning the preferences, behaviours and potential willingness to pay, of digital users. Finally, the online environment allows firms to extract the resulting data to personalise the informational choice environments of each user, including the digital 'shop floor' through which they make purchases.

[37] Digital breadcrumbs: they indicate the digital "crumbs" of information that we leave behind our passage when we use online services, interact on social networks, we use mobile devices, e-book readers, GPS, wearables, credit cards or loyalty etc. They also concern the so-called sensitive data, the data collected by security cameras, satellites and various types of sensors.

So, rather than products being displayed in a physical space offered at a single price to all customers, it is now technologically possible to offer the same goods to different consumers at different prices (Townley et al., 2017).

However, differential pricing, in the real word, still presents several practical challenges. First, sellers must figure out what consumers are willing to pay. This can be a complex problem, even for the companies with lots of data and computing power. A second challenge is competition, which limits a company's ability to raise the price, even if it knows that one customers might be willing to pay more than another. Third, companies need to prevent resale by customers seeking to exploit price differences. Finally, if a company does succeed in charging personalised prices, it must be careful not to alienate customers who may view this pricing tactic as inherently unfair (US Executive Office of the President, 2015).

6.1.2. Seller behaviour

In 2000, users discovered that Amazon.com was conducting price tests and complained about paying different price for the same DVD. Deleting the cookies that Amazon had left on their computer, or using a browser that did not accept cookies, some customers found they were getting much bigger discounts. The implication was that Amazon was offering bigger discounts to first time visitors to tempt them back. Amazon.com CEO promised that the company never would test prices based on customer demographics (Bourreau et al., 2017).

The fear of negative consumer reaction may explain why we have not yet entered an era of widespread personalised pricing. In any case, there are subtler 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 comparable, negative reaction from consumers seems less likely.

Second, a firm can engage in search discrimination or steering, that is the practise of showing different products to customers in different groups, based on the available information about consumers. For example, in 2012, the travel agency

OrbitzWorldwide discovered that Mac users showing, on average, a 30% higher spending propensity for a night at the hotel. The agency offered more expensive hotel to Mac users than to PC users (Bourreau et al., 2017). In the online environment, steering occurs when a web site alters its search results based on information about potential customer. Like third-degree discrimination, steering uses information about potential buyers, but not at the individual level.

6.1.3. Effects on welfare

Firstly, the analysis take into consideration a monopoly firm. With data, firm can improve its capacity to predict the willingness to pay of customers and increase its profit. When a firm sets personalised instead of uniform prices, a trade-off arises: some consumers with high willingness to pay can be worse off (appropriation effect), while some consumers with low willingness to pay can be better off (market expansion effect). The firm charges higher personalised price to consumers with high willingness to pay compared to uniform prices and charges lower personalised prices to consumer with low willingness to pay, and some consumers with low willingness to pay who could afford the good previously under uniform pricing can purchase it (Bourreau et al., 2017).

In general, personalised pricing increases consumer surplus if the demand expansion effect outweighs the appropriation effect. A necessary condition for a surplus increasing is that the total sold quantity increases.

As it is possible see in the following graph (fig.6), the half of the total area that lies under the downward-sloping demand line is the total welfare created by a particular good when all units are sold. When pricing uniformly, the seller will sell at the price represented by the dashed horizontal line and quantity represented by the vertical dashed line, earning profit equal to the square named “Monopoly Profit”. The DWL triangle is the deadweight loss, the value lost by the monopolist’s failure to sell units of the product in excess of those enclosed by the vertical dashed line. The consumer welfare is the value that the consumers who are able to buy at the uniform monopoly price enjoy in excess of what they pay. So, in the case of first-degree discrimination, the seller to capture both consumer welfare

and the deadweight loss and thereby to enjoy the entire value of the product (Ramsi W. A., 2016).

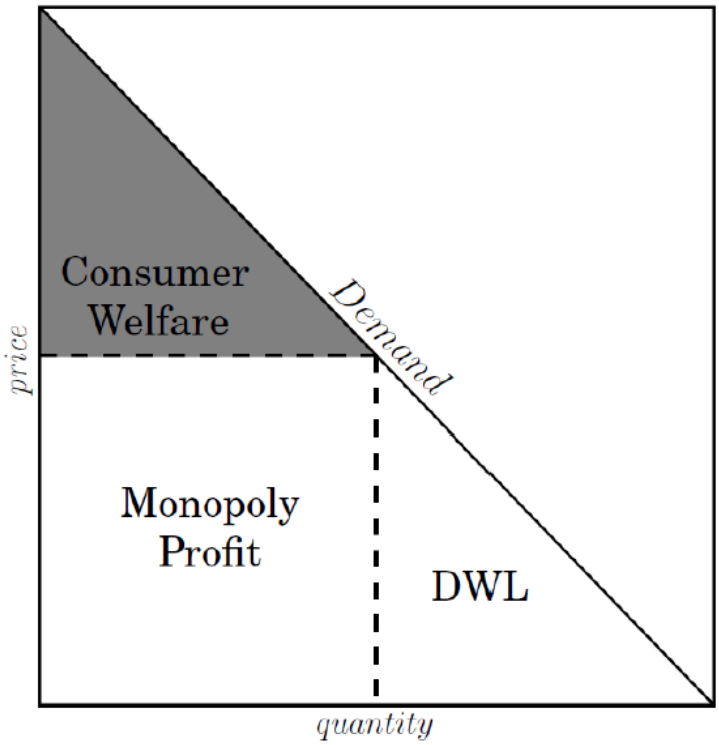

Fig. 13: demand graph for units of a particular good[38]

In a context of imperfect[39] competition, the relationship between information, price discrimination, and consumer surplus is considerably more complex. The economic literature says that, in these markets, price discrimination can either increase or reduce the intensity of price competition, depending upon the particular type of information used as a basis for offering personalised prices.

In monopoly markets, price discrimination is usually based on estimates of a consumers' willingness to pay. However, this form of price discrimination is not feasible where there is competition as consumers may simply switch to an alternative brand.

[38] Ramsi W. A. (2016), "Big Data, Price Discrimination, and Antitrust", Hasting Law Journal.

[39] In imperfect competitive market, there is rivalry between two or more firms, but individual firms may retain some degree of market power, i.e. an ability to raise price above marginal cost.

Price discrimination in imperfectly competitive markets has therefore to be conditioned on a different type of consumer information. Two main types may be distinguished:

1. The additional value a consumer receives from purchasing a good or service from one particular supplier over another (brand preference/ switching costs);
2. The cost incurred by consumers in comparing differing brands (search costs).

Access to both types of consumer information allows sellers the possibility of charging higher prices to those customers that are relative price inelastic[40]. This price insensitivity may arise for example where a customer has a strong preference for a particular suppliers' brand or where they would incur costs in switching to a rivals' brand (Townley et al., 2017).

Alternatively, price insensitivity may result from higher search costs. Consumers that do not search, for example, because they have high search costs or because they are impulsive, inattentive or otherwise economically unsophisticated purchasers, mark themselves out as being price insensitive. Accordingly, they may be charged a higher price than those that search more widely and make more considered purchases.

Although these two types of consumer information appear similar, the economic effects on competition and consumers can be very different: price discrimination based on brand preference tends to intensify competition (and increase aggregate consumer surplus), whereas price discrimination based on search costs tends to weaken competition (and reduce consumer surplus). The intuition underlying this is that with respect to brand preferences, each seller has a "strong" group of customers that prefer its brand, and a "weak" group of potential customers that prefers the brand of a rival seller. The strong customers for one seller are the weak

[40] Customer price inelastic: insensitive to changes in price

customers of its rival. Each seller will seek to charge higher prices to those that prefer its brand, while discounting heavily to attract the customers of rival brands. It is these targeted price cuts – aimed at winning customers from rival suppliers – that lead to an intensification of competition, lower average prices and higher aggregate consumer surplus compared to uniform pricing (Townley et al., 2017).

The opposite situation applies when price discrimination is based on information on customer search costs. Here all sellers want to charge low prices to ("weak") customers who search out the lowest price and high prices to "strong" consumers who have high search costs, those that do not shop around and those that make impulsive purchasing decisions. The "strong" customers are the same for both sellers. However, because these customers are price insensitive there is limited incentive to compete aggressively by offering lower prices. As a result, this type of price discrimination can undermine competition compared to uniform pricing.

In conclusion, in an imperfectly competitive market, competing firms do not generally gain from personalised pricing (Townley et al., 2017).

6.1.4. EU Regulation

The welfare effects of personalised pricing are ambiguous. In some case discrimination increases consumer surplus and the competition in comparison to uniform pricing, in others it reduce the concentration of the market.

In EU legislation, (Bourreau et al., 2017) there is not a large amount of rules on price discrimination. The Services Directive[41] prohibits discrimination based on the service recipient's nationality or residence. On this basis, the Commission open a probe against Disneyland Paris acting on the basis of allegations that cheap deals were only made available in for resident in France or Belgium. 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 no-discrimination.

[41] Services Directive: council on services in the internal market

The Commission (Bourreau et al., 2017) also proposed, in 2016, a regulation on geo-blocking. Geo-blocking occurs when 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 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 that is argued not to have sufficiently addressed discrimination of customers.

Price discrimination could be used as a monopolistic device. For example, an incumbent firm may pre-empty in a given market or consumer segment by setting very low prices in this market or segment. 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.

EU Commission, in these cases, usually applies Article 102 TFEU[42], which seek to ensure that particularly powerful firms do not abuse their market power within the EU:

"Any abuse by one or more undertakings of a dominant position within the internal market or in a substantial part of it shall be prohibited as incompatible with the internal market in so far as it may affect trade between Member States."

6.2. Targeted advertising

For many online firms, advertising is a major source of revenue. 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 to offer more targeted advertising[43] (Bourreau et al., 2017).

[42] TFEU: Treaty on the Functioning of the European Union

[43] Targeted online advertising refers to any form of online advertising that is based on information the advertiser has about the advertising recipient, such as demographics, current or past browsing or purchase behavior, information from preference surveys, and geographic information.

Online advertising involves a several players, including web publishers, advertisers and advertising network intermediaries who connect web publishers with advertisers seeking to reach an online audience.

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

Search advertising allows advertisers to target consumers based on their search queries. 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 and audio or video ads. Display advertising includes both non-targeted and targeted advertising.

Targeted display advertising, is becoming more and more important in the online advertising market.

Display advertising can take different forms:

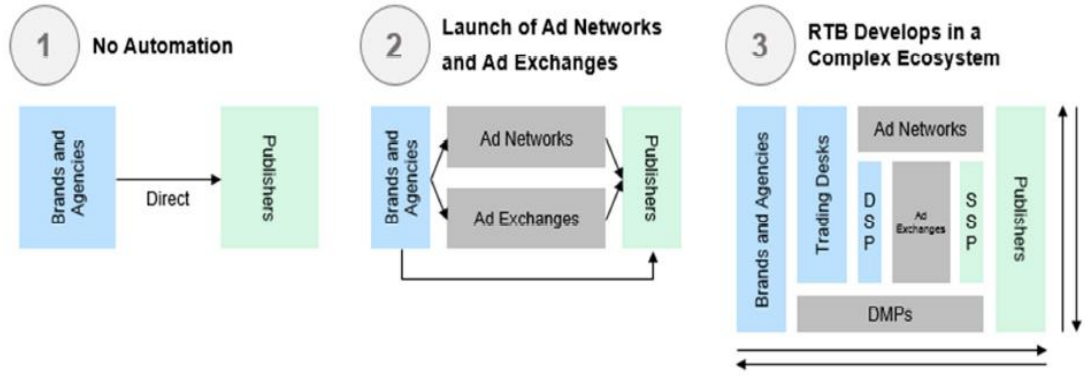

Fig. 14: Different type of target advertising

1. **No automation:** a publisher sells directly advertising space to advertisers. Publisher can uses the information that it has about its consumers to place its own targeted ads or those of third parties (Bourreau et al., 2017).
2. **Ad networks and ad exchanges:** the publisher contracts 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. Example include DoubleClick and AdMob (owned by Google). Ad network collect information about consumers' online behaviour, and can help advertisers target their ads for a specific audience. 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. An ad exchange is an automated platform that auctions advertising inventory from multiple ad networks (Bourreau et al., 2017).

3. **Real-time bidding:** 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, which help publishers sell their advertising space, and on the other side to demand platforms, which give access to buyers. Data management platforms sell data about users to the other players. When a user visits the websites of a publisher, the data management platform associated with the publisher identifies the user's profile through their cookies. Once the user connects to the server of the publisher, the latter sends the user profile to its supply-side platform, which in turn send the information to an ad exchange. The ad exchange provides the user profile to demand-side platforms. A two-stage auction follows. First, based on information provided by the data management platform on the user's profile, each demand-side platform starts an auction with its advertisers. Second, the winners at each demand-side platforms enter a second round auction at the ad exchange. The higher bidder of this second auction wins and their ads are then displayed when the website loads (Bourreau et al., 2017).

Targeted advertising takes place when firms place ads that target specific audience based on their estimated personal characteristics and interests.

Targeted advertising has two benefits for the advertisers relative to non-targeted advertising: targeted ads can be displayed only to the consumers potentially

interested in the product or service, in this way they reduce wasteful advertising, and the content of 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 click-through rate and the conversion rate of ads. While 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.

6.2.1. Market outcome

There some empirical evidence that targeted advertising ads are more effective than generic ads. Most of them concerns a specific from of targeting, called retargeting. A consumer visits a company's website, showing interest for the brand, later on, while visiting another website, the consumer is shown an ad from the first brand. In literature, it is possible divide between "dynamic retargeting", when an ad concern a specific product of a brand that the consumer may have looked at, and of "generic retargeting" when the ad is only generic, for example it show the brand's logo (Bourreau et al., 2017).

Because targeted advertising is more effective, it is expected to benefit advertisers and favour entry small niche advertisers-sellers, which would be excluded with regular ads. Targeted advertising may have differential effects: targeted advertising improved prospects of general outlets at the expense of tailored outlets. 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 (Bourreau et al., 2017).

6.2.2. 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. While consumers are more and better informed, they should benefit from targeted advertising.

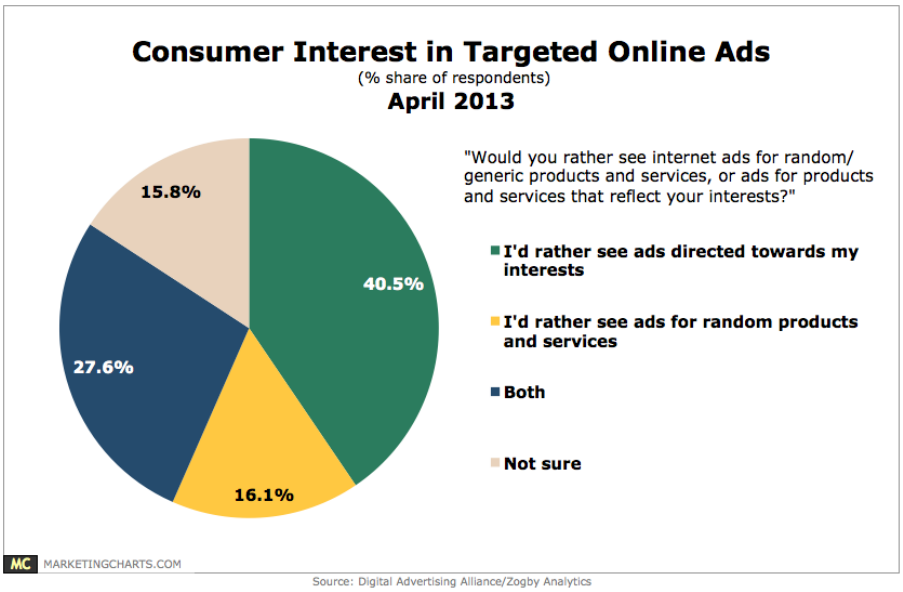

Fig. 15: chart of Zogby Analytics' study

A study, conducted by Zogby Analytics shows that 40.5% of respondents indicated that they prefer targeted to no-targeted advertisements, while less than 5% of respondents had an unfavourable opinion of behaviourally-targeted advertisements.

Furthermore, more effective targeted advertising reduces advertising costs for advertisers/sellers, which stimulates the demand for online advertising. Online sellers/publishers' revenues are expected to increase. These higher advertising revenues can be partially passed through to consumers in terms of lower subscription fees of higher quality service (Bourreau et al., 2017).

6.3. Personalised services and personalized recommendations

Web personalization is an automated process that identifies a user, collects his or her navigation patterns, analyzes known preferences of similar users, and estimates his or her specific preferences to tailor content for each user.

Personalization is automated and does not require the user's explicit input or control to generate individualized content. Many applications have incorporated aspects of web personalization. For instance, websites may place content relevant to each user's individual needs on their topmost page for easy navigation.

Personalized search engines are capable of capturing users' browsing histories and producing individualized search results. Among the various applications, product recommendation is the most widely used application of web personalization (Bodoff et al., 2014).

Recommendation tools play a crucial role in e-commerce. Online retailers offer considerably more products in their assortment than retailers in traditional sales channels. Huge assortments, however, are only beneficial for consumers if their search for appropriate products is supported by tools which help them to identify products that fit to their preferences (Hinz et al., 2010).

Recommender systems are used by e-commerce sites to suggest products to their customers and to provide consumers with information to help them decide which products to purchase.

Recommendation systems can have two different consequences for sales: first, decreasing search costs can lead to higher sales based on additional consumption; second, there can also be a shift in demand from blockbusters to niche products and vice versa, so that substitution effects can be observed. These two different consequences (additional consumption and substitution) are of high importance for online retailers: while additional consumption always leads to higher sales and potentially to higher profits, substitution is only advantageous if a product with a higher profit margin substitutes a low margin product. However, if providers know about margin differences between products, sales can systematically be shifted to products that are more profitable. Recommendation systems operate determine product recommendations that might be attractive to the specific consumers (Hinz et al., 2010).

According to Anderson (Hinz et al., 2010), just it is already said, recommendation systems shift demand from blockbusters away toward niches that better match consumer preferences. This effect of recommendation systems can be illustrated by the following example: Suppose that 90% of all consumers have a preference, for example, for the book "The Stand" by Stephen King, while 10% of all consumers prefer the book "Lost Symbol" by Dan Brown. If a recommendation system

correctly analyses that a consumer belongs to the smaller segment and would prefer the book “Lost Symbol”, then the system recommends the book “Lost Symbol” and the consumer will tend to buy this book. In the same situation without such an intelligent system the consumer would, based on probabilities, rather receive the recommendation to buy the book “The Stand” by Stephen King. Recommendation systems, therefore, lead to disproportionately reduced search costs for niches.

A distinction in these systems is made between the content-based approach, in which products with similar properties are proposed, and collaborative filtering, where similar consumers are searched and recommendations are made based on behavioural patterns.

Recommender systems enhance e-commerce sales in three ways:

Converting Browsers into Buyers: Visitors to a Web site often look over the site without purchasing anything. Recommender systems can help consumers find products they wish to purchase (Schafer et al., 2001).

Increasing Cross-sell: Recommender systems improve cross-sell by suggesting additional products for the customer to purchase. If the recommendations are good, the average order size should increase. For instance, a site might recommend additional products in the checkout process, based on those products already in the shopping cart (Schafer et al., 2001).

Building Loyalty: In a world where a site’s competitors are only a click or two away, gaining consumer loyalty is an essential business strategy. Recommender systems improve loyalty by creating a value-added relationship between the site and the customer. Sites invest in learning about their customers, use recommender systems to operationalize that learning, and present custom interfaces that match consumer needs. Consumers repay these sites by returning to the ones that best match their needs. The more a customer uses the recommendation system the more loyal he is to the site. Creating relationships between consumers can also increase loyalty, for consumers will return to the site that recommends people with whom they will like to interact (Schafer et al., 2001).

In addition, Service's personalization is the delivery of the right content to the right person at the right time to maximize immediate and future business opportunities.

Personalized services (Mela et al., 2003) attract customer attention and foster customer loyalty and lock-in. they reduces information overload, and highly relevant products yield satisfied customers. The customer loyalty that results from such personalization can translate into increased cash inflows and enhanced profitability.

Personalised services increase cross-selling[44], generated high levels of customer satisfaction, reduce transaction costs, and create a faster cycle times.

[44] Cross-selling: is the practice of selling an additional product or service to an existing customer

7. Big Data as a public good

Big tech platforms do not simply compete in a market. Increasingly, they are the market itself, providing the infrastructure for much of the digital economy. The platforms have become so dominant because they benefit from network effects. The Economist reported that Amazon captures over 40% of online shopping in America. With more than 2 billions monthly users, Facebook holds sway over the media industry. Firms cannot do without Google, which in some countries processes more than 90% of web searches. Facebook and Google control two-thirds of America's online ad revenues. Furthermore, users created large part of the technology and necessary data. The infrastructures of all tech companies were created collectively and they also feeds off network effects that are produced collectively.

In the public debate about Big Data some complain about mass information manipulation and the end of individuals' free will. Others fear that there will be a day when just a few firms, the "tech-giants", will dominate almost every market, thanks to their big data. Thus, to avert what would be the baleful end of economic pluralism, many hopes that these firms will be forced to share their big data and that antitrust law will help to achieve this result (Colangelo G. et al., 2017).

The Economist said that America's trustbusters have given tech giants the benefit of the doubt. They look for consumer harm, which is hard to establish when services are "free". The firms themselves stress that competition is just a click away and that they could be replaced by a new technology.

However, the barriers to entry are rising. Facebook not only owns the world's largest pool of personal data, but also its biggest "social graph"—the list of its members and how they are connected. Amazon has more pricing information than any other firm. Voice assistants, such as Amazon's Alexa and Google's Assistant, will give them even more control over how people experience the internet. China's tech firms have the heft to compete, but are not about to get unfettered access to Western consumers. If this trend runs its course, consumers will suffer as the

tech industry becomes less vibrant. Less money will go into startups, most good ideas will be bought up by the titans and, one way or another, the profits will be captured by the giants.

For this reasons, someone think that data should belong to all of users and data circulation should be viewed as a public good and data aggregators should become custodians of the public good. There is indeed no reason why the public's data should not be owned by a public repository that sells the data to the tech giants. The idea for a new way to think about privacy and data ownership also comes from the needs to encourage innovation and address the challenge of inequity in the digital age. Recognizing the potential of open data for fueling innovation and economic growth, many civic organizations and government officials, in the USA, have broadly called for "data for the public good.

However, for private-sector players, just it is already said, data is an important entry barrier to their businesses and one of most critical sources of competitiveness. Tech giants build their commercial prowess and financial valuations on the stickiness on their all-encompassing platforms and the huge amount of data these platforms yield. An important way for them to maintain a competitive and technological edge is to acquire the most competitive and innovative startups in the market. This threatens the future of tech innovation. With tech titans continuing to amass and hoard data through their platforms, private ownership and exploitation of data will continue to contribute to outsized financial valuations and skewed wealth distribution.

Therefore, from a policy point of view, recognizing the social and public good nature of data, it is possible to consider two general approaches to regulation. One is to ensure that data aggregators and owners serve the public interest by contributing a certain percentage of their annual income to public funds that can be used to support social and economic programs. The second approach suggested by Washington post, is to implement a system whereby private data aggregators may be granted a fixed term to exploit and extract value from data. After the term expires, all data should be made public.

8. Facebook/WhatsApp case

It is possible to analyse the case M.7217 WhatsApp/Facebook[45], to see how the European Commission evaluate a merger between two firms that possess data and how it assess the competition of the relevant markets. The transaction consisted of the acquisition of WhatsApp by Facebook for a purchase price of 19 billion of dollars. WhatsApp merged with and into wholly-owned subsidiaries of Facebook. As a result, Facebook solely control the entity into which WhatsApp have merged. The transaction contributed to Facebook's strategy of focusing its business on mobile development.

At first, we describe the firms involved. Facebook is a provider of websites and applications for mobile devices offering social networking, consumer communications and photo or video sharing functionalities. Zuckerberg's firm also provides online advertising space. In particular, it offers platform "Facebook", the consumer communications app "Facebook Messenger" and the photo and video sharing platform "Instagram".

WhatsApp is a provider of consumer communications services via the mobile app "WhatsApp". WhatsApp does not sell advertising space.

Secondly, to understand the Commission's decision, it is important to describe the relevant market in which the two firms operate and evaluate the competition before and after the transaction in these markets.

8.1. Consumer communications services

Consumer communications services are multimedia services communication solutions that enable people to reach out to their friends, family members and other contacts in real time. They can be offered as a stand-alone app or as a functionality that is a part of a broader offering such as a social networking platform. These services can be differentiated on the basis of various elements.

[45] European Commission (2017), "CASE M.8228 - FACEBOOK / WHATSAPP", http://ec.europa.eu/competition/mergers/cases/decisions/m8228_493_3.pdf

As regards Functionalities, consumer communications apps enable one to one and/or group real time communication in various forms, not all functionalities are available on all consumer communications apps.

Some consumer communications apps are available on only operating system (iMessage or FaceTime), while a large number of apps are offered for download on multiple operating system.

At last, some consumer communications apps are available for all type of device, while other are not: WhatsApp is only available on smartphone.

After the analysis, the Commission said that the segmentation that was most relevant for the assessment of this particular case was the segmentation based on platforms because WhatsApp was offered only for smartphones. Therefore, the Commission's assessment was done on the basis of a relevant product market including only consumer communications apps for smartphones. The Commission did not take into consideration the segmentation on the basis of operating system and functionalities because firstly, the majority of consumer communications apps were offered for download at least both Android and iOS which are the operating system installed on the greatest share of smartphones. Secondly, market investigation conduct by the Commission revealed that there were not functionalities that should be considered essential to a consumer communication services. The geographic market for consumer communications apps would be at least the European Economic Area (EEA). Market investigation revealed that no major differences exist in offering of consumer communications apps across the world. Therefore, all consumers with internet access are in principle free to download and install any app, irrespective of their geographic location anywhere in the world.

Now it is possible to show the competition assessment. The main drivers, defined by Commission, of the competitive interaction between consumer communications apps are the functionalities offered and the underlying network. First, consumer communications users have a broad range of choices when it comes to selecting and using consumer communications apps. Many of them use

more than one consumer communications app simultaneously depending on their specific needs (multi-homing). In this context, apps compete for customers by trying to offer the best communication experience. The functionalities offered are at the heart of the consumer communications apps' value proposition to customers and their improvement in order to gain the largest user base is a key innovation driver. In this regard, important areas of improvement include reliability of the communications service, which has a direct impact on the service's reputation and its appeal to users; and privacy and security, the importance of which varies from user to user but which are becoming increasingly valued. Second, a consumer communications service can offer utility to customers if the people they want to communicate with are also users of that service. Therefore, the relevance of the user base appears to be more important than its overall size. In this context, however, the size of the network of a consumer communications app can have a value for customers in two ways: a larger network implies that it is more likely that existing contacts will already be using a consumer communications app; and a larger network will afford greater opportunities for contact acquisition and discovery. Furthermore, perceived “trendiness” and "coolness" amongst groups of users is also an important factor in attracting new users and thus shaping the competitive landscape. Finally, price is one factor that influences the popularity of a consumer communications app. Indeed, the users of consumer communications apps tend to be very price-sensitive and expect a consumer communications app to be provided for free.

Facebook Messenger and WhatsApp differs in several elements:

1. The identifiers used to access the services (phone numbers for WhatsApp, Facebook ID for Facebook Messenger);
2. The source of the contacts (the user handset's address book for WhatsApp, all Facebook users in Facebook Messenger);
3. The user experience (which is richer in Facebook Messenger given the integration with the core aspects of Facebook social network);

4. The privacy policy (contrary to WhatsApp, Facebook Messenger enables Facebook to collect data regarding its users that it uses for the purposes of its advertising activities.

The only factors on the basis of which WhatsApp and Facebook Messenger were considered close competitors are the communications functionalities offered and the size of their respective networks. However, there is no feature offered by Facebook Messenger or WhatsApp that is not offered also by other market players. Moreover, there is a significant overlap exists between the networks of WhatsApp and Facebook which could rather point to a complementarity in the use of the two apps rather than to close competition. Furthermore, the EEA market for consumer communications apps features a significant degree of "multi-homing". In particular, WhatsApp and Facebook Messenger have been reported as being the two main consumer communications apps simultaneously used by the majority of the users in the EEA. This fact suggests that the two consumer communications apps are to some extent complementary, rather than being in direct competition with each other.

The consumer communications sector is a recent and fast-growing sector that is characterised by frequent market entry and short innovation cycles in which large market shares may turn out to be ephemeral. In such a dynamic context, high market shares are not necessarily indicative of market power and, therefore, of lasting damage to competition.

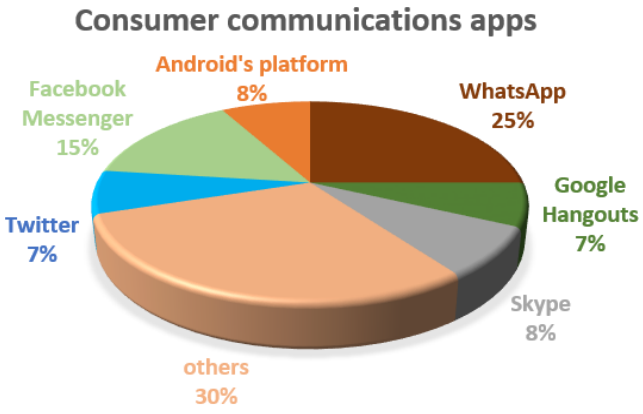

Fig. 16: Consumer communications apps market shares

Now it is possible to consider barrier to entry. There are no significant "traditional" barriers for a new consumer communications app to enter the market. First, as we have already said, the market for consumer communications apps is dynamic and fast-growing. In 2013, the use of messaging and social apps grew by 203%, more than any other type of apps. This growth is expected to continue in the future. New consumer communications apps are continuously offered for download to customers and are expected to be launched also in the future. Developing and launching a consumer communications app does not require a significant amount of time and investment. Second, there are no known patents, know-how or IPRs that would constitute barriers to entry, and the technologies implemented in consumer communications apps are increasingly standardised. Third, Facebook and WhatsApp do not have control over any element influencing entry. They do not have control over the operating system of smartphones, and are not therefore in the position to foreclose access to the final user of the consumer communications service. Secondly, email addresses, phone numbers and other elements that could be used as identifiers to access competing apps are ultimately controlled by the users. In the third place, the handsets' native address book for phone numbers or email addresses which could be used to build up a communication network is potentially available to all rival consumer communications apps providers.

In consumer communication apps market there are no significant costs preventing consumers from switching between different consumer communications apps. This happens for several reasons. First, all consumer communications apps are offered for free or at a very low price. Second, all consumer communications apps are easily downloadable on smartphones and can coexist on the same handset without taking much capacity. Third, once consumer communications apps are installed on a device, users can pass from one to another in instantaneously. Fourth, consumer communications apps are normally characterised by simple user interfaces so that learning costs of switching to a new app are minimal for consumers. Fifth, information about new apps is easily accessible given the ever-

increasing number of reviews of consumer communications apps on app stores. Telecom operators indicated, to the Commission, that switching costs for consumers would be represented by the loss of all data and interaction history when changing consumer communications app. In the present case, there are no evidences suggesting that data portability issues would constitute a significant barrier to consumers' switching in the case of consumer communications apps. Indeed, communication via apps tends to consist to a significant extent of short, spontaneous chats, which do not necessarily carry long-term value for consumers. Therefore, the transaction is unlikely give rise to an increase in switching costs.

However, a significant barrier to entry and expansion could be constituted by the presence of established players with a large user base and network effects in consumer communications apps. The existence of network effects as such does not a priori indicate a competition problem in the market affected by a merger. Such effects may however raise competition concerns in particular if they allow the merged entity to foreclose competitors and make more difficult for competing providers to expand their customer base. In this particular case, some factors mitigate the role of network effects in impeding entry or expansion. First, consumer communications apps, as we have already described, are a fast-moving sector, where customers' switching costs and barriers to entry/expansion are low. In this market, any leading market position even if assisted by network effects is unlikely to be incontestable. The market of consumer communications apps has a long track record of entry by new players. Also, competing consumer communications apps are able to grow despite network effects, both over time and following disruptions in the market. Such threat from new players constitutes and is likely to keep constituting a significant disciplining factor for the merged entity, regardless of the size of its network. Second, the use of one-consumer communications apps, just we have already said, does not exclude the use of competing consumer communications apps by the same user. Third, the merged firms do not control any essential parts of the network or any mobile operating system. Users of consumer communications apps are not locked-in to any

particular physical network, hardware solution or anything else that needs to be replaced in order to use competing products. Network effects could be strengthened if the transaction were to combine the separate user networks of WhatsApp and Facebook into one, substantially larger network. The integration between WhatsApp and Facebook would pose significant technical difficulties. This integration of two firms' networks would require matching WhatsApp users' profiles with their profiles on Facebook (or vice versa). This would be complicated without the users' involvement since Facebook and WhatsApp use different unique user identifiers. Furthermore, significant engineering hurdles would have to be overcome to enable cross-platform communications, reflecting the fundamentally different architecture of Facebook and WhatsApp. Even if some integration of WhatsApp with Facebook were to take place post-transaction, it would be reduced by the fact that there is already a significant overlap between the networks of WhatsApp and Facebook. Indeed, in the period between December 2013 and April 2014, 30-60% of WhatsApp users already used Facebook Messenger and between 70-90% of WhatsApp users were Facebook users and were therefore already within the reach of Facebook Messenger. Therefore, the net gain in terms of new members to the communications network would be much more limited than the addition of WhatsApp users to the Facebook user base would suggest.

8.2. Social networking services

Social networking services can be described as services, which enable users to connect, share, communicate and express themselves online, or through mobile app. Facebook social networking service consists of three core functionalities: user profile, newsfeed, and timeline. Facebook operates the world's largest social network that connects over 1.3 billion users worldwide and 200-300 million in the EEA.

User profile corresponds to user online identity, providing information on the user's jobs, school/university attended, relationship status, birthday, major life events, etc., as well as likes and interests (that is, music, movies, etc.). A user can

generally select to which degree the information in its profile is accessible to the public. Newsfeed is a regularly updating personalised display of stories from friends, pages, and other entities to which the user is connected. Timeline enables users to organise and display the events and activities that matter most to them (for example, interests, photos, education, etc.), such that they can curate their memories in a searchable personal narrative that is organised chronologically.

Social networking services can be segmented by platform, operating system or by intended use. Most social networks are accessible on multiple platforms and operating systems, so, they should not be further segmented according to a platform or an operating system. Commission market investigation, showed that a distinction could be drawn between social networking services promoting interpersonal contact for private and entertainment purposes (such as Facebook or Google+) and services which are used for professional purposes (such as LinkedIn). Nevertheless, there are overlaps between the purposes of intended use.

Because no competition concerns arise under any alternative market definition, the Commission left open whether social networking should be segmented according to the intended use. Furthermore, the Commission did not define the exact boundaries of the market for social networking services because this have no impact on the final decision.

The same social networking services is available throughout the world, or at least in most geographic regions, therefore, the Commission concludes that the geographic scope for the market for social networking services is at least EEA-wide, if not worldwide.

It is interesting to note that there is a certain overlap in the functionalities of consumer communications apps and social networking services. Both social networks and consumer communications apps enable users to exchange content with other people. However, social networking services tend to offer a richer social experience compared to consumer communications apps: the functionalities of consumer communications apps today are more limited and focus on enabling

basic communication between users rather than creating a richer experience around their digital identity. In addition, while both social networks and consumer communications apps enable communication between users, the communications functionalities and their usage differ. Hence, consumer communications apps facilitate instant real-time communication (with handsets ringing and notifications being pushed to recipients). Responses are generally sent promptly allowing a conversation. By contrast, messages in social networks, such as comments on a posting, are not normally expected to be responded to in real time. Moreover, social networks tend to enable communication and information sharing with a wider audience than consumer communications apps, which are more personal and targeted. For example, postings on a social network are generally shared with all contacts of a user (unless restricted), while communication on such consumer communications apps as Facebook Messenger and WhatsApp occurs mainly on a one-to-one basis. Because of considerable differences between functionalities, WhatsApp and Facebook are not close competitors in the potential market for social networking services.

If we consider integration between firms, it could take the form of cross-platform communication between WhatsApp and Facebook, enabling Facebook posts, status updates etc. to be delivered to WhatsApp, posting to Facebook from WhatsApp, or merging both services into one single platform. The integration of WhatsApp could strengthen Facebook's position in the potential market for social networking services by adding additional users and/or functionalities to the Facebook social network. Nevertheless, just we have already said, enabling cross-platform communication would necessitate substantial re-engineering of the services and re-writing of their code, given the differences in their architecture. Therefore, Facebook did not show any intention to proceed with the integration.

8.3. Online advertising services

Facebook's activities in the advertising sector consist of the provision of online non-search advertising services on Facebook's core social networking platform,

both on PCs and on mobile devices. By contrast, Facebook does not currently serve any ads on its Facebook Messenger app.

For the purpose of its online advertising activities, Facebook collects data regarding the users of its social networking platform and analyses them in order to serve advertisements on behalf of advertisers, which are as much as possible "targeted" at each particular user of its social networking platform. However, Facebook does neither sell any of the user data it collects nor provides data analytics services to advertisers or other third parties as a stand-alone product separate from the advertising space itself.

The Commission decided that online advertising constituted a relevant market separate from offline advertising.

The market investigation stated that advertisers typically purchase online advertising space and conduct advertising campaigns on a national basis, although global companies may also procure advertising space on a broader geographic scale (EEA-wide or even worldwide). Furthermore, prices for online advertising tend to differ depending on the country, based on a number of factors, such as demand and supply, local market conditions, Internet penetration rate, etc.

Therefore, the Commission concluded that the online advertising market and its possible sub-segments should be defined as national in scope or alongside linguistic borders within the EEA.

From the competition point of view, WhatsApp does not currently sell any form of advertising and does not store or collect data about its users that would be valuable for advertising purposes (for example, concerning age, verified name, gender, social group, activities, consuming habits or other characteristics). Moreover, messages that users send through WhatsApp are not stored in WhatsApp's servers, but only on the users' mobile devices or elected cloud.

Therefore, since WhatsApp does not currently collect any user data that are valuable for advertising purposes, the transaction does not increase the amount of data potentially available to Facebook for advertising purposes.

However, the transaction could nevertheless have the effect of strengthening Facebook's position in the online advertising market. The strength of this position could increase by introducing advertising on WhatsApp, and/or using WhatsApp as a potential source of user data for the purpose of improving the targeting of Facebook's advertising activities outside WhatsApp.

Firstly, WhatsApp would deviate from the "no ads" product strategy that it has followed so far, which may convinced certain users who feel that the ads disrupt their experience to switch to competing apps free of ads.

Moreover, even if the merged entity were to introduce advertising on WhatsApp, the transaction would only raise competition concerns if merger there were not to be a sufficient number of effective alternatives to Facebook for the purchase of online advertising space.

However, most of Facebook's advertising customers recognised the importance of advertising on Facebook, due to its large and highly engaged user base, its ad targeting opportunities and the generally high return on investment. Advertisers considered that there are a sufficient number of alternative providers of advertising services that compete with Facebook. These include Google, Yahoo!, MSN and local providers. More in general, customers did not raise any particular concerns with regard to the effect of the transaction on the online advertising market. Therefore, the Commission notes that, regardless of whether the merged entity will introduce advertising on WhatsApp, there will continue to be a sufficient number of other actual and potential competitors who are equally well placed as Facebook to offer targeted advertising.

Secondly, the integration would allow Facebook to have access to additional data from WhatsApp users to be monetised through advertising.

The collection of data from WhatsApp users who also have a Facebook account and use them for advertising on Facebook would require, first, a change in WhatsApp's privacy policy. Second, it would require Facebook, regardless of whether or not it would carry out some form of integration with WhatsApp, to match each user's WhatsApp profile with Facebook profile, provided she/he has

one. Third, some users could switch to different consumer communications apps that they perceive as less intrusive.

In any event, even if the merged entity were to start collecting and using data from WhatsApp users, the transaction would only raise competition concerns if the concentration of data within Facebook's control were to allow it to strengthen its position in advertising.

Post-Transaction will remain a sufficient number of alternative providers of online advertising services. In addition, there are currently a significant number of market participants that collect user data over Facebook. These include, first of all Google, which accounts for a significant portion of the Internet user data and, in addition, companies such as Apple, Amazon, eBay, Microsoft, AOL, Yahoo! among others. The graph below provides an overview of the estimated share of data collection across the web:

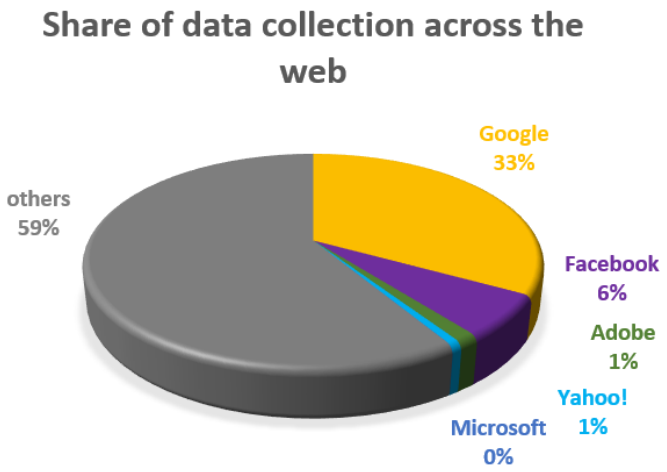

Fig. 17: Share of data collection across the web

Therefore, the Commission notes that, regardless of whether the merged entity will start using WhatsApp user data to improve targeted advertising on Facebook's social network, there will continue to be a large amount of Internet user data that are valuable for advertising purposes and that are not within Facebook's exclusive control.

8.4. Conclusion

It is possible to observe that the Commission analysed the three market involved. This analysis is done through document provided by Facebook and market

investigations. The Commission involved firms in the various markets and every person or firm that can help to describe the situation. After this analysis, the Commission is able to take the final decision.

For the above reasons, the European Commission has decided not to oppose the Transaction and to declare it compatible with the internal market.

In particular, the Commission considered that a large amount of internet user data that are valuable for online advertising are not within the exclusive control of Facebook. Therefore, even if Facebook would use WhatsApp as a potential source of user data to improve targeted advertising, this would not significantly impede competition on the online advertising market.

It is interested underlying that the Commission, on 20 December 2016, addressed a Statement of Objections to Facebook. This is happened because Facebook in 2014, when notified the acquisition of WhatsApp, informed the Commission that it would be unable to establish reliable automated matching between Facebook users' accounts and WhatsApp users' accounts. However, in August 2016, WhatsApp announced updates to its terms of service and privacy policy, including the possibility of linking WhatsApp users' phone numbers with Facebook users' identities. Therefore, contrary to Facebook's statements in the 2014 merger review process, the technical possibility of automatically matching Facebook and WhatsApp users' identities already existed in 2014, and that Facebook staff were aware of such a possibility.

This decision had no impact on the Commission's October 2014 decision to authorise the transaction under the EU Merger Regulation. Indeed, the clearance decision was based on a number of elements going beyond automated user matching. The Commission at the time also carried out an 'even if' assessment that assumed user matching as a possibility.

According to the Merger Regulation, the Commission imposed fines of up to 1% of the aggregated turnover of companies. The penalty is estimated approximately 110 million of euros.

9. Recent events

Now we describe briefly two events happen in the last months that involved big firms that collect and use data every day.

9.1. Google/Mastercard

The USA agency Bloomberg, on 30 August 2018, reported that Google and Mastercard made a secret agreement to obtain data about the shopping offline. Google paid millions of dollars for offline shopping data of Mastercard costumers. In this way, the Mountain View firm could realize a more sophisticated monitoring tool for their advertisers. Select Google advertisers have had access to this tool to track whether the ads they ran online led to a sale at a physical store in the USA.

Internal sources to Google said that the firm, before the launch of the new tool, built a new, double-blind encryption technology that prevents both Google and its partners from viewing their respective users' personally identifiable information. Therefore, Google does not have access to any personal information from its partners' credit and debit cards, and it does not share any personal information with partners. Furthermore, people could disable monitoring using Google's "Web and App Activity" online console.

Google announced the new service in 2017, called "Store Sales Measurement," and the company said it had access to approximately 70 percent of U.S. credit and debit cards through partners, without naming them.

This 70 percent could mean that the company has deals with other credit card companies, totalling 70 percent of the people who use credit and debit cards. Alternatively, it could mean that the company has deals with companies that include all card users, and 70 percent of those are logged into Google accounts like Gmail when they click on a Google search ad.

"Store Sales Measurement" had two components. The first lets companies with personal information on consumers, like encrypted email addresses, upload those into Google's system and synchronize ad buys with offline sales. The second injects card data.

It works like this: a person searches for "red lipstick" on Google, clicks on an ad, surfs the web but does not buy anything. Later, she walks into a store and buys red lipstick with her Mastercard. The advertiser who ran the ad is fed a report from Google, listing the sale along with other transactions in a column that reads "Offline Revenue". This happens only if the web surfer make a log in into a Google account online and made the purchase within 30 days of clicking the ad. The advertisers are given a bulk report with the percentage of shoppers who clicked or viewed an ad then made a relevant purchase. However, Mastercard does not view data on the individual items purchased inside stores. It is not an exact match, but it is the most powerful tool Google, the world's largest ad seller, has offered for shopping in the real world. Google knows that people clicked on ads and can now tell advertisers that this activity led to actual store sales.

Google is testing the data service with a small group of advertisers in the USA. With it, marketers see aggregate sales figures and estimates of how many they can attribute to Google ads, how much they spend or what exactly they buy. However, they do not see a shoppers' personal information. The tests are only available for retailers, not the companies that make the items sold inside stores. Early signs indicate that the deal with Mastercard has been a vantage for Google. The new feature also plugs transaction data into advertiser systems as soon as they occur, fixing the lag that existed previously and letting Google slot in better-performing ads. Beforehand, a firm received 5.70 dollar in revenue for every dollar spent on marketing in the ad campaign with Google, according to an iProspect analysis. With the new transaction feature, the return nearly doubled to 10.60 dollar.

It is important to say that most of the two billion Mastercard holders are not aware of this tracking. That is because the companies never told the public about the arrangement. The Electronic Privacy Information Center (EPIC) submitted a complaint about the sales measuring tack to the U.S. Federal Trade Commission[46].

[46] Federal Trade Commission: is the federal agency with both consumer protection and competition jurisdiction in broad sectors of the economy.

"People don't expect what they buy physically in a store to be linked to what they are buying online," said Christine Bannan, counsel with the advocacy group EPIC. "There's just far too much burden that companies place on consumers and not enough responsibility being taken by companies to inform users what they're doing and what rights they have."[47]

It is important to note that Google with this agreement want to fortify its primary business, the business of advertising. Indeed, this business, which hit 95.4 billion of dollars in 2017 sales, has maintained a growth rate of about 20 percent a year. However, many advertisers today are starting to purchase ads on Amazon, the company that hosts far more, and more granular, data on online shopping. Therefore, in response, Google has continued to push deeper into offline measurements.

9.2. Amazon

European Commission, on 19 September 2018, announced that it has opened a preliminary antitrust investigation into Amazon over the e-commerce firm's treatment of smaller merchants on its web site. The Commission want examine how Amazon uses data it gather through transactions, including those involving rival sellers, on its platform. Amazon sells its own product, and it, also, allow to third party retailers to sell their goods through it marketplace.

The most relevant thing that the commission has to determine is whether data is collected for legitimate purposes, like improve the service for the merchants, or for give Amazon's products a competitive advantage over the smaller merchants. Amazon can use data to understand what people want and the new tendencies and what the thing that makes people to buy is, and still, which kind of offer customers want to receive.

It is important to note that, last year more than 50% of products sold on the platform came from third party merchants. Services to third party generated 31.88

[47] https://www.bloomberg.com/news/articles/2018-08-30/google-and-mastercard-cut-a-secret-ad-deal-to-track-retail-sales

billion of dollars of revenues. This could represent an important information for the investigation.

9.3. Conclusion

This two recent events show how the authorities are worried that the big tech companies are becoming too dominant. Indeed the cited firms have strong positions in their own markets and they are able to exploit the advantage that they have created during the time. Google and Amazon has be able to understand the two sided market mechanisms and to exploit them; furthermore, today they make the competition on data to strengthen their dominant positions.

Google's news regard privacy concerns. Today the public opinion is too interested about these problems because they hit consumers that do not know or are not conscious about the treatment of their data.

Amazon's events consider the problem of abuse of dominant position. A company can restrict competition if it is in a position of strength on a given market. A dominant position is not in itself anti-competitive, but if the company exploits this position to eliminate competition, it is considered to have abused it.

Examples can include:

1. Charging unreasonably high prices;
2. Depriving smaller competitors of customers by selling at artificially low prices they can't compete with;
3. Obstructing competitors in the market (or in another related market) by forcing consumers to buy a product which is artificially related to a more popular, in-demand product;
4. Refusing to deal with certain customers or offering special discounts to customers who buy all or most of their supplies from the dominant company;
5. Making the sale of one product conditional on the sale of another product.

The Commission will understand if there is an abuse of dominant position only through a depth study on document provided by the firms and market investigations.

10. Conclusion

In literature, there are a lot of articles and reports that studies big Data and Big Data markets, experts try to give a definition to a new production input and they search to describe how markets are influenced by it and how they use it.

At the end of this work, I can say that Big Data world has singular characteristic. The presence of data network effect in markets, where data are relevant, is an element that reduce competition and create advantages for the big tech companies. In the Mitomo opinion, despite there are some competition's possibilities, it seems difficult to enhance competition in dominant platform business even where a potential entrant seeks to start a competitive service as far as there is no substantial product differentiation.

An important element is the particular structure of platforms: in those the presence of indirect network effect enhance by data, as I have presented in the Argenton and Prüfer model, tend to establish a monopoly structure the markets where data represent a fundamental element to improve quality of services and products.

The real cases have demonstrated how tech giants try to strength their market power increasing their network or their access to new data. In this way, they try to create string barriers to entry and to contrast other firm that try to damage their position.

The authorities open investigations against these firms to increases competition. Indeed the big firms tent to abuse of their dominant position to increase their market power. The authorities want also to protect people that ignore the power of their data. Indeed, as already said, concerns. Today the public opinion is too interested about privacy concerns problems because they hit consumers that do not know or are not conscious about the treatment of their data.

Bibliography

Ansari A., Mela C. F. (2003), “E-customization”, Journal of Marketing Research.

Argenton C., Prüfer J. (2012), “Search engine competition with network externalities”, Journal of Competition Law and Economics.

Autorità per le garanzie nelle comunicazioni (2018), “Big data. Interim report nell’ambito dell’indagine conoscitiva di cui alla delibera n. 217/17/CONS”.

Benlian A. (2015), “Web personalization cues and their differential effects on user assessments of website value”. Journal of Management Information Systems.

Bodoff D., Ying Ho S. (2014), “The effects of web personalization on user attitude and behavior: an integration of the elaboration likelihood model and consumer search theory”, MIS Quarterly.

Bourreau M., de Streel A. and Graef I. (2017), “Big Data and Competition Policy: Market power personalized pricing and advertising”, CERRE Project Report.

Colangelo G., Maggiolino M. (2017), “Big data as misleading facilities”, European Competition Journal.

Eisenmann T., Parker G. and Van Alstyne M. W. (2006), “Strategies for Two-Sided Markets”, Harvard Business Review, www.hbr.org.

European Commission (2014), “Case No COMP/M.7217 - FACEBOOK/ WHATSAPP”, http://ec.europa.eu/competition/mergers/cases/decisions/m7217

_20141003_20310_3962132_EN.pdf

European Commission (2017), “CASE M.8228 - FACEBOOK / WHATSAPP”, http://ec.europa.eu/competition/mergers/cases/decisions/m8228_493_3.pdf

EUROPE ECONOMICS (2017), “Big Data: what does it really mean for competition policy”, http://www.europe-economics.com/publications/mar_-_big_data.pdf.

Hinz O., Eckert, J. (2010), “The impact of search and recommendation systems on sales in electronic commerce”, Business & Information Systems Engineering.

Maffè Carnevale C. A. and Ruffoni G. (2009), “Two-sided markets: models and business cases”, http://didattica.unibocconi.it/

Mitomo H. (2017), “Data Network Effects: Implications for data Business”, https://www.econstor.eu/bitstream/10419/169484/1/Mitomo.pdf.

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

Park F. K., Seamans R. and Zhu F. (2017), “Multi-homing and Platform Strategies: Historical Evidence from the U.S. Newspaper Industry”, https://www.hbs.edu/faculty/Publication%20Files/18-032_4eb19781-7eb0-4da8-b345-9cb5294e8d1e.pdf.

Peppers D., Rogers M., Dorf B. (1999), “Is your company ready for one-to-one marketing”, Harvard Business Review.

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

Raguseo E. (2017), “Big data technologies: An empirical investigation on their adoption, benefits and risk for companies”, International Journal of Information Management.

Ramsi W. A. (2016), “Big Data, Price Discrimination, and Antitrust”, Hasting Law Journal.

Roson R. (2004), “Two-Sided Markets”, http://venus.unive.it/roson /papers/fusion2.pdf

Rochet J. and Tirole J. (2003), “Platform competition in two-sided markets”, Journal of the European Economic Association.

Rochet J. and Tirole J. (2006), “Two-Sided Markets: A Progress Report”, The RAND Journal of Economics.

Schafer J. B., Konstan J. A., Riedl J. (2001), “E-commerce recommendation applications”, Data Mining and Knowledge Discovery.

Sokol D. and Comerford R. (2016), “Does Antitrust Have A Role to Play in Regulating Big Data?”, https://ssrn.com/abstract=2723693

Tam K. Y., Ho S. Y. (2006), “Understanding the impact of web personalization on user information processing and decision outcomes”, MIS Quarterly

Townley C., Morrison E., Yeung K. (2017), “Big Data and Personalised Price Discrimination in EU Competition Law”, King’s College London

Turck M. (2016), “The Power of Data Networks Effects”, http://mattturck.com/the-power-of-data-network-effects/.

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

US Executive Office of the President (2015), “Big Data and differential pricing”, https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/docs/Big_Data_Report_Nonembargo_v2.pdf

Sitography

https://www.marketingcharts.com/digital-28825

http://bruegel.org/2016/11/search-engines-big-data-and-network-effects/

https://www.washingtonpost.com/news/theworldpost/wp/2018/05/02/big-data/?noredirect=on&utm_term=.676b8664073a

https://www.economist.com/leaders/2018/01/18/how-to-tame-the-tech-titans

https://www.bloomberg.com/news/articles/2018-08-30/google-and-mastercard-cut-a-secret-ad-deal-to-track-retail-sales

https://www.ftc.gov/about-ftc

http://ec.europa.eu/competition/consumers/abuse_en.html

https://www.nytimes.com/aponline/2018/09/19/world/europe/ap-eu-amazon-eu-antitrust.html

https://www.ilsole24ore.com/art/mondo/2018-09-19/amazon-ue-indaga-sull-uso-dati-piccoli-venditori-170446.shtml?uuid=AEi0ox1F

http://ricerca.repubblica.it/repubblica/archivio/repubblica/2018/09/20/leuropa-punta-i-fari-sui-dati-di-amazon33.html

https://www.nfx.com/post/network-effects-manual

RINGRAZIAMENTI

Giunti al termine di questi 5 anni è doveroso fermarsi un attimo a ringraziare coloro che hanno reso possibile questo percorso.

In primo luogo, voglio ringraziare il Professor Carlo Cambini che con pazienza mi ha guidato nella stesura della tesi finale.

Voglio poi ringraziare i miei genitori che hanno reso possibile tutto ciò, sostenendomi sia economicamente sia fornendomi tutto il supporto morale di cui avevo bisogno. Loro sono riusciti a farmi reagire nei momenti di sconforto e a darmi la carica per proseguire al meglio e soprattutto sono stati i primi credere in me e nelle mie capacità.

Un grazie particolare va alla mia gemella, che nonostante la distanza che spesso ci separa è stata sempre presente e pronta a spronarmi e a convincermi a dare di più, e a Simone che dopo dieci anni ha deciso di diventare una presenza fissa nella mia vita, lo ringrazio per aver sopportato le mie ansie e di essere stato sempre pronto a strapparmi un sorriso.

Un ringraziamento va a tutte le persone che ci sono state per una parola di conforto o per una chiacchierata di un'ora. Tra queste voglio in particolare dire grazie ad Alice e Francesca, due persone che nonostante la lontananza ci sono sempre state non abbandonandomi mai, a Giuseppe, conosciuto per caso tra il freddo norvegese, che è una presenza fissa nonostante ci dividano più di 600 chilometri, a Monica, che nonostante la divisione degli ultimi due anni ci è sempre stata e devo dire che è grazie a lei se oggi 24 ottobre mi laureo, e a Valentina, amica presente da una vita e sempre pronta ad aiutarti.

Infine ringrazio gli amici del "Little Italy" con cui ho condiviso un'esperienza di vita unica, e gli amici del "Baia" che riescono sempre a rallegrare le mie giornate estive e a essere presenti anche da Firenze.