UNIVERSITA' DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI "M.FANNO"

CORSO DI LAUREA MAGISTRALE IN BUSINESS ADMINISTRATION

TESI DI LAUREA

"ECONOMICS OF PRIVACY: THE ROLE OF DATA BROKERS"

RELATORE:

CH.MO PROF. MANENTI FABIO

LAUREANDO: BACCARO FEDERICO

MATRICOLA N. 1207022

ANNO ACCADEMICO 2020 – 2021

Il candidato dichiara che il presente lavoro è originale e non è già stato sottoposto, in tutto o in parte, per il conseguimento di un titolo accademico in altre Università italiane o straniere.
Il candidato dichiara altresì che tutti i materiali utilizzati durante la preparazione dell'elaborato sono stati indicati nel testo e nella sezione "Riferimenti bibliografici" e che le eventuali citazioni testuali sono individuabili attraverso l'esplicito richiamo alla pubblicazione originale.

The candidate declares that the present work is original and has not already been submitted, totally or in part, for the purposes of attaining an academic degree in other Italian or foreign universities. The candidate also declares that all the materials used during the preparation of the thesis have been explicitly indicated in the text and in the section "Bibliographical references" and that any textual citations can be identified through an explicit reference to the original publication.

Firma dello studente

Federico Baccaro

Abstract

As our lives become more digitized, as we interact with our digital devices on a more regular basis, as we see information about ourselves being tracked, recorded and distributed across all kinds of platforms, many of us are increasingly concerned about privacy.

While surfing the net, the user must give consent to a multitude of pages, accept cookies (most of the time without reading the terms) to access information and articles, but the subject of consent is often incomprehensible or misleading. Consumer choices are important to limit the amount of information they release in the online world on a daily basis, but they cannot be considered as the only ones responsible. For this reason, in this scenario legislation plays a crucial role in ensuring consumers a greater level of privacy protection.

In modern economies, especially digital economies, companies collect a huge amount of data in relation to our characteristics. This information is then used to implement well-planned strategies such as targeted advertising, price discrimination and personalized offers. In addition, there are companies (i.e. data-brokers) that enter the data market by favoring the collection, storage, organization, but also the sale of personal data, most of the time without the "data-subject", i.e. the person to whom the data refer, being aware of it.

How does the data-broker affect the strategies implemented by companies? How do data-broker strategies affect the outcome in terms of consumer surplus and total welfare? And in turn, how can the legislation intervene to limit potential anti-competitive behavior and favor greater social welfare?

In this paper we propose a model that encompasses the issue of economics of privacy with a focus on data-brokers and their strategic role in selling data to downstream firms, significantly influencing the equilibrium in terms of consumer surplus and total welfare.

Table of contents

Introduction

The rise of large digital platforms and technologies has significantly facilitated the collection and the commercial use of personal data. At the same time, user bases have released unprecedented amounts of data concerning almost every facet about their lives, including preferences, habits, sexual and political orientation. Companies collect data about users from varieties of sources, both directly and indirectly, and this allows them to create detailed profiling of customers to which they offer personalized and targeted products and services. This has inevitably given rise to privacy concerns. Legislation has tried several times to offer individuals greater control over their personal data and to provide consumers with greater transparency on the way in which personal data is collected and used by companies enacting regulation such as the European General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA).

Through the “Economics of privacy”, economists have analyzed the trade-offs associated with the use and protection of consumer data, both at the individual (e.g. data breaches of data from a social profile of a person) and at the societal level (e.g. influencing political election results, as happened with the Cambridge Analytica scandal).

In this paper we analyze a few critical issues in relation to the economics of privacy, dividing the discussion in three main chapters.

Chapter one covers the literature on the economics of privacy, revealing its distinctive features of complexity, interdisciplinarity and malleability, as well as its evolution over time. Moreover, we will analyze the consumer approach to privacy: how individuals make decisions related to privacy and how sometimes their decisions are counterintuitive. Choices of customers become relevant in the analysis of the trade-offs related to the interactions between consumers and companies collecting data to improve targeted strategies.

The second chapter describes the main features of the data market that reflects the evolution brought by the digital transformation. However, the market of data remains almost obscure and unexplored. In this context, a strategic role is played by data-brokers, intermediaries collecting, storing and selling data to companies. The practices conducted by data-brokers have been the subject of an in-depth analysis by the Federal Trade Commission, which in 2014 published a report entitled "Data Brokers: A Call for Transparency and

Accountability" to shed light on those companies, the amount of data they manage, and the lack of transparency characterizing their main activities.

In the third chapter we develop a model characterized by an upstream data-broker selling its data to two downstream firms that compete on the retail market. Access to the data supplied by the data-broker allows firms to discriminate final users. We characterize downstream firms' incentives to buy the data from the data-broker, and the optimal pricing strategy of this latter. The results of the model will be then translated into suggestions for the policy makers.

Chapter 1: Background

A lot of information is stored in the devices of users: people chat by using online applications, exchange messages and post pictures on social media, shop online, make research on browsers, and much more. These devices have made the life of humans simpler, but the same simple tasks users do with these devices help companies and organizations to create a very comprehensive picture about users themselves. The task done with these devices reveals information about financial data, location, pictures, ID, political interests, habits, the list of our contacts, health, and even more. Therefore, protecting the privacy of individuals has become a priority, especially in the information and digital age.

What do we mean by privacy? What forms does privacy take? What does it mean to violate privacy?

Privacy is a contested topic with a huge array of definitions. When talking about privacy, we must consider that the term "privacy" includes a multitude of meanings and areas of application, each with its own facets. Moreover, defining privacy also requires some abstract thinking. Generally speaking, privacy and the human rights associated with privacy can be seen as the imaginary barrier that prevents an individual from other people doing something harmful to him, and at the same time it allows a person to be open towards the people he trusts and to be close to those in the opposite case. This is why privacy can be seen as "the appropriate use of personal information under the circumstances. What is appropriate will depend on context, law, and the individual's expectations"[1].

Privacy has been analyzed by different perspectives and disciplines, including economics, law, sociology, and political science (Casadesus-Masanell and Hervas-Drane, 2015). Studies on privacy have concerned the autonomy, freedom, secrecy, anonymity and protection of personal data. As an example, in the past the issue of "privacy violation" was considered a type of violation with the potential to harm the dignity and integrity, but also the freedom and independence of an individual.

[1]International Association of Privacy Professionals, "IAPP Information Privacy Certification: Glossary of Common Privacy Terminology" 2011, privacyassociation.org.

1.1 Privacy concerns

Our lives are constantly characterized by innovative products and services. The digital transformation and the Information and Communications Technologies (ICTs) have dramatically changed the everyday life of people, from the way in which individuals interact between each other to the way in which business is done (Comino and Manenti, 2015).

The pervasiveness of the internet has allowed users to access products and services sold by companies online: online firms, in fact, can operate in virtually unbounded markets, and in this way they can reach even distant customers. According to Statista[2], in 2019 around 85% of people living in developed countries used the Internet, a percentage that amounts to almost 45% for the individuals living in non-developing markets. Overall, the global online access rate was around 50%, a result that is showing an increasing trend over time. Moreover, the e-commerce industry continues to grow rapidly. Statista reported that online shopping is one of the most popular online activities worldwide: in 2019, retail e-commerce sales worldwide amounted to 3.53 trillion US dollars and e-retail revenues are projected to grow to 6.54 trillion US dollars in 2022.

Online users are always connected, and they can use any application or platform they want and need to reach different purposes. But this connectivity also brings side-effects: these online activities create an enormous amount of data.

According to an estimate made by Visual Capital[3], data created on the internet in 2020 in one minute is terrific: 400,000+ hours or video stream by users, 2,500+ applications installed, around 7,000 packaged shipped, 200,000+ participants in meetings, 300+ new users and around 350,000+ stories uploaded on social media platforms, 500 hours of video uploaded by users, 50,000+ users connected, 40+ Ml messages shared. Despite the estimate, which includes the lockdown period caused by the pandemic, it is out of question the enormous amount of data shared by users and at the same time the amount of data collected by companies providing services.

Companies collect data about customers and these data can be used for different purposes. For instance, data can be used to provide useful recommendations to customers: Amazon recommends products according to previous purchases; online streaming platforms

[2] Statista (2020), Percentage of global population accessing the internet from 2005 to 2019, by market maturity

[3] https://www.statista.com/chart/17518/data-created-in-an-internet-minute/#:~:text=According%20to%20data%20compiled%20by,million%20messages%20shared%20via%20WhatsApp.

such as Netflix or YouTube recommend video on the basis of the previously watched videos. However, data can also be used to track the online behavior of customers and to ultimately price discriminate. In 2000, customers discovered that Amazon was charging different prices for the same item, a DVD movie[4]: unexpectedly, the company was charging higher prices to regular customers and lower prices for non-regular customers. Amazon tried to explain that the company was trying to figure out how much their loyal customers would have paid: this was possible because Amazon was tracking the online behavior of customers and therefore their online purchase histories, a possibility of gathering information to which earlier companies did not have access to. Through an online forum, customers compared their purchase experience and therefore Amazon was ultimately forced to refund them.

As early as 1998, Carl Shapiro and Hal Varian wrote in their book Information Rules: "[t]hose companies that are first, and best, at figuring out how to use the unique customer information available on the Web stand ready to reap substantial rewards" (p. 34). Thanks to the internet, online companies have information about their customers that they never had before, along with the technology to manage all the accumulated data: they have access to an enormous amount of data about user preferences, habits and characteristics. One of the recent events that has given rise to great concerns on the issue of privacy was the Cambridge Analytica scandal, which also involved Facebook. The 2018 Cambridge Analytica scandal has put under discussion the managing of data for billions of Facebook individuals. However, it has been an event with significant consequences on the concerns by people and users on how data is collected and managed not only by Facebook, but also by other big tech companies such as Apple, Google and Amazon and generally whoever company collects individual data.

Companies collect and manage varieties of data: data of different volumes and qualities, but also data collected through different tools and devices. Once managed, analyzed and properly combined, these data can reveal additional or inferred information and it can also be used by sophisticated technologies to test predictive outcomes. This is called "metadata", but Shoshana Zuboff, an Harvard professor, called it "behavioral surplus" in order to highlight the fact that additional information is extracted as a surplus of the reason by which initial data have been collected. Generally speaking, these terms indicates that data provides additional information which usually is generated through a proper use of technology. For example, knowing information such as phone numbers and IP addresses can

[4] The Washington Post, Streifeld, David, On the Web price tags blur: what you pay could depend on who you are, September 27, 2000

provide a starting point for compiling a picture of the online activities of users. Therefore, put together and sorted, data can reveal great added value.

The idea that data, when collected and aggregated, has more value was also examined by Acquisti et al. (2014), who published a study about facial recognition. The authors asked students of the University of Carnegie Mellon to stop for a study in front of a laptop inside the University buildings in order to answer some questions from a survey while at the same time the authors were taking some pictures of the student. As long as each student was filling the questionnaire on the laptop, the picture taken was sent to a cloud where the authors had previously downloaded public images from the Facebook profile of students: in this way, through a facial recognition software, the authors tried to find a match between the student at that moment in front of the laptop and the photos of the same taken from the social network. In this study, authors were able to match around 33% of the students. Not satisfied, the authors tried to push the study forward: the authors were in fact able to conclude that in trying to match an anonymous face with a Facebook profile by using facial-recognition software, it can be easily found a name for that face and retrieved more sensible data. In the experiment, they were able to partially predict the SSN (Social Security Number), an identity code used in the US for having access, for instance, to a credit card or to a mortgage. The study by Acquisti et al (2014) was really interesting because once again it confirms the idea behind the "data accretion": the fact that when you combine different databases, data increases, a concept somehow connected to that of "data linkability". The conclusions were that a face is not something anonymous, but is something that can be connected to more personal and more sensible data of the individual: there can be much value that can be extracted from data, and sometimes this value can be extracted only if data can be aggregated and compared among a multitude of individuals.

While it is true that almost all companies collect data, on the other hand it is also true that there are some particular companies that control a greater traffic of data and information. This applies for the so-called "Big Tech companies", that are the largest and most dominant companies in the information technology industry, not only in the United States, but globally. According to Forbes, Google is tracking an online user on 86% of the top 50,000 websites of the planet[5]. Moreover, according to a research by the University of Oxford in the UK, 90.4% of apps share data with at least one third party, in which they conclude that 35.3% of the apps

[5] Forbes, "Google is tracking you on 86% of the Top 50,000 websites on the planet", John Koetsier, March 11, 2020

share data with ten or more third parties[6]. Over the years, big tech companies have reached amazing market capitalization, creating higher barriers to entry and a very strong market power; moreover, they have been able to diversify their business into various profitable business niches, obtaining a widespread presence that allows them to collect varieties of data. These companies have a complete presence over different activities related to a day-to-day life of an individual. One example of a company having a presence in very different segments and sectors is Amazon. Besides of its e-commerce platform (Amazon.com), the company is present in other segments, such as: music (Amazon music), smart speaker (Amazon Alexa), audio-books (Audible), wallet and payments (Amazon Pay), logistic services (Amazon logistics), gaming streaming services (Twitch platform), cloud services (Aws), e-book (Amazon Kindle), film (IMDb), streaming services (Amazon Prime), and food (Whole Foods)[7].

The fact that few large companies have so many different applications and services around different activities (e.g. communications, chatting, web-browsing, searching, shopping, streaming, working, gaming) allow companies from one side to eliminate competition and stay competitive, while from the other side it allows "data linkability" of the digital life of online users, i.e. the possibility of cross-tracking online users across different activities, which ultimately constitutes a big issue to digital privacy. As an example, searching and browsing, if tracked, can reveal a lot about the inner thoughts and private moments or emotions and it can ultimately impact the health insurance fees of an individual.

The fact that these companies are gaining significant market power and dominance, together with enormous amounts of data and information about individuals and users has raised some concerns not only among people but also from the antitrust authorities. For example, in 2020 Google has announced the acquisition of Fitbit, the fitness tracking company. The antitrust hearing decided to launch an investigation into the deal because Google, through this acquisition, was potentially gaining even more market power and "increasing the already vast amount of data that Google could use for personalization of the ads it serves and displays"[8]. However, in the end, the acquisition was confirmed and

[6]Reuben Binns, UlrikLyngs, Max Van Kleek, Jun Zhao, Timothy Libert, and Nigel Shadbolt. 2018. Third Party Tracking in the Mobile Ecosystem. In Proceedings of the 10th ACM Conference on Web Science (WebSci '18). Association for Computing Machinery, New York, NY, USA, 23–31.
[7] https://www.amazon.jobs/it/business_categories/subsidiaries
[8]CNBC, EU approves Google's $2.1 billion acquisition of Fitbit, subject to conditions, December 17, 2020

ultimately allowed, but the direction of the antitrust hearing in recent years has focused on investigating the real market dominance of these companies.

Amazon leads the American largest online market, capturing approximately 70% of all sales in this market; Facebook continues to generate growing profits and users and appears not to have suffered the impact of sanctions and past privacy scandals; Google is the largest online search engine in the world, capturing over 90% of online searches; Apple has significant dominance in the app store[9]. These companies control key distribution channels, having access to user data as well as data from other companies. The enormous power and dominance consolidated by these companies over the years has become significant, therefore they have increased the concerns about their potential monopoly power they are leveraging on as well as the amount of the data they collect and treat.

In July 2020, the CEOs of Facebook, Google, Amazon and Apple testified in front of the Congress in a tech antitrust hearing with the aim to examine their dominance[10]. The committee argued that companies like Facebook and Google control how information is disseminated, and the same is true for Apple and Amazon but in relation to, respectively, the app store and the marketplace. For years, these companies have been investigated for their acquisitions, predatory pricing and potential anti-competitive behavior.

The Federal Trade Commission proclaims to act "protecting consumers and enhancing competition across broad sectors of the economy"[11]. As the United States, Europe is trying to limit big tech companies by imposing more transparency, trying to avoid monopoly positions and excessive powers, at the same time trying to create regulations that reflect the major changes taking place and the related repercussions from the point of view of consumers trying to protect their personal data and privacy. Another option that has been considered is to split Big Tech companies into smaller companies to limit their power, which instead could bring more value for investors, as was the case with Rockefeller which in 1911 was split into 34 smaller companies after the decision of the Supreme Court[12].

Also the EU followed suits in trying to limit the power of these companies. In December 2020, the EU proposed new rules and regulations in order to limit the power of Big Tech companies, such as Amazon, Facebook, Google and Apple with greater fines in

[9] U.S. House Judiciary Committee, “Online Platforms and Market Power, Part 6: Examining the Dominance of Amazon, Apple, Facebook, and Google”, July 29, 2020, (judiciary.house.gov)
[10] The Economist, Alphabet, Amazon, Apple and Facebook face an antitrust grilling, July 28, 2020
[11] Federal Trade Commission, Privacy and Data Security Update: 2019
[12] The Economist, Dismembering Big Tech, October 24, 2019

case of disrespect and violation of these rules[13]. These rules go in parallel with the important and modern regulations emanated by the European Commission, i.e. the "Digital Markets Act" and the "Digital Service Act": while the Digital Markets Act is aimed at "ensuring fair and open digital markets", the Digital Markets Act is aimed at "ensuring a safe and accountable online environment". The Digital Markets Act[14] proposes new rules that favor growth and a higher level of competitiveness, thus favoring the entry of smaller companies. Furthermore, the rules are aimed at putting the individual user at the center, protecting their online rights and establishing greater transparency in the interaction between the user and the company.

Another important issue is the presence of a data market and the related data-brokers. The fact that data, in addition to being collected, is also exchanged and marketed has given rise to great perplexities and has required intervention by the legislation. The subject of legislation does not only concern the protection of the individual, defined as the "data-subject", but also the trade in data and the competition or competitive advantage that a company can enjoy through the data at its disposal.

Given the great concerns, however, the responsibility should not lie entirely with the individual. Companies and businesses are asked to keep customer data safe, either because it is imposed by the law or because they want to build trust. Data collected by companies (e.g. location, online and offline monitoring, personal data) are of enormous importance to them and help companies to develop more efficient products and services, as well as advertising and marketing campaigns and the possibility of implementing discrimination practices. While customer data provides opportunities for companies to achieve better results and increase profits through personalized offers for customers, on the other hand, in order to access this data correctly, organizations must establish rules and actions of trust and transparency regarding what such data will be used for and with whom it will be shared[15].

In 1967, Alan Westin wrote a book called "Privacy and Freedom" in which he defined privacy as "the claim of individuals, groups or institutions to determine for themselves when, how, and to what extent information about them is communicated to others"[16]. By the 1990s, almost all countries followed suit. An example of this is that the Supreme Court embraced

[13] The New York Times, Big Fines and Strict Rules Unveiled Against 'Big Tech' in Europe, December 15, 2020

[14] https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-markets-act-ensuring-fair-and-open-digital-markets_en

[15] KPMG (2018), GDPR, data privacy and the consumer

[16] Alan F. Westin, Privacy and Freedom (New York: Atheneum, 1967), p.7

this view and in 1989 the Department of Justice stated that “privacy encompass[es] the individual's control of information concerning his or her person"[17]. This was not an US phenomenon and in fact Europe, Asia, and other countries have tried to adapt their rules and legislations. Europe, for instance, enacted the General Data Protection Regulation or GDPR (2016), which "strengthens data protection safeguards, provides additional and stronger rights to individuals, increases transparency, and makes all those that handle personal data more accountable and responsible"[18], or more recently California with the California Consumer Privacy Act (2018). These legislations were created with the aim of informing users on how their data are collected and treated, as well as providing them more control over the information released online.

After all, privacy is intrinsically linked to the individual, and the growing loss of control over personal information as well as the need for greater transparency with regard to data collection, data storage and data processing has made it necessary the intervention of legislation and regulators. Since it came into effect, the GDPR has attracted a lot of attention around the world and therefore there has been discussion about the possibility of obtaining a "global privacy consent". The evolving process regarding regulations about data privacy is continuing and it is bringing new proposals and regulations. For instance, another proposed extension of the GDPR is the “ePrivacy regulation”[19] (Directive on Privacy and Electronic communications) and the Digital Services Act which, but new regulations are emerging in countries such as South Africa, India, Singapore, everyone with its adaptations. In the end, the path being pursued is that of a common agreement on the importance of issues related to privacy, not only at a European level but also and above all at a global level.

1.2 Economic theories about privacy and the three waves

Economists have tried to analyze privacy from different points of view, considering that an agent aims to maximize its utility, but privacy can positively or negatively affect this utility, as well as it can also affect transactions and equilibria regarding data disclosure. The economics of privacy regards the “economic value and consequences of protecting and

[17] US Supreme Court, JUSTIA, Department Of Justice v. Reporters Comm. for Freedom of the Press, 489 U.S. 749, 763 (1989) (https://www.justice.gov/archive/oip/foia_guide09/exemption6.pdf)
[18] European Commission, Commission report: EU data protection rules empower citizens and are fit for the digital age, June 24,2020
[19] European Data Protection Board, Statement on the ePrivacy Regulation and the future role of Supervisory Authorities and the EDPB, adopted on 19 November 2020

disclosing personal information" and "the trade-offs associated with the balancing of public and private spheres between individuals, organizations, and governments" (Acquisti et al., 2016).

Economics of privacy is not a recent field. Actually, we refer to privacy as something that has evolved over time and as something that has been attributed to different meanings and definitions according to the period of reference.

While the first works on privacy were mainly focused on the topic of "privacy of information", which is a fraction of economics of privacy, subsequently, the study of privacy turned to the collection, processing and use of personal information, as well as the disclosure of data and information by individuals themselves. From this perspective, the economic studies concerning privacy have focused on the so-called "informational privacy", i.e. the "personal information and the problems and opportunities created by its collection, its processing, its dissemination, and its invasion" (Brandimarte and Acquisti, 2012). Another perspective describes privacy as "the policies, procedures, and other controls that determine which personal information is collected, how it is used, with whom it is shared, and how individuals who are the subject of that information are informed and involved in this process"[20]. At the center of this vision there is the individual, since the information revealed about a person is intrinsically a specific trait of the person. In the past two decades (2000-2020) the explosion of research on economics of privacy has significantly enlarged the boundaries of the study about privacy by economists.

The evolution of the concept of privacy should be related to the period of reference: as time changes, the concept of privacy has evolved accordingly. Acquisti et al (2016) analyze this evolution talking about three main "waves":

- First wave (early 1980s)
- Second wave (mid 1990s)
- Third wave (2000s and onward)

[20] Lauren Steinfeld and Kathleen Sutherland Archuleta, "Privacy Protection and Compliance in Higher Education: The Role of the CPO," Educause Review, vol. 41, no. 5 (September/October 2006), pp. 62–71

1.2.1 The first wave

Economics of privacy dates back to the early 1970s and 1980s (i.e. "first wave" period), as an issue treated by important economists in the field of economics and law, such as Richard Posner and George Stigler, who were part of the Chicago School of Economics, a school mainly known for its laissez-faire approach and the strong belief that when markets are left mostly to their own devices, they perform best.

In his articles of 1978 and 1981, Richard Posner touched on many interesting points. First of all, Posner admitted that there could be many different dimensions of privacy, confirming that privacy is a malleable concept. Among the things written by Posner, a relevant definition was that of privacy as a "concealment of information", and this idea was followed for a long time by economists. Moreover, Posner distinguished between "good information" and "bad information" (e.g. negative traits): according to this distinction, Posner stated that while individuals with positive traits have interests in showing them, people tend to hide bad information about themselves. For instance, an employee that is deficient in some characteristics has an incentive to conceal those deficiencies. Basically, the idea is that if an individual has "bad information", this can be a good reason why he may want to have his privacy protected.

Another important conclusion made by Posner in his work was that, by reducing information available to "buyers" in the market (e.g. in the example above, the employers), the efficiency of the market is reduced. According to this view, the "cost of privacy" or "concealment cost" is not incurred by the privacy subject, but by other people who cannot access that information. For instance, if an employer is not allowed to do drug-testing on employees, an employer may end-up having some deficient employees (e.g. drug-addicted employees): in the end, the buyer is "paying the price" for the privacy of the employee. Posner extended the argument also to non-market economic behavior. As an example, talking about the marriage, Posner said that there is an incentive to hide information before the marriage, because individuals tend to highlight their positive traits and at the same time hide their negative traits: once again, the cost of concealment is incurred by who receives misleading information (in this case the partner). Finally, Posner believes that privacy is re-distributive: privacy creates a reallocation of the costs from one party to another and this reduces efficiency.

According to Posner, privacy is depicted as something negative, since it can be seen as something that reduces market efficiency and therefore a possible solution to limit the negative effects of privacy is the usage of no regulations at all.

Following the idea proposed by Posner, George Stigler (1980), another member of the Chicago School of Economics, suggests that, independently of whether there is regulation that creates a right of privacy or not, just by following the economic incentives the exchange of information will lead to desirable economics outcomes independently of the ownership of data. For instance, consider the difference between a "good debtor" and a "bad debtor". For a good debtor, one that pays his debt in full, it is useful that the information about its reliability is well known, and therefore it seems reasonable for him to have a system that tracks the credit history of the debtor and to share his track-record with as many entities and credit institutions as possible. On the other hand, a bad debtor will push for having privacy on its credit history, but at the same time, by hiding this information, the debtor will end up paying higher rates on the debt. If people expect that good debtor shares information about his credit history and bad debtor hides these information, whenever an individual that hides information is encountered, it can immediately conclude that the individual is a bad debtor and therefore, regardless of whether there are or not information about the credit history of the individual, the same individual will end up paying higher rates on the debt.

Stigler (1980) proposed that data should not be owned by the person that owns the data, but by the person or the entity that incurs the cost of acquiring the information (i.e. the entity has the right over data). For instance, if Visa is building a large database and a big infrastructure to capture information about credit card transactions of an individual, that information belongs to Visa from a legal point of view and it is not owned by the individual. At the time, the idea proposed by Stigler was forward-looking, in the sense that his view was in contrast with the common legal view of the time, which instead was suggesting that the data-subject should own the data.

The issue related to the "data subject" and the idea that the subject is the owner of the data has been debated a lot, also in more recent times. In 2020, in the US some candidates of the Democratic party have proposed the idea of "data dividends", i.e. the concept that people should earn dividends and payments for the data that digital platforms (e.g. social media platforms) used about themselves[21].

[21] Source: https://www.datadividendproject.com

1.2.2 The second wave

After the "first wave", characterized by a period of growth in the topics related to economics of privacy, there was a relatively flat period on the issue regarding privacy and economics. In the 1990s, a new generation of economists, including Eli Noam, Kenneth C. Laudon and Hal Varian, started reconsidering economics of privacy and more specifically "privacy of data" in the light of the Information Technology (IT) revolution, defined as the "second wave".

Hal Varian (1996), among his books and articles, published a little paper about the economic aspects of personal data, in which he touched some important issues regarding personal data. First of all, he pointed out that positive or negative externalities are really what creates problems in terms of privacy management from an economic standpoint. For instance, a consumer can rationally decide to share or not some information with a company, but once data is shared with a company there can be many potential secondary uses of this data that the customer may not anticipate but for which he supports the consequences. These additional uses could create externalities, so some other trade-offs, both positive and negative, which end up affecting the consumer but over which the consumer has little or no control. For example, Amazon could generate positive externalities for the consumer by improving the recommendations for him, but at the same time the company can use these data and sell them to other companies and these other companies can ultimately use it in an harmful way, for instance they can be used for price discrimination practices against the customer.

In addition, Varian suggested that, during the 1990s, privacy was becoming a contested topic mainly because of the process of the digitization of information, which has collapsed the marginal cost of data collection and data storage. This has facilitated the access to resources technically difficult to access for the public, making them accessible with very low costs. Therefore, when the cost of collection, storage and of accessing information collapses, more and more people will be willing and able to access information, and this creates new privacy problems. The proposal of Varian was to assign property rights to private information in order to allow consumers to take control of how information about them is used, and for example making it costly for companies and organizations to access certain digital information.

Eli Noam (1997) suggested that in absence of transaction costs in trading data, initial assignment of privacy rights is arbitrary from the standpoint of economic efficiency, an idea that recall the theorem of Ronald Coase (1960). Eli Noam applied the idea of Coase to

privacy: the idea of Noam was that what really matters in terms of privacy outcome in the marketplace are the valuations of the data made by different entities. For example, if a consumer values its privacy more than what Amazon values the access and usage of consumer data, the consumer will be willing to pay Amazon an extra premium in order to get its privacy protected. Whereas if, on the contrary, Amazon is more interested in the data of the consumers than what the consumer value its privacy, the valuation by Amazon of the data of consumers is higher than the valuation about privacy by the consumer, and therefore even if there is a regulation that in theory protects the data by customer, Amazon will offer the consumer extra money (or a discount on the products) to get access to customer data. From this perspective, Eli Noam concludes that the use of data protection systems (i.e. cryptography) has no effect on the output of a transaction and on what ultimately remains to the party that has control over the data, but rather on the mere value exchanged between the two agents involved in the transaction. Essentially, the system of encryption does not end-up determining the final outcome of the system from an economic point of view, but it impacts on the issue related to "who has to pay whom?". Once again, if there is a regulation protecting data by a customer and Amazon really wants to, the company itself will have to pay the customer (i.e. the owner of the data), and so there is a redistribution of wealth from the service provider to the data subject. If there is no regulation, but the consumer really wants to protect its privacy, then in this case the customer will pay Amazon, so there is a redistribution of wealth from a privacy-conscious consumer to the organization. Therefore, the outcome is not affected inherently by the law, but it is affected by the economic incentives of the agents. However, the theoretical thinking of Noam revealed some problems from a practical perspective, because there are a lot of difficulties in constructing a system or a contract which allows consumers to freely trade ownership of their data.

Kenneth C. Laudon (1997), following somewhat the idea proposed by Hal Varian, tried to propose an information market in which the individual could exchange and transfer the rights to their personal data in exchange for a monetary counterpart. More precisely, the aim of his ideas was to "allow personal information to be bought and sold, conferring on the seller the right to determine how much information is divulged" (1996).

During this period, the interest among economists was on the economics of information. In 1998, Hal Varian and Carl Shapiro wrote a book called "Information Rules" about how to apply economics analysis to information markets and information goods. The point was that when we analyze information using economics we are analyzing specific goods and services that we call "information goods". For instance, a book can be considered

as an information good, because most people buy a book because they want the information contained therein. The same is true when a person buys a course or streams a video from the internet: it is not the physical experience of sitting and watching, but it is the information contained that matters the most. Under this perspective, there are a lot of goods and services that embody information and which are bought and sold on the market, but information itself is not an economic good: a person does not buy information, but books, lectures, consulting services and so on.

1.2.3 The third wave

The second wave set the foundation for an explosion of the research in the field of privacy economics that has taken place starting from the early 2000s, which we refer to as the "third wave". The third wave has been characterized by a huge interest in the economics of privacy, and in fact, for the two decades 2000-2020 it can be found an enormous amount of empirical and theoretical research of privacy economics. After the 2000s, because of the commercial expansions of the Internet and of the World Wide Web, there was a parallel expansion of the field of economics of privacy among its fragmentations, and in this way economics of privacy started to have more and more sub-fields. During this period, different things happened simultaneously:

1. An increased modeling sophistication: increasingly sophisticated theoretical game models and industrial organization models;
2. A diversification of focus of researchers and academics;
3. The emergence of empirical analyses: data to study in order to understand how customers value personal information, how companies value data, and so on;
4. The emergence of applied behavioral economic research, used as an attempt to study privacy decision making;
5. A parallel emergence of the economics of information security. The term "information privacy" itself refers to data privacy and data protection. This idea comes from the more recent relationship between data and technology: the need for protecting privacy is intrinsically related to technology and with the usage of technological devices. In the early 2000s, Ross Anderson wrote a paper called "Why

is information security hard?" (2001)[22], in which he blended information technology and cryptography and that led to the formation of a work on social and information security which since then has become very strong and more developed.

Graphically, the representation of the three waves can be depicted as:

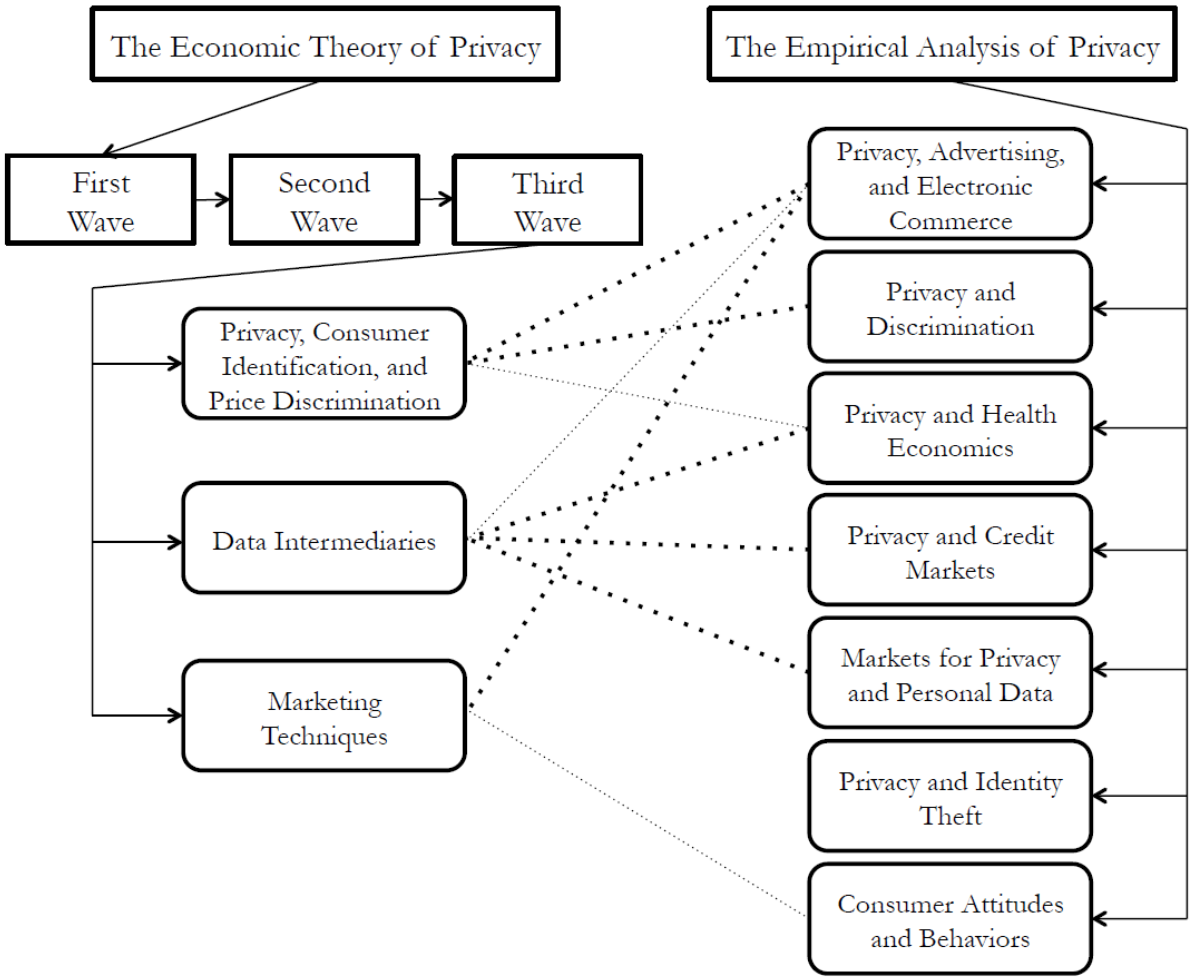


Figure 1: Alessandro Acquisti: Privacy, Policy and Regulation(2019) Atlanta Fed Financial Markets Conference

Following the distinction proposed by Acquisti et at (2016), the revision of the so-called "three waves" still confirms that the economy of privacy is a vast, complex, and malleable concept. The issue of privacy has been analyzed through different perspectives and fields of application; however, each of these has established the need to protect what is, after all, a fundamental human right. The problems that emerge from the violation of privacy, or in

[22]Anderson, R. "Why Information Security Is Hard - an Economic Perspective." Seventeenth Annual Computer Security Applications Conference, IEEE, 2001, pp. 358–65

general from the loss of this right, put the individual at serious risk. That said, legislation has also sought to protect individuals given the rise of numerous trade-offs associated with privacy

1.3 The consumer approach to privacy

In March 2020, Siân Brooke and Carissa Véliz, two professors from Oxford University, published the results of a questionnaire conducted online, concerning individuals with different nationalities and focused on issues related to privacy[23]. An important finding was that 92% of respondents had a negative experience related to privacy, whether it was identity theft, public humiliation, or even being a target of spyware.

What do people think about their privacy? Do they really care about protecting their personal data and information? To what extent are they willing to receive personalized offers in exchange for the disclosure of their data?

People react in different ways to privacy: some of them are really concerned about the current situation, while some others are not so worried about it. Human behavior towards the protection of personal data has been studied on one of the many branches of the privacy economy, called more precisely "behavioral economics of privacy".

There can be many issues in relation to how individuals take decisions regarding their personal data. A first problem concerns information asymmetry, i.e. the fact that individuals often do not know or cannot know what is happening to their personal data, due to a lack of transparency. A second problem, related in some way to the previous one, is bounded rationality: even if individuals were given as much information as possible about what happens to their online data, that information would be overwhelming. The analysis of these first two problems reveals a possible explanation: very often it is difficult to reach an optimal result in terms of privacy, not because of individuals, but because the ramifications and complexities of data sharing are not easy to detect and understand fully. A third problem is related to behavioral and cognitive biases: the decision-making process of individuals with regard to their data can differ from theoretical models of rationality commonly used in economics. For instance, in a series of experiments, Brandimarte et al (2013) found that giving people more control over their personal data can paradoxically make them more willing to take risks over their personal data by disclosing more information than what is

[23] Brooke, Siân&Véliz, Carissa (2020). Views on Privacy. A Survey. In Data, Privacy, and the Individual. PhilArchive copy v1: https://philarchive.org/archive/BROVOP-3v1

needed to strangers. This conclusion does not reveal the uselessness of personal data protection policies, but it highlights that the way in which consumers are given the possibility to control their personal data, which is related to the drafting of privacy policies, can influence and alter consumer decisions, exposing them to greater risks.

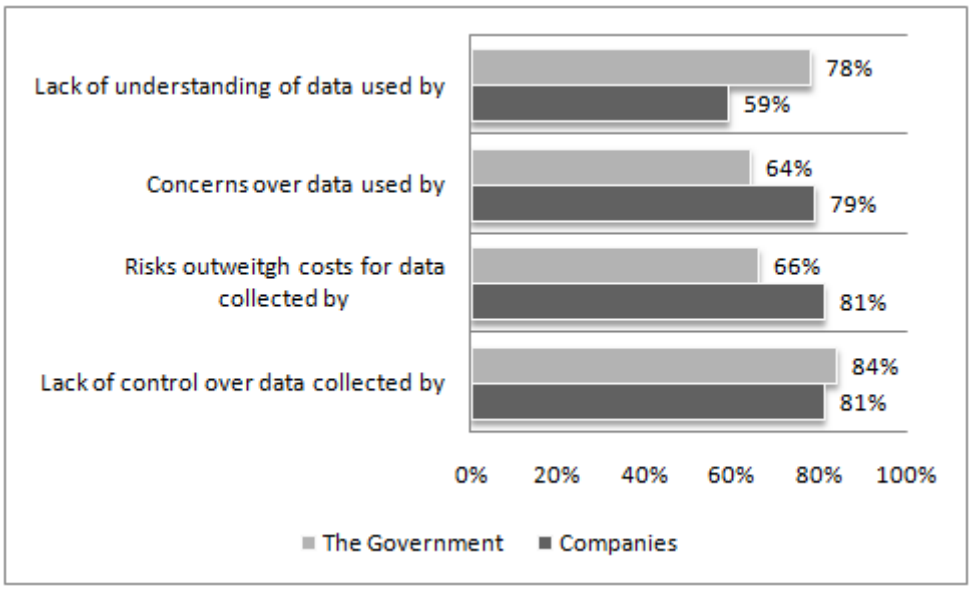
In the modern and digital economy, an important trade-off is between what customers offer about themselves (i.e. personal information and data) and what they get in exchange (e.g. recommendations and personalized offers). While people are willing to share personal information in exchange for tangible benefits or enhanced products or services, the disclosing of personal information and the way in which companies collect and use data is for them a real concern. However, the fact that these benefits are obtained in exchange for the disclosure of personal information has raised concerns about the net benefits. For instance, Alan Westin, which has significantly contributed to the research on privacy, in his 2008 survey concluded that "offering online users free email or free searches did not seem to a majority of our respondents to be a sufficient set of benefits or valued services to overcome the instinctive feeling of not wanting to be tracked and marketed to based on their online transactions and surfs" (Alan Westin, 2008).

Customers can be more or less aware of the opportunities and threats deriving from an economy based on data. There can be some efficiency gains from the willingness to share certain kinds of information by people. For instance, by revealing certain information they can more potentially find a beneficial match: the matching process works better when the algorithm has specific details about individuals. This happens, for example, with the recommendations of Amazon when an individual is shopping online, or even when he uses applications for dating. On the contrary, this helps companies to improve the personalization of their product and services, and facilitate innovation. However, revealing information can also have side effects: as an example, users can be the object of target advertising or price discrimination strategies. Sometimes, when revealing data, some customers can even adapt their behavior according to the information disclosed, in this way starting to regulate and to pay more attention to their actions (i.e. Hawthorne effect).

Said that, can customers fully protect themselves if they want to? Or, putting it in other terms, are there really market failures related to privacy? Among the multitudes of ways in which the term "market failure" can be used, in this context it is referred to the case in which there is a demand for a good and the market actually is not providing the proper supply for that good. If we analyze the issue from the perspective of demand and supply for privacy, it can be said that there is a real supply for privacy because there are some companies that

adopt privacy protection instances or that offer customers privacy technologies (e.g. Tor, ad-blockers services, etc).

In 2016 the Pew Research Center[24] conducted a study in the United States in which they found that around 91% of participants agreed that they had lost control of their personal and private information. Moreover, 86% of respondents said they have taken action to increase their level of online privacy protection, but much can still be done. A survey of McKinsey (2020) revealed that even if online users are aware of the online risks and practices regarding their data and their privacy, only a small portion of them are taking adequate countermeasures[25].



Source: Adapted from Pew Research Center (2019)

This graph, retrieved from the survey of the Pew Research Center (2019) shows the percentage of respondents who feel little or no control over the information collected by companies or the US Government. Overall, this graph shows that individuals suffer from different kinds of concerns regarding their data and the associated loss of privacy. More than 80% of respondents felt the lack of control over the information collected by companies and governments for data regarding varieties of details: physical location, activities on social media, private conversations, history of online and offline purchases, websites visited and searches done online. In addition, the lack of privacy and control over data is perceived as

[24] Pew Research Center, 2016, The state of privacy in post-Snowden America
[25] McKinsey & Company, The consumer-data opportunity and the privacy imperative, April 27, 2020

risky, associated with privacy concerns and lack of transparency on how data collected are then used by both companies and the government.

There are products and services that are used to embody information or to facilitate the enhancement of privacy. Therefore, like any other good or service, there is a market for privacy goods services, and the goods and services for protecting privacy vary according to the degree of privacy protection a person is looking for. Moreover, the market provides information and communication goods and services with different features, one of them being privacy: when looking for a product or service, a customer takes into consideration its various attributes, also in relation with its alternatives, such as the quality, the ease of use, the added value that a product or service brings, and so forth, but also the level of privacy protection as one of these attributes. People are choosing among different privacy goods and services they consume, and each of them provide different levels of privacy, and the willingness to consume them depends on the overall benefits and costs of those, relative to other goods and services that provide other features with more or less privacy protection.

In the market, companies offer goods and services that include a bundle of features, some of which relates to privacy, while some others do not. In a digital world, there can be many options an individual can take: encrypted or regular emails (2018 data reveals $123M extracted from Facebook and Google, McKinsey), chats, voice calls and video calls, browsers that block cookies or not, social media or search engines that store information or not, ad-blocking software (McKinsey stated that ad-blockers are used globally by more than 600 million devices) and incognito browsers (40% of internet users globally, McKinsey)[26]. For instance, a person can decide to use Telegram or Signal, instead of regular text messaging, or to use Tor or other peer-to-peer encryption systems like DuckDuckGo instead of Google in order to get a higher degree of privacy protection. McKinsey has reported a survey in which 14% of internet users adopt encrypted communication systems and only ⅓ of them change their online passwords on a regular basis[27].

While some people use alternative solutions and give up certain features in exchange for a higher level of privacy, not everybody is willing to make this cut, even if conscious that using certain tools leads individuals to be more subject to be traced or to a lower level of privacy protection. Let us consider DuckDuckGo, which offers a privacy-focus search engine, an alternative to Google: while Google dominates the search engine market (85%), DuckDuckGo has only the 1% of the market, despite having registered a significant past

[26] McKinsey, The consumer-data opportunity and the privacy imperative, April 27, 2020
[27] McKinsey & Company, The consumer-data opportunity and the privacy imperative, April 27, 2020

growth either in the market share and in its profits (The New York Times, 2019). Like Google and any other search engine, DuckDuckGo shows ads on the top of its search results, with the difference that it does not track the online behavior of users and therefore it does not show target and personalized ads. Despite the company offering an easy-to-use interface, users believe that by using it, they would obtain lower quality results despite the higher level of privacy protection. This example actually shows how people do not seem willing to "give up much to recover their privacy, and are easily overwhelmed when they decide to try to make a change"[28]. If people value privacy over other attributes, such as the quality of the search results, then we would expect a wider usage of services providing a deeper level of privacy protection such as DuckDuckGo, but actually only a small portion of the online users adopt it.

Even if it is not possible to entirely protect privacy in the online world, there are some countermeasures available for every online user. However, when dealing with privacy and privacy issues, there is a significant trade-off between stated preferences and revealed preferences: while on one side people feel like they are losing control over their personal data and information, on the other side they are not actually willing to use alternative tools, services and countermeasures that allow them to reach an higher level of privacy protection.

Generally speaking, there have been different proposals regarding tools that increase privacy protection, such as tools with more transparent privacy policies and applications increasing the level of default encryption. There are a lot of opportunities for consumers to reach a higher level of privacy or to choose bundled services that include levels of privacy at the expense of some other attributes, but stated preferences do not always correspond to revealed preferences. Despite people claiming to care about privacy, several times they end up making choices that are inconsistent with their stated preferences (Athey et al. 2017). This dichotomy is expressed by the "privacy paradox": when dealing with privacy, choices of customers can be in contrast with their actual behavior when they are incentivized to do so, e.g. when using a free or supposedly free product or service.

In the context of the behavioral economics of privacy, Brandimarte, Acquisti and Loewenstein (2013) study how people make privacy-related decisions about the protection of their data and how these decisions can sometimes be influenced by factors that in theory should not matter too much, such as the way in which requests are presented to them. The authors carried out a study on the control of personal data (i.e. "control over information

[28]The New York Times, A Feisty Google Adversary Tests How Much People Care About Privacy, 15 July, 2019

flows") trying to understand if, by giving people more control over their data, they feel more protected and paradoxically begin to take more risks, revealing more information about themselves. The authors conducted a random experiment involving two groups. A first group of people was shown a questionnaire with less invasive questions (e.g. sentimental situation) and others more invasive (e.g. use of drugs or sexual activity): the questionnaire provided for voluntary disclosure by respondents, meaning they could decide whether to answer a question or not, knowing that only the answers would have been taken into consideration. A second group of participants in the experiment was given a very similar questionnaire, but with one significant difference aimed at giving respondents more control over their information: an addition of a column in which, for each question, the candidate authorized or not the use of that particular answer. Paradoxically, the authors demonstrate that by making people feel more in control of their privacy, it makes them more willing to disclose personal information: in the second group, where people in addition to responding gave consent to use of information, the estimated response rate to the most intrusive questions was double that of the first group.

In conclusion, the privacy paradox reveals the issue behind the decisions affecting privacy made by users, a thing that can be taken into consideration when new regulations are drafted, avoiding regulations that "inadvertently lead consumers to be faced with additional effort or a less smooth experience in order to make a privacy-protective choice" (Athey et al., 2017).

Another thing to mention regards privacy policies terms and conditions, widely used by companies and online service providers to regulate the usage of personal data they collect, also thanks to regulations such as the European GDPR (2016) and other international proposals. These are very important documents in which users have the possibility to be aware of how their personal data are collected and treated. The problem is that not every person reads carefully these terms and conditions and this can be related to many factors (Steinfeld, 2016): e.g. the possibility to easily forego reading these documents, the length of the document, or the fact that people are more interested in only some paragraphs and not on the entire document[29].

Pew Research Center (2019) shows that about one-in-five Americans always or often reads privacy policies before agreeing to them, and moreover, only a minority of those who

[29]Nili Steinfeld, "I agree to the terms and conditions: (How) do users read privacy policies online? An eye-tracking experiment", Computers in Human Behavior, Volume 55, Part B, 2016, Pages 992-1000

read these terms and conditions say they read them all the way through[30]. In addition, the study by Pew Research refers that most Americans aim for stronger and stricter government rules and regulations about what companies can do with individual data, as well as they think they need, in almost equal terms, either better tools for allowing people to control their personal data by themselves or stronger laws governing companies.

The consent that people give to privacy policies is of particular interest, especially for few important aspects:

- Length and complexity of privacy policies consent, two factors that at the extreme make the privacy policy inaccessible or difficult to read;

- The way in which privacy policies are presented is somehow not useful at all. For instance, sometimes a user must forcibly give his consent in order to access a specific website.

- "Responsibilization": legal consent should not shift data responsibility to the user giving consent, instead it should give equal responsibilities and equal rights to both parties, with a focus on consumer data and privacy protection. Otherwise, the act of giving consent would have strong legal significance and implications mainly for the user.

In addition, some policies have constantly been changed over the past year: for example, it has been analyzed that the privacy policies of Google evolved from a two-minute read in 1999 to a peak of 30 minutes in 2018. For this purpose, regulations and legislations have intervened over these policies in order to push for a "concise, transparent and intelligible form, using clear and plain language"[31]. Despite the effort made by regulators and legislations to make them more accessible, privacy policies still remain full of legal jargon and opaquely explain how companies collect and manage data of individuals.

The problem is that when an individual gives consent to the processing of personal data without understanding how their data will be processed, which often happens when a user wants to quickly access digital content. To this regard, Jonathan David Leibowitz, who also served as the Chairman of the Federal Trade Commission (FTC), which represents the United States largest privacy regulator, said: "Initially, privacy policies seemed like a good

[30] Pew Research Center, 2019, Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information, November 15, 2019

[31] The New York Times, Opinion | We Read 150 Privacy Policies. They Were an Incomprehensible Disaster

idea. But in practice, they often leave a lot to be desired. In many cases, consumers don't notice, read, or understand the privacy policies. They are often posted inconspicuously via a link at the very bottom of the site's homepage – and filled with fine-print legalese and technotalk"[32].

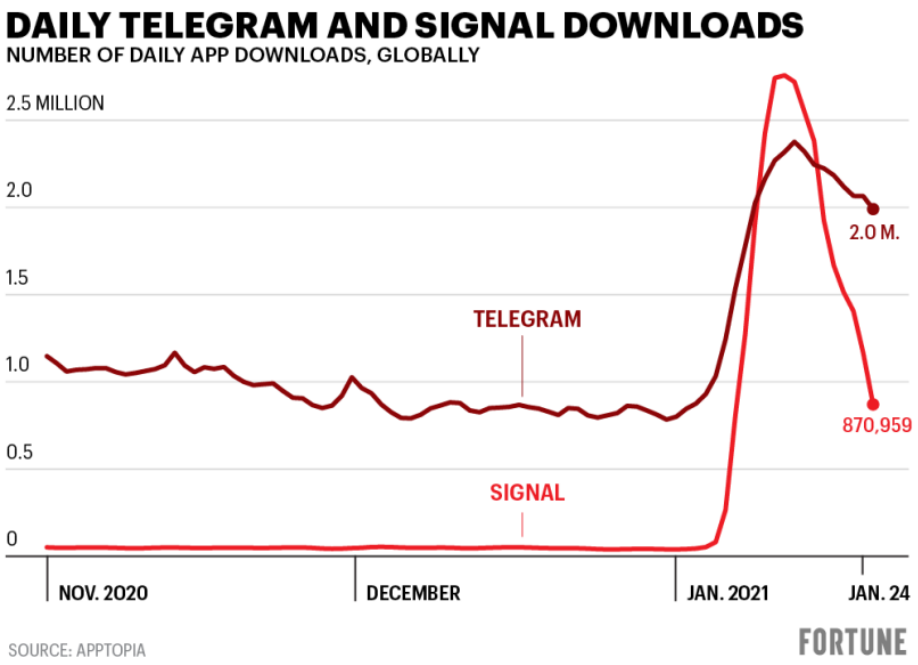
Privacy policy dictated by the United States and the OECD have focused on the idea that "with enough transparency and enough choice consumers would make better privacy decisions" (Athey et al., 2017). With the introduction of the Privacy Act of 1974 and the rules contained therein, "notice and choice" methods were used to safeguard privacy: first there is a notice, whose task is to provide users with information regarding the collection and use of personal information, then, on the basis of the information made available to them, users make a choice, deciding whether to agree or not to policies. Based on this notice and choice system and when incentivized to do so, Athey et al. (2017) demonstrate that users can be conditioned on whether or not to share their data and may even choose technologies that offer less privacy protection. Therefore, in the same way, privacy methodologies based on the notice and consent approach may not be efficient in their role of protecting digital users: the consent given by the user may not be the mirror of a true intention. It is therefore a certain complexity that consumer choice should be given to governments, companies and organizations that should keep their personal data safe.

Another important thing that influences the decision of users about privacy is what in the industrial organization field is called "network effect". Networks, especially in the online world, can become of bigger size also thanks to the so-called "data network effect"[33], i.e. the value of a product or service increases the more data goes into the system. The underlying idea is that, ultimately, all the users of the product or service benefit: the more users join the network, the more data they release from their usage of the product or service. Data put into the system is used to improve the value which is given back to users (very often through algorithms) and this additional value should end-up attracting even more users. That is a virtuous cycle that can significantly give companies a significant competitive advantage and for them this works better when everything is automated: the way in which it captures data, the way in which the product or service gets more valuable (especially if the system is capable of collecting data in real-time) and the way in which the product or service is provided to customers.

[32] Federal Trade Commission, So private, so public: individuals, the Internet & the paradox of behavioral marketing, Remarks of Commissioner Jon Leibowitz at the FTC Town Hall Meeting on "Behavioral Advertising: Tracking, Targeting & Technology" November 1, 2007
[33] The Economist, Data is giving rise to a new economy, May 6, 2017

Consider Waze, the application that provides information on road traffic thanks to the real-time contribution of the network of drivers: the more users join the network and the better it is for each one of them. Waze needs a minimum amount of information disclosed by customers to function properly, but it also collects data from the network and transmits it to the different users in real time. Therefore, the core value of the application is based mainly on the data collected and then transmitted to users, and users themselves can perceive the value created by data, either from those they introduce, but also from those that they receive back from the application, independently from the data introduced. This example shows how people can sometimes be incentivized and willing to introduce their information (e.g. name, location, etc) to benefit from a service that potentially gives them back useful information.

Individuals may be willing to join a network for many different reasons, either because companies are offering particularly good products or services, or because they want to join and feel part of a network or, in general, because they perceive value from being part of the network itself. The benefit of the individual also lies in the diffusion and pervasiveness that a particular product or service has reached. The network effect has been particularly important for platforms such as Facebook and WhatsApp, which over time have managed to have a significant user base, but when privacy issues affecting one of these platforms arise, such as in January 2021 with WhatsApp[34], the mass transition can be a difficult process to implement.



[34] Bloomberg, Why WhatsApp's New Privacy Rules Are Sparking Alarm, January 11, 2021

Telegram and Signal have been two applications that, differently from Messenger and WhatsApp, have been used mainly by users willing to obtain a much higher level of privacy protection. As the above graph shows[35], with the proposal at the beginning of 2021 to merge WhatsApp data with the rest of the advertising operations of its parent company Facebook, the number of daily downloads of Telegram and Signal have skyrocketed. This might suggest that people have decided to embrace the use of applications that offer them a greater level of privacy protection, but the adoption of Telegram and Signal did not coincide with a decline in the use of Messenger and WhatsApp. Therefore, this demonstrated the strength of the network effect and the fact that still a large number of users were not so worried about privacy concerns.

1.4 Privacy trade-offs

Everything done online releases some data about individuals. The American Civil Liberties Union reports that information derived from cell phone directories, e-mail information and Internet purchases can "paint a profoundly detailed picture of our lives"[36]. The profiling of each individual can be used, for example, for target advertising purposes and in some cases the user itself can benefit from some suggested products, but the phenomenon should be analyzed as a whole because the user itself can end up supporting the major risks. Moreover, the combination of different pieces of information can reveal a lot about the profile of a person (i.e. “data-linkability”).

When browsing online, or when using an application or a social network, people should be responsible and accurately decide what to share and what not, because once the content is shared, the platform gains a certain control over that information. The underlying idea is that social media are regulated by specific rules and norms. But, if we think about it, social media platforms and applications have gone beyond the scope of keeping people in contact with each other: they have become an integral part of our lives and the functions implemented over time allow users to do even more things.

Tucker (2014) analyzes privacy in relation to platforms, social networks and websites that typically receive significant revenue from advertising, which in turn is targeted based on user characteristics. During the conduct of his experiment, Tucker discovered that the social

[35]Fortune, Signal and Telegram downloads surge, passing Facebook chat tools, February 2, 2021
[36] Metadata and Privacy A Technical and Legal Overview (2014)

network (i.e. Facebook) changed some parameters regarding privacy, which allowed usersgreater protection over their personal data and greater control over the data exchanged with third parties. However, despite the change, advertisers could still use the same personal data available to them to personalize advertisements. More precisely, after the privacy policy change, users were almost twice as likely to react positively to personalized ad content and click on personalized ads, while no significant change was found in non-personalized ads. Despite the limitations of his work, Tucker concludes that greater scrutiny of personal information by consumers ends up benefiting the platforms and advertisers who invest in them.

A similar argument applies for e-commerce platforms: customers have to give away some data before making the purchase, but not all of them are necessarily related to the transaction. In concluding the transaction, customers can reveal credit card information but also other types of sensitive data, such as their behavior in the different websites or the "click path", i.e. the user clicks sequences within an internet page or across different pages. This can be turned into price discrimination strategies at the expenses of consumers.

1.4.1 The Model by Acquisti and Varian (2005)

Acquisti and Varian (2005) proposed a model considering an online monopolist that is able to condition prices on the purchase history of customers. In their model, the seller can register the purchase decisions made by consumers through cookies, IP address, credit card number, user authentication; while customers can put in place some countermeasures to maintain their anonymity (e.g. cookie deletion, private browsing, etc). If the online behavior by customers can be tracked, the monopolist can price discriminate on the basis of the purchase history or "click-stream".

The model consists of two consecutive periods: each customer visits the firm's website twice and in each period decides whether to purchase or not. In the first period, all customers visit the company's website for the first time: the seller offers its price and each buyer decides if to purchase or not. Customers have heterogeneous preferences: a proportion of customers (π) can have a high willingness to pay (v_h), while a proportion of consumers (1-π) can have a low willingness to pay (v_l). However, in the first period the seller cannot discriminate: it only knows the proportion π, but it cannot observe the identity of each customer, that is whether he/she has high or low willingness to pay. In the second period, things may change: thanks to

the information stored in cookies, the seller can discriminate. In fact, after having observed the decision made by consumers the seller can condition its price on the customer's purchase history.

The aim of their model is to determine the optimal pricing strategy for the monopolist in three different cases: i) absence of price discrimination; ii) cookies not disabled and iii) customers disable cookies.

The first case concerns the absence of price discrimination. In this case there is "privacy": the information regarding the first period interactions cannot be used by the seller in the second period to price discriminate. This means that in each period the seller has to decide whether to set a low price (v_l) selling to all customers or to set a high price (v_h) selling to only a few customers. In this case, the seller decides to set the low price if and only if $\pi v_h < v_l$; therefore its profits in each period are given by$\max\{\pi v_h, v_l\}$.

The second case is that of cookies not disabled. In this scenario there is "no privacy", meaning that the information regarding the first period interactions can be used in the second period to discriminate. Customers are "naive": they do not expect the first period information to be used by the seller to discriminate in the second period. Therefore, the seller decides its prices in order to maximize the sum of its profits in the two periods. In order to recognize customers, the monopolist can set a price in the range $[v_l; v_h]$: in this way customers with a high willingness to pay will end up purchasing the product, while customers with a low willingness to pay will not purchase. In the second period, having observed the customers' purchase decision, the seller will charge v_h to customers that have previously bought and v_l to others in order to incentivize them to acquire at least once. In this way, if customers are "naive" the seller will leverage on the absence of privacy, thus discriminating: each customer will pay a price according to its willingness to pay and the monopolist profits are $2v_h + (1 - \pi)v_l$.

The third case is that of customers disabling cookies: we are still in the "no privacy" regime but consumers are "rational". This means that customers can take some countermeasures: conscious of the potential discrimination, they decide not to buy in the first period in order to avoid to pay the high price in the second period. Let identify p_h and x_h as respectively the total price paid and total quantity acquired by the high-willingness-to-pay type of customer in the two periods; same reasoning for p_l and x_l for the low willingness-to-pay type of customer. The sellers chooses p_h, p_l, x_h, x_l in order to maximize $p_h + (1 - \pi)p_l$ subject to two constraints:

- *Participation constraint*: all customers prefer to buy rather than not to buy;
- *Incentive compatibility constraint*: high willingness-to-pay type of customers prefer the offer set by the monopolist for high willingness-to-pay type of customers, same for low willingness to pay type of customers.

In this case, the best strategy for the seller is to set $p_l = v_l$ and $p_h = v_h + v_l$in order to induce high willingness-to-pay customers to buy in both periods and low willingness-to-pay type of customers to buy only in the second period. Therefore, both constraints are satisfied: high willingness-to-pay type of customers buy in both periods, obtaining a surplus of $v_h - v_l$ and low willingness-to-pay type of customers do not buy at $p_h = v_h + v_l$. The fact that the seller has to deal with rational customers forces it to set a lower price in the first period in order to induce high willingness-to-pay types of customers to reveal themselves. In this third case, the monopolist seller ends up making profits equal to $(1 - \pi)p_l + p_h = \pi v_h + v_l$.
Which is the optimal strategy for the firm? In order to answer we need to compare profits in the "no-privacy" regime and the monopolist canprice discriminate, $\pi v_h + v_l$, with those with privacy, $2\max\{\pi v_h; v_l\}$, when the monopolist cannot price discriminate. Discrimination ("no privacy") is optimal for the seller if $\pi v_h + v_l > 2\max\{\pi v_h; v_l\}$ a condition which is never verified. This is due to the fact that the strategic behavior of high willingness to pay consumers forces the seller to set a lower price in the first period to induce them to buy (and to reveal themselves as high type consumers), and this make the discrimination strategy not profitable. The authors analyze extensions to this basic set-up and show that the "no-privacy" regime can become profitable when customers are less than perfectly rational.

In the wake of the Acquisti and Varian model, Taylor (2004) studies price discrimination in the form of dynamic prices, building a model that predicts the presence of two monopolists and a continuum of consumers. In his model, companies collect personal information from consumers and derive value from the possibility to implement price discrimination strategies on the basis of customers' preferences. Consumers can be of two types: either aware of the ways in which companies use personal data of individuals or naive. In the latter case, the surplus generated in a transaction is entirely absorbed by the company, and in fact only adequate regulation can intervene and change this balance. On the other hand, if consumers are aware of the use of their data made by companies, even in the absence of consumer privacy protection companies operate in the interest of customers, ending-up

protecting their privacy. In this way, Taylor demonstrates that consumer choices make deleterious for companies to practice privacy intrusion and violation of customers' data.

Villas-Boas (2004) proposes a work in which strategic consumers decide to be patient in the first period, learning about the product, and then deciding if or not to buy, thus anticipating future prices. This strategic action pushes companies to abandon price discrimination strategies in favor of the voluntary adoption of regulations aimed at protecting consumer privacy.

A study by McKinsey (2020) revealed that people are becoming "increasingly intentional about the types of data they share and with whom": customers are more likely to share data that is a necessary part of interactions with organizations with the organization, while, on the other hand, some other data is more sensitive and in this way the power of trust becomes crucial. For example, McKinsey's survey revealed that people are more likely to share their data with companies that operate in some specific industries, such as companies operating in the healthcare industry. Additionally, another interesting finding was that 87% of respondents admitted that they were unwilling to do business with a company concerned about security practices. Moreover, half of the interviewees stated that they are more inclined to trust companies that ask for the necessary information, thus limiting the amount of personal and sensitive information communicated and transferred to companies.

When a user is using the internet, browsing on their phone, subscribing to a certain social media platform, when completing an online transaction, when giving their consent or when using a supposedly free application, service or platform, companies are keeping track of the behavior of the user. Companies collect and use this information, and then they combine it with data about who the user interacts with, what store the user shops at, and they use that to figure out what the user might be interested in and build ads that they think are more likely to get us to click or buy. The information released make customers more vulnerable and exposed to a series of practices that may include target advertising, price discrimination, profiling. In addition, there is some data that is more sensible than others, such as data regarding health, opinions about politics, religion, or sexual orientation: these are all data that, if shared, can represent a serious threat for the individual. Through algorithms and more complex systems of data collection and analysis, even those data that are apparently less significant can be of enormous importance, because they can reveal a lot more about a person, i.e. attitudes, behaviors, lifestyle, habits, inner emotions. Information shared by users can also be transferred to third parties. Between 2017 and 2018, researchers at Oxford university analyzed approximately a third of the apps available in the Google Play Store and

found that the median app could transfer data to around ten third parties, with one in five apps able to share data with more than twenty third parties[37]. Their analysis showed that 88% of apps could have transferred data to third parties ultimately owned by Alphabet (i.e. Google), while 43% to businesses ultimately owned by Facebook. All of this data helps companies to create detailed profiles of customers.

In 2020, with the introduction of iOS 14.5, iPadOS 14.5 and tvOS 14.5, Apple[38] has introduced new and stricter privacy features for its applications. On one side, applications need to provide clear and transparent information on how they collect and utilize users' data in the Apple Store: this allows users to understand the privacy policies adopted by an app before downloading it. On the other side, Apple introduces the "opt-in" regime: though the AppTrackingTransparency (ATT) framework, the individual consumer must provide his affirmative consent to allow companies to track users and the consent for tracking allows companies to: i) understand which websites, other applications and offline places consumers have visited to put in place targeted advertising; ii) share user-related or device-related data with third parties and/or data-brokers. As reported by a study of AppFlyer, by the time early adopters of the new Apple operating system were asked, "most users (99%) choose not to allow tracking": consumers therefore revealed their tracking concerns by adopting an opt-in system that allowed them to obtain a greater level of privacy protection. However, as Apple itself expresses, there are still few but limited possibilities for applications to track users without obtaining their consent: for instance, data tracked can be shared with data-brokers for fraud detection, fraud prevention, security purposes, or to evaluate consumer's creditworthiness. In addition, there are still doubts about Apple's ability to verify the actual compliance with the new privacy rules by these applications.

Product customization requires possession of detailed customer information. In the digital world, for companies it is easier to recognize and collect data. Unlike the past, companies now have adequate tools to recognize which sites are searched for by online users, how they spend their time, and can even deduce additional information. When a product is personalized and oriented to the characteristics of specific customers or users, these same customers will end up assigning a higher value to those products. Product customization is also accompanied by a well-prepared pricing strategy, through which prices are established in such a way as to extrapolate the largest possible amount of this value.

[37] https://ig.ft.com/mobile-app-data-trackers/
[38]https://developer.apple.com/app-store/user-privacy-and-data-use/

At the fine line between public and private, the economics of privacy highlights tangible and intangible benefits and costs related to both the protection and disclosure of personal information. These can be advantages and disadvantages for the data-subject, i.e. the subject to whom such data refers, but also for companies and organizations collecting and using such information.

- From an economic perspective, the process of personalization of contents seems to be efficient because on one side consumers are getting relevant content or potential benefits, while on the other hand firms can easily target customers. Consumers search for products that meet their needs and this potentially increases their welfare and reduces the search costs, but this is not always the case. For instance, users can benefit from posting a photo on social media or from receiving a highly personalized service (e.g. discounts or offers), but that same information, on the other hand, can be used to carry out target advertising or price discrimination practices or they can be victim of privacy-related issues.
- Similarly, companies typically implement their strategies based on personal data at their disposal however they support some costs. For instance, through target advertising, companies can allocate the budget to the segment of customers which is theoretically more interested in their products. A comment that Hal Varian and Carl Shapiro make in their book is that when digitizing information, marginal costs of the good are collapsing, therefore this means that digital information is not costless. For instance, the cost that Facebook has to incur every day 48 just for running its network is significant, but the marginal cost of allowing one more participant to join the platform of Facebook is very small.

In this context, decisions taken by economic agents play a fundamental role. Consider the data-subject: he can decide whether to reveal his personal information to receive a more personalized product or service, but with the risk of being a victim of target advertising; or he could decide to keep his privacy protected avoiding price discrimination practices but giving up potential benefits. Similarly, a company can decide to guarantee greater customer data protection by renouncing to implement personalized strategies in this way gaining trust especially among privacy-oriented consumers; or it can try to implement targeted strategies with the limits imposed by regulation in terms of privacy protection: in fact, if the company does not respect these rules, it can support negative consequences, such as fines or loss of reputation.

In the context of economics of privacy, and in general when there is an interaction between two agents on the basis of data (either released or collected), the decisions taken by economic agents significantly determine the final allocations of costs and benefits. This is done by considering how the allocation of personal data and information has an impact over the individual and total welfare (Brandimarte and Acquisti, 2012).

In recent years, several papers have analyzed the consumer attitudes towards privacy and the trade-offs associated with the interaction between customers and firms willing to collect data about consumers in order to draw some conclusions and implications for policy-makers. Different models have analyzed the ability of companies to profile customers and its implications in terms of individual privacy.

Among others, Belleflamme, Lam and Vergote (2020) study a model of price competition between firms when they sell a homogeneous good and when they are able to profile consumers. However, in their model the profiling happens in an “imperfect" way: firms can identify consumers’ valuation for the product only imperfectly; therefore there is always a possibility that the consumer will remain anonymous. This means that companies end up with different profiling of customers, and therefore the nature of competition remains uncertain.

Being aware of discriminatory practices adopted by businesses, consumers can also behave strategically (so-called “Hawthorne effect”), seeking to become anonymous to protect themselves from discrimination practices. Among the models treating the issue, we recall the aforementioned model by Acquisti and Varian (2005) where consumers, involved in a strategic interaction with firms, can hide their "cookies" so as not to be recognized by companies. In addition, in the model by Conitzer, Taylor, and Wagman (2012) and in the model of Montes, Zantman and Valletti (2019), consumers support a cost to maintain their anonymity; while in the model of Choi, Jeon and Kim (2019) the collection of data requires the consent of consumers.

Casadesus-Masanell and Hervas-Drane (2015) proposed a model considering a mass of customers and two firms providing a homogeneous service but competing on two dimensions: price and privacy (or “disclosure”). Each firm enjoys two sources of revenue: one from the sales prices for the service they sell to consumers, the other from the sale of data on their consumers to third parties ("disclosure"). Consumers: i) are heterogeneous in their evaluation of the service; ii) provide their personal information only to the company from which they purchase the service (they may also not purchase any services, thus remaining out

of the market) to take advantage of any service personalization; and ii) do not want their data to be sold to third parties (i.e. they prefer the non-disclosure).

The authors solved their model by backward induction: in the first stage, firms simultaneously decide the consumer information disclosure and in the second stage firms simultaneously set their prices. In the third stage, having observed disclosures and prices, customers decide if to sign up for the service of one firm or to stay out of the market; finally, in the fourth stage consumers who have subscribed a service decide the amount of information to provide to the related company.

First of all, Casadesus-Masanell and Hervas-Drane analyzed the relationship between revenues and customer information (disclosure) which reveals two important trade-offs. The first one is associated with the revenue source: companies want to maximize their profits either by increasing the prices or the disclosure of consumers' information. However, on one side the increase in price ends-up determining a lower demand for their services: the lower consumer base reduces the data stock that enables firms to extract revenues from the disclosure. On the other hand, if firms increase the level of disclosure in order to sell information to third parties, this will decrease the consumers' willingness to pay for the service which ends-up reducing the revenues from the service. The second trade-off is associated with the level of disclosure set by companies: the higher the level of disclosure chosen by companies, the lower the amount of information disclosed by consumers; therefore each firm must find a balance between the stock of data obtained by consumers and the revenue from its disclosure to third parties.

Then, authors analyzed how the privacy level affects the competition at the firms' level. They concluded that privacy can actually soften competition when two conditions occur simultaneously: i) customers are heterogeneous so that firms can set differentiated privacy policies and ii) the consumers' willingness to pay is not too high so that the two firms can operate profitably.

The issue of privacy can be seen also analyzed by considering the effect of privacy and information disclosure and its societal consequences. Hillebrand and Hornuf (2021) analyzed the data donation process in the context of a "social dilemma", i.e. the situation in which individuals may be tempted to promote short-term benefits from non-cooperative behaviors, but found that in the long-term it would have been better to cooperate. The underlying idea is that individuals would be better off revealing personal information, thereby supporting a privacy risk (i.e. data leakage) but contributing to the social well-being, instead of not disclosing and freeriding on the contributions of others. This can happen because

individuals may be incentivized to do so if their underlying motivations lead them to prefer social well-being over personal utility. More specifically, their willingness to donate personal data (WDPD) increases when individuals perceive a strong moral obligation to donate and when they trust the institution to which people provide the data (i.e. they prefer academia or the government over private companies, so companies oriented not to profit, but towards social welfare).

1.5 Conclusions

With the enormous diffusion of cutting-edge technologies and the enormous connectivity achieved, people always leave traces of themselves in the form of personal data: therefore, individual privacy is in serious danger. For their part, companies have unprecedented possibilities to collect, store, use and manage huge amounts of data at low cost. Companies have great incentives to collect user data, because this allows them to put into practice personalization and profiling strategies, which then translate into target advertising and price discrimination.

Economists have dealt with the issue of privacy, revealing the many facets and fields of action in which the issue can be submitted. The economics of privacy, which is not a recent field, turns out to be a rather complex topic, which has undergone several evolutions in conjunction with the technological developments of the digital age.

In the face of enormous concerns related to privacy and personal information, the legislation has tried to intervene with regulations aimed at favoring greater control for users on their personal data and greater transparency in how data is collected and managed.

However, not only company collect data by themselves, but they also acquire customers' data from the data market where data, its main resource, is exchanged. In this market, there are specific agents who occupy strategic roles: among the agents who play a role of intermediation we find the so-called data-brokers, of which consumers very often ignore the existence.

In the next chapter we introduce the data market and then we will analyze in detail the data brokers.

Chapter 2: The market of data and data-brokers

In the first chapter we briefly discussed the economics literature on privacy. We have outlined the main characteristics and the multiple perspectives to which this broad topic can be subjected to. We have also identified the often negative consequences to which personal data and information of individual users can be subjected to. Not only is the privacy of individuals at risk, but what many people do not realize is that there is an ecosystem where data and personal information are traded and exchanged: the market of data.

The market of data has specific agents operating within it and a common vision in this market is that data is treated as a precious resource. In this chapter, we analyze the data market starting from a definition and its main characteristics and features, then we will focus on the operations of some specific types of companies that, within this market, play a crucial role: data-brokers.

2.1 The market of data

Why a data market? On the one hand, it could be argued that personal data should not be subject to buying and selling: there are certain things, like electoral votes, that should not be for sale because they could undermine the spirit of democracy. On the other hand, the presence of a data market with the proper rules and norms drafted by regulators can put the right limitations to a market of enormous value for companies. In order to understand its benefits and risks, it is important to first analyze the characteristics of the market of data.

The European Commission has defined the data market as "the market where digital data is exchanged as products or services derived from raw data"[39]. Despite a significant gap between Europe and the United States, the data market is expected to grow significantly in the coming years, mainly due to two main elements. First, the development and diffusion of data-based innovations also thanks to the contribution of artificial intelligence and other technological developments, which will have an ever-greater impact on economic growth. Second, by implementing international strategies promoting the benefits of the data market by giving responsibilities while protecting users. For example, at the European level, it can be

[39] European Commission (2020), The European Data Market Monitoring Tool

mentioned proposals such as the “European Digital Agenda”, the “Strategy for the Digital Single Market”, the “Europe 2020 Strategy” and the “European Union Action Plan for eGovernment”. Moreover, to foster a digital single market, the European Commission has created the "Free Flow of (non-personal) data" initiative with the aim of promoting the economy of data and related technologies, products and services that use data[40].

From the definition of data market, a first important consideration is that data is seen as the main resource, something that has some value and is treated as an object of exchange. For companies, having data does not always mean having access to the information contained within it: sometimes the content can be processed or extrapolated from the raw data, and in this case raw data has only potential value. The content, unlike the data itself, requires an interpretation based on knowledge, and most of the time this process involves the presence of specific companies in possession of the necessary skills. This distinction is important because at a functional level, the generation of economic value on data is crucial for evaluating and defining property rights on data from a legislative point of view.

Over time, data has changed its role and its contribution in an emerging data economy, especially in relation to the digital transformation.

Drexl (2017) analyzed the different contributions that data had in relation to technological developments. In the beginning, the Internet was used as an information and sales platform and an information society was emerging: essentially, the foundations for the creation and proliferation of data had been laid. In a subsequent phase, new types of services were offered to consumers, mainly financed through advertising: in this phase data were identified as an input of great potential for the emergence of business models based on data and information. Basically, the underlying idea was that the value of a service or platform increased the more they were attractive to consumers, which were willing to share personal data and information in exchange of a perceived value or benefit. With the advent of the Internet of Things (IoT), connectivity has become key, both in the physical and digital sphere but also among the two: this step led to an increase in data and an extension of the data collected. Data can now be used in different areas, or combined with each other to analyze and predict human behaviors, consumer habits, or general correlations (e.g. "data mining") and a lot of information can be inferred from data. Moreover, with the presence of a data

[40] https://ec.europa.eu/digital-single-market/en/free-flow-non-personal-data

market, data are seen as highly sustainable, non-rival, and traded in an active market at a low marginal cost[41].

Since data is characterized by a great variety, it is necessary to distinguish data not only in terms of types of data, but also according to the source[42].

A first important distinction is between raw data and processed data. This distinction implies that raw data is not always of value to companies. Sometimes, companies and organizations need to work on data that has already gone through a primordial process that has given it shape and meaning, or a usable content. As we will see later, there are companies that deal precisely with organizing and processing data, adding value to the data and allowing it to be processed, which is essential if we think of the huge volumes of data nowadays generated.

A second distinction is between the spheres of personal and non-personal data, which gave rise to a debate on privacy issues and the subject of property law. Personal data are data that, directly or indirectly, refer to an identified or identifiable natural person, while non-personal data are not related to an identified or identifiable natural person, or whose link to the natural person was present in the past but now the relationship does not hold anymore. This is important because the application of the rules within the GDPR (i.e. the European legislation on data protection) depends precisely on this distinction.

A third distinction is between static and dynamic data. Static data is data that cannot be changed during its processing. Examples of static data include a newspaper article, which can be read but the data source cannot be changed, or a CD ROM, which is not changed during processing. On the contrary, dynamic data is data that changes during its processing: when the data is re-entered they are never the same, but they are constantly changing. For example, a CD RW that can be rewritten or edited. Generally speaking, dynamic data are better suited to contexts in which there is a need to quickly update information, and where there is a need for receiving contributions from different agents, while statistical data has more limitations.

Another interesting difference is about the data utilization: data can be used both as an input and as a output. Generally, data are used as input for improving marketing and targeting campaigns, as well as to improve the personalization of products and services. On the contrary, data are sold as output in the forms of data packages or data products: this happens,

[41] European Commission (2017), The economics of ownership, access and trade in digital data, JRC104756

[42] Cambridge International AS & A Level Information Technology (2017). Types of data

for instance, when a data-broker or a marketing agency sell these data to other companies or third parties.

For what concerns the source, a big distinction is between direct and indirect data source. Direct data sources represent sources where data are collected for a particular or specific purpose. These are also called “original sources”, since there is no need to gather the information from other third parties; such as questionnaires or interviews. On the contrary, indirect data sources are those sources that collect data for a specific reason, but that can use them for another. This happens for example when a company or organization collects data on individuals of a commercial nature, and then decides to resell this data to third parties, to identifies specific categories of individuals with specific characteristics, or to use it to implement strategies such as price discrimination or target advertising.

Generally, if on the one hand direct sources personally take responsibility for the processing, storage and use of data on the basis of the consent of each individual, on the other hand indirect sources represent a greater source of risk for the consumer. With the possibility for companies to collect data quickly and easily, companies can collect information about consumers indirectly, without even knowing about these practices. This information is of great quality for businesses, but it seriously endangers individual privacy.

These characteristics have reflected the proliferation of digital devices, sensors and services as a result of the expansion of the digital economy. In fact, the overall impacts of the data market on the economy as a whole are measured in the value of the “Data Economy”, which is actually expected to grow faster than the data market (European Commission, 2020).

Drexl (2017) identifies the main characteristics of a data market, in relation to its primary source, data:

- "Volume": huge volume of data produced by several sources are of enormous volume, so large that it dominates the capacity of the storage and treatment systems;
- "Velocity" which recalls the dynamic nature of big data, which changes constantly, almost in real time.
- “Veridicity”, since there is a need for data to be reliable, also from a legal standpoint.
- "Variety" or the wide variety of types and formats of data, which can also come from very different sources, that can be combined to find conclusions, correlations and to give sense to the aggregated information.

The more data a company has at its disposal, the more power the company can exercise: for example, the possession of data by a company could allow it to decide who to sell to and at what price. Moreover, a lot of data is not exchanged and there are no alternative sources for obtaining it: this undoubtedly makes competition less fierce. In addition, there are other strategies companies can use to limit competition. First, strategic data-driven mergers to leverage economies of scale and cost efficiency. Second, continuous data collection in their own platforms to remain competitive and creating barriers to entry. Third, the creation of a strategic presence in different segments in order to exploit the interoperability of data and to be able to collect more information about the same user. However, the fact that some data is not exchanged does not always prevent new players from overcoming the entry barriers: for instance, the large amount of data held by a big company like Facebook has not prevented the development of other platforms such as Snapchat. In conclusion, the competitive advantage derives from the use made of data and from the organizational skills that this use entails for companies.

Before the advent of digital and information technologies (ITCs), the cost of production, the marginal cost of storage, use, distribution and transposition were particularly high. With the introduction of digital technologies, on the other hand, information and data can be stored, replicated and transmitted electronically, quickly and economically and sustained with a significantly lower energy cost than though analog information systems. Moreover, unlike the past, modern digital information reduces information to its most basic expression, made up of a minimum number of distinguishable states necessary to detect information, that is, a two-state binary format, 0 and 1. The binary digital system constitutes a universally shared information format, and this greatly facilitates data transposition and connectivity between different digital devices.

The significant proliferation of digital devices, sensors and services as a result of the expansion of the digital economy has resulted in a considerable volume of data. Moreover, in the data market, information technologies have significantly lowered the costs of collecting, using and distributing data, as well as reducing search costs. The use of algorithms has also helped companies and data brokers to overcome the problem of noise deriving from an increasing need by companies to collect, store and manage huge amounts of data. Therefore, companies and organizations can leverage economies of scale and scope.

The processing and content extraction of data and their information can be performed in different ways and with different degrees of complexity. One tool often used is machine learning. With the introduction of constantly-updated data and information, these models are

constantly improving. Economies of scope are closely related to these types of models: when two sets of data overlap each other, even partially, the cost to extract knowledge from the two together is lower than the cost to do it separately for each data set. Therefore, economies of scope resulting from joint learning produces more benefits and fosters the creation of more insights. Additionally, algorithms and information learned from one dataset can in some cases be transposed to other datasets or extended to adjacent areas of analysis. Conversely, avoiding economies of scope by separately applying algorithms or machine learning systems could lead to greater cost and less significant results in terms of combinations of data and insights and insights into information.

In general, economies of scope allow companies and data intermediaries to obtain lower costs for collecting and analyzing data. Furthermore, they allow us to better understand why companies are led to collect an increasing amount of data on individuals, overcoming the noise that could be generated by an excessive amount of information. On the other hand, however, economies of scope are subject to returns that decrease over time and therefore do not last forever.

Another important aspect of the data market is interoperability, i.e. the process by which information and data is made useful between systems, applications or components. Information interoperability is quite interesting in the context of technological information systems as it promotes trade, innovation, reactivity to market challenges, and it lowers costs (especially communication costs), adds flexibility to the decision-making process and it also helps to lower the barriers between market players. Generally speaking, in an increasingly digitized world, the benefits of interoperability significantly outweigh its costs and challenges (Palfrey and Gasser, 2012).

The interoperability has provided significant characteristics to the data market: barriers to entry are significantly reduced, an easy access to data, and the possibility for different actors to extract value from the content contained therein. Moreover, compared to traditional economy, in the data economy value is generated differently and in a way that reflects digital transformation (Drexl, 2017): while the traditional economy generally refers to vertical value chains in which every step of the chain, from inputs to output, is adding economic value sequentially, in the data economy the ideal framework involves a complex and dynamic paradigm in which several contributors can simultaneously add value. This is why in the data market, through collaborations and interconnections, new products, services and firms can easily arise.

The development of cutting-edge digital technologies has also made it increasingly easier to observe user behavior. Users are perennially connected: they do research, communicate and entertain. Basically, they can use any application or platform they want and need to achieve very different purposes. All of this has side effects: this connectivity and these online activities create a huge amount of data, which can be also traded among firms. In the same way, the data market has created some concerns linked to the processing, storage and treatment of personal data and information of individuals.

Privacy executives have sought, through a pragmatic approach, to manage the sharing of personal data through guidelines on collection, processing and sharing, in both national and public interest. Despite this, OECD countries have not always embraced these directives and most member states have declared that they have adopted them only in more recent years[43]. Digital security incidents are potentially very risky, not just for businesses and governments, but for individuals as well.

Since there can be lots of problems in terms of data protection, this has required the intervention of legislation. A first distinction between property on the data set in the form of bits and bytes, and property on the information that a dataset contains. When data is stored in a company server, from a legislative point there is a first problem of intervention. A second distinction is between the "syntactic" and the "semantic" level: while the former concerns the representation of information in different ways and is generally accessible to all (e.g. video or a digital book), the semantic level refers to the meaning that can be extracted from a representation of information, something not necessarily accessible to everyone (e.g. the meaning of a book results accessible only to whom is able to understand the language). A third distinction is between the protection of each individual data information and between the protection of the entire dataset in its entirety and composition.

The identification of a data market implies that there are specific agents operating within it[44]. Among these agents, there can be identified data professionals, whose core business is related to the analysis, management, organization, visualization and storage of data and information. Data professionals are operators capable of handling huge amounts of data, and are characterized by being at the forefront of database technologies. For this reason, data professionals are characterized by particular resources, capabilities and skills to carry out their work and the demand and supply for these skills has been growing strongly in recent times both at the European and international level.

[43] OECD digital economic outlook 2020
[44] The European Data Market Monitoring Tool (2020)

In addition to data professionals, in the data market there can also be identified data companies, so organizations "directly involved in the production, delivery and/or usage of data in the form of digital products, services and technologies". The two main types of data companies are data suppliers, which continue to generate an increase in their revenue year-on-year, and data users. The main activity of data suppliers concerns the production and the supply of products, services and technologies related to digital data, while data users are those companies and organizations that improve their businesses leveraging on the collection, organization and proper usage of data and the information contained within.

This classification of the different roles occupied by different agents, however, is not unique: in fact, there can be agents who are both data professionals and data providers, such as data brokers. Data-brokers therefore may have the necessary skills to process and analyze data, but also the ability to organize and resell the collected and arranged data. This in fact denotes a certain complexity in the data market, where different market agents can cover multiple roles, occupying in this sense a more strategic role.

2.2 Data-brokers

In general, companies collect information about customers on different aspects: preferences and hobbies, habits, health, purchase history and transactions, friends and other contacts, credit score, location. Companies not only collect data about customers, but they can buy these data from other companies called "data-brokers", which are "companies that collect consumers' personal information and resell or share that information with others" (Federal Trade Commission)[45].

According to a 2020 research by NATO Strategic Communications Centre of Excellence, there are around 5,000 data brokers worldwide, registering an industry of around $178 billion of revenues. These include companies such as Acxiom, Experian, Equifax, CoreLogic, Lifelock and TowerData. Moreover, there are around 10 million open datasets and around 4,8 billion internet users globally. However, the global data economy is projected to reach $400 billion with 175 zettabytes of data produced worldwide by 2025[46]. One of the main data-brokers is the American company Acxiom, which has recently been renamed LiveRamp. The company has over 20,000 servers for collecting and analyzing data on over 700 million people around the world. In 2018, in its website, the company reported that

[45] FTC (2014), Data Brokers: A Call for Transparency and Accountability
[46] Nato Strategic Communications Centre of Excellence (2020): Data Brokers and Security

“Acxiom has the most expansive and compliant data offering in the world, which now encompasses more than 62 countries, 2.5 billion addressable consumers and more than 10,000 attributes—for a comprehensive representation of 68 percent of the world’s online population”[47].

Data-brokers tend to specialize in certain sectors or market niches with the aim of obtaining a competitive advantage. Data can be collected from a variety of sources, mainly commercial, government, publicly available information and online tracking. Data collected by data-brokers are sorted and used for a variety of purposes, e.g. marketing and advertising or commercial aims. In addition, other additional information can be deduced from the data that the data brokers have at their disposal. Data-brokers usually exchange data even among themselves or they can buy data from other companies collecting data in order to have very accurate datasets. However, with the introduction of the European GDPR, the sharing of data between data-brokers without the authorization of customers is considered illegal (Gu, Madio and Reggiani, 2021).

These characteristics reveals an industry characterized by strong complexities and different layers of activities.This complexity is significant if we also consider that there are multiple layers of data-brokers in the process starting with the collection of raw data and finishing with the sale of organized data to its end customers. Bergemann and Bonatti (2012) distinguish data providers into financial data providers (e.g. Bloomberg and Thomson Reuters), credit rating agencies (e.g. Equifax, Moody's, Standard & Poor’s), data brokers (e.g. LexisNexis and Acxiom) and online aggregators (e.g. Spokeo and Intelius).

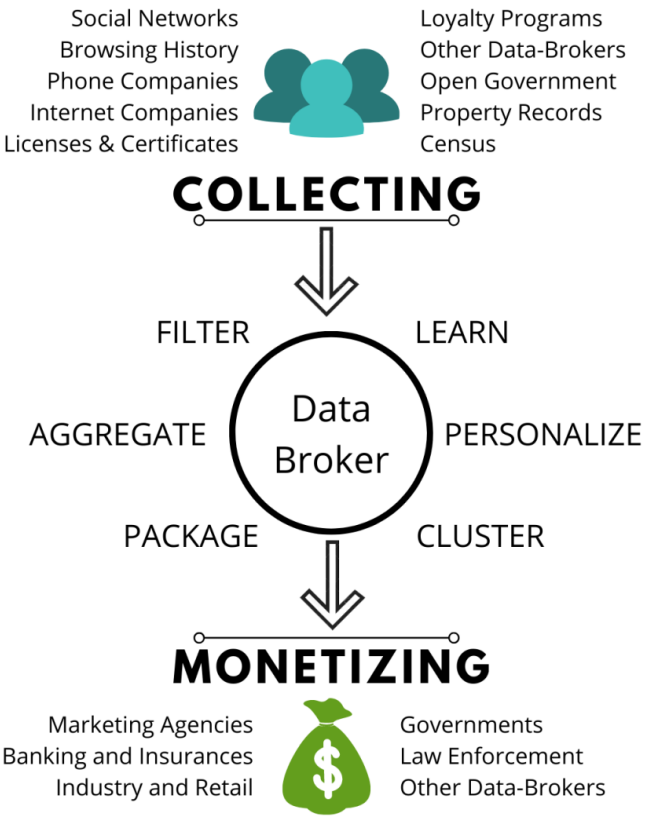
Data-brokers aggregate data and create marketable products, such as access to their database, lists based on observed data, data about subgroups of consumers with similar characteristics and behaviors and whose future behavior is predicted based on past actions. For instance, data-brokers can assign a score that can be used to predict the likely behavior of current or potential consumers. These companies typically sell their products to downstream companies, which acquire them to improve the positioning of their products and their strategies in general.

The data collected and analyzed by data-brokers can be used to make further inferences about them, even (and in a more risky way for the user) at the level of sensitive data. Subsequently, on the basis of these inferences, sub-categories of users can be created. Some categories may seem more harmless, such as "pet owner" or "pizza lovers", while

[47]https://www.acxiom.com/news/acxiom-launches-global-data-navigator-tool-offering-marketers-visibility-into-global-audiences/

others may be based precisely on sensitive or partially sensitive data (usually based on ethnicity, income or conditions of health, according to the 2014 Federal Trade Commission report). An example of sensitive categories can be "expectant parent", "diabetes interest" and "cholesterol focus". When data and information are combined, the added value that is obtained is significant. Each small piece of data that is combined with other information related to the same individual allows companies to create rather precise profiles of individuals. These profiles can also be organized into a large multitude of categories, each of which has different characteristics and behaviors, some of them more wide and general, while others more narrow and even sensitive. Ultimately, these aggregated data are then used to predict behaviors of individuals and therefore are of particular interest to many firms that can acquire them to implement or improve their strategies, e.g. target advertising, personalized services but also to adopt price discrimination strategies. Data-brokers can also combine data online with offline data, and then offer organized data packages to market online, then sell them to companies. This is often done through tools such as websites that use registration functions (e.g. cookies) to "find online consumers and target them with Internet advertisements based on their offline activities" (FTC, 2014). Then, as long as the cookies remain in the browsers of customers, data-brokers can continuously offer targeted offers every time the consumer surfs on the Internet. Most of the time, this happens in consumer non-awareness.

Graphically, the activity of data-brokers can be represented in this way:



Source: Adapted from Birckan, Dutra, Macedo and Godoy Viera (2020).

This graphical representation shows how data-brokers typically work. The process starts with the data collection from varieties of sources: customers release a lot of detailed information about themselves. Then there is a process aimed at organizing and sorting the information collected: the consumer information is put together and combined and the end result is a very detailed profiling of the individual or sub-groups. Finally, the information collected and organized is typically sold as a data product to downstream customers of the data-brokers, typically firms willing to obtain and access these data.

The ability of data brokers to collect information has also been favored by the advancement of new and increasingly sophisticated technologies, such as the use of algorithms, machine learning and artificial intelligence, but also by increasingly lower storage costs. Technology has also allowed data-brokers to easily store information and this can happen, for instance, in view of future business strategies, but data stored is not always secure, and in fact this data can be the target of malicious people aiming at stealing this information for less benevolent purposes.

Some customers may claim to benefit from having personalized enhanced services or lower transaction costs, but the risk to which customers and their data are exposed is not

irrelevant. Depending on the work of data-brokers, consumers can have benefits in terms of offering products and advertising that consumers can find relevant because they need, prefer, or are looking for them. Some data broker products can even help to prevent fraud, for example there are data-brokers such as ID Analytics whose product is related to risk mitigation.

While on one side data-brokers may offer some benefits, on the other hand they can create significant risks for the users concerned. For example, associating a consumer with a specific category or subcategory could make them more vulnerable to higher payment rates when taking out insurance policies: a person suffering from diabetes could be targeted advertising specific products at higher prices. In the same way, the possibility to store consumer data at lower costs and for indefinite periods of time, as well as the possibility of updating them continuously, could expose consumers to security and privacy risks and potentially leading to identity theft, theft of information related to their habits, personal passwords, information related to credit cards, security codes and financial frauds.

Consumers can be at risk as these companies sell or share their data. The problem is that this often happens without the consumer being aware of these practices and therefore potentially unable to request what information about them is held and with which companies it is exchanged. There are few reasons for which data-brokers represents a threat for customers (Twetman et al., 2014):

1. Violation of Privacy. Most of the times, the activity of data-brokers happens without a direct and transparent interaction between customers and data-brokers: these companies treat data of individuals without their consent, or without their knowledge or awareness.
2. Data exposure. One of the activities of data brokers concerns the storage of information, which makes them potential victims of hackers or cybercriminals. Lack of security practices exposes consumers to many risks, especially if sensitive information and data are exposed. In the past, data-brokers such as Acxiom, Epsilon and Experian have been hacked.
3. Advertising and targeting. Among the various categories, there are data brokers who sell data for the purpose of targeted advertising. These practices often occur in violation of privacy and consumers can be heavily influenced by them.
4. Exploitation. In this case we refer to the fact that personal data can be used in harmful and unethical ways, for example data breaches, identity theft, phishing attacks, credit

card skimming. For example, in 2019, Marriott reported a record 383 million guests exposed to data breaches[48].

Consumers may be unaware of the presence of these companies, or they may not know exactly how their data is collected and processed. Moreover, the problem is that, even if they wanted to, consumers cannot easily consult and verify the data that these companies have at their disposal.

One example of issues in relation to activities done by data-brokers is represented by what has happened with the Cambridge Analytica scandal.

In 2014 Aleksandr Kogan, at the time a researcher at the University of Cambridge, developed an application called "This is your digital life" with the aim of creating "psychological profiles" based on the activities carried out by online users, e.g. manifested preferences, comments, involvement in some specific groups and communities. With the consent of Facebook, this application could have access not only to the profiles of single users, but also to their contact list: this made it possible to collect data for an amount of approximately 50 million total users[49]. Data was then transferred to Cambridge Analytica, a British data mining and political strategy firm.Cambridge Analytica entered in possession and then analyzed these data also with the help of very sophisticated algorithms to finally define an accurate psychological profile for each user with specific information related to their interests, behaviors, and even emotions. The aim was to create and show, for each of these profiles, highly targeted and personalized political contents.

In 2016, Cambridge Analytica was accused of having obtained and misused the data collected to work on the presidential campaign on Donald Trump, adopting practices and strategies aimed at leveraging the targeting of voters, also thanks to an improper use of Facebook. This happened in the absence of transparency for the individuals involved and their data. In fact, the strategy of Cambridge Analytica was aimed, by using Facebook data, to identify subgroups of voters to create and design messages and content aimed at persuading and influencing their opinion: the company has essentially tried to identify target profiles which, through well calibrated strategies, were able to influence their choices effectively. In 2012, something similar had already happened during the presidential campaign of Obama, but at

[48] Bloomberg, Marriott Says Only 383 Million Guests Exposed in Breach, January 4, 2019

[49] The Guardian, "Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach", March 17, 2018

that time the voters had been informed about those practices and it was happening in accordance with the rules of Facebook.

When the affair emerged, Facebook was accused of not paying enough attention to user data and protecting their privacy. This resulted in a campaign called "#DeleteFacebook", which resulted in the deletion of numerous profiles from the platform. Mark Zuckerberg, CEO of Facebook, was finally subjected to an investigation by government institutions[50].

The multitude of issues that arose in relation to the practices conducted by data-brokers have required the intervention of regulators. In fact, the activity of data-brokers has been put under investigation, together with the need to protect privacy of customers and individuals in relation to their activities.

In 2014, the Federal Trade Commission reported some interesting results about an in-depth analysis of nine data-brokers collecting information about customers from different sources (e.g. government, commercial and public available sources), and offering products and services for a variety of purposes (e.g. marketing, risk mitigation and people search). The conclusion of this analysis was that, despite data-brokers have at their disposal increasingly sophisticated tools for collecting general and sensitive data about customers, together with the possibility to infer new information about them and the sub-groups created, the data-brokers industry and the commercial data market continue to present various problems. In particular, a general lack of transparency towards individuals and a lack of clarity on the methods used to manage the data collected on them. Another remarkable result of the report was that, in considering the nine data brokers examined, "one of the nine data brokers has 3000 data segments for nearly every U.S. consumer" (Federal Trade Commission, p. V).

Differently from ordinary goods, when dealing with the economics of data trading a buyer must evaluate the information before purchasing the information and ultimately decide his willingness to pay. This means that the seller needs to disclose some of the information it possesses. But when this disclosure occurs, the buyer no longer has any incentive to pay for what is ultimately shown to them. This problem is referred to as the "Arrow Information Paradox". To solve this paradox, intellectual property and property rights can be used: when information is protected through the use of patents or copyrights, it can be disclosed without the risk that the seller will not receive any compensation. However, this paradox may not be always valid: there are heterogeneous data that are composed of several parts, and the partial

[50] BBC, Cambridge Analytica: Facebook 'being investigated by FTC', March 20, 2018

disclosure does not necessarily reveal the whole information. Moreover, the use of contractual relationships or relationships based on trust allow the exchange of information[51].

The cost structure of data sellers is typical of the markets for information products. In fact, it is generally characterized by high fixed costs (e.g. costs to implement the necessary infrastructures and data processing technologies) and by low marginal costs.These sellers must then decide on pricing strategies as well. Setting a price for information is no easy task: typically, an "information good" is priced based on the value it has for its consumers, and not as an increase in unit cost. However, “since people have widely different values for a particular piece of information, value-based pricing leads naturally to differential pricing” (Shapiro and Varian, 1998). Pricing can vary according to some factors, such as the competition among sellers, the interaction between data providers and customers, the setting considered (e.g. monopoly, Hotelling setting, or two-sided platform) and also according to the demand by buyers. Among the interaction between buyers and sellers of data, we have to consider the possibility of the data-broker to discriminate among buyers, therefore the possibility of offering an exclusivity to only one or more firms.

In recent years, thanks to increasingly sophisticated tracking tools and data analysis capabilities, data-brokers have found it easier to collect and organize consumer information. The lack of transparency in the practices conducted by data brokers and the concerns associated with it has given rise to specific regulations for data brokers (Federal Trade Commission, 2014). In the 2012 "Privacy Report"[52], the US Commission discussed the privacy concerns raised by the practices of data-brokers and identified different uses for the information they collect (i.e. entities subject to FRCA, companies related to marketing practices, companies related to non-marketing practices outside the FRCA) and identified two main recommendations to increase transparency of data brokerage firms:

1. Providing consumers access to the information that data-brokers collect about them;
2. Providing data-brokers with guidelines to improve their transparency, for example by describing how they collect information about consumers, informing consumers about the types of companies to which the data is transferred, etc.

[51] European Commission, The economics of ownership, access and trade in digital data, 2017

[52] Protecting Consumer Privacy in an Era of Rapid Change: Recommendations For Businesses and Policymakers (2012); available at: https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf

In 2013, the United States Government Accountability Office released a report on the practices conducted by data brokers, concluding that the US Commission should intervene in a more marked way thanks to legislative intervention. In 2014, two bills entitled "Data Broker Accountability and Transparency Act" and "The Data Accountability and Trust Act of 2014" were introduced with the aim of improving the transparency of data intermediaries and their practices, leading them to make the information collected available on each consumer.

However, the activities of data brokers are not limited only to the US borders: today these companies operate globally. Therefore, the activities of data brokers have been subjected to the careful analysis of international legislation.

More recently it has been issued the California Consumer Privacy Act (2018) that, together with other rules, norms and legislation, has also treated issues related to data-subjects. Among other things, from January 1, 2020 the CCPA allows a consumer the "right to request a business to disclose the categories and specific pieces of personal information that it collects about the consumer, the categories of sources from which that information is collected, the business purposes for collecting or selling the information, and the categories of third parties with which the information is shared"[53]. In addition, California law requires data-brokers to register annually with the Attorney General and to "provide information on how consumers can opt out of the sale of their personal data".

2.2.1 Literature on data-brokers

The literature on data intermediaries has tried to analyze the main characteristics of the data market and the strategic role that data brokers play as data providers. Typically, the setting proposed by these models presents an upstream data market with the presence of one or more data-brokers, and a downstream firm market with one or more firms operating within it. Companies buy data from the upstream market to get more precise information on consumers' willingness to pay, then implementing discriminatory strategies, mainly price discrimination and target advertising. These models also offer important conclusions for the drafting of privacy policies, as they allow to shift attention to the still unclear and transparent practices that are at the basis of the data market and the work of data intermediaries.

A first strand of literature analyzes a data market characterized by a single upstream monopolistic data-broker, who enjoys decision-making power when offering their data to

[53] California Legislative Information (2019), AB-1355 Personal information

downstream companies. For instance, when selling its data, the data-broker can decide to offer its data exclusively to only one downstream firm.

Clavorà Braulin and Valletti (2016) build a model by considering an upstream data broker selling its information about consumers' preference to two firms competing downstream. In particular, the data intermediary can decide whether to sell their data to both downstream firms or to offer its data through an exclusivity agreement to only one firm. The authors conclude that the vendor always has an incentive to sell their data exclusively, however this creates allocative inefficiencies. The first best of their model is in fact obtained when the data-broker sells to both companies, however this result never emerges in equilibrium. The conclusions are then converted into policy suggestions aiming at regulating the exclusive sale of data by data intermediaries.

Montes, Zantman and Valletti (2019) illustrates a model considering three agents: consumers, firms and a data supplier. In their setting, companies are willing to obtain data from the data-broker in order to implement price discrimination practices, while consumers can avoid to be the target of these strategies by supporting a "privacy cost". However, the authors conclude that, in equilibrium, the data-broker ends up selling its customers' data to only one of the firms by offering the exclusivity. Their result is then translated into suggestions for the drafting of privacy policies: similarly to Clavorà Braulin and Valletti (2016), policymakers should discourage exclusivity agreements and ensure consumers greater privacy protection.

In addition to selling its own data exclusively, the monopolistic data-broker can also choose the amount of data to offer to downstream companies, acting on the quantity.

Bounie, Dubus, and Waelbroecker (2020) create a model in which a monopolistic data-broker can strategically decide the amount of information to sell to competing firms, eager to obtain it for price discrimination practices. The setting is represented by an Hotelling line: the amount of information sold partitions this line into segments of consumers' information and firms buying these segments of information can set specific prices to target customers. The aim of this model is to understand how the competition at the firms' level is affected by the quantity of information provided by the data-broker to the market. The data-broker can weaken (no or little customers' information sold) or strengthen (all customers' information sold) the intensity of competition at the level of companies. The authors conclude that the data-broker strategically sells partial (incomplete) information about consumers, thereby weakening competition at the firm level.

Another strand of literature does not consider a monopolistic data-broker, but the presence of more than one data-broker in the upstream data market. This implies that there may be two alternatives: i) the data brokers compete with each other, ii) the data-brokers decide to cooperate with each other.

Ichihashi (2020) proposes a competition model between data intermediaries in the data market. Data brokers collect personal consumer data under compensation and then resell it to downstream companies. The assumption is that the data provided by consumers is non-rival: consumers can offer the same data to different intermediaries, thereby obtaining compensation from each intermediary. The compensation given to consumers by data intermediaries is important: if this is sufficiently high, intermediaries will offer the data to downstream firms at a lower price. For this reason, data-brokers will offer a low reward for the data offered by consumers. In his model, the author argues that upstream competition benefits consumers. However, if the data purchased by downstream firms is used in a harmful way towards customers (i.e. firms use data to extract the maximum possible consumer surplus), consumers end up supporting the negative impacts that would occur with the presence of a single monopolist data-broker.

Gu, Madio and Reggiani (2021) propose a model concerning the role of data-brokers supplying information to downstream firms (i.e. data-buyers). The authors distinguishbetween "sub-additive" and "super-additive" data, according to the lower (sub-) or higher (super-) value of the merged data in respect to the sum of the separated dataset. Depending on the nature of data (i.e. sub- or super-additive) and on the cost for merging data, data-brokers can decide whether to compete or to share their data between each other. The authors conclude that in some circumstances data-brokers can be incentivized to share their data between each other: more specifically, data sharing happens when data-brokers are more efficient than data-buyers in merging datasets. However, data-sharing practices among data-brokers have been subject to regulation: for instance, the European GDPR has enacted more stringent rules deciding, among other things, that data-brokers cannot exchange data without consumers' permission. Therefore, even if there can be positive externalities and pro-competitive effects deriving from data-sharing, their analysis reveals that regulators should focus more on regulating the co-opetition practices and effects of these practices.

2.3 Conclusions

In this section we have presented the data market, highlighting its main characteristics. The main resource of this market is data which is treated, transformed, stored, managed and traded among specific market actors. Within this context, a crucial role is played by data-brokers. The data market still has some dark sides, and many practices within it remain unclear. Furthermore, there is still little transparency in the way consumer data is collected, processed and stored by data-brokers. In 2014, the Federal Trade Commission intervened with an in-depth study on data-brokers entitled "Data-brokers: a call for transparency and accountability", which highlighted the main issues concerning data intermediaries and their practices.

The literature on data brokers has also analyzed the structure of the data market by presenting models with data-buyers and data-sellers, finding interesting ideas for drafting privacy policies.

In the next chapter we analyze a model involving a monopolist data-broker that can decide the price for its data and two downstream firms willing to obtain its data. The price set by data-brokers significantly affects the outcome of the equilibrium and on the level of privacy protection of customers. Finally, we will also propose some conclusions and suggestions for the policy-makers. According to the literature on data-brokers shown in this chapter, our model will find a lot of similarities with the strain that considers a monopolistic data-broker that has a certain power when selling its data to downstream firms, especially because it is able to set the price for the data it possesses. Among the literature previously presented, our model will be close to the models of Clavorà Braulin and Valletti (2016) and Bounie, Dubus, and Waelbroecker (2020).

Chapter 3: The role of data-brokers in a vertically related market

Thanks to technological progress, companies are able to collect, store, share or sell specific customer information, which usually can be used to implement targeted advertising and differentiated pricing. This has given rise to growing concerns regarding the protection of personal data, which require an ever more significant intervention in terms of regulations and legislation, not only for what concerns data protection, but also in relation to competition between companies, also considering that the role of data has become crucial in establishing competitive power. Companies are incentivized to obtain data on users and consumers, since they can implement highly personalized strategies. Among their possibilities, companies can also buy data from data-brokers, data intermediaries operating within the market of data.

We therefore condensate these elements into a model characterized by a vertical related industry with an upstream data-broker who collects and sells data to two downstream firms competing and using data to price discriminate consumers. Our aim is to determine not only the incentives for the two downstream firms to use the data provided by the data-broker, but also the incentive for the data-broker to set a price for its data that enables only one or both firms to buy it. We then conclude discussing the impact of such strategies on market efficiency, and some implications for policy-makers involved in the drafting of privacy policies.

The starting point of our analysis will be a recent model by Shy and Stenbacka (2016), where the authors analyze and compare the impact of different privacy regimes on firms' profits and social welfare. In the Shy and Stenbacka model, privacy means the impossibility for firms to use data about users' preferences to engage in price discrimination strategies; on the contrary, the market is characterized by no privacy protection if firms can freely use customers data and price discriminate. Overall, the two authors find that without privacy protection firms are better off than with strong privacy protection but also that some degree of protection is desirable for the firms. Moreover, they show that the consumer surplus and the total welfare increase with the level of privacy protection, and they refer to this property as "monotonicity".

3.1 The model

Following the setup of the Shy and Stenbacka model, we will assume that:

- There are three companies: an upstream data-broker and two downstream companies (A and B). The two downstream firms compete producing differentiated products or services labeled as A and B.
- There are *2n* consumers: *n* consumers are A-oriented (i.e. they have a preference for A) and *n* consumers are B-oriented (i.e. prefer B over A, prices equal). Formally, A-oriented customers evaluate v_h product A and v_l product B, with $v_h > v_l$, for B-oriented customers the opposite applies.
- Consumers in the previous period (t_0) bought from one of the two firms, and in the current period (t_1) they have to decide whether to confirm their purchase decision or not.
- In t_0, a portion $(1-\mu)$ of customers has purchased their preferred company; these customers are named as "matched" customers. Clearly, a portion μ is "mismatched" that is customers who have purchased the least preferred product. All throughout the paper, we assume that $0 < \mu < 1/2$.
- Consumers who decide to change product have to bear a switching cost, s. In addition, the parameter σ, where $\sigma > 0$, measures the heterogeneity of switching costs: high values of σ generate a greater differentiation of switching costs among all buyers with $s \in [0, 1]$.

The timing is as follows: in the first stage, the data-broker offers its data to the two downstream firms (it sets the price of the data, that we indicate with *t*), and in the second stage firms decide whether to buy the data or not and then compete; access to the data allows a firm to price discriminate customers. The model is solved by backward induction.
In addition to collecting more and more data, companies also have an interest in buying data from data-brokers. The purchase of data packages allows them to have a detailed profiling of consumers: the more data they have, the more accurate the information on consumers (current and potential) will be, the more they will be able to easily implement targeted strategies (especially targeted advertising and price discrimination) and the more they will be able to play a dominant role in the reference industry.

3.1.1 The last stage: the downstream firms

In the last stage of the game, there are three possible scenarios:

1. Both firms have purchased data from the data-broker;
2. None of the companies has purchased data from the data-broker;
3. Only one of the two firms (be it A or B) has purchased the data from the data-broker.

These three scenarios differ from the amount of information each firm has access to. As explained above, when one firm buys the data is able to discriminate not only between matched and mismatched customers, but also to discriminate against customers in relation to their previous purchase decision.

In what follows, we analyze and identify the equilibrium (i.e. optimal prices) for each of these possible scenarios.

1. Both companies do acquire data (i.e. "no privacy" regime)

When both firms do acquire and use data (we refer to this as the "no privacy" scenario), firms have both the same and highest amount possible of information about customers; specifically, the data allow firms to ascertain which customers have purchased their preferred products in stage t_0 and from which firm. In this case they can charge up to four different prices:

1. A price for the matched customers of the firms that have bought their preferred product, p_{ih} (with $i = A, B$);
2. A price for the matched customers of the company that have bought from the rival company, p_{il};
3. A price for the mismatched customers that have bought from the company they prefer, q_{ih};
4. A price for the mismatched customers that have bought from the company that prefer the less, q_{il}.

In this way, companies can price discriminate customers according to their type (i.e. i-oriented or j-oriented) and according to their past purchases (i.e. previous purchase).

Formally, the utility function of the generic customer$c$ who has a relation with firm $i = A, B$is:

$$U_i(c)\begin{cases} v_h - p_{ih} & i-oriented\ and\ continues\ to\ buy\ i \\ v_l - p_{il} & j-oriented\ and\ continues\ to\ buy\ i \\ v_l - q_{jl} - s\sigma & i-oriented\ and\ switches\ to\ j \\ v_h - q_{jh} - s\sigma & j-oriented\ and\ switches\ to\ j \end{cases} \quad (1)$$

where s (with $0 < s < 1$) indicates the cost of switching brand and σ ($\sigma > 0$) measures the heterogeneity of switching costs, and where v_h is the higher valuation and v_l is the lower valuation with $v_h > v_l > 0$.

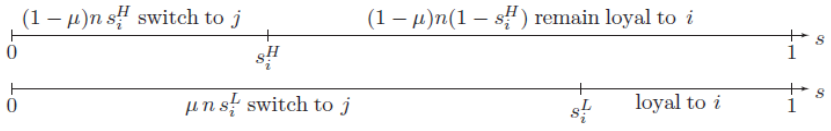
Note that for a customer having a preference for a company *i* (i.e. A or B), switching to competition *j* (i.e. the other company) can be particularly onerous, because the customer has a preference for the company *i* and moreover he has to sustain switching costs s.

Switching costs are particularly important in the model, either for matched and mismatched customers. If we consider the correctly matched customers of company *i*, some of them switch to competition *j*: in fact there might be customers that despite having a preference for company *i* may decide to switch to company *j*. This can happen, for example, because company *j* offers an advantageous price: in this case, those that have a really low switching cost can decide to switch, while those that have a high switching cost do not switch and remain with their preferred company.

How many customers switch brand and how many do not switch? In order to answer these questions we need to find, both for the matched and for the unmatched users, the indifferent customer, indifferent between switching and not switching.

For "matched" customers, i.e. i-oriented customers (where *i* is A or B) that have a preference for company i, this means finding the level of s that solves the condition $v_h - p_{ih} = \mathrm{v}_l - q_{jl} - s\sigma$: all of the customers that have a switching cost higher than this value, that we indicate with s_{ih}, remain loyal to company *i*, while the others switch to competition, *j*. For "mismatched" customers, i.e. the customers that has a preference for company *i* but that have previously bought from company *j*, this means finding the *s* by solving $v_l - p_{il} = \mathrm{v}_h - q_{jh} - s\sigma$, that we indicate with s_{il}.

Graphically:



As in the model of Shy and Stenbacka, we define the difference in individual evaluations, $\Delta = v_h - v_l$where the higher valuation is v_h and the lower valuation is v_l, with Δindicating the utility loss (gain) associated with a customer mismatch (match).

From the indifferent conditions, the threshold levels of the switching costs are defined as:

$$s_{ah} = \frac{p_{ah} - q_{bl} - \Delta}{\sigma} \quad and \quad s_{al} = \frac{p_{al} - q_{bh} - \Delta}{\sigma} \qquad (2)$$

$$s_{bh} = \frac{p_{bh} - q_{al} - \Delta}{\sigma} \quad and \quad s_{bl} = \frac{p_{bl} - q_{ah} - \Delta}{\sigma} \qquad (3)$$

Firms can set four different prices to maximize their profits, that for A and B respectively,can described as:

$$\pi_A = p_{ah}(1 - s_{ah})(1 - \mu)n + p_{al}(1 - s_{al})\mu n + q_{ah}s_{bl}\mu n + q_{al}s_{bh}(1 - \mu)n - t \qquad (4)$$

$$\pi_B = p_{bh}(1 - s_{bh})(1 - \mu)n + p_{bl}(1 - s_{bl})\mu n + q_{bh}s_{al}\mu n + q_{bl}s_{ah}(1 - \mu)n - t \qquad (5)$$

In these two functions, the first two terms are related to the portion of "loyal" customers, i.e. customers that keep buying from the same firm: $(1 - s_{ih})(1 - \mu)$ represents the proportion of matched customers that remains loyal, while the proportion $(1 - s_{ij})\mu$ represents the proportion of mismatched customers that remain loyal. The last two elements represent matched and mismatched "non-loyal" customers. Finally, as we are in the scenario characterized by both firms buying the data, the profit functions include the cost of purchasing the data from the upstream data-broker t: if a firm does acquire data it pays the cost, vice versa it does not. For now, the value of *t* is considered as an exogenous cost; while subsequently, the cost *t* will become an endogenous variable.

Substituting (2) and (3) into the profit functions in (4) and (5) and then maximizing profits, it is possible to obtain the profit maximizing prices charged by the two firms in this subgame:

$$p_{ah} = p_{bh} = \frac{2\sigma + \Delta}{3};\ p_{al} = p_{bl} = \frac{2\sigma - \Delta}{3};\ q_{ah} = q_{bh} = \frac{\sigma + \Delta}{3}$$

$$and\ q_{al} = p_{bl} = \frac{\sigma - \Delta}{3} \quad (6)$$

The switching costs threshold are obtained by substituting (6) into (2) and (3):

$$s_{ah} = s_{bh} = s_1(a,a) = \frac{1}{3} - \frac{\Delta}{3\sigma} \quad and \quad s_{al} = s_{bl} = s_2(a,a) = \frac{1}{3} + \frac{\Delta}{3\sigma} \quad (7)$$

Where $s_1$is the equilibrium threshold for the correctly matched customers $(1 - \mu)$, and s_2 is the equilibrium threshold for the mismatched customers μ, where (a, a) indicates that we are in the subgame where both firms use the data to price discriminate (i.e. "no privacy" regime). Finally, substituting the prices found in (6) and the two switching cost thresholds (7) into (4) and (5) we have that if both downstream firms acquire data from the upstream data-brokers, the profits for the two firms A and B are:

$$\pi_i(a,a) = \frac{n[5\sigma^2 + 2\sigma(1-2\mu)\Delta + 2\Delta^2]}{9\sigma} - t \qquad \forall i = A, B \qquad (8)$$

When the two downstream firms do acquire data from the upstream data-broker, their profits are affected by the difference between the higher and the lower valuation (Δ), by the heterogeneity of the switching costs (σ) and by the proportion of mismatched customers (μ).

2. None of the companies acquire data from the upstream data-broker (i.e. "privacy" regime)

If none of the two companies do acquire data, we are in the (n, n) subgame ("privacy" regime). This means that, differently from the previous scenario, companies have less information about customers. Following Shy and Stenbacka, we assume that in this case, companies can only identify their own previous customers; therefore they end up setting two prices: one for the customers that have previously bought, that we indicate with p, and one for those customers that have previously bought from competition, q. Moreover, the fact that firms do not have access to the data provided by the data-broker also means that, differently from before, each company is unable to distinguish between matched and mismatched customers.

Substituting $p_a = p_b = p$and$q_a = q_b = q$ into the utility function (1) and into the profit functions (4) and (5), and then maximizing for the values of p and q it is possible to obtain the equilibrium prices in the subgame:

$$p_a = p_b = p = \frac{2\sigma + (1 - 2\mu)\Delta}{3} \quad and \quad q_a = q_b = q = \frac{\sigma - (1 - 2\mu)\Delta}{3} \qquad (9)$$

Substituting (9) into (2) and (3), the threshold levels of the switching costs are defined as:

$$s_1(n, n) = \frac{1}{3} - \frac{(4\mu + 1)\Delta}{3\sigma} \quad and \quad s_2(n, n) = \frac{1}{3} - \frac{(4\mu - 5)\Delta}{3\sigma} \qquad (10)$$

Substituting (9) and (10) into (4) and (5), the amount of profits obtained by the two firms are:

$$\pi_i(n, n) = \frac{n[5\sigma^2 + 2\sigma(1 - 2\mu)\Delta + 2\Delta^2(1 - 2\mu)^2]}{9\sigma} \qquad \forall\, i = A, B \qquad (11)$$

Clearly, differently from the previous case, in this case the equilibrium profits do not depend on t, as both firms do not buy the data from the data-broker.

3. One of the two firms (be it A or B) does acquire data, while the other does not (i.e. "mixed regime)

The third scenario is the one in which one company (be it A or B) acquires data and the other does not. The following discussion will assume that company A acquires data and B does not (a, n); however, since the model is symmetric, the reasoning is the same for the case in which B does acquire data and A does not (n, a).

If company A acquires data it can set four different prices (as in the case of "no privacy"), while company B, that does not acquire data, can set "only" two prices (as in the case of "privacy"). Therefore we end up respectively with $(a)$for company A and $(n)$for company B, which means a case of (a, n).

Formally, the utility function of the generic customer c having a relationship respectively with A and B becomes:

$$U_A(c)\begin{cases} v_h - p_{ah} & A-oriented\ and\ continues\ to\ buy\ A \\ v_l - p_{al} & B-oriented\ and\ continues\ to\ buy\ A \\ v_l - q_b - s\sigma & A-oriented\ and\ switches\ to\ B \\ v_h - q_b - s\sigma & B-oriented\ and\ switches\ to\ B \end{cases} \quad (12)$$

$$U_B(c)\begin{cases} v_h - p_b & B-oriented\ and\ continues\ to\ buy\ B \\ v_l - p_b & A-oriented\ and\ continues\ to\ buy\ B \\ v_l - q_{al} - s\sigma & B-oriented\ and\ switches\ to\ A \\ v_h - q_{ah} - s\sigma & A-oriented\ and\ switches\ to\ A \end{cases} \quad (13)$$

From the indifferent conditions, the threshold levels of the switching costs are defined as:

$$s_{ah} = \frac{p_{ah} - q_b - \Delta}{\sigma} \quad and \quad s_{al} = \frac{p_{al} - q_b + \Delta}{\sigma} \quad (14)$$

$$s_{bh} = \frac{p_b - q_{al} - \Delta}{\sigma} \quad and \quad s_{bl} = \frac{p_b - q_{ah} + \Delta}{\sigma} \quad (15)$$

The two companies want to maximize the profit functions, respectively for company A and B, described as:

$$\pi_A = p_{ah}(1 - s_{ah})(1 - \mu)n + p_{al}(1 - s_{al})\mu n + q_{ah}s_{bl}\mu n + q_{al}s_{bh}(1 - \mu)n - t \quad (16)$$

$$\pi_B = p_b(1 - s_{bh})(1 - \mu)n + p_b(1 - s_{bl})\mu n + q_b s_{al}\mu n + q_b s_{ah}(1 - \mu)n \quad (17)$$

In this case, since only one company acquires data, only company A incurs the cost of acquiring data, t.

Substituting (14) and (15) into the profit functions in (16) and (17) and then maximizing profits, it is possible to obtain the profit maximizing prices charged by the two firms:

$$p_{ah} = \frac{2\sigma + (1 + \mu)\Delta}{3}; \quad p_{al} = \frac{2\sigma - (1 - \mu)\Delta}{3}; \quad q_{ah} = \frac{\sigma + (2 - \mu)\Delta}{3};$$

$$and \quad q_{al} = \frac{\sigma - (1 + \mu)\Delta}{3} \quad for\ company\ A$$

$$p_b = \frac{2\sigma + (1 - 2\mu)\Delta}{3} \quad and \quad q_b = \frac{\sigma - (1 - 2\mu)\Delta}{3} \quad for\ company\ B \quad (18)$$

Switching costs are obtained by substituting (18) into (14) and (15):

$$s_1(a,n) = \frac{1}{3} - \frac{(\mu+1)\Delta}{3\sigma} \quad and \quad s_2(a,n) = \frac{1}{3} - \frac{(\mu-2)\Delta}{3\sigma} \tag{19}$$

The model of Shy and Stenbacka assumes that the switching cost thresholds are in between 0 and 1, therefore the two values s_1 and $s_2$exist in this range. For these values, the model exists and there is an internal solution.

Finally, substituting the prices found in (18) and the two switching cost thresholds (19) into (16) and (17) we have that if A does acquire data from the upstream data-broker, while B does not, the profits for the two firms A and B are:

$$\pi_A(a,n) = \frac{n[5\sigma^2 + 2\sigma(1-2\mu)\Delta + 2\Delta^2(1+5(1-\mu))]}{9\sigma} - t$$
$$\pi_B(a,n) = \frac{n[5\sigma^2 + 2\sigma(1-2\mu)\Delta + 2\Delta^2(1-2\mu)^2]}{9\sigma} \tag{20}$$

Clearly, given the symmetry of the model, we have that $\pi_A(n,a) = \pi_B(a,n)$ and $\pi_B(n,a) = \pi_A(a,n)$.

Data or not? Finding the Nash Equilibrium in the firms' choice

The above profits represent the pay-offs firms enjoy in the various possible scenarios regarding the acquisition of data from the data-broker, given the price t. We are now in the position to determine the equilibrium of the game played by the two firms who simultaneously decide about data acquisition.

The normal form representation of the game is the following:

		B	
		Acquire	**NotAcquire**
A	**Acquire**	$\pi_A(a,a); \pi_B(a,a)$	$\pi_A(a,n); \pi_B(a,n)$
	NotAcquire	$\pi_A(n,a); \pi_B(n,a)$	$\pi_A(n,n); \pi_B(n,n)$

where:

$\pi_i(a,a) = \frac{n[5\sigma^2+2\sigma(1-2\mu)\Delta+2\Delta^2]}{9\sigma} - t,$

$\pi_i(n,n) = \frac{n[5\sigma^2+2\sigma(1-2\mu)\Delta+2\Delta^2(1-2\mu)^2]}{9\sigma}$ with $i = A, B,$

$\pi_A(a,n) = \pi_B(n,a) = \frac{n[5\sigma^2+2\sigma(1-2\mu)\Delta+2\Delta^2(1+5\mu(1-\mu))]}{9\sigma} - t,$

and

$\pi_A(n,a) = \pi_B(a,n) = \frac{n[5\sigma^2+2\sigma(1-2\mu)\Delta+2\Delta^2(1-2\mu)^2]}{9\sigma}$.

Both companies acquire the data is a Nash equilibrium if:

$$\pi_i(a,a) \geq \pi_i(a,n) \qquad \forall\, i = A,B$$

Formally, Using expressions (6) and (12), it follows that $(a,a)$is Nash equilibrium if$t \leq t_1$where $t_1 = \frac{8}{9} \cdot \frac{n\Delta^2\mu(1-\mu)}{\sigma}$

Alternatively, $(n,n)$is Nash equilibrium if:

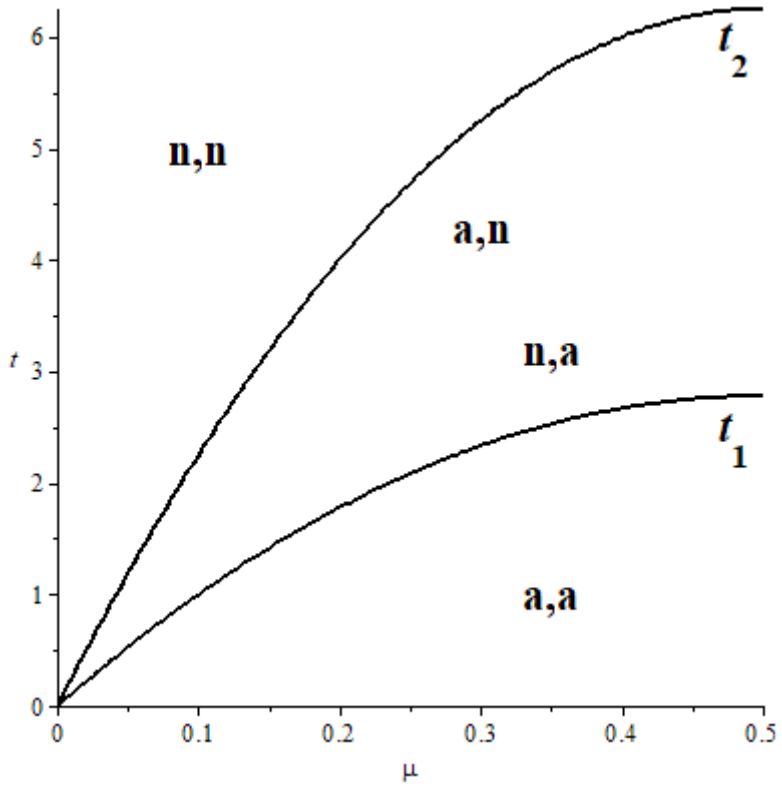
$$\pi_i(n,n) \geq \pi_i(a,n) \qquad \forall\, i = A,B$$

Using the expressions (9) and (12),(n,n) is Nash equilibrium if $t \geq t_2$where$t_2 = \frac{2n\Delta^2\mu(1-\mu)}{\sigma}$,with$t_2 > t_1$.

Finally, using expressions (8), (11) and (20), it is easy to check that (a,n) or(n,a) are the equilibria if $t_1 < t \leq t_2$. Therefore, for the area $t_1 < t \leq t_2$, the Nash equilibrium is when one company (be it A or B) does acquire data while the other does not.

Result 1*: Given the price of the data, t, the equilibrium in the second stage game is* (a,a) *if* $t \leq t_1$,(a,n) *or* (n,a) *if* $t_1 < t < t_2$ *and* (n,n) *if* $t \geq t_2$ *where* $t_1 = \frac{8}{9} \cdot \frac{n\Delta^2\mu(1-\mu)}{\sigma}$ *and* $t_2 = \frac{2n\Delta^2\mu(1-\mu)}{\sigma}$.

The two thresholds $t_1$and t_2 are both increasing in μ; graphically in a (μ,t) space, the equilibrium of the game can be represented as follows:



This diagram identifies the Nash Equilibria of the model. When the price of the data is very large, $t > t_2$ both firms do not acquire and (n, n) is a Nash equilibrium; alternatively, when the price is low, $t \leq t_1$ both firms do acquire data and in this case (a, a) is a Nash Equilibrium. If t takes intermediate values, $t_1 < t \leq t_2$, the equilibrium is the mixed one, whereby only one firm purchases the data.

The diagram reveals that the area with the mixed equilibrium gets smaller the lower the fraction μ of mismatched customers. Interestingly, if we make the difference between t_1 and t_2:

$$t_1 - t_2 = \frac{10}{9} \cdot \frac{\mathrm{n}\Delta^2 \mu(1-\mu)}{\sigma}$$

It is possible to see that, given μ, the area with the mixed equilibrium gets larger the larger Δ and the smaller σ.

3.1.2 The first stage: The data-broker

Going backward, we can define the equilibrium in the first stage, where the upstream data-broker decides the price of data, t.

The data-broker decides the price that guarantees him the greatest profit; there are two alternatives:

a) $t \leq t_1$; in this case we know that both firms purchase the data; the highest price the data-broker can make is t_1 and its profits are equal to $2 \cdot t_1 = \frac{16}{9} \cdot \frac{n\Delta^2\mu(1-\mu)}{\sigma}$;
b) $t_1 < t \leq t_2$ with only one firm buying the data; in this case the highest price that the broker can make is t_2 and its profits are just $t_2 = \frac{2n\Delta^2\mu(1-\mu)}{\sigma}$.

Since $t_2 > 2 \cdot t_1$, the second option is preferred by the data-broker and in equilibrium the data are sold only to one company.

Result 2*: The subgame perfect equilibrium of the two stage game is characterized by the data broker setting* $t_2 = \frac{2n\Delta^2\mu(1-\mu)}{\sigma}$ *and only one firm purchasing the data.*

Interestingly, this is the same equilibrium that the data-broker would obtain if it offered an exclusivity agreement on data to only one of the two downstream firms. However, the data-broker does not explicitly discriminate or offer the exclusivity when it sells its data as happens for instance in the model of Clavorà Braulin and Valletti (2016), but it is the downstream market itself that when the price for data is t_2, selects the equilibrium characterized by only one firm purchasing the data.

3.1.3 Welfare Analysis

After analyzing the equilibrium of our vertically integrated market, one might wonder what the socially optimal configuration of the market is. We answer this question by analyzing consumer surplus, producer surplus (considering the price for data, t, which also affects the profits of the data-broker), and total welfare (i.e. the sum of consumer surplus and industry profits) in the three possible cases that may arise:

- Case 1 - Both companies acquire data (a, a) at the price of t_1;
- Case 2 - Only one company acquires data (a, n) or (n, a) at the price of t_2.
- Case 3 – No one purchases the data (n, n).

Case 1 - Both companies acquire (a,a)

In this scenario both companies acquire data from the upstream data-brokers at t_1.

Theconsumer surplusis the sum of individuals' net surpluses; formally:

$$CS_i = (1-\mu)n\int_{s_1}^{1}(v_h - p_{ih})ds + (1-\mu)n\int_{0}^{s_1}(v_l - q_{jh} - s\sigma)ds + \mu n\int_{s_2}^{1}(v_l - p_{il})ds + \mu n\int_{0}^{s_2}(v_h - q_{jh} - s\sigma)ds \quad \text{with } i = A,B \qquad (21)$$

where the first two terms indicate the surplusesof matched customers: the first term represents the surplus of i-oriented customers which continue to buy from *i*, the second term represents the surplus i-oriented customers that switch to *j*. Similar interpretations for the third and fourth term for the portion of mismatched customers, μ.

Substituting the equilibrium prices in (6) and the corresponding switching cost thresholds (7) into (21) for A and B yields aggregate consumer surplus:

$$CS(a,a) = CS_A(a,a) + CS_B(a,a) = \frac{n}{9\sigma}\{\Delta^2 - 11\sigma^2 + 2\sigma[v_h(5-\mu) + v_l(4+\mu)\} \qquad (22)$$

In this case, the consumer surplus decreases with μ, the proportion of mismatched customers.

When both firms acquire data, the profit function for each company is the same as in (8), with $t = t_1$. Formally, firms' equilibrium profits are:

$$\pi_i(a,a) = \frac{n}{9\sigma}\{5\sigma^2 + 2\sigma(1-2\mu)\Delta + 2\Delta^2(1-2\mu)^2)\} \quad \forall\, i = A,B \quad (23)$$

Finally, the profits enjoyed by the upstream data-broker selling at t_1 to both firms, are:

$$\Pi = \frac{16}{9}\cdot\frac{n\Delta^2\mu(1-\mu)}{\sigma} \qquad (24)$$

The producer surplus (P_s) for the case of are the sum of the profit of the data-broker (24) plus twice the profits for each firm found in (23), $\Pi + 2\pi_i(a,a)$. Formally:

$$P_s(a,a) = \frac{2n}{9\sigma}\{5\sigma^2 + 2\sigma(1-2\mu)\Delta + 2\Delta^2\} \qquad (25)$$

Total welfare is described as the sum of the total consumer surplus in (22) plus the producer surplus found in (25), which yields to:

$$W(a,a) = \frac{n}{9\sigma}\{5\Delta^2 - \sigma^2 + 2\sigma[v_h(7-5\mu) + v_l(2+5\mu)]\} \qquad (26)$$

Case 2 - "Mixed case" (a,n) or (n,a)

Now we analyze the case in which one company acquires data and the other doesnot. We consider the case in which A acquires and B does not acquire (a,n), but the same is true for the opposite case, (n,a).

Total consumer surplus, i.e. the sum of the two consumer surplus for company A and B is defined by substituting the equilibrium prices in (18) and the corresponding switching cost thresholds (19) into (21), which yields to:

$$CS(a,n) = CS_A(a,n) + CS_B(a,n) = \frac{n}{9\sigma}\{(1+5\mu-5\mu^2)\Delta^2 - 11\sigma^2 + 2\sigma[v_h(5-\mu) + v_l(\mu+4)\} \quad (27)$$

In this case the upstream data-broker is selling to only one firm at t_2. Therefore, the profits are the same as in (20) but for the company acquiring data (in this case company A) we have that $t = t_2 = \frac{2n\Delta^2\mu(1-\mu)}{\sigma}$. Formally, this yields to:

$$\pi_A(a,n) = \frac{n}{9\sigma}\{5\sigma^2 + 2\sigma(1-2\mu)\Delta + 2\Delta^2(1-2\mu)^2\}$$

$$\pi_B(a,n) = \frac{n}{9\sigma}\{5\sigma^2 + 2\sigma(1-2\mu)\Delta + 2\Delta^2(1-2\mu)^2\} \quad (28)$$

The two firms make different gross profits and in fact the firm that buys the data (in this case A) makes higher gross profits than the company that does not acquire (in this case B). However, since firm A buys the data it has to pay the price t_2, the data-broker is able to extract entirely as much surplus as it can. For this reason, the final result is that in equilibrium the net profits of the two firms in the asymmetric case are the same.

In this case, the data-broker is selling to only one company at t_2; therefore its profit equals to:

$$\Pi = \frac{2n\Delta^2\mu(1-\mu)}{\sigma} \quad (29)$$

Producer surplus (P_s) is described as the sum of the two downstream firms' profits found in (28) and the profit for the upstream data-broker in (29), which yields to:

$$P_s(a,n) = \frac{2n}{9\sigma}\{5\sigma^2 + 2\sigma(1-2\mu)\Delta + \Delta^2[2+\mu(1-\mu)]\} \quad (30)$$

Total welfare is therefore defined by summing the producer surplus found in (30), and total consumer surplus found in (27):

$$W(a,n) = \frac{n}{9\sigma}\{(5+7\mu-7\mu^2)\Delta^2 - \sigma^2 + 2\sigma[v_h(5\mu-7) + v_l(5\mu+2)]\} \quad (31)$$

Case 3- None of the companies acquire data (n,n)

Suppose that the data broker sets $t > t_2$; in this case, no firm purchases the data. Alternatively, one may think to the case in which there is no upstream data-broker selling its data to the two downstream firms.

Total consumer surplus is defined by substituting the equilibrium prices in (9) and the corresponding switching cost thresholds (10) into (21), which yields to aggregate consumer surplus.Formally:

$$CS(n,n) = CS_A(n,n) + CS_B(n,n) = \frac{n}{9\sigma}\{(1+32\mu-32\mu^2)\Delta^2 - 11\sigma^2 + 2\sigma[v_h(5-\mu) + v_l(\mu+4)\} \quad (32)$$

For the case of (n,n), since both firms do not acquire data, the upstream market of the data-broker and its related profits are not considered. Therefore, producer surplus equals the sum of the profits for each firm found in (11). Formally:

$$P_s(n,n) = \frac{2n}{9\sigma}\{5\sigma^2 + 2\sigma(1-2\mu)\Delta + 2\Delta^2(1-2\mu)^2\} \quad (33)$$

The total welfare is described as the sum of the total consumer surplus in (32) and the profits of the two downstream firms in (33). Formally, thisyields to:

$$W(n,n) = \frac{n}{9\sigma}\{(5 + 16\mu - 16\mu^2)\Delta^2 - \sigma^2 + 2\sigma[v_h(7 - 5\mu) + v_l(5\mu + 2)\} \quad (34)$$

3.2 Results comparison

Producer surplus

Producer surplus (P_s) is described as the sum of the profits of the two downstream firms plus, eventually, the profits of the data-broker.

Formally, the difference between the producer surpluses in the three different cases is:

$$P_s(a,n) - P_s(a,a) = \frac{2}{9}\frac{n\Delta^2\mu(1-\mu)}{\sigma}$$

$$P_s(a,a) - P_s(n,n) = \frac{16}{9}\frac{n\Delta^2\mu(1-\mu)}{\sigma}$$

$$P_s(a,n) - P_s(n,n) = \frac{2n\Delta^2\mu(1-\mu)}{\sigma}$$

As $0 < \mu < \frac{1}{2}$, we have that $P_s(n,n) < P_s(a,a) < P_s(a,n)$. Therefore the following result holds:

Result 3*: The producer surplus is larger when one firm (either A or B) does acquire data and the other does not,* (a,n) *or* (n,a)*, with the upstream data-broker making a profit equal to* t_2.

In addition, the difference between the producer surpluses in the three cases is increasing as a function of μ, i.e. the fraction of mismatched customers: as the fraction of mismatched customers (μ) increases, the benefit of a firm to have knowledge of customers' preferences increases as well.

Consumer surplus

Formally, the difference between the consumer surpluses (CS) of the three different cases is:

$$CS(n,n) - CS(a,n) = \frac{32}{9}\frac{\Delta^2\mu(1-\mu)}{\sigma}$$

$$CS(n,n) - CS(a,n) = \frac{3\Delta^2\mu(1-\mu)}{\sigma}$$

$$CS(a,n) - CS(a,a) = \frac{5}{9}\frac{n\Delta^2\mu(1-\mu)}{\sigma}$$

Therefore, we have that $CS(a,a) < CS(a,n) < CS(n,n)$. As before, the difference between the consumer surplus of the three different cases is increasing as a function of μ, the portion of mismatched customers.

Result 4*: The consumer surplus increases with the level of privacy protection.*

This results reflects the "monotonicity" property already described in Shy and Stenbacka: therefore, consumer surplus is higher when both firms do not acquire data. The less privacy protection, the higher the amount of information at firms' disposal, the higher the possibility for firms to set a higher number of different prices, the lower the consumer surplus. The fact that firms can price-condition customers based on their brand preferences is not beneficial for consumers, since firms can extract more consumer surplus as the privacy protection level decreases.

Total welfare

The total welfare (W) is described as the sum of total consumer surplus (CS) and producer surplus (P_s). Formally the total welfare differences between the three cases are defined as:

$$W(a,n) - W(a,a) = \frac{7}{9}\frac{n\Delta^2\mu(1-\mu)}{\sigma}$$

$$W(n,n) - W(a,n) = \frac{n\Delta^2\mu(1-\mu)}{\sigma}$$

$$W(n,n) - W(a,a) = \frac{16}{9}\frac{n\Delta^2\mu(1-\mu)}{\sigma}$$

Since$0 < \mu < \frac{1}{2}$, we have that $W(a,a) < W(a,n) < W(n,n)$. In fact, we register the lower total welfare in the case of (a,a), when both firms acquire data about customers from the upstream data-broker; while the higher level of total welfare is represented by the case of

(n, n), when none of the firms acquire data. As for the consumer surplus (CS),also the total welfare (W) reveals the "monotonicity" property, and the following result holds:

Result 5: *Total welfare increases with the level of privacy protection.*

Therefore, the more information companies possess about customers and the easier it is for them to discriminate against consumers on the basis of their preferences. The less a firm knows about customers, the lower its possibility to price discriminate and the higher the total welfare. This means that the enhancement of privacy policies aiming at reducing the possibility for firms to discriminate on the basis of customers' preferences ends up increasing the total welfare.

In conclusion, the consumer surplus and the total welfare increase with the level of privacy protection (i.e. "monotonicity"). On the contrary, the producer surplusdo not follow the "monotonicity" property, because it shows higher results for the "mixed" regime, when only one company (be it A or B) acquires the data and not in the case where the level of privacy protection of customers is higher.

From this discussion, we end up with the following Corollary describing the social optimum:

Final Corollary:

A. *The socially optimal outcome is* (n, n). *This means that the best possible outcome form the social perspective is the one in which there is not upstream data-broker (absence of data market); companies do not buy any data and cannot discriminate;*
B. *The second best is the "mixed" regime,* (a, n) *or* (n, a)*, which is actually the market equilibrium where the data-broker offers its data at* t_2 *and only one company (be it A or B) purchases it.*

3.3 Extension of the model: the dual-approach

We analyzed the setting that includes an upstream data broker who sells its data to two downstream firms. What if the data broker and one of the two downstream companies are vertically integrated?

This setting is particularly interesting when looking at real world digital markets which are dominated by tech giants, such as Amazon. Amazon follows a so-called "dual approach": it controls the platform where transactions take place and, in many cases, competes with other vendors in the retail market. The issue of our interest is that being the operator of the platform, Amazon has access to all the relevant information about the transactions taking place on the platform, also those regarding its rivals in the retail market. Amazon knows the identity of rivals' customers, it observes transaction prices, it knows about sellers and their commercial strategies and so on. This feature of platform markets has attracted the attention of regulators and policy makers as significantly contributes to maintain the parket power of dominant operators.

We can use our model to try to represent this situation, by assuming that one of the two vendors has also access to the data, that is it is vertically integrated backwards and controls the data-broker. In this scenario, the vertically integrated operator can decide whether or not to sell the data to the rival firm with which it competes in the retail market. The question then is the following: is this firm willing to sell its data to the rival? And is the rival willing to purchase such data?

Suppose company A (e.g. Amazon) controls the data, while B is the rival in the retail market. A collects its data but it does not pay the price to get the data. Using our previous notation, if firm A does not sell the data to B it gets $\pi_A(a, n)$, gross of t that in this case it does not pay as it controls the data, while if it sells the data to B it gets $\pi_A(a, a)$. Hence, A has an incentive to sell its data to B if and only if the price it receives from the sale of the data more than compensates the smaller profits:

$$t > \pi_A(a, n) - \pi_A(a, a)$$

On the other hand, if firm B purchases the data it gets $\pi_B(a, a)$, gross of the payment, while it gets $\pi_B(a, n)$ if it does not acquire the data. Hence, firm B purchases the data from A if and only if the payment is not too large:

$$t < \pi_B(a, a) - \pi_B(a, n)$$

Therefore, A sells the data and B buys them if and only if:

$$\pi_A(a,n) - \pi_A(a,a) < t < \pi_B(a,a) - \pi_B(a,n)$$

Hence, there is room for the sale of the data only if:

$$\pi_A(a,n) - \pi_A(a,a) < \pi_B(a,a) - \pi_B(a,n)$$

that is if (remember, the model is symmetric therefore, $\pi_A(a,a) = \pi_B(a,a)$):

$$2\pi_A(a,a) - [\pi_B(a,n) + \pi_a(a,n)] > 0$$

Using firms' equilibrium profits in the various scenarios, this inequality becomes:

$$-\frac{2n\Delta^2\mu(1-\mu)}{\sigma} > 0 \qquad with\ 0 < \mu < 1/2$$

Which is clearly impossible. Hence the following result:

Result 6*: It does not exist a price for the data that A is willing to accept and that B is willing to pay.*

With all the limitation of our simple model, this extension reveals that a vertically integrated firm which acts both as a data-broker and compete in the retail market has not incentive to sell its data to the rival firm.

Today, few digital platforms cover a central role in specific segments (e.g. Amazon, Facebook, Google, Booking, etc) and keep maintaining a dominant position in the market of reference also thanks to the possession of valuable information about customers. With this data, companies can more accurately assess their willingness to pay and more easily set differentiated prices or personalized services which ultimately make consumers more willing to make the purchase. Therefore, there are some companies that occupy a more strategic position in their reference market. According to our model, these companies do not sell the data, rather they prefer to keep it for themselves. Our model is static, so a dynamic extension

of the model would ensure that the market continues to remain in a condition of dominance by the firm that controls the data. However, in recent years, regulators are pushing to limit the competitiveness of these giants, enacting legislation in favor of a limitation of their market power, favoring lower barriers to entry and higher competitiveness. And this is happening both at the European and at the international level. For instance, in December 2020, a regulation proposal called "Digital Markets Act" was presented with the main idea of putting order to the digital market, mainly from an antitrust point of view, and numerous antitrust hearings involving the CEOs of the big tech companies have taken place.

3.4 Conclusions

This paper deals with issues regarding the economics of privacy, which analyses “the trade-offs associated with the balancing of public and private spheres between individuals, organizations, and governments” (Acquisti et al., 2016).

Privacy is a complex and malleable concept. It can be identified both as the need to isolate oneself from others by taking refuge in one's "private sphere", as well as the need of "hiding information" (Posner, 1978). Dealing with a “data-subject”, privacy involves also the identification of some "property rights" (Stigler, 1980). Subsequently, towards the end of the 90s, with the digitization of information and the reduction of marginal costs of collection and storage of data and information (Shapiro and Varian, 1998), economists have studied the consequences from the transfer and exchange of data: positive and negative externalities (Varian, 1997), the issue of transaction costs, assignment of ownership and control rights (Noam, 1997; Laudon, 1997). With the advent of cutting-edge and increasingly sophisticated information technologies, the economics of privacy has developed in its different ramifications, increasing its complexity. However, in light of the possibility for companies to more easily identify consumers, as well as their characteristics, behavior or "clickstream", the issue of privacy violation has been studied considering the "informational privacy" (Brandimarte and Acquisti, 2012) and the related "information security".

Are consumers worried about their privacy? On the one hand, it can be argued that consumers receive personalized and enhanced offers: they can obtain a benefit given by an offer that is more relevant to them, which reduces their search costs and increases their welfare (Acquisti et al. 2016). However, the possibility of easily identifying consumers has exposed them to risky practices such as price discrimination, target advertising and personalization, but also identity theft and the aforementioned privacy violations. In deciding

their level of privacy protection, individuals face the so-called “privacy paradox”: they claim to be concerned and willing to protect their privacy, but they end up giving up privacy when they are incentivized to do so (Athey et al. 2017).

Profiles of customers are constantly updated on the basis of new information companies can collect both directly and indirectly from data-brokers. Moreover, companies are also helped by data-brokers. Data-brokers are described as "companies that collect consumers' personal information and resell or share that information with others" (Federal Trade Commission, 2014). These companies can be of various types: financial data providers such as Bloomberg, credit rating agencies such as Moody's, "pure" data brokers such as Acxiom and online aggregators such as Spokeo (Bergemann and Bonatti, 2012). Data-brokers play a strategic role concerning the collecting, managing, organizing and selling data products or packages, but these practices usually happen in consumer non-awareness and customers end up supporting the major risks. This is why the legislation has intervened in order to put some limitations on how data should be managed by intermediaries, third-parties and retailers through regulation such as the European GDPR.

In the third chapter we presented a model with a data-broker, analyzing its strategic role as suppliers of worthful data and information to two downstream firms. Starting from the model of Shy and Stenbacka (2016), we have considered two downstream firms that have to decide if to acquire or not the data from an upstream data-broker: buying data means supporting a cost t, which also is crucial in determining the profits for the data-broker.
This model is solved by backward induction: in the last stage companies have to simultaneously decide if to acquire (supporting the cost) or not acquire (avoiding the cost) the data. After having analyzed the decision made by downstream firms, by backward induction we analyze the upstream data-broker operating in the data-market and deciding the price for its data. Our model reveals that, in equilibrium, the data-broker sells its data to only one firm at a larger price. In the last phase of our analysis we perform a welfare and we show that despite total producer surplus being higher for the “mixed” regime (where one of the companies acquires data), consumer surplus and total welfare are larger in the privacy regime (where none of the companies acquire data). The “first best” of the model is therefore the scenario where there is no upstream data market and none of the firms acquire data (“privacy” regime), and the “second best” is the “mixed” regime, i.e. when only one of the two downstream firms (be it A or B) acquires data from the upstream data-broker. Despite the limitations of the model, we can comment on its interesting results.

In the model, we have shown that the price for data set by the upstream data-broker t significantly affects the equilibrium at the firms' level. In fact, according to the price for data chosen by the data-brokers, each of the two downstream firms can buy or not the data, but this does not mean that the data-broker is offering the exclusivity to only one firm. On the contrary, it is the market itself that determines this outcome.

As a "second-best" result, the model predicts that in equilibrium only one firm (be it A or B) acquires the data while the other does not. In reality, what we actually observe is that data ownership is not widespread among all the companies on the market, rather a few companies control huge amounts of data. Our model is static, therefore a potential extension of the model could be the dynamic version of the model, considering for instance that the firm acquiring the data in the long-run could exclude the rival from the market. Therefore, it can be assumed that a policy intervention could be aimed at imposing the sale of data to all downstream companies, in order to put them all on the same level, thus preventing a tendency to monopolization.

A comment can be done also considering the "first best" outcome, which corresponds to the case where none of the companies do acquire data from the data-broker. From a social welfare perspective, this result suggests that it would make sense to promote the maximum level of privacy protection, even if this means the closing of the data-market. Actually, for consumers, this corresponds to the scenario that provides them with a higher level of data protection. This conclusion suggests that privacy policies should aim at ensuring higher consumer surplus, which in our model is increasing with the level of privacy protection. This result embraces the idea that in the age of digitization there is a need to push for greater privacy protection, especially for what concerns customers and online users, which usually are the most vulnerable. Even with the limitations of the model, this result supports the idea of the current policies, regulations and laws aimed at granting the consumer a greater level of data protection, such as the European GDPR or the California Consumer Privacy Act.

Overall, our results are in line with the idea of creating a "global consensus on privacy" by standardizing global and international laws aimed at protecting more customers and at the same time favoring the competition between firms while downsizing the monopolist positions held by few players in specific markets. For example, the European GDPR has found counterparts such as the Lei Geral de Proteção de Dados (LGPD) in Brazil and the California Consumer Privacy Act (CCPA) in California that both entered into force in 2020. However, it must also be considered the dynamic nature of the legislation that could have influences in the years to come, not only on consumer protection, but also from the

point of view of the data collection by companies, ultimately leading to significant changes in the data market.

There are still many open questions in the field of the economics of privacy. The first question is linked to the definition of the cost of privacy intrusions, i.e. the harmful consequences from the disclosure of personal information, which can be treated both from an economic (i.e. quantitative) approach, and from a psychological (i.e. abstract) approach. The second question seeks to understand whether there is an optimal amount of privacy protection, both from an individual point of view and for the society as a whole. This is a complex issue that also includes the analysis of the value extracted from data, that most of the time is combined between them. The third question, related to the previous one, tries to go further: if and once an "optimal" amount of privacy can be identified, who should be responsible for achieving that certain amount? Should the individuals be responsible through their behavior and informed choices? Should the corporate market self-regulate itself and compete on privacy? Or should it be the government through its regulation?

For these reasons, research on the economics of privacy still has ample room for growth in the years to come.

Bibliography

Acquisti, A. and Brandimarte, L. (2012). *The Economics of Privacy* (Editors: Peitz, Martin and Waldfogel, Joel), The Oxford Handbook of the Digital Economy, Oxford University Press

Acquisti, A., Gross, R. and Stutzman, F.D. (2014). Face Recognition and Privacy in the Age of Augmented Reality. *Journal of Privacy and Confidentiality*, 6(2), 1.

Acquisti, A., Taylor, C. R. and Wagman, L. (2016). The Economics of Privacy *Journal of Economic Literature*, Vol. 52, No. 2.

Acquisti, A., and Varian, H.R. (2005). Conditioning Prices on Purchase History. *Marketing Science* (Providence, R.I.), vol. 24, no. 3, INFORMS, pp. 367–81.

Anderson, R. (2001). Why Information Security Is Hard - an Economic Perspective. *Seventeenth Annual Computer Security Applications Conference*, IEEE, pp. 358–65.

Athey, S., Catalini, C. and Tucker, C. (2017). The Digital Privacy Paradox: Small Money, Small Costs, Small Talk. *National Bureau of Economic Research Working Paper Series*. No. 23488.

Belleflamme, P., Lam, W. M. W., and Vergote, W. (2020). Competitive imperfect price discrimination and market power. *Marketing Science*, 39(5).

Binns, R., Lyngs, U., Van Kleek, M., Zhao, J., Libert, T. and Shadbolt, N. (2018). Third Party Tracking in the Mobile Ecosystem. In Proceedings of the 10th ACM Conference on Web Science (WebSci '18). *Association for Computing Machinery*, New York, NY, USA, 23–31.

Birckan, G., Dutra, M., Macedo, D. and Godoy, V.A. (2020). Personal data protection and its reflexes on the data broker industry. *1st EAI International Conference on Data and Information in Online Environments.*

Bergemann, D. and Bonatti, A. (2012). Markets for Data. Meeting Papers 538, *Society for Economic Dynamics.*

Bounie, D., and Dubus, A., and Waelbroeck, P. (2020). Selling Strategic Information in Digital Competitive Markets. *RAND Journal of Economics*. Forthcoming.

Brandimarte, L., Acquisti, A., and Loewenstein, G. (2013). Misplaced Confidences Privacy and the Control Paradox. *Social Psychological and Personality Science*, 4 (3), 340-347.

Brooke, S. and Véliz, C. (2020). Data, Privacy & The Individual. *Madrid: Center for the Governance of Change*, IE University.

Casadesus-Masanell, R. and Hervas-Drane, A. (2015). Competing with privacy. *Management Science*, 61(1):229–246.

Cabral, L.M.B. (2000). *Introduction to Industrial Organization*. The MIT press.

Choi, J.P., Jeon, D.S. and Kim, B.C. (2019). Privacy and personal data collection with information externalities. *Journal of Public Economics*, Volume 173, Pages 113-124.

ClavoràBraulin, F. and Valletti, T. (2016). Selling Customer Information to Competing Firms. *Economics Letters*, vol. 149, Elsevier B.V, pp. 10–14.

Comino, S., and Manenti, F.M. (2014). *Industrial Organisation of High-Technology Markets: the Internet and Information Technologies*. Edward Elgar.

Conitzer, V., Taylor, C. and Wagman, L. (2012). Hide and Seek: Costly Consumer Privacy in a Market with Repeat Purchases. *Marketing Science* (Providence, R.I.), vol. 31, no. 2, INFORMS, pp. 277–92.

Drexl, J. (2017). Designing Competitive Markets for Industrial Data – Between Propertisation and Access. 8 *JIPITEC* 257 para 1.

European Commission, (2017). The economics of ownership, access and trade in digital data. *JRC Digital Economy Working Paper*, JRC104756, ISSN 1831-9408.

European Commission (2020). The European Market Monitoring Tool, Key facts & figures, first policy conclusions, data landscape and quantified stories. D2.9 Final Study Report, ID: N- 30-CE-0835309/00-96.

Federal Trade Commission (2014). Data-brokers: a call for transparency and accountability. https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf

Gu, Y., Madio, L. and Reggiani, C. (2021). Data Brokers Co-opetition. Available at SSRN: https://ssrn.com/abstract=3308384.

Hillebrand, K. and Hornuf, L. (2021). The Social Dilemma of Big Data: Donating Personal Data to Promote Social Welfare. *Max Planck Institute for Innovation & Competition* Research Paper No. 21-08.

Ichihashi, S. (2020). Competing data intermediaries. *Mimeo*.

Laudon, K.C. (1996). Markets and privacy. *Commun. ACM* 39, 9, 92–104.

Laudon, K.C. (1997). Extensions to the Theory of Markets and Privacy: Mechanics of Pricing Information. *NYU Working Paper* No. 2451/14166.

Lavagnino, M.B. (2013). Privacy Revealed. EDUCAUSE Review, 48, no.1.

Montes, R., Zantman, W. and Valletti, T. (2019). The value of personal information in online markets with endogenous privacy. *Management Science*, 65(3):1342–1362.

Noam, E. (1997). Privacy and Self-Regulation: Markets for Electronic Privacy. *U.S. Dept. of Commerce, Privacy and Self-Regulation in the Information Age.*

OECD (2017). OECD Digital Economy Outlook 2017. *OECD Publishing*, Paris, https://doi.org/10.1787/9789264276284-en.

OECD (2020). OECD Digital Economy Outlook 2020. *OECD Publishing*, Paris, https://doi.org/10.1787/bb167041-en.

Palfrey, J.G. and Gasser, U. (2012). Interoperability in Information Systems in the Furtherance of Trade. *NCCR Trade Regulation Working Paper* No. 2012/26, Berkman Center Research Publication No. 2012-21.

Pew Research Center (2016). The state of privacy in post-Snowden America, https://www.pewresearch.org/fact-tank/2016/09/21/the-state-of-privacy-in-america/

Pew Research Center (2019). Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information, November 15, 2019

Posner, R. (1978). The Right of Privacy. *Georgia Law Review* 12(3), pp. 393–422.

Posner, R. (1981). The Economics of Privacy. *The American Economic Review*, vol. 71, no. 2, The American Economic Association, pp. 405–09.

Shapiro, C. and Varian, H. R. (1998). *Information Rules: a Strategic Guide to the Network Economy*. Oxford University Press.

Shy, O., and Stenbacka., R. (2016). Customer Privacy and Competition. *Journal of Economics & Management Strategy*, vol. 25, no. 3, Wiley Subscription Services, Inc, pp. 539–62.

Steinfeld, N. (2016). “I agree to the terms and conditions”: (How) do users read privacy policies online? An eye-tracking experiment. *Computers in Human Behavior*. 55. 992-1000.

Stigler, G. (1980). An Introduction to Privacy in Economics and Politics. *The Journal of Legal Studies*, vol. 9, no. 4, The University of Chicago Law School, Dec. 1980, pp. 623–44.

Streifeld, D. (2000). On the Web price tags blur: what you pay could depend on who you are. *The Washington Post*.

Taylor, C. (2004). Consumer Privacy and the Market for Customer Information. *The Rand Journal of Economics*, vol. 35, no. 4, RAND, pp. 631–50.

Tucker, C. (2014). Social Networks, Personalized Advertising, and Privacy Controls. *Journal of Marketing Research*, vol. 51, no. 5, Oct. 2014, pp. 546–62.

Twetman, H., Bergmanis-Korats, G., Biteniece, N., Fredheim, R., Bertolini, G. and Bay, S. (2020). Data Brokers and Security. Risks and Vulnerabilities Related to Commercially Available Data. *NATO Strategic Communications Centre of Excellence*.

US Supreme Court (1989). DOJ v. Reporters Comm. for Free Press, 489 U.S. 749, *JUSTIA* (https://www.justice.gov/archive/oip/foia_guide09/exemption6.pdf)

Varian, H. R., (1996). Economic Aspects of Personal Privacy. In: Privacy and SelfRegulation in the Information Age. *National Telecommunications and Information Administration*, US Department of Commerce.

Villas-Boas, M.J. (2004). Consumer Learning, Brand Loyalty, and Competition. *Marketing Science*, vol. 23, issue 1, 134-145

Westin, A. (2008). How online users feel about behavioral marketing and how adoption of privacy and security policies could affect their feelings. *Report, Privacy Consulting Group*, Washington, DC.