POLITECNICO DI TORINO

Corso di Laurea in Ingegneria Gestionale

Tesi di Laurea Magistrale

Big Data, Competition and Privacy

Relatore:

prof. Cambini Carlo

Correlatore:

prof.ssa Abrardi Laura

Candidato:

Sgaramella Antonella

Ottobre 2019

Index

Introduction 6

1. A general overview of the online ecosystem 8

1.1 Big Data 9

1.2 The Value of Personal Data 10

1.2.1 Market Valuation 12

1.2.2 Individuals' Valuation 14

1.3 Use of Data 17

1.3.1 Target Advertising 17

1.3.2 Price Discrimination 18

1.3.3 Other Forms of Discrimination 20

1.4 Digital Platforms 21

1.4.1 Common Features 22

1.4.2 Types of Platforms 24

1.4.3 Market outcomes of the Economy of Platforms 26

1.5 Third Parties 28

1.5.1 Third-Party Tracking 28

1.5.2 Third-Party Market Power 29

2. A critical review of the literature 32

2.1 Natural monopolies and market failures 32

2.1.1 Definitions of Natural Monopoly 33

2.1.2 The Social Costs of Monopoly 34

2.2 Taxation of Digital Monopoly Platforms 36

2.2.1 Problems of The Traditional Taxation System 36

2.2.2 Taxation on Data Collection 38

2.2.3 Taxation on Advertising 39

2.3 Data Ownership 40
2.3.1 Data Ownership Fragmentation 41
2.3.2 Proposal Solutions in Literature 43
2.4 Data Portability 46
2.4.1 The Theoretical Literature on Switching Costs 47
2.4.2 Multi-sides Market Implications of Data Portability 50
2.4.3 An Economic Analysis of Data Portability 52
2.5 Data Access 54
2.5.1 Wider Access in Consumers' Perspective 55
2.5.2 Pitfalls in Data Markets 58
2.6 Data Trade 59
2.6.1 Information Selling in Financial Markets 61
2.6.2 Pricing of Information 63
2.6.3 Structure of Information 65
3. A Model of Competition with Endogenous Data Extraction 67
3.1 Linear prices 69
3.1.1 Asymmetric Information 69
3.1.2 Symmetric Information 71
3.2 Two-part tariffs 73
3.2.1 Asymmetric Information 74
3.2.2 Symmetric Information 76
Conclusions 79
Bibliography 83

Introduction

Nowadays, everyone is talking about Big Data. Every day individuals engage in transactions involving their personal information. New technologies are aggregating and collecting more data than ever before. Businesses are aware they can use the vast amounts of data that they already have within their organizations to achieve better performances. Lower costs, efficiency improvements, and increased sales are direct consequences of a deeper understanding of the customer journey and the augmented customer experience. According to a MIT Sloan Management Review partnered with IBM Institute for Business Value[1], top-performing companies use data analytics five times more than the others in their industry. Thus, senior managers are looking for better ways to extract full value from their information abut consumers and obtain a greater competitive advantage.

The world's most valuable resource is no longer oil, but data. Google, Amazon, Facebook, and Apple (GAFA, for short) are the most profitable companies in the global economy and their market power is based on the vast amount of information collected. Moreover, most of these digital-economy giants and many high-growth start-ups are multisided platforms. Thereby, the attention is focused on multi-sided markets and their inner characteristics as the presence of multiple actors, network externalities, and cross subsidisation. A data-driven digital platform economy is emerging, creating a new unique competitive scenario among platforms where ability to build scale is essential and concerning for a fair competition. The larger platforms are, the more valuable they tend to be for their users. In such a scenario, individuals benefit from disclosing their personal information in the form of target advertising, personalized prices or reduced searching costs. However, such disclosure implies a privacy cost suffered by consumers increasingly concerned about their privacy and locked-in in platform ecosystems. Hence, the phenomenon of big data and digital platform raise interdisciplinary concerns ranging from competition to consumer protection and privacy.

After a general overview of the online ecosystem offered in the first chapter, the thesis aims to conduct a critical literature analysis over the use of data made by dominant platforms, while creating linkages with traditional branches of economic literature. Monopoly, perfect competition, switching costs, entry barriers, commons and anti-commons literature, models of

[1] For instance, see Lavalle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big Data, Analytics and the Path from Insights to Value. *MIT Sloan Management Review, Vol 52. No. 2*, 21-31.

data trade have been discussed in the second chapter. Finally, once reviewed the traditional literature on data competition, we have tried to develop a two-stage game with an endogenous extraction of data. Indeed, the traditional models typically find that the monopolist prefers to sell data to one retailer only, so as to maximize the value that it can extract from the downstream market. While this result is concerning from a policymaking point of view, as it practically reinforces the monopolistic trends of vertically related markets, it strongly depends on the assumption made in traditional textbooks that data are an indivisible amount. Thus, in our model, we have made the quantity of data purchased affecting the price paid for it or the cost for its extraction. Therefore, new interesting insights have been obtained.

1. A general overview of the online ecosystem

The rise of the so-called Web 2.0 (blogs, social media, online social networks, search engines) has emphasized user-generated contents, participatory culture, and interoperability in the virtual community. In contrast to the first generation of Web 1.0, where people were limited to passively viewing contents, the role of end-users has changed. Individuals are no more only consumers of information, but public producers of high volumes of personal data. In this data-driven economy, we are observing the emergence of dataopolies: companies that control a key platform, which attracts users, sellers, advertisers, software developers, apps and accessory makers to its ecosystem. These major companies in the digital economy are as of today the most profitable firms in the global economy: Google, Apple, Facebook, and Amazon (or GAFA for short). Their market power is based on the vast amounts of collected information, which have substantial economic value as they can either sell data or use it to enable advertisers to target and personalize ads.

Individuals engage daily in transactions involving their data by using social networks, search engines, e-commerce website, web browsers and they can directly benefit from sharing their data. It could be in the form of having contacts with other peers via an online social network; or in the form of coupons, discounts, or personalized services after joining a merchant's loyalty program; or in the form of reduced search costs and increased accuracy of information displayed one experiences when a search engine tracks them. However, all those benefits have a privacy cost of disclosing personal information. With the advent of data analytics and the increasingly sophisticated capabilities to obtain value from data, public concerns over personal privacy are increasing. Thus, the difficult trade-off associated with protecting or sharing personal information is nowadays a crucial theme of economic and policy debate.

Social media are creating a culture of disclosure of one's activity, location, emotions, work history, and political opinions. The spread of IoT devices has blurred the distinctions between digital and physical, online and offline. Consumers are welcoming these new devices and services into their homes and daily lives, which provide benefits through increased connectivity. However, these trends are expanding the potential for data collection and use by digital platforms with increasing risks to the privacy and autonomy of users.

As all these developments will have potentially significant implications for society, therefore governments must have a role in monitoring them and considering how they can be best managed through public policy and, when required, regulations.

1.1 Big Data

Every day data is recorded, stored, and analyzed but what exactly is big data and what is the difference between regular data? The answer is in the 4Vs:

- *Volume* of data processed, which are frequently larger than terabytes and petabytes. Thus, they require advanced processing technologies than traditional capabilities;
- *Variety* of data aggregated, which come from different sources from traffic patterns and video downloads to web history and medical records. Moreover, they are of different types ranging from structured, to semi-structured or unstructured;
- *Velocity* at which data is generated and collected;
- *Value* of the information found in data made possible by technological progress. Indeed, companies can leverage data to adapt their products and services to better meet costumers' needs, optimize operations and infrastructure, and find new sources of revenue.

Defined the main characteristics, it is important to clear the different typologies of data as different legal constraints apply to their collection and use. UE data protection legislation's main goal is to protect the fundamental rights of natural people and in particular their right to privacy and protection of personal data. To achieve this objective, data are divided into different categories related to their protection risk.

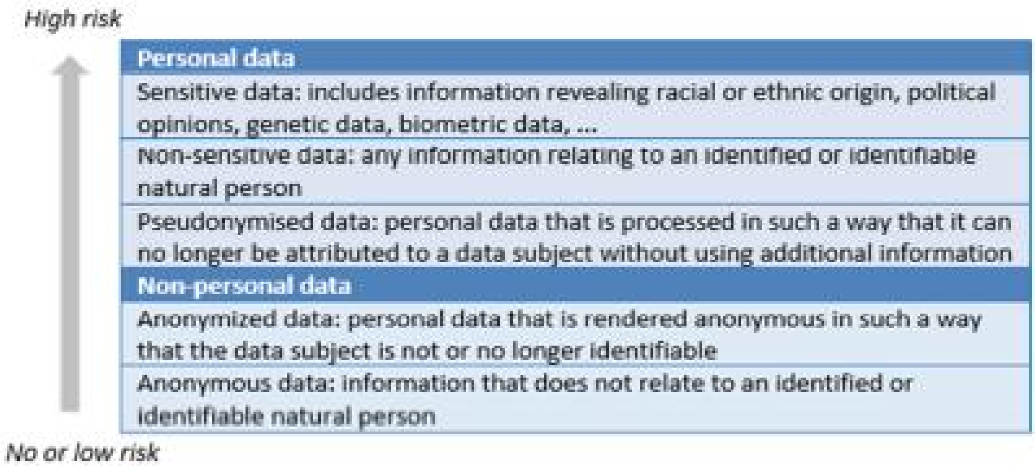

Figure 1: Typology of data with the associated data protection risk

As shown in Fig.1 the most sensitive type is represented by personal data defined as *"any information relating to an identified or identifiable natural person (data subject)"*. [2]Thus,

[2] Article 4(1) of the General Data Protection Regulation

the notion includes directly identifying natural persons such as names and contact details (ethnic origins, ages, genres, political opinions) but also, IP addresses as they can be linked to one individual. However, a distinction can be made between different subcategories of personal data. Stricter rules applied to the processing of personal data referred to as **sensitive data** which includes information about racial origins, political opinions, religious beliefs, biometric data, health data or data concerning sexual orientation. **Pseudonymized data** is personal data processed in such a way that can be attributed to an individual only with the use of additional information. Since data that has undergone this process of pseudonymization may still identify a natural person, the data protection rules fully apply to this category.

The second macro group is represented by non-personal data, information that is not under the data protection law. A distinction can be made between anonymous and anonymized data. **Anonymous data** is information which does not relate to an identified or identifiable natural person taking into account of all the means likely reasonably to be used. **Anonymized data** is personal data that is made anonymous in such a manner that data subject is no longer identifiable. It is excluded from the scope of application of the UE data protection rules if the process of anonymization is irreversible.

1.2 The Value of Personal Data

Data is considered to be the gold of the digital era, business models that rely on personal data as a key input are increasingly common. The rapid emergence of personal data as an asset in business processes justify the numerous attempts to measure and to estimate the monetary value of personal data.

Mapping the value chain can help identify the roles played by different actors and to point out the main stages in which monetary value is added. According to the OECD report,[3]the personal data lifecycle can be presented as a four-step value chain: data collection, storage and aggregation, analysis and distribution, and usage.

Collection/ access

Personal data is collected in a variety of ways. Data can be volunteered entered by individuals who explicitly share information about themselves (e.g. when someone creates a social network profile or enters credit card information for online purchase). Data can be legally

[3] OECD (2013) *Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value,* OECD digital economy Papers, No.220, OECD Publishing

observed, captured by recording users' activities (e.g. location data when using mobile phones). Finally, data can be inferred, based on the analysis of other personal data or of similar anonymous data.

A firm may collect data directly, usually having direct contact with the person or the object from which the data is collected, or indirectly, usually by buying the data from data brokers. Recently, the collecting activity has experienced significant transformations with the growth of IoT and different communication technologies that are generating new sources of data.

Storage and aggregation

Once data has been collected, it can be stored and aggregated. Because of the volume of these datasets and for cost reasons, data are increasingly stored remotely and accessed online. Moreover, different actors in the value chain usually keep this information increasing the magnitude of the effect of data disclosure (e.g. financial records are kept by retailers, banks and other financial institutions and taxation agencies; location information is kept by mobile phone operators, ISPs, utilities, transport operators and others).

Analysis and distribution

The third step takes the collected and stored data and combines it with other information to develop detailed profiles or infer information about macro trends that can be used for various purpose. Thus, it represents a key value enhancing step as the value of an individual record could be much more efficiently leveraged as the number of records to compare it increases. This work may be done by firms who progressively built more detailed profiles of their customers or by data analytics firms, who often resell the combined profiles in the market creating a real data exchange marketplace.

Usage

Once the data have been collected, stored and analyzed, they are used in many ways by businesses, the public sector, and end-users. The information generated can be used to better understand customers and offer them targeted advertisements or services, to improve business operations efficiencies, as well as to identify macro trends in different sectors, including healthcare, transportation, and safety.

Once understood the value chain of data, different methodologies can be used for estimating the monetary value. According to OECD (Organization for Economic Co-operation and Development),[4] there are two main categories of valuations as shown in Fig 2: the former is based on market valuation and refer to values of personal data that can be observed or derived from the market. The latter is based on individuals' valuation.

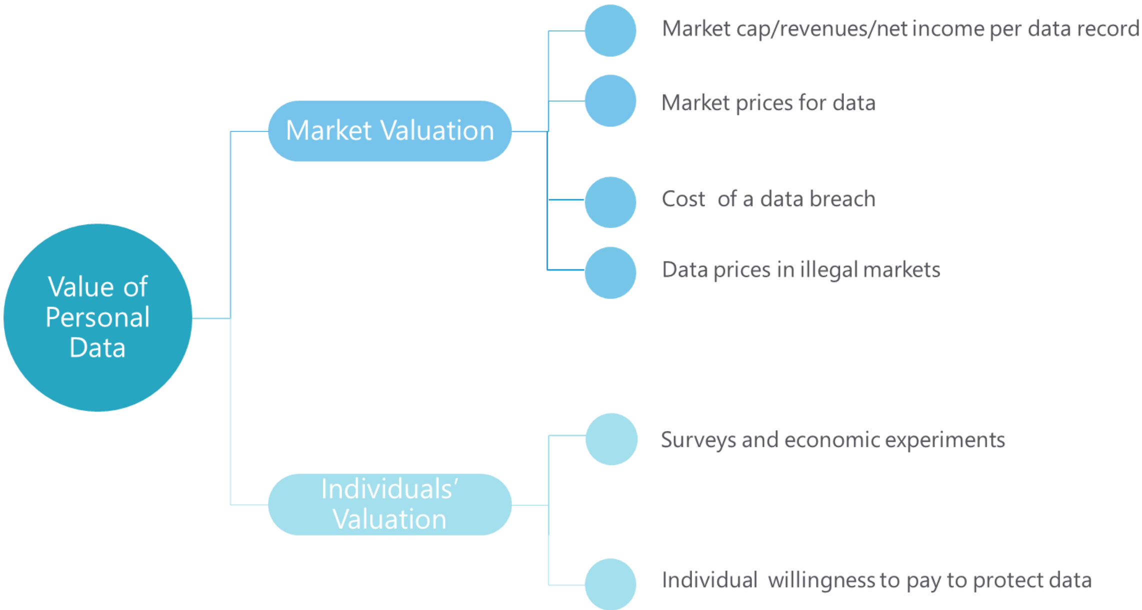

Figure 2: Estimates of the value of personal data

1.2.1 Market Valuation

Observable measurable proxies of the value of personal data include market capitalization/ revenues/ net income per data record, market prices for data, cost of a data breach and data prices in the illegal market.

Financial results per data record

This method looks at financial results such as market capitalization, revenues and net income to asses the value of data records of a company, which derives most of its revenues from selling or buying personal data such as a data broker. However, each of these financial indicators as a proxy of value presents some weaknesses.

The market capitalization is an approximation of the market value of the company. In reality, this value fluctuates with general market sentiments and other economic shocks that may not be tied to the effective real value of personal data. The second and the third indicators may be considered a more realistic measure of the monetary value of data as they linked to each data record the average amount of revenues or, even better, of net incomes (revenues minus

[4] See OECD (2013), "Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value", OECD Digital Economy Papers, No. 220, OECD Publishing.

costs) held by the firm. Certainly, they are both less volatile with market sentiments and external shocks. However, the final indicator represents the most robust proxy for data economic value derived from financial information because it excludes costs which may vary and not be directly related to the process of data valuation. Nevertheless, there are numerous components other than personal data that impact the financial results of a company. Hence, these measures could be misleading.

Market prices for data and illegal markets

Another way to asses the monetary value of data is to link it at the price in the market, where data brokers or third parties sell information. However, personal data is a “non-rival” good in economics, meaning that its use by one customer does not exclude the others and the same record can be sold many times to many costumers and be used multiple times by the same customer. As a result, the market price does not reflect the full monetary value of that record subject to multiple transactions, but it’s only an indication of the price paid for one copy of the data and do not reflect the total “earnings” over time.

Related to this market approach there is the valuation of personal data in the illegal cyber-crime market. These are markets in which cybercriminals buy and sell software, DoS, malware code and personal data (e.g. credit card numbers, users accounts, email address list, bank accounts, and others) see Table 1. Illegal data may be considered a "rival good" since the

Overall Rank 2009	Item	2009	2008	Range of Prices in USD
1	Credit card information	19%	32%	0.85 - 30
2	Bank account credentials	19%	19%	15 - 850
3	Email accounts	7%	5%	1 - 20
4	Email addresses	7%	5%	1.70/MB - 15/MB
5	Shell scripts	6%	3%	2 - 5
6	Full identities	5%	4%	0.70 - 20
7	Credit card dumps	5%	2%	4 - 150
8	Mailers	4%	3%	4 - 10
9	Cash-out services	4%	3%	0 - 600 plus 50%–60%
10	Website administration credentials	4%	3%	2 - 30

Source: Symantec (2010)

Table 1: Goods and services for sale on cyber-crime markets

stolen information will be less valuable to criminals if it has already been used by others. This would imply that the price of illegal data is likely closer to the full market value than would be the case with legal data that can be sold multiple times without a loss of utility. However, given the illegal nature of the market, prices of transactions will never be fully transparent and thus are difficult to collect and cannot be used as a proxy of value for data.

Costs of a data breach

This methodology tries to look at the value of data measuring the cost to individuals who have their identities stolen or the costs to firms when there is a data breach. These costs associated with the loss and misuse of information may give some indications about value, but do not represent a real estimation.

1.2.2 Individuals' Valuation

Other ways to put an economic value on data is either to run economic experiments and surveys or understanding the willingness to pay of consumers to protect their data in the form of insurance.

The first method allows extracting the amount of money sufficient for customers, according to their individual valuation of personal data, to give away their information. Moreover, individuals differ for their valuation of privacy: the amount of money they are ready to spend to protect their data from disclosure. Both valuations of personal data and privacy are extremely sensitive to contextual effects. (e.g. brands and industries are trusted differently: social security numbers and credit card information are perceived as most important to protect, while less importance is attached to names or telephone numbers). Frog (2011) shows with experiments run in the United States, India, and China that the highest value (USD 150- 240 per entry) is given by individuals to their national identity numbers and credit card information. Browsing history, location, and health information have a middle valuation of around USD 50. Finally, the lowest value is attached (USD 3-6) to data about online advertising click history or purchasing history.

The second method can reveal the willingness to pay to protect data as a proxy of the value of personal data, only in markets offering insurance policies to protect against identity theft. For example, Experian sells an identity-theft protection service ProtectMyID for USD

155 per year in the US[5]. However, the weakness in this scenario is that the value attached to data is only a pure individual perception of the economic value and not a realistic measurement.

Many other studies have tried to determine individuals' valuation of personal information and they can be roughly divided into two groups. The first set of papers derives the value of data from individuals' willingness to accept payments (WTA) in exchange for disclosing private information. Whilst, the second set focuses on individuals' willingness to pay (WTP) to protect the same data. Thus, the WTA represents the lowest price a person would be willing to accept giving away a good (protection of data) she initially owned. The WTP instead is the maximum price a person would be willing to pay to acquire a good (protection of data) she did not own. Studies of the first type are much more numerous. Huberman, Adar, and Fine (2005), for example, developed a reverse second-price auction model to extract the value everyone placed on disclosing personal information. In this auction, each participant had to anonymously submit his data and the price required for disclosure. The one demanding the least was paid the second-lowest demanded price and in exchange for this gain, the winner had to reveal his personal information to the other auction participants. Therefore, the financially competitive nature of the reverse auction won by the lowest offer, let the authors extract the WTA of each subject involved. They used auctions to study the average price demanded by participants for age and weight. It emerges that age information is less sensitive than weight, with average prices of USD 57.56 for age versus USD 74.06 for weight. In addition, they try to collect prices of more controversial information such as salaries and savings. In this case, the percentage of individuals demanding more than USD 100 was 48% for salary and 38% for savings. Gkatzelis, Aperjis, and Huberman (2015) worked on a *Baseline Mechanism* in which a buyer is interested in obtaining access to a representative sample of a subset of data, thus she must set a price ensuring that a good portion of the individuals chooses to reveal their information. The mechanism first asks each seller (individual) the minimum price for which he would allow the buyer to access his information. Then, among the sellers with the highest price (Cmax), one of them is randomly expelled from the set. From the remaining set is selected the sample, and each sampled seller is paid Cmax in return for his information. For example, let's consider a company with 1000 subscribers. The firm has collected attributes of each of the subscribers with their purchasing and browser history. A third party is interested in buying this information over a sample of 100 consumers. 300 of the subscribers have required at least USD

[5] https://www.experian.com/consumer-products/identity-theft-and-credit-protection.html

10 to be part of the sample, 300 at least USD 5, and the remaining 400 did not care about their privacy. Thus, Cmax= USD 10 and one of the 300 subscribers, requiring the highest price, has been expelled from the set. The sample of 100 people has been randomly selected from the 999 non-discarded ones for a total cost of USD 1000 to the third party (each seller being paid USD 10). The aforementioned mechanism is similar to a lottery that extracts the WTA of sellers, who will be paid with a probability equal to the proportion of individuals sampled or will receive no payment otherwise.

Both experiments confirm Frog's results. Indeed, Huberman, Adar, and Fine (2005) tested the valuation of health (weight, age) and more sensitive information (salary, savings) and confirmed the range of USD 50 and more than USD 100 estimated by Frog for these categories. Gkatzelis, Aperjis, and Huberman (2015) studied the value of lower value data (purchasing history) and the prices founded by the authors among USD 10 and USD 5 are in line with Frog's USD 3-6 estimations.

Instead, studies on the WTP, in which consumers are asked to consider paying money to protect their data, are scarcer. Tsai et al (2011) find that when privacy information is displayed in a more salient way, participants take the privacy policy into account and they become willing to pay a moderate premium price to purchase from online merchants with better privacy protection. If data practices and privacy policies are not in an easy-to-digest form, they might not influence the decision of consumers, who would not pay one Euro more for their privacy. The results of laboratory, hybrid and field experiments conducted by the European Network and Information Security Agency (ENISA)[6] go in the same direction. The aim of these tests, selling online cinema tickets, was to find out if individuals value their privacy enough to pay a mark-up at the firm which collects less information. The results show that a higher share of sales is obtained by the privacy-friendly firm when no price differences occur. However, when the privacy-unfriendly firm charges a lower price on tickets of 0.50 €, it obtains a greater market share with no more of a third of consumers willing to pay this mark-up to the privacy-friendly firm. Thus, consumers are not disposed to pay more than 0.50 € (around USD 0,55) for increased privacy but they might ask around USD 3-6 to give away the same information.

Finally, from all these researches emerge a WTA-WTP gap, with WTA tending to be larger than WTP. It predicts that individuals, despite valuing their privacy, when asked to pay to increase it they might be deterred from doing so by the prospect to lose money. This

[6] See Harasser Andreas, Jentzsch Dr Nicola, Preibusch Sören *Study on monetizing privacy. An economic model for pricing personal information* ENISA, 2012

discrepancy on how much people will pay to protect their data and how much they will accept to give the same data away states the failure in finding a truthful estimate of the value that individuals assign to data. Nonetheless, data is to the digital economy, what oil is to the industrial economy. If information is power, control over personal information affects the balance of economic power among parties. Thus, personal data is an essential resource, source of competitive advantage.

1.3 Use of Data

Analyzed the value in owning personal data, firms can make different uses of information. The two most common practices are the possibility to set personalized prices to costumers and address them with target advertisings. However, there are other forms of discrimination a consumer may incur in, as explained in the following sections. Hence, disclosure of personal information can benefit both data holders (efficiency gains, surplus extraction, increased revenues through consumer tracking) and data subjects (personalization, fewer search costs, target offers and promotions). At the same time, such benefits can turn into disadvantages and be costly for both firms (costs incurred when occurs a data breach, reduced trust in online markets) and consumers (different forms of discrimination, identity theft or psychological discomfort).

1.3.1 Target Advertising

Online advertising is perhaps the most common example of how firms use the huge amount of data about users. The collection of consumer information allows creating an individual profile based on which sites the visitor has looked at or what he has searched for. The practice of target advertising takes place when a firm creates ads addressing a specific audience of customers based on their estimated profiles. The most famous form of target advertising is the so-called 'retargeting': when a consumer surfs on the net, showing interest for a brand, later on, while visiting other websites, an ad from the first brand will be shown to the consumer. Thus, targeted ads have the benefit to be displayed only to customers potentially interested in the product or service, thereby increasing the possibility of a good match and the overall advertising efficiency. On the customer point of view, target advertising reduces search costs, providing consumers with the information they are interested in.

However, White et al. (2008) argue that consumers may find intrusive and annoying these ads[7]. Hence, is necessary to avoid a negative reaction and to fully exploit the potential of target advertising over the regular one, to expose consumers to fewer and more relevant ads. Moreover, firms that engage in target advertising has to comply with the European ePrivacy Directive, which states that cookies or similar methods for user tracking cannot be used unless under explicit consumers' consent[8].

1.3.2 Price Discrimination

A firm price discriminates when it charges different prices to different customers for similar goods where this price difference does not reflect cost differences[9]. Price discrimination (PD thereafter) is a practice to extract as much as possible of consumers surplus, taking into account their different willingness to pay for the product or service. The efficiency of these discriminatory practices is linked to some conditions:

- The firm should have some market power so that can impose prices over marginal costs (p>Cm);
- There is no possibility of arbitrage or resale (otherwise, consumers who benefit from lower prices would have an incentive to resell the good at higher prices)
- Availability of information about individual preferences.

The economic literature distinguishes between three types of PD (Pigou, 1920).

I-degree PD: the firm has perfect knowledge and is able to set individual prices and fully extract consumers' WTP. Each consumer has its own offer take it or leave it.

III-degree PD: the firm observes aggregate preferences and can price discriminate between groups, but not within groups.

II-degree PD: heterogeneity of consumers cannot be observed by the firm, which can offer conditions (such as menus of options or packages) and consumers self-select these options revealing their type and preferences.

---

[7] (White, Zahay, Thorbjørnsen, & Shavitt, 2008) In their article they talk about a "personalization reactance" experienced by consumers, which may react negatively to highly personalized messages when the customization fit is not explicitly justified.

[8] See the Data Protection Directive (195/46/EC) and the Privacy and Electronic Communications Directive (2002/58/EC), also known as ePrivacy Directive which regulates cookies and other similar devices through its amendments such as Directive 2009/136/EC, the so-called EU Cookie Directive, and the Privacy and Electronic Communicatiions (EC Directive) (Amendment) Regulations 2011.

[9] Strictly, by a more rigorous definition, that of Stigler, price discrimination is "the sale of two or more similar goods at prices which are in different ratios to marginal costs". See George J. Stigler, *The Theory of Price,* 3rd edition 1966, pp 209-215

Certainly, the huge amount of data collected on consumers are facilitating price discrimination. Websites' abilities to track consumers and to dynamically update and personalize prices for each visitor are bringing online markets closer to the theoretical scenario of I-degree PD. Thus, Internet-based price discrimination practices have become with the passing of time increasingly effective and sophisticated. Valentino-Devries et al. (2012) suggest customers may be subject to different prices based on the ability to estimate the online visitor's physical distance from a rival brick-and-mortar store. Mikians et al. (2012-2013) find evidence of 'search discrimination' or 'steering': showing costumers different sets of products, with different prices following their research for a certain category. In particular, it emerges price differences of 10% to 30% based on browser configurations of the different online visitors. For example, on the Wal Street Journal was published in 2012 an article about a travel agency Orbitz, which was showing more expensive travel options to Mac users than to Windows visitors saw. Hence, the case demonstrates how tracking people's online activities can use even seemingly innocuous information (e.g. the fact that customers visiting Orbitz.com from a Mac) to start predicting their tastes and spending habits.

However, online websites and platforms practicing price discrimination have to be aware of negative customer reactions. Well known in September 2000, the case of Amazon.com which outraged some customers when its own price discrimination was revealed. One buyer complained that, after deleting the cookies on his computer identifying him as a regular Amazon customer as a result, he watched the price of a DVD offered to him for sale drop from $26.24 to $22.74. The company said the difference was the consequence of a random price test and offered to refund customers who paid the higher prices. Thus, these events could potentially threaten and undermine the reputation of a brand, reducing customers' trust in online markets, in particular, if PD is realized in a non-transparent way. However, there are finer and more acceptable, from a customer point of view, way of acting for a firm to obtain the same result. For example, a firm can fix the same prices for all the consumers applying personalized offers. Costumers will end pay different net prices; hence the strategy is equivalent to setting personalized prices. However, discounts are more difficult to compare, and this will reduce the number of negative reactions.

In sum, the Internet allows shoppers to easily compare prices across thousands of stores. But, with the advent of big data, it also enables businesses to collect detailed information about a customer's purchasing history, preferences, and financial resources and to set prices accordingly. As follow, price discrimination is an increasing trend, though firms may have to adopt indirect methods (personalized discounts) in order to avoid upsetting their customers.

1.3.3 Other Forms of Discrimination

Price discrimination is probably the most known form of discrimination involving the use of personal data. However, there are new forms of discrimination that A.Acquisiti et al. (2015) mention in their article.

Consider, the new role played by social networks and all the information available about job candidates. Nowadays, publicly provide personal information about himself/herself through social networks such as sexual orientation, political opinions or religious beliefs. Employers are supposed to not ask about such information during the hiring process. However, Bertrand and Mullainathan (2004) prove that recruiters can't stop screening candidates' social media profiles and using the information collected for 'hiring discrimination'.

Also, consider online platforms that allow tenants to find landlords or platforms that enable car owners to share their vehicles with other drivers or passengers. In all these cases of IT sharing economy, for sure the availability of data about peers improve efficiency in how resources are shared. However, the disclosure of personal information fosters the possibility of sexual, social or racial discrimination. Using data from Airbnb, Edelman and Luca (2014) denounce a case of race discrimination, as the race on the platform can be inferred from the profile photo of the accounts. They find that in New York City landlords who were African American pay approximately 12% less than not African American for the same rent.

Hence, these examples highlight that data does not only have an important economic dimension, but also a social impact; which is incredibly stronger in every sphere of people's daily life from their leisure time to their professional career.

The discussion above about big data and its usage underlines a lack of awareness and myopia among consumers regarding the extent, nature, and depth of the consequences of sharing their personal data. A. Acquisiti et al. (2015) talk about a 'reversal of informational asymmetries' generated by disclosing personal data: at first, data subjects have information that data holders don't know. Later on, data subjects may not know how data holders will use their data and with what consequences. Hence, disclosing data often brings an immediate benefit, intangible (likes under a post) or tangible (discount from a merchant). However, the costs of doing so are often not sure and incurred at a more distant time in the future (an HR recruiter won't like the post as much as the friends did; the merchant will use the information for price discrimination). Thus, the role of governments and antitrust authorities should guarantee more transparency, solving the information asymmetries regarding the usage and subsequent consequences of shared information.

1.4 Digital Platforms

Most digital-economy giant (e.g. Google, Facebook, Amazon, eBay, PayPal) and many high-growth startups (e.g. Uber, Airbnb, Netflix) are multisided platforms. There has been a rapid growth of digital platforms explained by several distinct factors: the transition to the online world and the rapid increase of internet users in past decades, the innovative, user-friendly services the platforms provide. Thus, a digital platform economy is emerging and radically changing a wide range of human activities including how people work, socialize and create value in the economy. Innovation is no more in the product but in the business model, in the new ways value is delivered to the final consumer. This seems to be the key for success in the digital era, thus great corporate attention is given to the choice of the business model: the different types of agents to connect, the architecture and governance of the platform as well as the intellectual property protection regime, the relative prices charged to the different sides.

These realities are mainly multisided platforms, infrastructures that facilitate the connection and interaction between several types of actors who have complementary resources and needs and therefore create value for each other. Basically, multisided platforms consist of groups of individuals who use the platforms for different reasons (see Figure 3). For example,

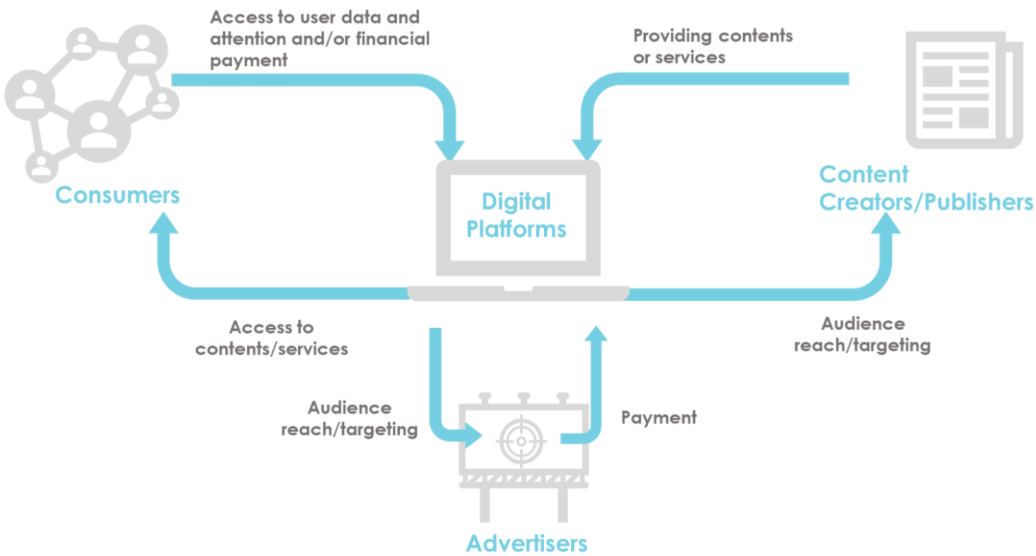

Figure 3: Interactions of digital platforms with their actors

users can take advantage of using these platforms for the possibility to communicate with other peers, find and consume content or services, transact with merchant businesses, or share their own generated contents. At the same time, merchants, content publishers or service providers on one hand and advertisers, on the other hand, can benefit from the usage of digital platforms easily reaching online audiences.

1.4.1 Common Features

The rise of this phenomenon and its relevance in the digital economy has encouraged researches in literature for classifying these platforms, define their properties and identify criticalities and new market outcomes. There is not a clear and widely accepted definition of multisided platforms. However, they can be characterized by some common properties.

Multiple actors

There is a consensus on the idea that one of the inner properties is the presence of distinct types of users or parties ('economic agents') who need each other in some way and who rely on platforms to intermediate transactions between them. Evans (2003) suggested a broad definition, based on this feature*: "Multisided platforms coordinate demand of distinct groups of customers who need each other in some way".*

In its role as a coordinator among parties, the platform may act as a merchant or a broker. In a **broker model** content creators, services providers, or merchants choose the price of their products or services. Instead, in a **merchant model,** the platform determines prices over transactions among parties. For example, Apple functions as a merchant for music on iTunes Store setting the price of music; while it acts as a broker for applications in the App Store, where independent developers fix the price for their applications. Hagiu and Wright (2013) explain the strategic decision made in Cupertino: there are few music publishers, but thousands of application developers. Thus, the management costs of dealing with so many different suppliers would have been greater; moreover, as the application market is much more competitive, Apple had little to gain from trying to control the prices of applications. Finally, every choice such as the one about the governance of the platform more centralized or decentralized plays a pivotal role in the success of the business model.

Indirect Network effects

The essential role of indirect network effects was underlined by Rochet and Tirole (2003), who build a model of platform competition with two-sided markets. These markets with network externalities are characterized by the presence of two distinct sides, which benefit from same-side and cross-side effects.

A **same-side network effect** takes place when an increase in the number of users of one side of a platform affects the value of the service to a given user on that side of the platform. An example of this kind of effect is a traditional social network. If there are only a few users of the social network, the platform will be not valuable for any users. While, large scale social media platforms, such as Facebook and Instagram, may have a greater ability to attract users as a great portion of users' social group or family already are on the platform, which is perceived as a 'must-have'.

A **cross-side network effect** occurs when an increase in the number of users of one side of a platform affects the value of the service to a given user on the other side of the platform[10]. For instance, on a credit-card platform, an increase in the number of consumers of the card increases the value of the card to a merchant. Moreover, an increase in the number of merchants accepting the card makes it more valuable to the consumers. Thus, the cross-side network effect operates in both directions creating a positive feedback loop in favor of a large-scale platform, which may create a barrier to entry or expansion.

Although this situation of network effects will power a virtuous circle once an attractive offer to all the economic agents has been created, each platform to succeed must address the celebrated "chicken-and-egg problem" to get both sides on board. Thus, the resolution of this strategy problem occurs whenever the value proposition to a group of stakeholders is dependent on another stakeholder penetration.

Cross Subsidisation

Finally, a third common aspect of multisided platforms refers to their pricing structure, which usually involves cross-subsidization practices. They usually derive revenues from one type of users and subsidize another type of users, who are charged a very low price, potentially less than the marginal cost. Frequently, platforms offer their services to end-users for a zero monetary price to obtain consumers' attention and data, which they monetize. Indeed, as the number of consumers increases because of cross-side network effects, the value of the platform for advertisers is consequently higher. For example, the larger user-base implies that the number of customers exposed to the advertising campaign is higher, which may increase the advertiser's

[10] Strictly, Rochet and Tirole (2003) define the presence of indirect network effects *"the net utility on side 'i' increases with the number of members on side 'j'"*, as a fundamental element in the definition of two-sided markets.

return from the campaign. Moreover, more users equal more data and the possibility to better address consumers with target advertising, increasing the efficiency of ads too. As a consequence of the cross-side network effects and the higher value of the platform to advertisers, revenues earned from advertisers will be higher and sufficient enough to justify this pricing structure.

1.4.2 Types of Platforms

Once highlighted which definitions are used in literature, how researches have identified multisided platforms and their features, the following step is to classify them and assess their market power. Platforms are diverse in functions and structures, but regardless of the specific type, they all depend on the digitalization of value-creating human activities. Some of the most salient types of digital platforms are:

- *Search engines:* software systems designed to search for information on the World Wide Web. Through the usage of sophisticated algorithms, they are able to systematically show the findings of the research as a set of links. Google Search is the world's most popular search engine, with a market share of 92% as of June 2019 (StatCounter, 2019). Other examples include Bing, Yahoo!, Baidu, DuckDuckGo, etc. (see Figure 4)

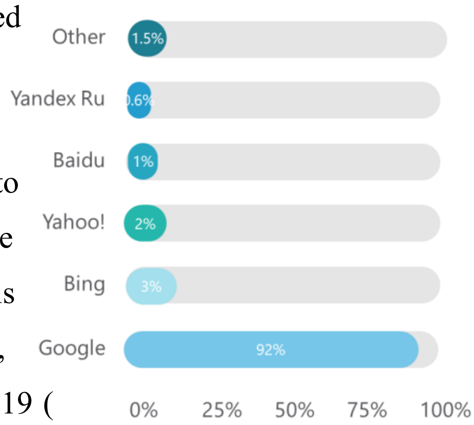

Figure 4: Search Engine Market Share Worldwide

- *Social networks:* they primarily allow users to communicate with others, share and consume content generated by others, participating in social networking. Usually, they offer additional instant messenger services, and the most famous examples are Facebook and its recent acquisition Instagram, Pinterest, LinkedIn, Snapchat. The most widely used social network is still Facebook with about 68% of the market share (StatCounter, 2019), even though

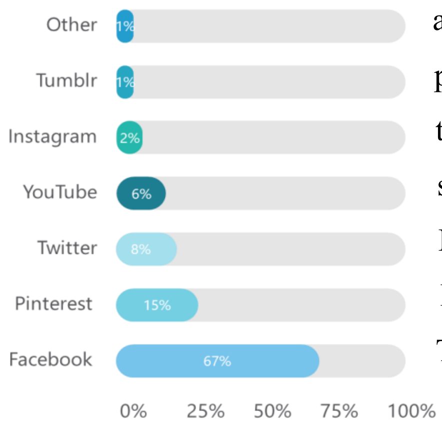

Figure 5: Social Media Stats Worldwide

the recent spread of primary photo-based socials (see Figure 5).

- *Retail platforms:* platform-based multisided marketplaces like Amazon, Alibaba, eBay, JD.com are the outcomes of a process of transformations of the retail segment. From the traditional brick and mortar stores and the development of big shopping centers, malls and retail chains in the past decades, nowadays retail has become a digital reality. These ongoing changes are creating more personalized, convenient, and speedy shopping experiences in a service system perspective.

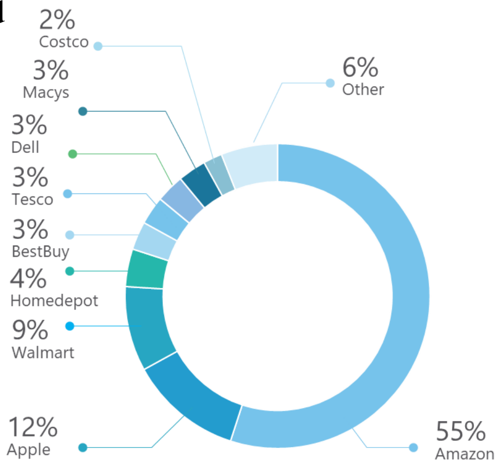

Figure 6: Top 15 Online Shops Worldwide

- *Service-providing platforms:* online intermediaries that offer a wide variety of services, which are disrupting traditional markets and different types of businesses. Spotify provides an online music service offering users the ability to stream audio music files on-demand, allowing them to enjoy the benefits of unlimited music access. Wikipedia aims to realize a democratic idea of knowledge, building a free online encyclopedia, created and edited by volunteers around the world. Airbnb promotes the notion that vacant rooms in one's house can be a source of income by renting them. Uber and Lyft are entrepreneurial initiatives that facilitate the unlocking of the commercial value in underused personal assets such as automobiles. Thus, in an increasingly service-oriented society focuses on making easier customer's experience a bunch of other examples can be found in this category.

The pattern is clear. Even if new tech markets seem to be very competitive because of the volatile nature of technologies themselves and the huge number of innovative start-ups in the online world, there is one winner per market. Once a tech company achieves clear market leadership, it is difficult to displace or eclipse it. In this scenario of a winner-takes-all, platforms are so dominant in their own markets that the biggest competitive threat is that a newer, bigger, adjacent market emerges, dominated by another player.

1.4.3 Market outcomes of the Economy of Platforms

Some new concerns raised by two-sided markets are specifically related to the emerging phenomenon of platform dominance. Data can reinforce network, feedback, scale and scope effects creating a unique scenario of competition between platforms. An important factor in explaining the growth of dominant platforms is, doubtless, that the larger platforms are, the more valuable they tend to be for their users. Thus, the ability to build scale is a key factor success for platforms because of the existing relationship between value delivered to end consumers and their sizes. Moreover, the scale reached enables the possibility to collect, use and process huge amount of data in order to optimize user experience or create and shape new markets. Advantages of scope are the direct effect of data accumulation thanks to the dual role played by data in leveraging market power. On one hand, data power allows successful expansion to other markets. On the other hand, this expansion to other markets enables to collect more data from related services, strengthening the platform or ecosystem dominance. Hence, incumbents with a dominant position can leverage their market power to conquer "connected markets" in a data-driven domino effect, which helps in explaining why web giants are increasingly invading markets that often seem far from their core business.

Economies of scale and scope related to network effects are key determinants of a platform's success. However, there are other factors facilitating market dominance. High switching costs, lack of interoperability, limited data portability increase customer locks-in, while enhancing the data advantage of incumbents. This incumbency advantage is partly mitigated by the possibility of multihoming[11], which shapes the competition between two-sided platforms. Indeed, prices on one side of the market depend on the extent of multihoming on the other side of the same market. Generally, multihoming on one side of the platform intensifies price competition on the other side as lower prices are used in the attempt to create exclusive relationships on the latter side. For example, nowadays an increasing number of consumers have in their pocket both Amex and Visa cards. When Visa started to lower the transaction per fee paid by merchants, more sellers decided to turn down payments with Amex card, knowing that a great portion of Amex buyers were also owners of a Visa card.

Finally, incumbency advantage in the platform economy is preserved by exclusionary behaviors in the form of mergers or foreclosures. In the last 10 years, dominant digital platforms have made more than 400 acquisitions globally, acquiring their potential competitors of complementary technologies (e.g. Facebook/ Instagram, Google/ DoubleClick,

[11] The practice of users to adopt new platforms without quitting the old ones.

Google/YouTube, eBay/PayPal). When a dominant platform can't acquire a competitor, it may have an incentive to block it via foreclosure impacting rivals in an adjacent market or limiting them in its core market. Thus, monopoly tying[12] of a complementary product or service can prevent entry in its primary market and leverage market power in the secondary market.

Although competitive outcomes of the reality of two sides markets are the ones more studied in the literature of platform economy, increasing attention is given to privacy issues and consumer protection. Many digital platforms can collect a huge amount of data on users 'activities, even beyond what is actively provided by them during the usage of online services. Information asymmetries and the bargaining power held by platforms vis-à-vis consumers lead to uninformed choices and difficulties in assessing the current and future costs of providing their user data. The lack of enough control over data generated bring to a disconnection between consumers' privacy concerns and their actual behaviors, usually referred to as 'privacy paradox'[13]. It regards choices made by consumers that do not prioritize privacy, even if they are paradoxically concerned about it. One possible explanation is that consumers care about privacy in theory but, in practice, the value derived from using a digital platform's services offsets the privacy 'price' they pay in allowing the collection of their personal data. Generally, consumers have different levels of privacy awareness; therefore, different data collection practices aligned with their privacy preferences should be desirable. However, digital platforms use clickwrap agreements with take-it-or-leave-it terms that bundle a wide range of contents, leaving consumers with no meaningful control over the collection, use, and disclosure of their data. Besides, these privacy policies are usually long, complex, vague and difficult to navigate with additional links to separate web pages. These features are likely to increase the amount of navigation and reading time for users, contributing to consumers' tendency not to read online terms of services and creating increasing information asymmetries over privacy policies. As such, it may be appropriate to impose more requirements over privacy policies to make them clearer and ensure that consumers can engage in the digital economy in an informed way, making decisions that are in their own best interests.

[12] Tying of two products (or services) occurs when a seller sells one good (tying good) on the condition that the buyer buys the other good (tied good) from that seller or imposes on the buyer the requirement that she will not purchase the other good from another seller.

[13] See P. Norberg, DR Horne, and DA Horne *The privacy paradox: Personal Information Disclosure Intentions versus Behaviors*, The Journal of Consumer Affairs 41(2007)

User data represents the input of a multitude of markets in the digital economy, their volume is growing exponentially[14] and therefore their collection, use, and disclosure are of increasing importance for consumer protection and privacy, but also for the online general competitive and innovative environment. Hence, data-driven markets open questions at the intersection of three different subjects: privacy, competition, and consumer protection.

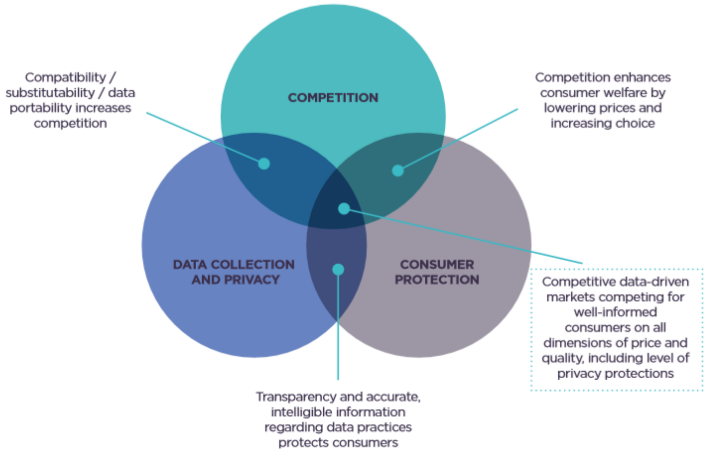

Figure 7: Overlapping issues in data privacy, competition, and consumer protection

1.5 Third Parties

Power in the digital economy is linked to the use and collection of personal information. 2.5 quintillion bytes of data are generated every day at the actual pace (Forbes, 2018). The monetization of data is becoming a significant business in the digital era and a large share of these revenues is generated by the third-party tracking industry.

1.5.1 Third-Party Tracking

Third-parties are companies who have access to user data through technology which is embedded on multiple first-party websites or mobile applications. Generally, when a user installs an app or visits a website, the third-party code collects information and associates it with a unique identifier sent to a remote server, controlled by the third-party. The more the third-party's code is integrated with multiple websites and apps, the greater the tracker's power to collect data from different sources and combine it into a single behavioral profile.

Third-party tracking serves different purposes broadly divided into three categories:

[14] According to IBM report 90% of all data that exists in the world today has been created in the last two years – Bernand Marr *How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read* Forbes, 21 May 2018

- *Market research services:* that consist of reports on customer behavior, their buying habits and purchase attitudes;
- *Marketing information services:* that imply the creation of more comprehensive profiles on individual consumers for direct marketing purposes through target ads;
- *Media measurement services:* that entail audience measurements on how an app or website is used by matching website visitors with external information such as their demographics, interests or geolocation data.

Clearly, third-party trackers have commercial advantages over first-parties; not only they can more easily create efficiencies of scale, but they can also generate economies of scope aggregating data from different first-party services. Thus, they can access a richer pool of data. By contrast, a first-party only has access to the data of its own users as they engage with its particular service. However, third-party tracking potentially may raise more privacy concerns than first-party data collection. They can access multiple records from different websites and apps of a single consumer, not constrained by contractual policies with an end-user that usually limit the data processing activities. Hence, the increasing power in the hand of third-parties has attracted the attention of regulators. At first much of this attention has come from privacy and data protection regulators. More recently, competition authorities have begun to be concerned too about the impacts of the consolidation of tracking capabilities in the hands of a small number of firms.

1.5.2 Third-Party Market Power

Traditional measures used by antitrust regulators, based on the concepts of market share and definition of the relevant market, failed when comes to the third-party industry. Indeed, a firm's market share is estimated as the percentage of revenues or units sold of the total industry. However, the power of a third-party is related to its capacity to collect, analyze and elaborate data which may be not directly related to the sales and revenues of the firm. Moving to the measure of the market concentration by defining a relevant market, it’s even more meaningless in the digital ecosystem. The tracking activity of apps and web sites can occur in such different environments, that the only attempt to define a single market will be unsuccessful in truthfully representing the real capability of a tracker.

Binns Rebeun et al. (2018) [15]in their work discuss the third-party power in terms of *prevalence* and *prominence*. The former is related to the number of first-parties on which a third-party is present. The latter refers to the number of users of the first-parties. Both dimensions are important: a third-party with high prevalence will have access to multiple sources of data. However, combined with low prominence (first-parties with very few users) the data collected will not be so valuable. On the other hand, a tracker with high prominence which is present on a very popular first-party, combined with a low prevalence (the third-party in question operates only on that popular platform) will not have more value than the first-party itself.

Mergers and acquisitions are a common occurrence in data-related sectors and consolidation between third-party trackers has the potential to merge personal data from multiple sources into one profile. Well-known the acquisition of Google of DoubleClick in 2008, second only to Google Analytics for the highest prevalence among third-party services. The strategic move of the board of directors has affirmed Google's tracking capability around 70% of prevalence on websites and 80% on mobile apps. However, large data-mergers third-party related are often missed. This is because those firms collect data indirectly via first-parties, and the estimation becomes harder for the volume and variety of data that will be combined as a result of mergers between them. Hence, the tendency of competition authorities to focus on first-party rather than on third-party data combinations.

M&A decisions in the third-party industry arise privacy issues from the consolidation of tracking capabilities. However, data protection and competition might go in opposite directions regarding these decisions. Consolidation of tracking capabilities could be both positive and negative for privacy. On one hand, mergers increase market concentration reducing the number of distinct tracking entities the consumers are exposed to. On the other hand, M&As raise the number of distinct ways in which a consumer could be tracked by a single entity. Hence, there's a privacy trade-off between reducing the number of tracking firms and the amount of data collected by a single tracker.

From the analysis conducted so far, it emerges that antitrust authorities must be ready to embrace greater complexity, changing traditional thresholds with new measures. Moreover, they should embrace complexities from the overlapping of competition, data protection, and

[15] See Reuben Binns, Jun Zhao, Max Van Kleek, Nigel Shadbolt. 2010. *Measuring third party tracker power across web and mobile* arXiv preprint arXiv:1802.02507 (2018)

privacy, instead of hiding them behind the appearance of disciplinary coherence. Competition law cannot exclude the human and social aspects of data-driven operations from its analysis. It is necessary to start to think beyond disciplinary boundaries, looking for solutions at least dynamic as the evolving digital sector.

2. A critical review of the literature

This second chapter aims to conduct a critical review of literature about data, organized by main topics. The goal is to offer insights into research to be transformed into practical actions by antitrust authorities and governments. The belief is that a deeper knowledge of the ongoing changes will create a society in which the advent of technology, so much part of our daily life to be taken for granted, will be faced more consciously and themes like the fundamental right to privacy and consumer protection won't be forgotten.

As emerged from the first chapter, the rapidly growing volume and economic importance of data in the digital economy are creating concerns related to the giants (Facebook, Amazon, Alphabet, Microsoft) that deal in data, the oil of digital era. They may be broken up, as Standard Oil was in the 20th century. However, their success is much more resilient than size. It is related to consumers, few want to live without Google's search engine, Amazon one-day delivery or Facebook's community. They look unstoppable.

Should we be concerned about these titans?
In which ways can regulators intervene to maximize social welfare?
Would more data sharing and open access improve efficiency and reduce data market failures?

Thus, such dominance has raised questions used as guidelines of this review. In the following chapters, they are analyzed and studied in detail, even if there are no easy answers to these largely empirical questions. Moreover, the discussion will be linked with traditional branches of economic literature, including monopoly and perfect competition, switching costs, the commons and anti-commons literature, and models of data trade.

2.1 Natural monopolies and market failures

Natural monopolies occur when a firm can serve the entire industry at a cost lower than it would be if there were several firms. This generally because of the nature of costs in the involved industry. Typically, there are very high fixed costs and low marginal costs, with large economies of scale. For example, commodities such as the supply of water or electricity involve building a big network infrastructure. As a result, fixed costs are enormous but the marginal cost of adding an extra user is very low and the average total cost will continue to fall as extra users are added to the network. Recently, the digital platform businesses have become the modern-day equivalent or the old-style utilities in terms of being examples of natural monopoly. Indeed, they are characterized by high fixed cost to create their product and services, e.g. costly implementation of IT infrastructure and processing technologies, and near-zero marginal costs.

2.1.1 Definitions of Natural Monopoly

In the attempt to find in literature, the definition of the term "natural monopoly", Sharkey (1982) in his work [16]offers a complete and detailed overview of the intellectual history of the topic. John Stuart Mill and Alfred Marshall were the first to start to speak about natural monopolies in the late '80s. However, with Richard Posner (1969) the attention about the theme was shifted from the number of sellers in the market to the cost characteristics of the technology of supply of the industry. Finally, a few years later Sharkey fully defined the conditions to talk about natural monopoly distinguishing the single product case from the multiproduct one.

Single product case

In a single product context, the presence of economies of scale is a sufficient, not necessary, condition to meet the definition of natural monopoly. When a firm's technology of production is characterized by economies of scale its average cost of production declines as the output expands. Such cost function, for a single-product firm, will have the property of subadditivity.

Consider a market for a homogeneous product in which there are n firms. Each of them produces an output q^i and has an identical cost function C (q^i). The total output of the market is given by $Q = \sum_1^n q^i$and the cost function is said subadditive if:

$$C(Q) < C\,(q^1) + C\,(q^2) + \ldots + C\,(q^i) \qquad (1)$$

According to (1), it is less costly if the market is supplied with the output Q by a single firm rather than by two or more competing firms. Hence, cost functions that have this property of subadditivity meet the definition of natural monopoly. However, in the single product case, firms' cost functions may have ranges in which economies of scale are over, but they are still subadditive. That's why the condition of economies of scale is sufficient, but not necessary.

Multiproduct case

[16] Sharkey, W.W. (1982), The Theory of Natural Monopoly, Cambridge: Cambridge University Press.

Moving to the multiproduct setting the previous condition is neither sufficient nor necessary. In this scenario, given the vector of quantities of k products $\boldsymbol{q}^i = (q_1^i, q_2^i, \ldots, q_k^i)$ and the presence of N vectors of this type, the cost function is subadditive if:

$$C(\sum q_1^i, \sum q_2^i, \ldots) = C(\sum \boldsymbol{q}^i) < \sum C(\boldsymbol{q}^i) \qquad (2)$$

The sufficient cost conditions that will lead to multiproduct subadditivity are:

- the presence of some form of economies of scope, when the costs of producing two or more goods jointly, is less than that of producing each separately:

$$C(q1, q2) < C(q1, 0) + C(0, q2) \qquad (3)$$

- the presence of some form of multiproduct economies of scale, tested by the presence of declining average incremental costs[17]. Defined the incremental cost of producing the product q_1, holding q_2 as constant:

$$IC(q1|q2) = C(q1, q2) - C(0,q2) \qquad (4)$$

where C(0, q2) is the stand-alone cost of product 2. The average incremental cost of q1 is indeed obtained by dividing (4) by q1:

$$AIC(q1|q2) = [C(q1, q2) - C(0,q2)]/q1 \qquad (5)$$

If (5) declines as q1 increases (holding q2 constant), this is a measure of a single product economy of scale in a multiproduct context.

2.1.2 The Social Costs of Monopoly

Although the importance of the historical evolution of the concept of natural monopoly, greater attention is given to the consequences of these market outcomes. It is well-known that the presence of monopoly power results in higher prices, lower volumes and other market inefficiencies. However, the presence of subadditive costs per se does not naturally imply a monopoly power, as explained by Baumol, Panzar and Willig's theory of *contestable markets*.

[17] See for instance Baumol, Panzar and Willig, 1982 *Contestable Markets and the Theory of Industry Structure.*

The authors disrupt the myth that economies of scale and high concentration of the market will turn to the classical textbook monopoly price. Indeed, it depends on the nature of fixed costs, that cannot represent a sunk cost. In this case, they will not represent a barrier to entry or exit for new entrants, which otherwise, once entered the market, must be worried about fixed commitment costs. Thus, new entrants can leave the market without incurring in any mobility barriers and practice a “Hit and Run” strategy [18]if the monopolist raises prices. The only threat of potential new entrants, which may offer a lower price, disciplines the incumbents forced to respect a break-even constraint. This equilibrium price is equal to the average total costs in the single product scenario, and Ramsey-Boiteux’s price in multiproduct context. It is not the first-best of perfect competition since is higher than marginal costs but is the so-called second-best efficient price. Thus, thanks to the literature on contestable markets, the presence of a high concentration of the industry does not exclude that prices way still be competitive in the second-best sense. A monopoly naturally emerges, but it may have small social costs with the impossibility by regulators to do any better.

If it is not this the case, there is a clear need for regulation from governments to mitigate the distortions caused by high barriers to entry and unregulated prices sustained above both marginal and average costs. Back to dataopolies represented by digital platforms, the owner of data (the first party itself or a data broker) will fix a price to sell or license the access to its dataset. The grey rectangle represents revenues for the data owner. The blue triangle is the consumer surplus derived from data or the difference between their willingness to pay and the actual price paid. This price is obtained maximizing the data owner's profit and not the aggregate welfare of society. Indeed, the dark grey triangle represents the deadweight losses for society associated with this price. (See Figure 8)

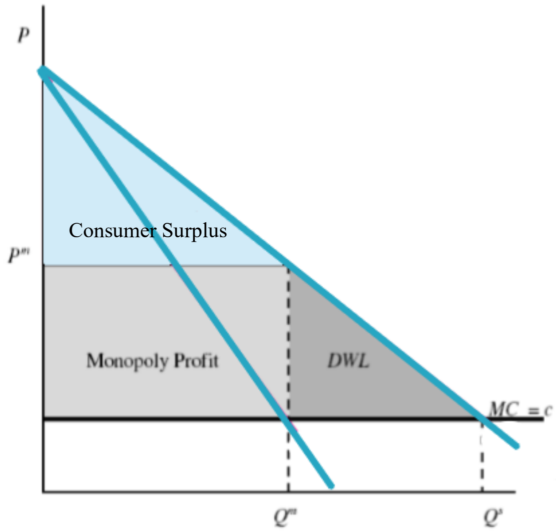

Figure 8: Profit-Maximizing Monopolist

[18] A ‘Hit and Run’ strategy is used by firms to enter a market when profits are high, and then leave when the profitability reduces. For example, if the incumbent practices the monopolistic price the new entrant may enter the market with a lower price, take all the sales and then leave it before any retaliation takes place. Such a strategy makes sense where there are very low barriers to entry and exit.

The task of policymakers is clearly to minimize the deadweight losses for society. The application of this economic mechanism shows the failure in data-driven markets of having one single entity with strong data ownership and the right to exclude the others from the access of data. Thus, as concluded in the Working Paper (2016) by the European Commission's in-house science service "wider access to data is required to maximize benefits". Open data, widely accessible by everybody, may prevent a data economy dominated by few monopolies.

2.2 Taxation of Digital Monopoly Platforms

With a market value of USD 961.3 billion, Apple leads the ranking of the world's largest companies (Statista,2019)[19]. Amazon, Alphabet, and Facebook follow closely behind, representing the most valuable listed firms in the global economy. However, they are well known for their low levels of taxation and ability to design worldwide fiscal strategies to take advantage of the difficulties of the national and international tax law in facing the new challenges of the digital revolution. Thus, taxation is having a hard time to keep up with the pace of change in the digital economy and to capture the value created. This represents a source of losses for tax policy, one of the most important instruments of industrial policy. It emerges the need for a 'call for action' to regain the lost tax incomes that governments need to promote fair competition in data-driven markets against monopoly tendencies and to support growth and innovation.

2.2.1 Problems of The Traditional Taxation System

The difficulties national and international laws are having, while facing the digital revolution, are evident both in terms of direct taxes (corporate income tax, local business tax) and indirect taxes (ad valorem tax). The reasons are in both the characteristics and dynamics of the digital economy emerging reality.

Value creation mismatch

One common feature of all tech-companies is their intense use of data from their users. This leads to the phenomenon called by Collin and Colin (2013) of *free labor.* Consumers, while using the services provided by platforms, disclose their personal data becoming virtual non-remunerating workers for these companies. Indeed, data is collected, stored and processed

[19] The 100 largest companies in the world by market value in 2019 (consulted on August 25th, 2019). Retrieved from https://www.statista.com/statistics/263264/top-companies-in-the-world-by-market-value/

to produce a better service to end-users, who have the double role of consumers and production auxiliaries. Blurring the distinction between production and consumption, giving no monetary compensation for the activity users perform it is certainly not fair that companies do not contribute with taxes to countries where their users "work" for free. Hence, the digital transformation is systematically disconnecting the place of business from the place of value creation: platforms can supply their digitals services where they are not physically presented or can scale without mass. Because of the contributes of remote citizens to revenues, there is a value mismatch between where profits are currently taxed and where they are created. Consequently, it is increasingly difficult to fix the location of value creation. The application of the traditional law, which taxes profits in the country where the company's head office is located and not where the company does business, appears with no doubts aged and inadequate. Besides, the lacks in the regulatory frame cause opportunistic behaviors of tech companies. They establish their physical presence in tax havens and pay virtually no taxes in other countries, despite doing a lot of business in those places.

No dividends policy

The second common feature, which can explain the low level of taxes on platforms' profits, is related to tech companies' dividend policy: profits are usually not paid out in dividends, but they are saved and reinvested in the business. Indeed, innovative firms tend to invest as much as possible in future growth, by financing through their own earnings projects of innovation. Historically in the tech world, except Apple and Microsoft, all major companies (Facebook, Netflix, Amazon.com, Alphabet) have never paid dividends to shareholders despite their undiscussed market cap. Seen as negative signals of the end of a period of growth, something for mature sectors and not for tech pioneers, dividends represent another form of tax savings for digital giants.

Intermediary's business model

Finally, digital companies are disrupting all the value chains of traditional sectors (travel, telecommunications, healthcare, urban services, banking). Acting usually in all different industries as intermediaries, the main effect is to reduce the share received in the past by other players in the value chain. As the digital revolution continues to spread in the entire economy, profits from different sectors will be sent offshore captured by tech giants, disappearing from the GDP of countries and stopping their growth.

Nonetheless, the digital revolution has brought growth and productivity gains. What is needed is a tax policy to support this transition, ensuring that these benefits lead to the organic development of all society. Therefore, they should contribute their fair share to this effort paying taxes. It is no more sustainable a model in which digital companies are the only ones who continue to take advantages of data to either scale their businesses or increase their own profitability. Correcting the tax advantage enjoyed by digital economy companies must become an urgent theme to be addressed, on the agenda of governments' industrial policy.

2.2.2 Taxation on Data Collection

To overcome such impasse as platforms cannot be taxed based on their profits, one possible solution is linking taxation to the collection and use of data in a given country. The idea has been studied in detailed by Bourreau et al. (2018), who develop a model in which a monopolistic platform provides personal services to users and sells advertising spaces to online sellers. It is a standard two-sided business model in which users disclosing personal data benefit of being targeted with relevant ads, on the top of having personalized services. Instead, online sellers gain access to the large audience of the platform, increasing their advertising campaign efficiency as more personal data is been collected from each user. Clearly, data represents an essential input for the monopolistic platform obtained for free. Thus, the proposal to put a tax on data that would allow capturing part of the dominant platform's profits.

Bourreau et al. (2018) in their model consider a world in which each economic transaction is subject to a value-added tax (VAT). However, they distinguish two different situations. In the first scenario, the platform charges only the online sellers for ads. This is the typical business model of platforms like Google or Facebook, which offer free access for users and generate revenues only with advertisements. Instead, in the second scenario, the platform charges also users with a subscription or monthly fee. This is the case of Spotify or Netflix, which extract revenues from both advertisers and end-users.

A free-for-users platform

In this case, the introduction of a tax on data collection brings the platform to raise prices for the online sellers, the only side to be charged under this business model. The increased prices reduce the number of advertisers, accordingly reducing the number of users because of cross-sides network effects. Less volume of data collected results in lower targeting efficiency and a reduced number of ad-generated transactions, thereby reducing the VAT returns. This tax base interdependence limits the effect of the introduction of the tax on data collection.

Nevertheless, if the introduced tax *t* is lower than the VAT rate the total fiscal revenues increase. But if t > VAT, fiscal revenues, together with consumer surplus and sellers' profits, also decrease with a general reduction of social welfare. Finally, when the VAT rate is high enough the effects of taxation on data on the overall social welfare involve a trade-off between larger fiscal revenues and lower consumers surplus and sellers' profits.

Two-sided financing platform

In the second scenario, not worried about competition, the monopoly platform distorts again its prices in response to the tax increase. The increase in the subscription fee for consumers and the price of ads for online sellers reduces the number of both sides of the platform, thereby reducing the ad-generated transactions. In terms of social welfare, consumers and advertisers are hit by the tax on data collection. On one hand, the total fiscal revenues increase because of the small tax on data and the VAT higher gains from the increased subscriptions. On the other hand, they are lowered by the decreased in number ad-generated transactions. If t > VAT fiscal revenues also decrease, and social welfare is unambiguously reduced. When the tax on data is low enough to be t ≤VAT, the first effect on fiscal revenues dominates and their total increases, with the aforementioned trade-off between larger tax incomes and lower surplus for consumers and sellers.

Thus, Bourreau et al. (2018) conclude that creating an ad-valorem tax on data is a week way for governments to try to capture a big share of the digital value creation from platforms. Indeed, when the introduced tax is small enough, the increased fiscal revenues come mainly from a higher fiscal burden bore by consumers and sellers because of the monopolist ability to distort its prices. Certainly, interesting would be the case with competition in which the platform will not just raise its prices, but it will bear itself the increased fiscal burden.

2.2.3 Taxation on Advertising

In the attempt to find an alternative way to raise additional fiscal revenues from the dominant platforms, Bourreau et al. (2018) in their work explore the possibility to introduce an ad valorem tax on advertisings, which represents the major source of incomes.

In the first scenario, where the platform charges only the advertisers, an ad valorem tax generates additional fiscal revenues without tax base interdependence effects. Indeed, if marginal costs are negligible the platform has no incentives to raise advertising prices as this will imply higher taxes to be paid. Thus, with the introduction of such ad valorem tax on ads and under a hypothesis of rigidity of the demand, the volume of ad-generated transaction won't decrease, adding to the VAT proceeds new fiscal incomes.

In the second scenario, as a small ad valorem tax is introduced the platform will react by changing the balance of its business model. As revenues from advertising are taxed, it will decrease the price of ads for online sellers and charge a higher subscription fee for users. This will turn in higher participation of sellers and a lower number of users. Yet, the reduced number of users will not shrink the number of transactions, thereby reducing VAT proceeds, as more sellers on the platform will create benefits for them eager to pay more for the subscription. Once paid the fee, they will transact more because of the enhanced value delivered to them by the cross-side network effects of the greater participation of sellers. Thus, the small ad valorem tax on advertising will unambiguously increase tax revenues by adding its own proceeds to increased VAT proceeds, which consist of the VAT collected on subscriptions and the VAT collected on transactions. The former increased by the value of the fee, the latter by the number of economic operations.

Finally, given that the most of business models involved in digital economy rely on target advertising based on data collection, an ad valorem tax on advertising seems to be a preferable policy for governments to extract more fiscal profits from the giants of the digital era.

2.3 Data Ownership

Most goods in economics are rival. Material goods, by nature, can't be used at the same time by different individuals. If a person eats an apple, the same fruit cannot be consumed by someone else. Thus, the consumption by one subject prevents simultaneous consumption by others. However, with the advent of the digital era and the central role played by data in the economy, non-rivalry has become increasingly common and easier. Indeed, data and information are non-rival: any number of firms, people, or machine learning algorithms can use the same data simultaneously without reducing the total amount of data available to anyone else. Moreover, non-rivalry has been made easier because of the reduction in the cost of copying and transmission of digital contents. In the analog era, data ownership was not a big issue because of the natural barriers of excludability conferred by the significant costs to copy and re-use analogic datasets. Nowadays, low-cost digital information technologies have raised questions about the protection of ownership. Clearly, ensuring the excludability of intangible and non-rival goods require a well-developed legal system. Yet, this is far from the actual status of data ownership rights' framework.

The regulation about data basically rests on two pillars. First, the Database Directive (1996) that attempts to harmonize the copyright-rule applied to original databases and to create a new sui generis right to protect non-original databases. Second, the General Data Protection Regulation GDPR (2016) that establishes constraints on the collection, aggregation, and processing of data, although it does not offer further clarifications about a more precise allocation of ownership rights over data. Thus, the legal debate about data ownership has been focused on a double face problem:

- granting exclusive ownership rights to data;
- granting access to exclusively owned data.

Indeed, ownership gives the exclusive right of usage, but also the right of exclusion. The data owner establishes an actual monopoly on decision-making regarding the use of data. He may block the others from access, thereby destroying the potential gains of sharing data, given its property of non-rivalry. Finally, the ownership theme opens up on one hand concerns about privacy and personal data protection; on the other hand, raises considerations about competition and losses in social welfare.

2.3.1 Data Ownership Fragmentation

Exclusive ownership rights will transform a common good as data, characterized by the property of non-rivalry, into private property. Privatization may go too far if ownership is excessively fragmented, transforming the tragedy of the commons into a new anti-commons problem.

The two tragedies are symmetric, where anticommons property can be considered as the mirror image of commons property. They indeed represent two sides of the same problem: an excess of rights over the same resource. However, the tragedy of the commons arises when multiple individuals have the rights of usage. While the symmetric anti-commons problem emerges as more than a single agent has the right of exclusion. In the anticommons tragedy, the waste of potential economic value takes the form of underutilization, consistent with the overutilization waste in the conventional commons tragedy. Moreover, in both cases, the size of these losses is greater as the number of agents with simultaneous rights increased.

This symmetry is graphically shown in Fig 9[20]. The social optimum is obtained with the figure of the monopolist or benevolent planner. As the number of users (in the commons tragedy) or excluders (in the anticommons tragedy) increases, there is an equal reduction of generated value, which derives from the strategical interactions of multiple agents. Each individual has the so-called free-riding behavior, generating negative externalities and diseconomies to the others who hold similar rights. Such egoistic conduct produces an

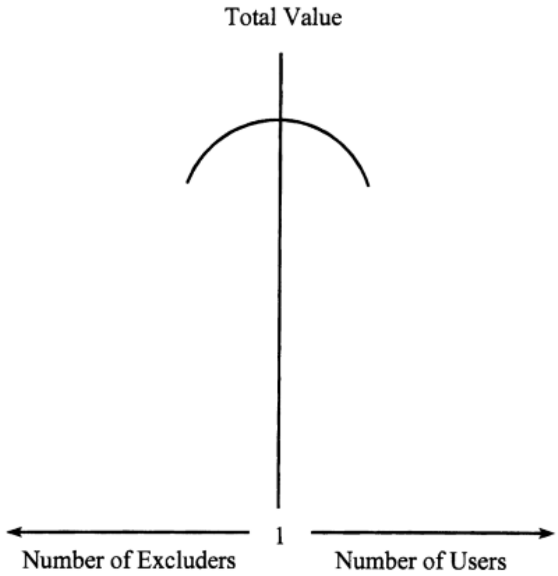

Figure 9: Commons and Anticommons' value symmetry

allocative inefficiency in which each subject tries to take advantage over the others, thereby damaging himself and the total value generated.

Thus, the logic of the two models is basically the same. In the commons side of the model, the right owner by adding one unit of the input of the common resource reduces the productivity of the others and their gains. Conversely, in the anticommons side of the model, the right owner by reducing one unit of the input (via prices) of the common resource reduces gains available for the others. Despite all the emerged symmetries between the two tragedies, in the past economists have concentered much more attentions on the common side of the problem. In 1968, Hardin with his work had already placed the focus of public attention on this theme. While, only in late 1998 Michael Heller introduced the concept of anticommons to explain the building situation of Moscow in the post-socialist period, where all the empty storefronts were a canonical example of underuse. However, the urgency and pervasiveness of

[20] The picture is taken from Buchanan, J., and Y. Yoon (2000), *Symmetric tragedies: commons and anti-commons*. Journal of Law and Economics, 43(1), 1–13.

the anticommons setting in many areas of modern economic interaction have given new importance to the theme.

Back to our discussion about data, the attempts of policymakers of clearly defining ownership rights in the past years seem to forget the related tragedy of anticommons. The uncoordinated exercise of exclusion rights leads to under-utilization of data. This is particularly problematic in case where economies of scope cannot be realized because of ownership fragmentation. Each dominant platform tries to exclude the others from realizing the potential benefit of data aggregation. The reduced interoperability and data sharing will turn in social welfare losses for all society, which is not fully exploiting the value of data.

The emergency for policymakers to rethink the legal framework of data ownership seems clear. A model in which tech giants try to take advantage, in their own myopic interests, of data utilization by reducing the potential benefits of data aggregation and thereby the growth opportunities and innovation advances for the society as a whole, needs to be reviewed.

2.3.2 Proposal Solutions in Literature

Studies in literature have been focused on who should own the property rights over collection and distribution of data between firms and consumers. The concern clearly shows that Coase theorem [21](1960) does not apply. However, to reach Coase's economically efficient solution, regardless of the initial distribution of property rights, the markets should be perfectly competitive with no transaction costs, perfect information and no market power differences. These assumptions are hardly verified in the real world, thus the concerns about data ownership are well-founded and deserve further research.

Jones and Tonetti (2018)

In their article, the authors develop a model in which data is a byproduct of consumption. They emphasize the non-rivalry of data, which is formulated by their ‘learning by doing’ and is captured by the equation:

$$D_{it} \leq \alpha x_{it} J_{it} + (1-\alpha) B_t \tag{6}$$

[21] The Coase Theorem states "that when there are conflicting property rights, bargaining between the parties involved will lead to an efficient outcome regardless of which party is ultimately awarded the property rights, as long as the transaction costs associated with bargaining are negligible.”

The amount of data D_{it} used by the firm i to help in its production is the sum of two terms. The first term J_{it} is the data that the firm gets from its business, which may be restricted if consumers own this data (x_{it}<1). While the second term B_t is the shared data from other firms, aggregated into a bundle, that can be used by firm i and any other firms simultaneously without being depleted. Finally, the weights α and $(1-\alpha)$ express the importance of own versus others' data. In such an economic environment, where the starting point is the non-rivalry of data, different allocations of property rights are analyzed to study the effects on production output, consumers privacy, and general welfare.

The first allocation examined is the one where firms own data. As we have seen from (6) data can be used by the firm that has generated it, but it can also benefit the other firms, because of the non-rivalry. Indeed, when firms have property rights, they tend to limit the data sharing, with losses for social welfare. Moreover, they may not properly respect the privacy of consumers. In the second allocation considered the government controls the ownership right. Jones and Tonetti observe in this case the improvement of consumers' privacy protection. However, the costs associated with the anticommons tragedy of underutilization of data are even bigger than in the first scenario.

Finally, it is analyzed the case in which consumers are the property owners of rights on their data. They result good at balancing privacy and data sharing with market outcomes nearly close to the optimal case of the social planner, who shares the amount of data which maximize the gains in productivity for all the firms, taking into account the consumers' privacy concerns. Jones and Tonetti suggest to policymakers a legal arrangement in which consumers own the data associated with their purchases, and they are the one responsible for selling it to data brokers or third parties generating gains from their personal information for themselves and the entire society.

Dosis and Sand-Zantman (2019)

The authors focus on the same topics, however reaching different conclusions. They develop a two-stage model in which a monopolistic platform offers a service to its customers, who paid a fee to access the service. While interacting users disclose their personal information, which should be used and monetized in the second period by the platform. As a matter of facts, consumers suffer a privacy cost. However, they can be divided into two groups:

- Users with a *low willingness to pay* for the service, they have a low privacy cost and they have no problems if their data are collected and monetized by the platform; generally, they are the ones who opt for low usage of the service.
- Users with a *high willingness* to pay for the service, they have a high privacy cost and they are not willing to disclose their data. They opt for high usage of the platform.

Thus, the platform has two strategies: either it relies on usage and subscription fees, thereby reducing revenues from data collection or it relies on data monetization by reducing incomes from usage and subscription fees.

It is a world where property right matters because of Coase's theorem inapplicability. Indeed, there is a commitment problem where the data owner is not able to commit ex-ante the level of ex-post data collection and monetization. Thus, if firms own the rights, they will exploit all the data collected from users, who will opt for low usage. While, if users are the owners, as suffering a privacy cost, they won't allow any data collection and thus they will opt for high usage. Under these conditions, if the value of data is high is more convenient for both users and firms, if platforms own the rights. Indeed, the larger the value of data the more websites have incentives to lower subscription fees to obtain a larger customer base and compensate the users to opt for usage. Consequently, the firm's profit will be higher because of the gains obtained by high-value data and consumer's surplus will be higher as well because of the economic incentives obtained for joining the platform and their increased usage. However, if the value of data is low the opposite holds. The best solution for both consumers and firms is the scenario where consumers own the rights. Indeed, the cost of distortion in usage offsets any benefit from monetizing data because the platform does not experience large gains from collected data. Indeed, the website will have no incentives to lower its prices, customers will suffer privacy costs and they both will be worse-off if the platform were the rights' holder.

Sand-Zantman et al. go further with their analysis. They imagine a market for data, which opens after the generation of data but before the data monetization.

Figure 10: Timing of the game

They demonstrate that when users own the data, low types with low privacy costs negotiate to opt for low usage, whereas high types do not and opt for high usage. Instead, when the platform owns the data, low types opt for low usage and do not trade because they are not willing to pay for a low privacy cost; whilst high types trade and opt for high usage. Thus, the presence of a market for rights benefits both users and firms. It partly addresses the non-commitment problem, allowing customers to follow their natural tendencies: low types toward low usage and high types toward a high one. Despite increasing the general welfare, the presence of the market for rights does not change the conclusions. The authors confirm their value-based allocation of rights with consumers as rights' owners for low values of data and firms in case of high data values.

Hence, the two proposed solutions from literature seem to come up with conflicting results. Tonetti et al (2018) with their idea of customers as data owners, regardless the value of data versus Sand-Zantman et al's (2019) different data ownership rights allocation according to the value of data. Despite these apparent discrepancies, a deeper comparison shows the different boundaries conditions of the two models. Sand-Zantman is focused on a monopoly scenario, where the absence of competition and multiple dominant platforms as owners of rights over data turns into myopia in considering the anticommons tragedy related to the underutilization of data. Thus, probably for future research, it will be interesting to merge the two analysis studying into Tonetti's competitive environment Sand-Zantman's ownership considerations according to the value of data.

2.4 Data Portability

In the era of digitization and digitalization of virtually all areas of business, society, and life, the theme of data portability has acquired a central role in the public debate. In the EU, the right of data portability has been formally stated in GDPR Art. 20[22]and it basically allows transferring personal data from one organization to another one. Before moving on the market outcomes of this free flow of personal data, it is important to take a deeper look over it. A first point should be the assessment of which data falls under the right of data portability. Obviously, it concerns personal data which are actively provided by users and observed data, indirectly provided by data subjects when using the services or devices (e.g. location, web search history, activity logs, etc.). Instead the inferred data, obtained by subsequent analysis, usually does not fall under applicability.

[22] The first paragraph of Art. 20 GDPR states that "The data subject shall have the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format and have the right to transmit those data to another controller without hindrance from the controller to which the personal data have been provided"

The second paragraph of GDPR Article 20[23] contemplates having a direct transmission of information between organizations, when technically feasible. Thus, for real efficiencies, data portability requires systems interoperability, as the possibility of exchange information due to similar technologies used by controllers. Finally, the right to data portability does not imply the right of erasure or to be forgotten. Once exercised the right to port their data, users can still access the services of the original data controller, without having lost any information. Indeed, the exercise of this right does not necessarily imply a switch of services, but there are different reasons why users would like to take advantage of the presence on several platforms at the same time. Thus, the emerging phenomenon of multihoming in the digital era.

2.4.1 The Theoretical Literature on Switching Costs

The main goal of data portability is to give subjects more control over their personal data, yet the direct economic outcome is lowering the switching costs. Indeed, personal data entered to benefit one particular online service offered by a given platform may create a lock-in effect for users, as switching to another service provider will imply the cost of re-entering all the data. Thus, one could wonder why public policy should be so focused on promoting data portability? Do switching costs make markets less competitive? Mainly, answers to these questions are offered by Klemperer's several studies, one of the most important names for literature on switching costs.

Firstly, Klemperer (1987 b) distinguishes between three types of switching costs:

- *Transaction costs*, that represent the costs to move from one data controller to a functionally identical online service offered by another firm. General examples of this category of costs are offered by online storage services where consumers' personal data, photos, contacts need to be reuploaded or banking accounts where switching implies to reinsert credit card information or online streaming services where personalized contents need to be recreated. For instance, users cannot move their data from Facebook to Google +, transfer documents from Dropbox to OneDrive, or move playlists from Spotify to Deezer;

[23] "In exercising his or her right to data portability pursuant to paragraph 1, the data subject shall have the right to have the personal data transmitted directly from one controller to another, where technically feasible."

- *Learning costs* are related to the time-consuming activity for users to adapt and understand how a specific service or cloud computing environment works. Usually, the knowledge required to use one service provider may not be transferable. This, in turn, leads rational consumers to display brand loyalty when faced with the choice between two functionally identical services;
- *Contractual costs* are the only one, among the three aforementioned categories of costs, to be artificial and entirely created by firms' discretion. Indeed, transaction and learning costs represent real costs of switching between service providers, whose size can be influenced by firms. Conversely, contractual costs are the result of strategic choices of firms, who decide to create 'frequent-flyer' programs or discount coupons valid for the next purchases, generating incentives to remain with the same firm.

Each of these types of switching cost concurs to transform products that are ex-ante homogeneous into ex-post purchasing heterogeneous, giving to firms some forms of market power over their existing customers, and thus creating the potential for monopoly profits (Klemperer, 1995). Therefore, in his work, he deepens the effects on the market competitiveness in the presence of switching costs.

Higher prices

The starting point of Klemperer's work in 1995 was his previous research on a two-period model of duopoly for homogeneous products (Klemperer, 1987). In this model, consumers initially choose among the two firms, in the second period they will face costs associated with changing supplier. Therefore, higher switching costs will turn in more customers locked-in to repeat purchasing from their first choice. The reduction in users' sensibility to prices, as they are less tempted by a price cut to change supplier than in a market without switching costs, and thereby in firms' elasticities of demand creates an incentive for firms to raise their prices. Thus, in the second period, firms usually will fix higher prices than in the first one to exploit revenues from their locked-in customers. Because of the second-period exploitation dependence from the firm's ability to gain market share in the first period, at this stage of the game firms will compete more aggressively. They will tend to fix lower prices to invest in market share, that will be valuable for them in the future. However, rational customers will be able to anticipate that the lower the prices fixed in the first period, the higher the market share the firm will have in the second period and a firm with a high market share will raise prices, focusing on exploiting its existing customers instead of attracting new ones. Thus, even in the first period consumers are less price sensitive. The presence of switching costs makes demand less elastic in both periods, boosting incentives for firms to fix higher prices than they would be without switching costs in every period.

However, a two-period model is not realistic and not so useful to analyze real competition which usually is repeated over many periods and characterized by the presence of an initial period where consumers have no ties to any particular firms, and following periods in which firms faced old and new users without the possibility to price-discriminate among them. In general, in period *t*th, each firm maximizes its total future profits:

$$V_t^F = \pi_t^F + \delta V_{t+1}^F(\sigma_t^F) \tag{7}$$

They are the sum of current profits π_t^F at period *t* and the discounted profits δV_{t+1}^F from period *t+1*, that will depend on period t market share σ_t^F.[24] Maximizing for its decisional variable, the price *t*th, the firm's first-order condition is:

$$\frac{\partial \pi_t^F}{\partial p_t^F} + \delta \frac{\partial V_{t+1}^F}{\partial \sigma_t^F} \frac{\partial \sigma_t^F}{\partial p_t^F} = 0 \tag{8}$$

Therefore, the firm in each period must balance two opposite incentives: the "harvesting incentive" and the "investing incentive". If greater importance is given to current profits, it will fix higher prices to exploit its current locked-in customers ($\partial \pi_t^F / \partial p_t^F > 0$). Nevertheless, there is a reverse incentive to lower its prices for attracting new customers and increasing the current market share ($\partial \sigma_t^F / \partial p_t^F < 0$), thereby boosting the future profits ($\partial V_{t+1}^F / \partial \sigma_t^F > 0$). These two opposite incentives may be exactly balanced, so that the optimum price, obtained by maximizing the first-order condition, will result not higher than the scenario without switching costs. However, several other effects bring the firm to raise its prices despite the potential losses in future profits.

First, the discounting effect $\delta < 1$ reduces the attractiveness of future profits compared to the present ones. Second, if the firm raises its prices today, competitors will gain more market share and they will fix higher prices next period with a less aggressive price strategy and a softened competition. Third, as already mentioned, consumers will have a less elastic demand

[24] Of course, V_t^F and π_t^F are themselves functions of the market share of the previous period σ_{t-1}^F. Yet, it is not shown explicitly to not abuse with formalisms.

than in absence of switching costs and thus, they will be less influenced by prices and more focused on products characteristics than in case they could switch providers without costs.

These three reasons, in addition to the previous discussion, bring Klemperer to conclude that a direct consequence of the presence of switching costs is increased prices and thus, a general reduction of the competitiveness of the whole market.

Product differentiation

Moreover, the presence of switching costs reduces product variety. Indeed, firms have fewer incentives to functionally differentiate their products: if they do that, some consumers may decide, despite the switching costs, to go by another supplier for these functional differences. If instead, products are only artificially differentiated by switching costs, functional reasons are never a reason to change supplier (Klemperer, 1995). Thus, the market outcome is that with switching costs prices are higher with homogeneous products than with differentiated ones, regardless of the traditional literature on differentiation as a way to reduce price competition.

Finally, Klemperer concludes that switching costs by raising prices, reducing incentives to product differentiation and creating deterrence to the entrance may reduce the market's competitiveness. In such a less competitive industry, firms may not be motivated to improve their efficiency using costly activities to produce and thus, generating more social losses. Therefore, this will justify policy interventions to reduce switching costs and promote competition. Nevertheless, the results obtained will depend on whether the switching costs are reduced proportionally or by lump-sum[25]. As observed by Bouckaert et al. (2012), only a proportional decrease in switching costs increases competition and both consumer and social welfare. Conversely, a lump-sum reduction decreases consumer welfare and if increases social welfare, it does that by increasing firm profits.

2.4.2 Multi-sides Market Implications of Data Portability

The standard single-sided market conclusions may change in the case of online platforms that have the characteristics of multi-sided markets. Data portability, by lowering switching costs and the number of users who are locked in into one particular service provider,

[25] A proportional decrease will imply that consumers with high switching costs will enjoy higher reduction than others. While, in a lump-sum decrease all consumers will enjoy the same cut, regardless of their initial level of switching costs.

means more competition, increased choices and lower prices. However, in the platform economy, other forces will reduce the effectiveness of lowering switching costs and will impact on the market's general competitiveness.

First, the central role played by network effects which represent one of the main sources that block-in consumers in a given platform. Even though data can be ported to join a different platform with a right to data portability, consumers may still be deterred in their decision because of the presence of network effects. Indeed, the utility the users derive from the already installed network on the actual platform may overcome even a potentially higher base utility offered by the other platform, if not enough users from the same side (direct network effects) or the other side (indirect network effect) are active. Thus, the right to data portability foster market entry, but the presence of network effects weakens its potential. Moreover, it exists a data context-specificity that is not solved by the right to data portability. For instance, let's think about the system of reviews or ratings used by the major platforms to build users' reputation e.g. Uber, or Airbnb, or selling reputation on marketplaces like Amazon or eBay. The process of building a reputation can be seen as a long-run investment that requires time and a certain number of transactions honestly completed in order to be achieved. Moreover, it is a user's investment platform-specific, that cannot be ported or transferred even in the presence of the right to data portability. Finally, even if not mentioned in our discussion so far, when it comes to data portability it is likely to raise issues of privacy and data security. For instance, some data transactions may include multiple subjects. Despite some of them could agree on porting their data, the others may not. Allowing the transfer, will turn into a privacy violation of the subjects involved and not compliant. Furthermore, more data restrictions could be evaded. If the limit is imposed by one platform, by porting data to another platform not having the same restriction, it could be easily bypassed.

In other words, multi-sided platform outcomes of data portability should be approached with more caution than single-market implications of reduced switching costs. The effects on foster competition and innovation need to be interpreted carefully and the rights involved such as privacy and security are fundamental ones, that must be protected.

2.4.3 An Economic Analysis of Data Portability

Wohlfarth (2017)[26] develops a Hotelling's model of horizontal differentiation to study the competitive effects of data portability in the digital economy. It is a two stages game with a unit mass of consumers uniformly distributed between zero and one (Hotelling, 1929) and two content providers CPs located at the extreme of this linear city (assume CP_A in x=0 and CP_B in x=1). However, CP_A is active in both periods, whereas CP_B enters only in *t*=2. Both the incumbent (CP_A) and the entrant (CP_B) have a business model, which monetizes on data collected about its own consumers. On the other hand, disclosing data implies a disutility for consumers, who suffer a privacy cost. Hence, Wohlfarth analyzes the incentives to the market entrance and the effects on social welfare whether there is a regime of data portability or not.

In *t=1,* a general user faces the choice of either becoming active on CP_A or remaining inactive on both CPs. Thus, in the first case a user located in x, who becomes active at CP_A, will derive a first-period utility of:

$$U_A^{t=1}(x) = \upsilon_A - \tau x - r_A \tag{9}$$

where υ_A is the base-utility offered by the platform, r_A is the disutility resulting from entering personal data and τ is the transportation cost to reach CP_A.

In *t=2*, CP_B enters the market so that users have to choose, assuming the full coverage of the market, between the two competing CPs based on the greater utility derived. A user located in x and active at CP_A will derive a second-period utility of:

$$U_A^{t=2}(x) = \begin{cases} \upsilon_A - \tau x - \theta r_A\,, & if\ U_A^{t=1}(x) \geq 0 \\ \upsilon_A - \tau x - \theta r_A - r_A\,, & else \end{cases} \tag{10}$$

where the strategical variable for CP_A is now θrA, the additional information required to users for being active in *t=2*. Among consumers, the ones already active in t=1 on the platform ($U_A^{t=1}(x) \geq 0$), they do not suffer a further disutility for the already disclosed information. Conversely, who join the platform A in *t=2* will have to provide both additional information and the previous one r_A.

[26] For further details see Wohlfarth, Michael (2017). *Data Portability on the Internet: An Economic Analysis,* 28th European Regional Conference of the International Telecommunications Society (ITS): "Competition and Regulation in the Information Age", Passau, Germany, July 30 - August 2, 2017, International Telecommunications Society (ITS), Passau

While a user located in x who decides to become active at CP_B, will derive a second-period utility of:

$$U_B^{t=2}(x) = \begin{cases} \upsilon_B - \tau(1-x) - r_B + r_A\,, & if\ U_A^{t=1}(x) \geq 0\ with\ d = P \\ \upsilon_B - \tau(1-x) - r_B\,, & else\ (d = NP) \end{cases} \tag{11}$$

Where the utility function shows that users need to insert all data required by the platform B, either if they were not active in period *t=1* or there is no data portability. Whilst, if users were active at CP_A in *t=1* and there is the possibility to port their data, they just have to reveal the net amount of information $(r_B - r_A)$ required by CP_B.

By solving $U_A^{t=1}(x) = 0$ for *t=1* and $U_A^{t=2}(x) = U_B^{t=2}(x)$ for *t=2,* it is possible to find $x^{*t=1}$ the indifferent consumer between joining platform A or not using any of the CPs. While the second equation gives the position of the two indifferent consumers between CP_A and CP_B, in the two different regimes with data portability or not $(x_P^{*t=2}, x_{NP}^{*t=2})$. Solving backward the game by maximizing platforms' profits, it is possible a comparison of the equilibriums in the two regimes.

What emerges from Wohlfarth's analysis is that under a regime of data portability the incumbent CP_A requires fewer data. Indeed, without data portability, a higher amount of information is used to create lock-ins. However, the possibility to port data vanishes this effect lowering the total amount of information required. Conversely, the entrant CP_B increases the collected data for users. Indeed, it is aware that under a regime of data portability users will have to disclose only the net amount of information $r_B - r_A$, due to the possibility to port the already entered data r_A. Thus, users will always suffer a lower disutility comparing with the scenario without data portability $r_B \geq r_B - r_A$, else being equal. As data represents the key production input and the source of revenues, clearly the incumbent reduces its profits while new entrants raise their profitability.

Data portability results in an efficient way to foster market entry, innovation and service variety as small entrants are more willing to enter the market due to the expectation of higher profits. Moreover, as the incumbent suffers from data portability lowering its profits, in case of a single firm in a dominant position this should be a way to foster a fairer competition in the market. However, the model suggests that users might be worse off with a right to data portability. Indeed, if both services offer a comparable quality then consumers' surplus increases. Instead, if their offer is asymmetric and the entrant has a superior value proposition, then the entrance will be profitable even without a data portability right that may be harmful to consumers. Nevertheless, a right to data portability unambiguously increase the total surplus. Finally, policymakers when introducing the right to data portability must carefully ask themselves if they want to foster market entry and innovation by making data more available, or they are purely focused on consumers' surplus. Moreover, it implies costs to be implemented for data-driven services already present in the market. Therefore, data portability requirements may have anti-competitive effects on SMEs and startups, who may reduce their investments into innovation because under obligation to fulfill requirements for portability.

To sum up, there are some trade-offs to be considered that suggest approaching the right to data portability with more caution before invoking for mandatory enforcement of the law.

2.5 Data Access

In the digital era of the data-driven innovation (DDI), where the potential from data collection, use, and analytics is evident in both its economic and social aspects, the access to and sharing of data has become a critical issue. The European Commission, in its studies about the European data market published in 2017, has forecasted that data economy is expected to impact for the 4% on the EU GDP by 2020, with approximately 10 million data workers.[27]

Indeed, data represents an essential resource for economic growth, job creation, and societal progress. It boosts labor productivity, fosters innovation in products, services, and business practices, thereby changing the rules of how economic value is created. These benefits are particularly evident in data-rich sectors such as science, health care, and transportation. However, it can be easily predicted, with the increasing use of artificial intelligence (AI), that the role of data will further gain more importance, even in traditional less data-intensive industries such as agriculture or construction. Thus, in a smart, inclusive economy data may play a central role in generating sustainable growth and innovation, thereby helping in the address of some global and social challenges. Enhanced access to data is seen as a way to disclose and maximize this value (OECD,2017).

A remarkable example comes from the field of science, on which in 2012 the European Commission focused its energy towards the purpose to reach an 'Open Science' by 2020. The initiative was part of the wider project Horizon 2020[28] , aiming to create a single more

[27] Retrieved from: https://ec.europa.eu/digital-single-market/en/news/final-results-european-data-market-study-measuring-size-and-trends-eu-data-economy

[28] Horizon 2020 is the biggest EU Research and Innovation program ever, with €80 billion invested in 7 years (2014 to 2020) to ensure the global competitiveness of Europe. For further details see: https://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020

competitive market for research, knowledge, and innovation. At that time, on the Digital Agenda for Europe has been set the ambitious goal to adopt an 'open data' policy covering all the scientific information[29]. The Commission in its communicate stresses the potential of enhancing access to scientific publications and data for boosting the benefits of public investments in research in terms of greater efficiency, faster growth, improved quality, and transparency. Indeed, easier access to previous researches would avoid duplications, thereby producing from the collaborative efforts faster to market and qualitatively better results. Furthermore, improving access to scientific information will increase openness and transparency towards citizens, more science-literate and aware of challenges in the 21st century.

Despite the evidence of the economic and social benefits coming from open access, it exists a reluctance to share data and legitimate reasons for keeping data "closed", that need to be addressed in order to realize this superior program. First, individual incentives and society interests are rarely aligned: data creators, bearing some costs, usually tend to privately take advantage of the value of data created. However, the most important barriers to data sharing are privacy concerns and the risks associated with disclosing confidential information. Stakeholders when sharing their information face some risks associated with data breaches, digital security, and violation of other private interests like trade secrets or personal data.

Finally, policy agendas need to create a governance framework, which may foster best practices to enhance wider access to data for the benefits of the economy and society as a whole, while addressing individuals and organizations' legitime concerns about data sharing.

2.5.1 Wider Access in Consumers' Perspective

We have already mentioned the positive consequences arising from increased data sharing for the entire society. Yet the existing perplexities of individuals about disclosing their information make the use of data still under its potential. However, considering the privacy cost suffered by consumers, is this reluctance well-founded and justified from an actual decrease in their welfare?

Valletti et al. (2019) develop a model in which there is a data broker that holds information about consumers, used for price discrimination in the retail markets. Their reluctance in disclosing personal information is endogenized in the model with a cost c they can pay to not reveal their information, thereby avoiding being in the database. Thus, consumers can be divided into two groups (see Figure 11), priced differently by retailers:

[29] See 'open data' package adopted on 12 December 2011, COM (2011) 882

- *anonymous market*: formed by all the new consumers plus the old ones who have paid the privacy costs;
- *personalized market*: formed by old consumers, who have not paid the privacy cost, whose preferences and willingness to pay is therefore known to the firms, able to make personalized offers to them, thus extracting all their surplus.

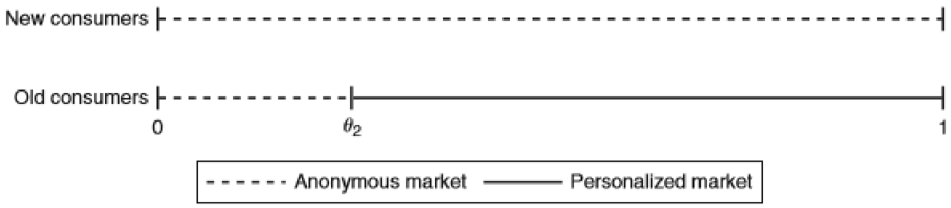

Figure 11: Division of consumers in the two markets

To obtain relevant considerations about consumers' surplus, which may be useful to address the opening question of the paragraph, Valletti et al. study two market scenarios, investigating the prices fixed by firms for their products, according to the availability of information about consumers. In the first case, the retailer firm is a monopolist. Whilst, in the second scenario there's a more competitive market with two competing firms who can have access to information. Under duopoly, the authors analyze different informational structure, depending on how many firms acquire access to consumers' data, thereby offering important insights. Fundamental assumptions, that holds in both scenarios, is the role of the data broker as a monopolist, who can sell its data in one unique block and the preferences θ of consumers uniformly distributed in [0,1], as in a standard Hotelling model of horizontal differentiation.

Monopoly Scenario

Assuming a monopolist located in x=0, who has bought information about consumers from the data seller, he will fix prices both in the anonymous and personalized market so as his profits will be maximized.

$$\pi(p) = \int_0^{\theta 1} p \, d\theta + \int_0^{\min\{\theta 1, \theta 2\}} p \, d\theta + \int_{\theta 2}^{1} p(\theta) \, d\theta \tag{12}$$

The first two terms represent revenues from the anonymous market, which can be split in sales to new consumers and sales to old consumers, who have paid the privacy cost. Instead, the last term represents profits from tailored offers $p(\theta)$ made in the personalized market. The price p, fixed in the anonymous market, is set by the monopolist as an estimation of the average willingness to pay of consumers in this market. Clearly, the presence of old consumers (who have paid the privacy cost) raises the average WTP. Thus, the price p fixed by the monopolist is increasing in the number of old consumers in the anonymous market, which corresponds to a low privacy cost c. As c increases, fewer consumers will pay that cost and the price p will consequently decrease. Notwithstanding a lower basic price p, monopolist' profits will generally increase in c. Indeed, as fewer consumers will afford to pay the high c, the revenues made in the bigger personalized market, where the firm is able to capture all the consumers' surplus, will overcompensate losses in the anonymous market.

Duopoly Scenario

In this case, Valletti et al. (2019) compare the results of two different informational allocations between the two competing firms. First, they study the situation in which just one of the firms, let's say firm A, has access to information about consumers. They show that if privacy costs are low enough (c$\leq \bar{c}$), the uninformed firm will find not profitable to compete in the personalized market. Thus, the profits of the two firms will be:

$$\pi_A = \int_0^{\theta 1} p_A \, d\theta + \int_0^{\theta_2} p_A \, d\theta + \int_{\theta 2}^{1} p_A(\theta) \, d\theta \tag{13a}$$

$$\pi_B = \int_{\theta 1}^{1} p_B \, d\theta \tag{13b}$$

As the cost of privacy c starts increasing, fewer consumers will pay for it. Thus, the informed firm A, that fixed its price in the anonymous market p_A as an average WTP between the WTP of new consumers and of the old ones who have paid the privacy costs, will thereby lowering p_A due to the smaller number of consumers with higher WTP. As a matter of fact, the uninformed firm B will be forced to reduce its price p_B to continue competing in the only market entered. Since p_B decreases, the tailored price $p_A(\theta)$, a function of p_B, decreases as well. Indeed, firm A has to leave more surplus to consumers lowering its price also in the personalized market, who otherwise will have incentives to buy from B. The fierce competition caused by increasing costs of privacy and the consequent reduction of prices, benefits consumers while reducing firms' profits. When the cost of privacy becomes high enough to exceed the threshold $c > \bar{c}$, firm B will enter also in the personalized market making the competition even stronger.

However, the best scenario for consumers is when both firms have access to information. Under this assumption, even in the personalized market firms will have the same basis to compete for every consumer. The enhanced competition will reduce the potential for rent extraction lowering tailored prices, that will be so low to not make any economic sense for no consumers to pay privacy costs. Therefore, an efficient allocation is achieved in terms of welfare. First, transportation costs are minimized since market shares gained by the firms in the two markets are symmetric. Second, there are no market inefficiencies generated from privacy costs.

In a nutshell, evidence from Valletti et al. (2019)' s research shows that policymakers should promote a non-exclusive allocation of consumers information among competing firms, demonstrating benefits of increased access to data. Reassuring consumers concerned about the consequences of sharing their preferences and personal data, where are the obstacles then to the realization of such informational allocation?

2.5.2 Pitfalls in Data Markets

The problem in Valletti et al. (2019)'s model comes from the data market, where the owner of information sells data about consumers to the two competing firms. It has full bargaining power acting as a monopolist in the data market and its optimal selling strategy consists to grant exclusive rights of access over the database to only one of the two competing firms. Indeed, it can either choose to sell the dataset to only one firm fixing the price T_1 or charge both firms the price T_2. In order to maximize its profits, equal to:

$$\pi_{DS} = max\,\{\,T_1, 2T_2.\} \tag{14}$$

the data seller DS sets prices with a reverse auction mechanism, in which the highest rent, extractable from participants, is the difference between winning and losing the auction for the subject. Thus, T_1 is fixed by the DS as the difference between the profits the retailer firm will have, being the informed firm or the uninformed one in the Hotelling game. Whilst, T_2 represents the difference in profits for one of the two firms between the case in which both have access to consumers information and only the other can use this data. As $T_1 > 2T_2$, the DS will pursue its own interests abetting an asymmetric allocation of information.

However, there is a strong assumption behind Valletti et al. (2019)'s modeling of the data market which is represented by the non-divisibility of data that can be sold only as one unique block. Therefore, making data a variable of the model to be set together with its price, is this informational allocation still optimal for the data seller? Are there any existing models in literature at odds with Valletti's selling strategy of data? In the attempts to answer these questions, we will concentrate further on literature about information acquisition and data trade.

2.6 Data Trade

The starting point for literature about data trade is Arrow's Information Paradox (Arrow, 1962). In order to complete the purchase, the buyer of information must have enough details to decide whether or not to buy. However, once the customer has this detailed knowledge an effect transaction of information has occurred without any compensation, and there is no longer any reason to pay for it. The chief point in his work is the difficulty in creating a market for information since the efficient allocation would be the use by its original possessor. Despite Arrow's noteworthy studies, his conclusions seem outdated and far from nowadays big data market. Moreover, in 1962, Arrow referred to "a piece of information as an indivisible commodity" and clearly this represents another old-fashioned assumption in today digital era. For instance, the process of digitalization has made data much more excludable, making the paradox no longer fully applicable, as data can be split up in packages and displayed in bits.

Currently, selling data is no more only granting access to a given database, but it takes multiple and much complex forms. Thus, personal data is increasingly conceived as a tradable asset, due to its potential in creating value-added for companies and costumers. A complex ecosystem of entities collecting, analyzing and trading this new asset has emerged. They mainly collect information about users from a variety of sources, to create detailed profiles to be sold to companies that in turn, use this knowledge to refine their market segmentation strategies. It is well known that Uber, Lyft and other ride-share companies have pricing algorithms, which dynamically change prices accordingly to demand and supply conditions. However, in 2018 an article on The Guardian denounces the case of two friends, who quoted a different price after calling an Uber to go to the same place at the same time.[30]While the case of Uber has been brought to public attention, every day an increasing number of companies engages in these practices. Consumers' purchasing stories are aggregated into scores that quantify and predict their willingness to pay, health risk, or creditworthiness. "Every consumer has at least one CLV score, more likely several" (The Wall Street Journal, 2018). The Customer Lifetime Value (CLV) measures the potential financial value of customers and once deployed internally by big

[30] In the same article the case of Dan, who noticed decreased prices after switching from his personal credit card to his corporate card in the Uber app. Evidently, for pricing algorithms, dad-Dan was not as a desirable passenger as corporate-Dan. For further details see: "Is your friend getting a cheaper Uber fare than you are?". The Guardian, 2018

data brokers, these scores are sold in data markets. The transmission of such information critically creates links across past transactions and future consumer's interactions with other firms. Moreover, scores are not regulated and not available to consumers, who thereby ignore the impact of their actions and links among transactions. In the absence of regulation, it is difficult for even more sophisticated consumers to behave strategically in apparently unrelated transactions. On these themes, Bonatti and Cisternas (2019) develop a model in which they show how the use of scores influences the consumers' willing to reveal their information, generating a *ratchet effect*[31]. It usually takes place when parties, involved in a long-term relationship with informational asymmetry, are afraid that the disclosure of its private information may worsen the future terms of trade. Hence, a strategic consumer, driven by ratcheting forces and aware of the consequences generated by the level of her consumption, manipulates future prices by reducing the present quantity demanded and thus lowering her score. To avoid losses in quantities demanded, profits and information collected firms need to provide incentives to consumers for their information disclosure.

Hann et al. (2002) present an interesting empirical analysis where they show that individuals' willingness to disclose personal information systematically vary with incentives. However, as consumers can be divided into three different segments -convenience seekers, information sellers, and privacy guardians- in terms of trade-offs between benefits of disclosing information and privacy costs, firms must address each of these segments with tailored incentives. Indeed, convenience seekers will be attracted by future convenience and simplified site navigation, information sellers by offering monetary rewards, and privacy guardians by enhanced procedural fairness and privacy protections. With consumer increasing awareness about businesses practices, and thereby concerns about disclosing personal information rising at a fast pace; privacy, together with lack of consumers' trust, has been identified as one the biggest impediment to the full deployment of data market potential. As the profitability of trading personal information is directly linked to the availability of such information, privacy policies and reward structures unequivocally play a fundamental role in trading personal information.

[31] First introduced in literature by Freixas, X., Guesnerie, R., & Tirole, J. (1985),they explain the ratchet effect with the problem of central planning production. Usually, the firm has more information about its productivity than the Central Planner (CP). Thus, the CP uses incentive schemes, reviewed over time, for acquiring knowledge about the firm's productivity. What emerges is that firms tend to underproduce in order to avoid more demanding scheme in the future.

2.6.1 Information Selling in Financial Markets

The issue of how to optimally pricing information has been firstly addressed in finance literature. The general problem faced in the financial context is how a monopolist can sell information to agents, who in turn can use it in the subsequent market to buy stocks. The focus on the given market analyzed introduces some specificities (e.g. considerations about risk tolerance of agents, the reduced value of information as more traders have access to information). Yet, the general mechanisms of information sales can be used in many other contexts, reading the results in more general terms.

Admati and Pfleiderer

The first contributions come from Admati and Pfleiderer's researches. In one of their earlier work "Selling and Trading on Information in Financial Markets" (Admati, & Pfleiderer 1988) they study the case of an information owner who wishes to sell it in a financial market, finding the optimal selling strategy to maximize its profit. Evidence from their analysis demonstrates that the results depend on the amount of risk the seller is willing to bear. Indeed, if the seller is a risk-neutral subject then his profits will decrease as the number of informed agents increased. Instead, if the trader is a risk-averse then he may benefit from the increased number of informed agents. Intuitively, the spread of information brings to increased competition among traders, therefore resulting in lower profits for the industry. However, in some cases, the benefits of risk-sharing can compensate for the losses caused by competition.

Two years later (Admati, & Pfleiderer 1990) they imagine a different scenario, where there is a monopolist that faces two ways of selling information. As the only owner of the information, the seller can either provide buyers informative signals about the stock or create a portfolio based on her private knowledge. The former is a direct sale of information, after observing the signal, the buyer can buy herself the stock. The latter is an indirect sale, the data is never transferred, buyers can invest in shares of the portfolio being guided to take better actions in the financial market based on the seller's knowledge, without anyway giving away any data. The optimal selling strategy is the result of the seller's trade-off between surplus extraction and control over the usage of information. Indeed, the monopolist in the direct sale can extract all the surplus the information generates by fixing a fee paid by buyers. However, because this surplus is related to the difference in expected utility between informed and uninformed buyers, as the number of informed agents increased, due to negative externalities, the value of information decreases[32]. Indeed, to maximize such surplus the seller would generally like to control the usage of information. This might be done in an indirect sale, where the owner can fix a price for ‘usage of information’ in the form of the per-share price of the fund. Indeed, by charging each investor a price that is a linear function [33]of the number of shares the investor buys, the information owner effectively charges for the usage of his own knowledge.

As a result, in situations where negative externalities are strong, it is more valuable for the monopolist to restrict the usage and control it efficiently with indirect selling of information, to avoid a reduction of the total surplus by excessive usage of data. Whilst, when the usage of information has less effect on the total surplus the direct sells represent a more profitable selling strategy for the monopolist.

Garcia and Sangiorgi (2011)

Garcia and Sangiorgi, more recently have studied the same topic of selling information in financial markets focusing on risk considerations. Indeed, in their work, the main trade-off faced by the seller of information is between risk minimization and profits maximization.

They imagine a modern security market in which information is sold by a brokerage firm to investors. The optimal selling mechanisms are represented by two corners solutions. Either sell to as many agents as possible a noisy and imprecise version of information or sell to a single agent (or a very small group) a more precise version of it. The first solution minimizes the risk by splitting information among a great number of agents, who in turn will be owners of a very small portfolio. This is consistent with newsletters buy by small investors, who usually are more risk-averse. Instead, the second scenario of exclusive disclosure of information maximizes the expected trading profits by allowing a few well-informed agents to take advantage of such knowledge. Typically, these are hedge funds or mutual funds, who can afford expensive researches obtaining higher profits, yet bearing doubtless level of risks that cannot be easily afforded by single individuals. Finally, their results seem to confirm earlier findings of Admati and Pfleiderer’s works, thereby allowing us to extract some general considerations.

Summing up, what can be concluded from the analysis of information trading in financial markets is that wider access to information increased competition among informed

[32] Reasonably, as more agents will have the same information about stocks, they will tend to buy or sell the same stocks, thereby vanishing the competitive advantage given by acquisition of data.

[33] For instance, linear prices are the most commonly observed way in which mutual funds are priced

agents, thereby leading to a reduced industry profit. The loss of data value from usage in such markets derives from a major rivalry nature of financial information, that cannot be expanded to the market of personal information. Nevertheless, the impasse of Valletti's selling model still holds: the monopolistic seller of information, in the attempt to maximize its profits, will tend to an exclusionary allocation of information, thereby reducing competition and boosting revenues for the few informed agents.

2.6.2 Pricing of Information

A different vein of the literature has examined the sale of information to competing firms operating in an interdependent product market. For instance, Iyer & Soberman (2000) develop a model in which the sale of information corresponds to the sale of heterogeneous signals about valuable product modifications, to firms competing in a duopoly of differentiated products. Taylor (2004) studies the effects of selling a customer list that allows the acquiring firm to price discriminate among consumers, based on their purchasing history. Among these papers, this critical review focuses on Bergemann and Bonatti's work (2012), which is the most similar one to Valletti's selling mechanism of information.

Bergemann and Bonatti, 2012

The two professors from Yale and MIT University in their paper "Markets for Data" deepen the topic of data provision and pricing in an environment in which there is a monopolist data vendor, who sells access to his database to downstream firms. Competing firms can, in turn, use the information acquired to improve their product positioning engaging in price differentiation practices.

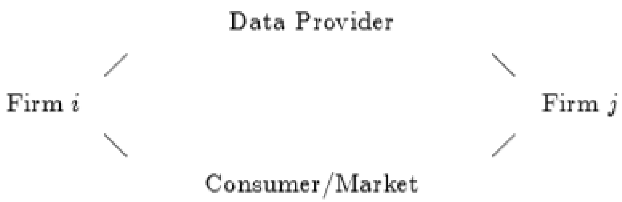

Figure 12: Upstream and Downstream Interactions

More specifically, they develop a Hotelling model of horizontal differentiation in which they study the firms' profits under different information structures. In particular, three regimes of information are analyzed:

1. The database provides no information to any competing firms, that will play a standard Hotelling game;

2. The database provides asymmetric information to only one of the two competing firms. The informed firm will be able to take advantage over the uninformed one, with tailored offers to consumers;
3. The database provides a symmetric, but noisy version of the database to the competing firms. For instance, Bergmann and Bonetti imagine a data seller's capacity to add some noises to its database, that will offer precise information about consumers located close to the acquiring firm, but only a very imprecise version of information about consumers located far away from it. With this bipartite information structure, each of the firms will be able to construct local monopolies extracting all the surplus from the closer costumers, while acting as a weak competitor for remote consumers not having any information about them.

It is a two-stage game, where in the first stage the data seller offers the database and the two competing firms decide whether to buy or not the information and in the second stage they compete under one of the three different information structures, which represents three different subgames following the data purchase. The normal form of the game at stage 1 is given by:

	buy	not
buy	$\Pi_F - p_D,\ \Pi_F - p_D$	$\Pi_I - p_D,\ \Pi_U$
not	$\Pi_U,\ \Pi_I - p_D$	$\Pi_H,\ \Pi_H$

Figure 13: Normal representation of the game at stage 1

In the matrix π_H represents the Hotelling profits realized when no one of the firms has access to the database (subgame 1). π_I, π_U denote the profit of the informed and uninformed firm with asymmetrical information (subgame 2). Whereas π_F represent the profits realized under a symmetric, but biased distribution of information.

Contrarily to Valletti's auction strategy price, in this model the data company can make sequential offers to the firms, thereby exploiting the strategic complementarity of the purchasing decisions of the two firms. Supposing the first firm reject to pay a general price p_D. Thus, the second firm as long as $\pi_I > \pi_H$ will be willing to pay the price $p_D = \pi_I - \pi_H$ to obtain the information advantage. Therefore, now the first firm will be willing to pay $p_D = \pi_F - \pi_U$ to avoid the resulting informational disadvantage. Thus, Bergemann and Bonatti conclude that the optimal selling strategy for the data seller, if it is able to make sequential offers, is to charge both firms with the price $p_D *= \pi_F - \pi_U$ (i.e. the price corresponding to the entire value of information when the other competitor is fully informed).

Finally, as firms have increased access to the same database in a perfect symmetric allocation of information, this will lead to fierce competition and reduced rents for competing firms. For this reason, the optimal selling strategy for the data provider, under this model assumptions, can never be the symmetric allocation of information among firms. Indeed, it is in the interest of the data seller to leave the firms with the higher profits as possible, that he will be able to capture with the price p_D, given his monopolistic bargaining power. Thus, the proposed solution of the subgame 3 is the most efficient, minimizing the transportation costs for consumers and allowing full coverage of the market, and the most profitable for firms, who will be able to extract monopolist rents from the closer consumers.

2.6.3 Structure of Information

The previous discussion about a symmetric but biased distribution of information opens another more recent strand of literature. Indeed, information is rich and can be modified in different ways creating from the same data different informational products. For instance, even one-dimensional data like a score about consumers can be turned into two different sub-products as some firms may be concerned about targeting the ones with excellent scores, while others may want to identify very low scores. This introduces the problem of versioning information and the design of the optimal information structure to be sold. Bergemann and Pesendorfer (2007) consider the problem of the design of an optimal information structure in an auction context. Later, Gentzkow and Kamenica (2011) discuss the same topic in a competitive environment where agents take actions based on the received information by the principal, who may influence the eventual agents' decisions changing the information structure. However, in this critical review, the more recent work of Bergemann, Bonatti, and Smolin (2018) is analyzed in more detail.

Bergemann, Bonatti, and Smolin (2018)

In their model information buyers faces a decisional problem: given a state of the world ω, from a finite set $\Omega = \{\omega_1, \dots, \omega_I\}$, and an action a from the set A = $\{a_1, \dots, a_J\}$, he seeks to match the state and the action with an ex-post matrix of payoffs as shown in Figure 14.

$$\begin{array}{c|ccc} u & a_1 & \cdots & a_I \\ \hline \omega_1 & u_1 & & 0 \\ \vdots & & \ddots & \\ \omega_I & 0 & & u_I \end{array} .$$

Figure 14: Matrix of payoffs

The potential buyers have some prior information; however, they desire to acquire more information from a data seller to increase the match probability state-action. Thus, their willingness to pay is determined by their prior information level (unknown by the data seller), which represents a new aspect of horizontal differentiation among consumers to be exploited with practices of price discrimination by the data provider. Besides, considering data as the input for the decision problem of individuals, the same information can generate different actions taken by subjects due to the prior information beliefs.

In such a scenario, the revenue-maximization information policy should allow to discriminate among consumers with a different prior level of information and then charge them accordingly. Indeed, buyers can be divided into two groups:

- low types: the ones ex-ante more informed, thus with a lower willingness to pay for more information;
- high types: the ones ex-ante less informed and with a higher willingness to pay, as the possibility to undertake the correct action without additional information is very low.

In a nutshell, the seller's problem consists of screening types in order to design and price different versions of experiments, which represent different information products from the same database.

Bergemann et al. demonstrate that the optimal menu consists of at most two experiments: a fully informative experiment for high types and a partial informative experiment for low types. The degradation of the quality of information allows extracting positive rents from low types while charging higher prices for high types. Indeed, high types would not attach a positive value to partial information because the noisy signal is too weak to correctly guide their action. Contrariwise, low types can improve the quality of their decision-making, paying a lower amount of money for the experiment. Finally, bundling the two quality of experiments the data seller is able to screen the buyer's initial information levels and to extract more surplus from both types, offering an informational product tailored to their needs.

Finally, we can state that in the literature about data selling some common traits are emerged: the presence of a monopolist data owner, the representation of data either as one unique dataset or as signals, the absence of costs in the activity of data collecting, and the general conclusion that an exclusive allocation of information is pursued by the data seller. Thus, once depicted the traditional data market's competitive scenario, we have tried to change the rules.

3. A Model of Competition with Endogenous Data Extraction

The literature on data competition, reviewed in the previous chapters, typically finds that a monopolist prefers to sell data to one retailer only, so as to maximize the value that it can extract from the downstream market. While this result is concerning from a policymaking point of view, as it practically reinforces the monopolistic trends of vertically related markets, it strongly depends on the assumption that data are an indivisible amount. The quantity of data purchased does not affect the price paid for it or the cost for its extraction. In this chapter, we assume that the monopolist can contract with the downstream firms over the quantity of data, as well as on the price for each unit of information.

We develop a two-stage model with three players: a data seller and two competing firms (firm A, firm B). The data seller represents the third-party data firms, those entities collecting, analyzing and matching thousands of attributes then used to create detailed consumers' profiles to be sold in the data market. Instead, the two competing firms represent the first-party data holders, who directly interact with users and can decide to buy additional information to improve the offered services. The data seller acts as a monopolist selling information to the two competing firms, that will play a competition game à la Bertrand[34]. The French mathematician assumes that competition is over prices and production follows the realization of demand. This seems reasonable in an industry environment, as the market of data, where there are no strong capacity constraints, but production can be easily adjusted to meet the effective demand.

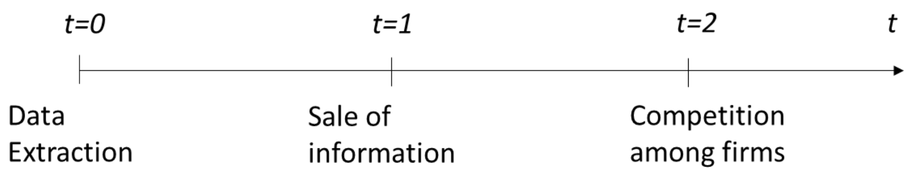

Figure 15: Timing of the game

The timing of the game is as shown in Figure 15. In period *t=0* the data seller collects, analyzes and stores information about consumers. This entails a cost of data extraction

[34] The model, first formulated by Bertrand in 1883 and the formalized mathematically by Edgeworth in 1889, was a critique to the previous Cournot's work (1838). In the Cournot's model producers simultaneously and independently make production quantity decisions, and then bring what they have produced to the market, with the market price determined as the price that equates the total supply with the demand. In Bertrand's model, on the other hand, producers simultaneously and independently name prices rather than quantities. This will lead to harsh competition, driving profits of firms to zero through a series of price undercutting.

$$C = F + cd\,, \tag{15}$$

where F is the fixed cost and c is the unit cost. Usually, data sellers are characterized by a cost structure with relatively high fixed costs F and low marginal costs c (Jullien, 2006). Indeed, costs are mainly related to IT infrastructure, computing capacity, and energy, which are all expenditure examples not very sensitive to the volume of generated data. Once extracted and aggregated, in *t=1* information is sold in the data market. In this first stage of the game, the data seller can decide either to sell information to both firms or to grant exclusive access to just one of the two buyers. Therefore, under these two possible informational structures, in *t=2* we have studied the market outcomes of the competition game in terms of bought quantity of data and realized profits by the two firms. The amount of available information is strictly related to the surplus delivered by the firm to end-consumers. Indeed, assuming the first-party as a service provider, the more information will collect about users the better will tailor its services. Thus, the utility derived by consumers will be an increasing function at a decreasing rate of the quantity of personal data collected about them i.e. $\frac{dS}{dd} > 0$, $\frac{d^2S}{dd^2} < 0$. The first amount of data will allow to segment different classes of consumers, based on general information such as sex, age, and location. The recognition of consumers will enable firms to reward them with personalized offers or target advertising, notably increasing their derived value. However, as the volume of collected data increased, more attributes will be added to the profile of consumers, who will certainly continue enjoying a tailorization of their online experiences, but in a lower measure. Formally, S is defined by the relation:

$$S = \upsilon + \sqrt{d} - p \tag{16}$$

where υ is the base-utility offered by the first-party and p is the price charged to users.

Finally, we have developed the model under two alternative hypotheses. First, data is sold by the monopolist with a linear tariff. Second, the data seller charges a two-part tariff[35]. Under both scenarios, assuming a perfect information game and a "common knowledge of rationality"[36] (Aumann, 1995) , we have solved the model with the backward induction logic.

[35] For instance, see paragraph 3.1 and 3.2.

[36] To let the backward induction work, each player must be rationale i.e. one always chooses the action maximizing his payoff. However, this is not enough. Players have to believe that all other players are also rational. Thus, common knowledge of rationality is a scenario in which players ascribe rationality to each other.

First used by Zermolo in 1913 as a cornerstone of his proof that chess has optimal pure strategies, backward induction has been long used in the game theory. It consists of first considering actions taken in the last period of time. The last player makes a choice that maximizes his payoff. Taking this in mind, the previous player makes his choice maximizing his payoff; and so on. This process continues backward until the beginning of the game is reached. Therefore, in our model, we start by considering the actions taken in t=2 by the two competing firms. Next, reasoning backward in time, the monopolist using this information may determine what to do in t=1.

3.1 Linear prices

Vertical relationships between manufacturers and retailers impact crucially on consumers' welfare, as the competition among competing firms is a result of the whole industry structure. Pricing strategies adopted by manufacturers, or in our case by the seller of information, are among the instruments used to build such vertical power relationships. We first start our analysis considering the scenario in which a data seller extracts a quantity d and sells it downstream at a uniform unit price z. Hence, the data broker may decide either to allow an exclusive allocation of his information in favor of just one between the two competing firms or to opt for a non-exclusive grant.

3.1.1 Asymmetric Information

In this section, we describe the case in which the data broker sells the data to one downstream firm. Suppose firm A is the only one to have access to the upstream market of information.

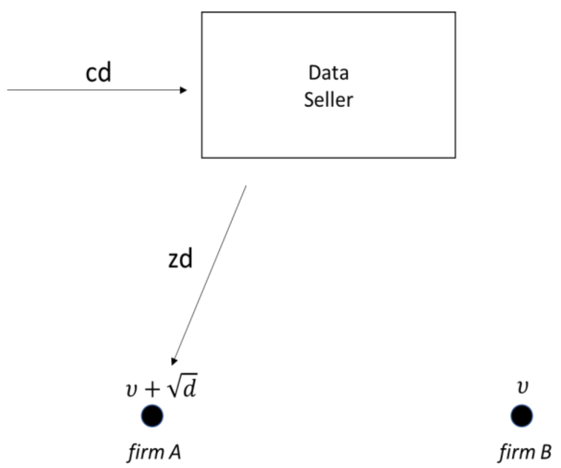

Figure 16: Data market under an asymmetric distribution of information

Under such an asymmetric distribution of information, the two firms will deliver a different utility to end-consumers. Firm B, excluded from the acquisition of information, will be able to only offer the base-utility. While firm A's surplus will be similar to expression (16). Thus, the provided surpluses by the two firms are respectively:

$$S_A = v + \sqrt{d} - p_A; \tag{17a}$$
$$S_B = v. \tag{17b}$$

However, to not vanish the value added by data with an excessively high price, firm A aims to be at least as competitive as firm B, matching the value delivered to consumers by its competitor. Hence, equating (17a) to (17b), we obtain the price p_A fixed in the downstream market to final consumers by firm A

$$p_A = \sqrt{d}. \tag{18}$$

That said, firm A will maximize its profits finding the optimal amount of data to buy from the data owner. Formally, π_A is defined by the relation

$$\pi_A = p_A - zd. \tag{19}$$

Writing $p_A = \sqrt{d}$ in (19) and calculating the FOC $\frac{d\pi_A}{dd} = 0$, it follows that the required amount of data is

$$\frac{1}{2\sqrt{d}} = z \rightarrow d = \frac{1}{4z^2} . \tag{20}$$

Moving backward to the first stage of the game, the monopolist has to fix the price of information maximizing his rent extraction. Given the data broker's profit

$$\pi_M = (z - c)d - F, \tag{21}$$

and putting (21) together with (20), the result is $\pi_M = (z - c)\frac{1}{4z^2} + F$. By solving $\frac{d\pi_M}{dz} = 0$, we obtain the price charged by the data seller to the downstream firm:

$$-\frac{1}{4z^2} + 2\frac{c}{4z^3} \rightarrow -z + 2c = 0 \rightarrow \; z = 2c. \tag{22}$$

Now substituting the optimal price in the quantity of data acquired by the retailer firm A, i.e. $d = \frac{1}{16c^2}$, we can finally express all the profits as the only function of c. Indeed, under these assumptions, we obtain data seller's profits equal to

$$\pi_M = \frac{2c - c}{16c^2} - F = \frac{1}{16c} - F, \tag{23}$$

firm A's gains equivalent to

$$\pi_A = \sqrt{\frac{1}{16c^2}} - 2c\frac{1}{16c^2} = \frac{1}{8c}, \tag{24}$$

and total profits of the industry satisfying

$$\pi_T = \pi_M + \pi_A + \pi_B = \frac{1}{16c} - F + \frac{1}{8c} + 0 = \frac{3}{16c} - F. \tag{25}$$

3.1.2 Symmetric Information

In this section, we consider the case in which the data broker sells the data to both the downstream firms. Under a symmetric distribution of information, both competing firms have access to the market of data.

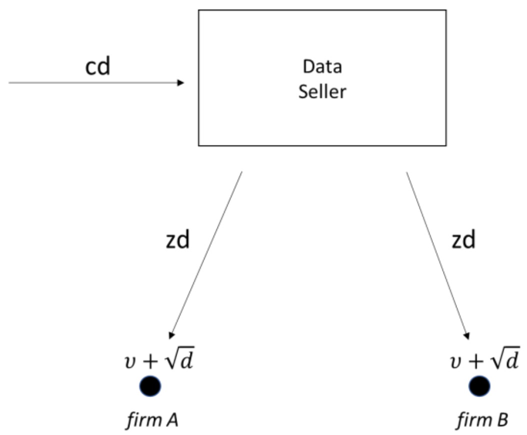

Figure 17: Data Market under a symmetric distribution of information

Assuming the downstream market as a duopoly of homogeneous products, so that consumers' decisions are solely based on prices, with an infinitely price-elastic demand, neither firm A nor firm B will set a higher price than its competitor. Indeed, this would yield the entire market to the rival. As in a classical Bertrand's paradoxical scenario, the two players reach a Nash equilibrium when they both charge prices equal marginal costs, driving their profits to zero. Thus, we have:

$$\pi_A = \pi_B = 0 \rightarrow p - zd = 0 \rightarrow p = zd\,. \tag{26}$$

Moreover, under a common information knowledge the two competing firms will provide the same surplus to the final consumers $S = v + \sqrt{d} - p = v + \sqrt{d} - zd$. As one would aspect, the optimal amount of bought data by each firm satisfies $\frac{dS}{dd} = 0$ and is equal to

$$\frac{1}{2\sqrt{d}} - z = 0 \rightarrow d = \frac{1}{4z^2}\,. \tag{27}$$

Proceeding as in the asymmetric case, backward it is solved the monopolist's pricing problem. However, in this market scenario, the data broker has double returns over the same costs of production c for a certain quantity of collected and aggregated data d

$$\pi_M = (2z - c)d - F. \tag{28}$$

After replacing in (28) the amount of data bought by each one of the two firms, we obtain $\pi_M = (2z - c)\frac{1}{4z^2} - F$. Therefore, maximizing with respect to the monopolist's decisional variable $\frac{d\pi_M}{dz} = 0$, we hereby derive:

$$\frac{2}{4z^2} - \frac{2z - c}{2z^3} = 0 \rightarrow z = c. \tag{29}$$

Hence, the price fixed by the data seller is lower than in the previous scenario. For instance, see expression (22). The direct consequence is a higher quantity of sold data $d = \frac{1}{4c^2}$, that allow us to calculate the final profits. As a result, the monopolist's gains are

$$\pi_M = \frac{2c - c}{4c^2} - F = \frac{1}{4c} - F. \tag{30}$$

Furthermore, as $\pi_A = \pi_B = 0$ under the symmetrical information structure, the total profitability of the industry will solely consist of the data seller's profit $\pi_T = \pi_M$.

Finally, the discussion about the stated market outcomes with linear prices is summarized in the next proposition.

Proposition 1. *With a regime of linear prices, the data seller is eager to sell data to both firms, promoting a symmetric allocation of information. Even though the monopolist sells information at a lower price than in the asymmetric scenario, the higher sold quantity offsets losses, boosting his profits*.

What emerged comparing the results of the two informational structures of our model, on one hand, confirms the traditional assumptions from literature. Indeed, wider access to information implies an increased competition for firms, which is so rough to vanish the profits in the downstream market. Furthermore, the general profitability of the industry benefits of the enhanced available information in the market. On the other hand, the increased competition does not turn in a lower potential surplus extraction by the data seller. Yet, the profits realized under a symmetric allocation of information are higher, as a consequence of the higher amount of sold data. Here it comes the breakdown with related literature and the disruptive potential of the idea to endogenize data, paid as a function of the purchased quantity and no more treated as an indivisible whole.

3.2 Two-part tariffs

A second application of the model of endogenous data extraction concerns the use of a different pricing strategy adopted by the monopolist to charge access to data: a two-part tariff.

Traditional theories have referred to two-part tariffs as price-discrimination practices, employed by firms with market power. Well-known is Oi's work, in which the author discusses Disneyland pricing dilemma. Indeed, Walt Disney as the owner of an amusement park can either charge lump sum admission fees and gives the rides away, or it may let the people enter inside the park for free and impose high prices on the rides. However, consumers derive no utility from going to the park itself and too high prices on each ride will deter the consumption. Thus, the best solution for Micky Mouse's monopoly is to set a price per ride equals the marginal costs, maximizing consumers' welfare and then extracting such welfare with the fixed part (Oi,1971). Since Pigou's discussion about the use of two-part tariffs by a monopolist for second-price discrimination (Pigou, 1920), there are now several examples of such practice also in more competitive markets. A very common example comes from the world of clubs and bars, that frequently charge a fee for the entrance as well as a price per drink[37]. Other examples include telecommunication companies, health clubs, and so on.

Thus, following the standard textbook models, we have decided to apply the pricing strategy of the two-part tariff in our scenario. The monopolist is the data seller, whose product is data, who charges a lump sum fee T for the right to access the database and then a variable price z function of the level of consumption. Assume (15) and (16) still hold. Hence, no differences occur in terms of the cost function of data extraction and the surplus delivered by consumers. What is changed is just the pricing strategy, analyzed under the two different informational structures.

3.2.1 Asymmetric Information

In this section, we describe the case in which the data broker sells the data to one downstream firm. Suppose firm A is the only one to have access to the upstream market of information, being subject to a two-part tariff.

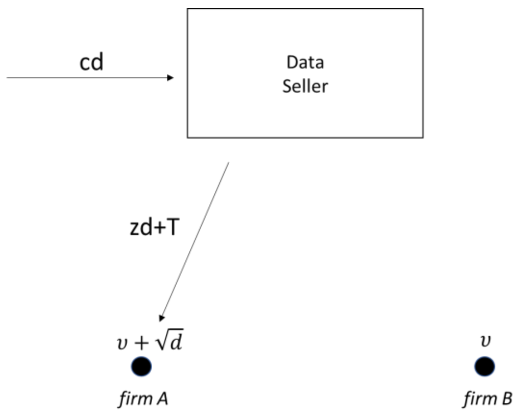

Figure 18: Data market under an asymmetric distribution of information and a two-part tariff regime

In the second stage of the game, when the two downstream firms compete, there are no consequences of the newly adopted pricing strategy by the data seller. Hence, firm A fixes the same price $p_A = \sqrt{d}$ and buys the same amount of data $d_A^* = \frac{1}{4z^2}$, as with linear prices. For instance, see expressions. (18) and (20).

[37]For instance, see Berglas, E. (1976). On the theory of clubs. *American Economic Review 66 (May)* 116-121.

However, moving backward in time to the first stage of the game, we face the introduced differences. Indeed, the new monopolist's profits are

$$\pi_M = (z - c)d + T - F \,. \tag{31}$$

Following the literature about two-part tariffs, the monopolist extracts all consumer surplus with the fixed part. Now suppose the two firms being the data seller's consumers and their profits, net of the fixed part, being the extractable rents. Formally, T satisfies

$$\mathrm{T} = \pi_A = p_A - zd \,. \tag{32}$$

Under "the common knowledge of rationality" assumption of the backward induction logic, the monopolist will anticipate the optimal amount of data bought by firm A in *t*=2. Moreover, writing (32) in (31) his profits may be rewritten as

$$\pi_M = (z - c)d + p_A - zd - F = p_A - cd \ - F = \sqrt{d} - cd - F \,. \tag{33}$$

Given (33), the first-order condition $\frac{d\pi_M}{dd} = 0$ for the data seller's optimal amount of data is

$$\frac{1}{2\sqrt{d}} = c \to d_M^* = \frac{1}{4c^2} \,. \tag{34}$$

As it is in the best interests of the monopolist to incentive the use of the optimal amount of data required by the firm, being able to capture its consequent profits, the data seller will match his offer with the required data by firm A i.e. $d_M^* = d_A^*$, thereby having

$$\frac{1}{4c^2} = \frac{1}{4z^2} \to z = c \,. \tag{35}$$

In line with the aforementioned literature about two-part tariffs (Oi,1970), expression (35) proves the data broker's convenience in charging a price equal to the marginal cost. Indeed, this is equivalent to Pigou's first-degree price discrimination, which maximizes monopolist profits by capturing all consumer surplus. Now using (34) in (33), we derive the monopolist's profits as the only function of c

$$\pi_M = \frac{1}{2c} - c\frac{1}{4c^2} - F \rightarrow \pi_M = \frac{1}{4c} - F\,. \tag{36}$$

Finally, due to the full rent extraction from the downstream firms $\pi_A = \pi_B = 0$, the total industry profitability only consists of the data seller's profits $\pi_T = \pi_M$.

The results of the direct comparison with the asymmetric scenario under linear prices are recorded in the following proposition.

Proposition 2 *The two part-tariff, under an asymmetric distribution of information, makes the monopolist better off, while the informed firm A worse off. Nevertheless, the total profitability of the industry is higher than under linear prices.*

Proposition 2 can be interpreted in the following manner. As a consequence of the monopolist's support to the consumption of the optimal quantity of data, a greater value is created in the market. However, due to his ability to extract rents with the fixed part of the two-part tariff, the increased value generated within the industry will be entirely captured by the data seller. In addition, the informed firm A will be completely deprived of any profit, further increasing the monopolist gains.

3.2.2 Symmetric Information

In this section, we consider the case in which the data broker sells the data to both the downstream firms. Under a symmetric distribution of information, both competing firms have access to the market of data, being subjects to a two-part tariff.

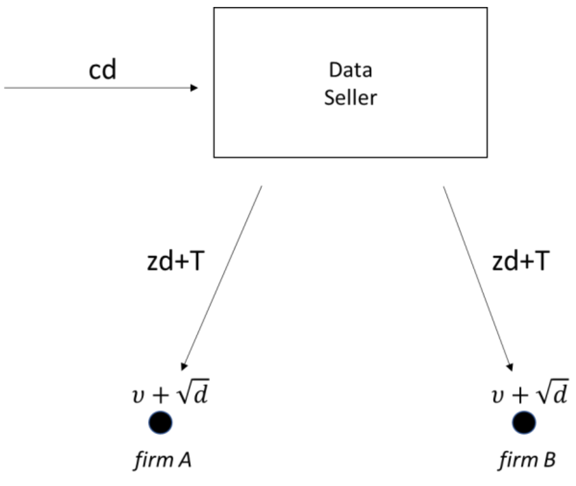

Figure 19: Data market under a symmetric distribution of information and a two-part tariff regime

The case of a symmetric distribution of information under a two-part tariff regime is straightforward, but still allows to draw interesting insights when compared to the linear-tariff case. Indeed, as we have stated in the asymmetric case in the second stage of the game there are no differences caused by the change in the pricing strategy of the monopolist. Therefore, firm A and firm B compete à la Bertrand with price wars and lowering their profits to zero $\pi_A = \pi_B = 0$. Moreover, they purchase exactly the same amount of bought data under linear prices and symmetric distribution of information $d = \frac{1}{4z^2}$.

Thus, proceeding back to the first stage of the game, the data seller's payoff function is given by

$$\pi_M = (2z - c)d + 2T - F. \tag{37}$$

However, the rough competition among equally informed firms makes the monopolist unable to extract any rent with the lump sum fee. It is easily verified that T is given by

$$T = \pi_A = \pi_B = 0. \tag{38}$$

Replacing T=0 in (37), we obtain the monopolist's payoff function under symmetric information and linear prices. For instance, see expression (28). Therefore, the equilibrium price maximizing the data seller's profits can be obtained by (29), while the value of his profits is given by (30). Finally, the discussion is summarized in Proposition 3.

Proposition 3. *Under a symmetric distribution of information, the model with endogenous data gives the same market outcomes with linear prices or a two-part tariff.*

In addition, some other interesting insights may be derived. By comparing the principal variables of the game resulting from paragraph 3.2.1 and 3.2.2., it can be noticed that the monopolist fixes z=c in both scenarios and realizes the same profits. For instance, notice (36) = (28). Besides, neither under an asymmetric distribution of information nor a symmetric one, the two competing firms are able to realize profits $\pi_A = \pi_B = 0$. However, in the first case is the monopolist to bring to zero the informed firm's profit by extracting all its rents. While, when both firms have access to information, the increased competition is responsible for driving to zero their profits. As a result, the total industry profitability always consists of the only data seller's profits $\pi_T = \pi_M$. In a nutshell, all these considerations are recorded in the following proposition.

Proposition 4. *The use of a two-part tariff makes the monopolistic seller indifferent between a symmetrical and an asymmetrical allocation of information, being his profits the same.*

If Proposition 1 and Proposition 4 hold, then we can summarize the overall results of our proposed model as follow.

Proposition 5. *Under a model with endogenous data, the seller of information is either in favor of a symmetric distribution of information (linear prices) or indifferent between the two information structures (two-part tariffs). Thus, the preference to sell data to one retailer firm only is confuted.*

The general findings stated by Proposition 5 underline a new breach into traditional literature. First, this fosters the awareness of the importance of assumptions in mathematical models. Tatsuhiko, a professor from Tokyo University, in 2006 has published an article[38] on these topics where he considers assumptions as a bridge that connects the mathematical world with the real world. Hence, the use of a different path may bring to a very different world's scenario. In our case, the choice of endogenize data, instead of treating information as an indivisible amount, represents the different bridge used to come up with new insights that may be turned in actions to be undertaken. Indeed, Proposition 5 would lead us to conclude that a policy ensuring greater access to data and democratization of information can be applied by the antitrust authorities, without any friction on the part of the big data broker companies. Furthermore, it will be an efficient way to best promote competition and get over market inefficiencies related to the anticommons tragedy of underutilization. Notwithstanding, reshaping antitrust for the digital economy will not be easy. It will imply new risks associated with more data sharing, threatening privacy.

Yet, if governments and policymakers do not want a data economy dominated by a few dominant platforms, they will need to act soon.

[38] For instance, see Tatsuhiko, S. (2006). Understanding the Role of Assumptions in Mathematical Modeling: Analysis of Lessons with Emphasis on 'the awareness of assumptions'. *Tokyo Gakugei University*.

Conclusions

From the covered topics what is emerged so far is that data will play a central role in the 21st-century economy. Techno-optimist may be enthusiastic, arguing that new digital services and technologies will raise productivity in unforeseen ways. Techno-pessimists may be more concerned about the potential risks in terms of privacy and unfair competition brought by the actual pace of innovation. However, this is the moment to answer to the raised questions at the beginning of the critical review about the emerging dataopolies.

Should we be concerned about these tech giants?

Yes, we should. First, the digital economy raises concerns about the necessity to promote fair competition. Indeed, digital platform businesses have become the modern-day equivalent of natural monopolies. Because of their costs structures mainly based on fixed costs, together with network effects, economies of scale and scope they are creating a concerning competitive scenario, dominated by the formula of a 'winner-takes-all'. Other aforementioned factors, which facilitate the market dominance, are the presence of high switching costs, lack of interoperability and limited data portability. The direct consequence of such a market scenario is the reduction of the market's general competitiveness in the form of higher prices, greater market power held by incumbents, and increasing entry barriers for new entrants. In such a less competitive industry, dominant platforms may not be motivated to improve their efficiency, thereby generating social losses, market failures, and hampering the process of innovation. Nevertheless, a not negligible relationship exists between data-driven innovation and log-term productivity: investments in data have positive spillovers across a wide range of industries. However, the digital economy has more to contribute to the global economy than what it currently does. Nowadays, the use of data is principally limited inside large platform ecosystems. Indeed, each dominant platform tries to exclude others from realizing the potential benefit of data aggregation. On one hand, such uncoordinated exercise of exclusion rights leads to under-utilization of data, as in the textbooks tragedy of anticommons. Thus, unexploited opportunities exist. On the other hand, policymakers have few instruments to try to capture a big share of the digital value created by platforms, limiting their ability to scale.

Second, other concerns are more related to privacy issues and customer protection arguments. Consumers while surfing on the net become producers of a high quantity of personal information. Usually, they enjoy some benefits of such disclosure in terms of price discrimination, target advertisings and tailored offers. However, such benefits may turn into potential online harms for users[39]. Digital platforms may be a tool for abuse, bullying and other forms of either racial or sex discrimination. Greater awareness about data practices and third party industry, a higher individuals' valuation of their personal information will create more conscious citizens, who better understand risks of their online activities enjoying more safely the offer of services.

In which ways can regulators intervene to maximize social welfare?

A right to data portability, wider open access to information and more data sharing may be effective steps to be taken by policymakers. As confirmed by traditional literature, a non-exclusive allocation of consumers' information among competing firms increases fair competition. Lowering the firm's potential for rent extraction, a greater surplus is left to consumers and a more efficient allocation is realized in terms of welfare i.e. no market failures generated by privacy costs paid by consumers, boosted profitability of the whole industry.

However, in traditional textbooks, the realization of such superior welfare allocation has been hampered by monopolistic data seller's interests. Indeed, the monopolist's convenience to sell data to one retailer firm only has been long claimed. Our model, changing the traditional assumption of data treated as an exogenous variable, refutes this result opening the possibility for further research on the topic. A natural extension would be to introduce data sellers' competition within our framework.

Finally, associated adjustments to embrace the full benefits for the overall economy could be significant. Obstacles and barriers to the access and sharing of data should be carefully analyzed, and a global data governance policy should be developed. Leaders in business, education, and government must be ready to work on policy programs and prepare their countries for future growth and development, beyond typical election cycles. The long-term vision is essential. Reflection is also required on how some initiatives to tax the web giants can't be taken by a single economic block, as the UE alone, to be fully effective. Global action in terms of G20 countries or OECD is indeed required.

In conclusion, what is sure is that we are facing a next product revolution (NPR), which implies a confluence of different digital technologies to new materials and new processes, some of which are not even available in the near future (EOCD;2017). The more governments will

[39] For instance, see UK Government. (2019). Online Harms White Paper. *Presented to Parliament by the Secretary of State for Digital, Culture, Media & Sport and the Secretary of State for the Home Department by Command of Her Majesty.*

understand how digital innovation is radically changing and reshaping the rules of competition, the better they will be prepared for the risks and could proactively reap the benefits. The opportunity exists to influence technological development and is now.

Better policies for better lives. *(OECD)*

Bibliography

Acquisiti, A., Leslie, K. J., & Loewenstein, G. (2013). What Is Privacy Worth? *Journal of Legal Studies 42, no. 2*, 249–274. .

Acquisiti, A., Taylor, C., & Wagman, L. (2016). The Economics of Privacy. *Journal of Economic Literature 54(2)*, 442-492.

Admati, A., & Pfleiderer, P. (1988). Selling and trading on information in financial markets. *American Economic Review, 78(2)*, 96-103.

Admati, A., & Pfleiderer, P. (1990). Direct and indirect sale of information,. *Econometrica*, 901-928.

Aperjis, C., Gkatzelis, V., & Huberman, B. A. (2015). Pricing private data Electronic Markets.

Arrow, K. (1962). Economic Welfare and the Allocation of Resources for Invention. In C. o. Universities-National Bureau Committee for Economic Research, *The Rate and Direction of Inventive Activity: Economic and Social Factors* (p. 609-626). Princeton University Press.

Aumann, R. J. (1995). Backward Induction and Common Knowledge of Rationality. *Games and Economic Behavior (8)*, 6-19.

Australian Competition & Consumer Commission. (2018). *Digital Platforms Inquiry.*

Barwise, P. (2018). Nine Reasons why tech markets are winner-take-all . *London Business School* .

Baumol, W. J., Panzar, J., & Willig, R. D. (1982). *Contestable Markets and the Theory of Industry Structure.* New York: Harcourt Brace Javanovich .

Bergemann, D., & Bonatti, A. (2012). Markets for Data. *Discussion Paper*.

Bergemann, D., & Bonatti, A. (2015). Selling Cookies. *American Economic Journal: Microeconomics, 7(3)*, 259–294.

Bergemann, D., & Bonatti, A. (2018). Markets for Information: An Introduction. *Cowles Foundation for Research In Economics Yale University*, Discussion Paper No. 2142.

Bergemann, D., & Pesendorfer, M. (2007). Information Structures in Optimal Auctions. *Journal of Economic Theory, 137*, 580-609.

Bergemann, D., Bonatti, A., & Smolin, A. (2018). The Design and Price of Information. *American Economic Review 108(1)*, 1-48.

Berglas, E. (1976). On the theory of clubs. *American Economic Review 66 (May*, 116-121.

Bertrand, J. (1883). Theorie Mathematique de la Richesse Sociale. *Journal des Savants* , 499-508 .

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *The American Economic Review 94(4)*, 991–1013.

Biglaiser, G., Calvano, E., & Crémer, J. (2018). Incumbency Advantage and its value. *Journal of Economics and Management Strategy*.

Binns, R., & Bietti, E. (2019). Acquisitions in the third-party tracking industry: competition and data protection aspects. *Working Paper* .

Binns, R., Zhao, J., Van Kleek, M., & Shadbolt, N. (2018). Measuring third party tracker power across web and mobile . *arXiv preprint arXiv:1802.02507* .

Bonatti, A., & Cisternas, G. (2019). Consumer Scores and Price Discrimination. *Discussion paper, MIT*.

Bouckaert, J., Degryse, H., & Provoost, T. (2012). Enhancing market power by reducing switching costs. *Economics Letters. 114(3)*, 359–361.

Bourreau, M., Caillaud, B., & De Nijs, R. (2018). Taxation of a digital monopoly platform,. *Journal of Public Economic Theory*.

Buchanan, J., & Yoon, Y. (2000). Symmetric tragedies: commons and anti-commons. *Journal of Law and Economics, 43(1)*, 1–13.

Coase, R. H. (1960). The Problem of Social Cost. *The Journal of Law and Economics*, 1–44.

Collin, P., & Colin, N. (2013). Task Force on Taxation of Digital Economy. *La Documentation Française*.

Cournot, A. (1838). *Recherche sur les Principes Mathematiqlles de la TIeuorie des Richesses.* Paris.

Crawford, G. S. (2019). The Problems with (Dominant) Platforms. *11th Paris Conference on Digital Economics.* Paris.

Dosis, A., & Sand-Zantman, W. (2019). The Ownership of Data.

Duch-Brown, N., Martens, B., & Mueller-Langer, F. (2016). The economics of ownership, access and trade in digital data. *Digital Economy Working Paper, JRC Technical Reports.*

Edelman, B. G., & Luca, M. (2014). Digital discrimination: The case of airbnb.com. . *Harvard Business School NOM Unit Working Paper*, 14-54.

Edgeworth, F. (1889). The pure theory of monopoly. In *Collected Papers relating to Political Economy 1925* (p. vol 1). Macmillan.

Engels, B. (2016). Data portability among online platforms. *Internet Policy Review, 5 (2)*.

European Commission . (2012). Towards better access to scientific information: Boosting the benefits of public investments in research. *European Commission.* Brussels.

European Commission. (2017). *Final report of the European Data Mrket study measuring the size and trends of the EU data economy.* Available at: https://ec.europa.eu/digital-single-market/en/news/final-results-european-data-market-study-measuring-size-and-trends-eu-data-economy.

Evans, D. S. (2003). Some empirical aspects of multi-sided platform industries. *Review of Network Economics 2 (3)*, 191-209.

Fast, V., Schnurr, D., & Wohlfarth, M. (2018). The Value od Data in Digital Markets: Short-Term Competitive Advantage or Long-Term Market Power? *29th European ITS Conference.* Passau, Germany.

Freixas, X., Guesnerie, R., & Tirole, J. (1985). Planning under Incomplete Information and the Ratchet Eﬀect. *Review of Economic Studies, 52(2)*, 173–191.

Frog. (2011). The Value of Personal Data: A Global Perspective. *Presentation at Mobile World Congress*.

Garcia, D., & Sangiorgi, F. (2011). Information Sales and Strategic Trading . *Review of Financial Studies, 24(9)*, 3069-3104.

Gentzkow, M., & Kamenica, E. (2011). Bayesian Persuasion. *American Economic Review, 101*, 2590-2615.

Gerardi, D., & Maestri, L. (2016). Dynamic Contracting with Limited Commitment and the Ratchet Eﬀect . *Discussion paper, Collegio Carlo Alberto.*

Hagiu, A., & Wright, J. (2013). Do you Really want to be an eBay? *Issue of Harvard Business Review* .

Hann, I.-H., Hui, K.-L., Lee, S.-Y. T., & Png, I. P. (2002). The Value of Online Information Privacy: An Empirical Investigation. *Mimeo, Marshall School of Business, University of Southern California*.

Harasser, A., Jentzsch, D. N., & Preibusch, S. (2012). Study on monetizing privacy. An economic model for pricing personal information . *ENISA*.

Hardin, G. (1968). The tragedy of Commons, . *Science, 162* , 1243-1248.

Heller, M. A. (1998). The Tragedy of the Anticommons: Property in the Transition from Marx to Markets. *111 Harv. L. Rev. 621.*

Huberman, B. A., Adar, E., & Fine, L. (2005). Valuating Privacy. *IEEE Security and Privacy 3*, 22-25.

Iyer, G., & Soberman, D. (2000). Markets for product modication information. *Marketing Science, 19(3)*, 203-225.

Jones, C., & Tonetti, C. (2018). Non rivalry and the Economics of Data . *Meeting Papers vol. 477 (Society for Economic Dynamics)*.

Joskow, P. L. (2005). Regulation of Natural Monopolies . *MIT, Center for Energy and Environmental Policy Research* .

Jullien, B. (2006). Pricing and other business strategies for e-Procurement platforms. *Handbook of Procurement*.

Kenney, M., & Zysman, J. (2016). The rise of the Platform Economy. *Issues in Science ad Technology*, 60-69.

Klemperer, P. (1987a). Markets with Consumer Switching Costs. *Quarterly Journal of Economics, 102*, 375-394.

Klemperer, P. (1987b). The competitiveness of markets with switching costs. *The RAND Journal of Economics (18:1)*, 138-150.

Klemperer, P. (1988). Welfare effects of entry into markets with switching costs. *The Journal of Industrial Economics, 37*, 159-165.

Klemperer, P. (1995). Competition when consumers have switching costs: An overview with applications to industrial organization, macroeconomics, and international trade. *Review of Economic Studies, 62 (4)*, 515-539.

Lavalle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big Data, Analytics and the Path from Insights to Value. *MIT Sloan Management Review, Vol 52. No. 2*, 21-31.

Mahdawi, A. (2018). Is your friend getting a cheaper Uber fare than you are? *The Guardian*.

Marr, B. (2018, May 21st). How much data do we create every day? The mind-blowing stats everyone should read. *Forbes*.

Marshall, A. (1890). *Principles of Economics.* London: MacMillan.

Mattioli, D. (2012). On Orbitz, Mac Users See Costlier Hotel Options. *The Wall Street Journal*.

Mikians, J., Gyarmati, L., Erramilli, V., & Laoutaris, N. (2012). Detecting price and search discrimination on the Internet. *In Proceedings of the 11th ACM Workshop on Hot Topics in Networks*, 79-84.

Mikians, J., Gyarmati, L., Erramilli, V., & Laoutaris, N. (2013). Crowd-assisted search for price discrimination in e-commerce: First results. . *In Proceedings of the Ninth ACM Conference on Emerging Networking Experiments and Technologies*, 1-6.

Montes, R., Sand-Zantman, W., & Valletti, T. (2019). The Value of Personal Information in Online Markets with Endogenous Privacy. *Management Science 65(3)*, 1342-1362.

Norberg, P., Horne, D. R., & Horne, D. A. (2007). The privacy paradox: personal Information Disclosure Intensions versus Behaviours. *The Journal of Consumer Affairs 41*.

OECD. (2013). Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Vlaue. *OECD Digital Economy Papers, No. 220*.

OECD. (2016). Enhanced access to data- Reconciling risks and benefits of data re-use. *OECD expert workshop.* Paris.

OECD. (2017). *The next product revolution: A report for the G20.* Paris: OECD.

Oi, W. (1971). A Disneyland dilemma: Two-part tariffs for a Mickey-Mouse monopoly. *Quarterly Journal of Economics 85 (February)*, 76-96.

Pigou, A. C. (1920). *The Economics of Welfare.* London: MacMillan.

Posner, R. A. (1969). Natural Monopoly and Regulation. *Stanford Law Review 21: 548643.*

Ramasastry, A. (2005). Web sites change prices based on customers' habits. *CNN.com*.

Rochet, J.-C., & Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the European economic association 1 (4)*, 990-1029.

Safdar, K. (2018). On Hold for 45 Minutes? It Might Be Your Secret Customer Score. *The Wall Street Journal*.

Sharkey, W. W. (1982). The Theory of Natural Monopoly. *Cambridge University Press.*

Statista. (2019). *The 100 largest companies in the world by market value in 2019.*

Stigler, G. J. (1966). *The Theory of Price, 3rd edition.* New York: The Macmillan Company.

Tatsuhiko, S. (2006). Understanding the Role of Assumptions in Mathematical Modeling: Analysis of Lessons with Emphasis on 'the awareness of assumptions'. *Tokyo Gakugei University*.

Taylor, C. R. (2004). Consumer privacy and the market for customer information. *Rand Journal of Economics, 35(4)*, 631–650.

The Economist. (2017, May 6th). The world's most valuable resource is no longer oil, but data.

Tsai, J. Y., Egelman, S., Cranor, L., & Acquisti, A. (2011). The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study. *Information Systems Research 22:*, 254-268.

UK Government. (2019). Online Harms White Paper. *Presented to Parliament by the Secretary of State for Digital, Culture, Media & Sport and the Secretary of State for the Home Department by Command of Her Majesty.*

Valentino-Devries, J., Singer-Vine, J., & Soltani, A. (2012). Websites vary prices, deals based on users' information. *Wall Street Journal*.

White, T. B., Zahay, D. L., Thorbjørnsen, H., & Shavitt, S. (2008). Getting too personal: Reactance to highly personalized email solicitations. *Marketing Letters 19(1)*, 39-50.

Wohlfarth, M. (2017). Data Portability on the Internet: An Economic Analysis. *28th European Regional Conference of the International Telecommunications Society (ITS).* Passau, Germany.

Zermolo, E. (1913). Ueber eine Anwendung der Mengenlehre auf die Theorie des Schachspiels. In *Proceedings of the Fifth International Congress of Mathematicians, Vol. II* (p. 501-504). Cambridge: Cambridge Univ. Press. .

Acknowledgments

Above all, I would like to thank the help, availability, and guide offered by my principal supervisor, Prof. Cambini Carlo. Furthermore, I would like to acknowledge the contribution of my second supervisor for this thesis, Prof. Abrardi Laura. Your keen eyes for all the details in the realization of the model have been indispensable.

I also would like to take this opportunity to thank Politecnico di Torino, which has helped me broaden my horizons and deepen my knowledge. It has taught me to learn and be curious, to dive deep, to deliver results, and to insist on the highest standards raising the bar of personal achievements higher and higher. Particularly, I feel grateful for the award of the Erasmus+ Scholarship that gave me the opportunity to study for one year in Canada. It was one of the best experiences of my life, for my career and my personal growth. Finally, it was, weird to say, my happy place for 5 years where I have met colleagues, or more properly, friends for life.

Last, but definitely not least, I'd like to thank my family. Without their tireless support, I would not have been able to complete my university path. Thank you for the time spent listening to my insecurities and giving me advice. Thank you for the freedom you have always left me in making my decisions because there is no greater act of love than liberty. Thank you for the trust you have daily placed in my future projects. Thank you for being my angel investors, who have helped me to get off the ground. I hope you will see the made investments as a positive NPV project. I am greatly indebted to you, it was your unconditional support, love, and care that made my dreams come true.

Finally, I extend my last thanks to all my friends, who are an essential part of my life. There is no great solo player, without a good team.

Antonella Sgaramella