Master Thesis U.S.E

"Market failures and economic theory, an insight on their effects on data-driven companies"

Student: Giacomo Mercuri – 1198130

g.mercuri@students.uu.nl

First Supervisor: Luigi Pinna

Second Supervisor: Timo van Balen

Abstract

This research is aimed at exploring how data-driven companies have become so dominant in the tech industry, often thanks to the effects of market failures; in particular, how their growth has been fostered by Network Effects and Asymmetric Information. The main question addresses the benefits that data-driven companies profited from market failures; the two complementary questions are aimed to explore how to determine a 'real' economic value of users' personal data and, lastly, how premium or freemium pricing models can reduce information conflicts between users and platforms. To answer these questions, a process of reviewing previously implemented research, papers and studies is performed, along with an analysis of some relevant anecdotal evidence, and combined with a critical review of concepts taken from economic theory.

JEL-codes: D43, D82 & C81

Keywords: data-driven companies; market failures; asymmetric information; economic value of data; Freemium pricing model

For the writing of this thesis, I would like to thank in the first place my supervisor Luigi Pinna, a source of inspiration and stimulus for the choice of the topics covered, my second supervisor Timo Van Balen for his commitment and general supervision. Finally, I would like to thank my friends and family for helping and supporting me, especially in the final period of the year.

Table of contents

1. Introduction

This research is aimed at exploring how data-driven companies have become so dominant in the tech industry thanks to the effects of market failures, in particular, how their growth has been fostered by Network Effect and Asymmetric Information.

Nowadays, the digital world is increasingly breaking into the real world, influencing and shaping many aspects of our lives. Data-driven companies are the companies that are currently and also in the future driving this change, with growth rarely comparable in past economic history. Many data-driven companies make a huge volume of profits by collecting, analysing, and selling their users' personal information. Various entities are involved in this context, including: online platforms, users, data brokers, advertising agencies and governments that decide on legislation regarding the limits of exploitation and commodification of data. Furthermore, the largest of these such as Meta, Google, and Amazon, among others, have become among the biggest firms (each of the three have exceeded a market value of $1 Trillion at a recent point in time), with a global reach that has been attained by very few other firms in other market segments.

Increasingly, microeconomic theory seems to have failed in explaining the mechanism for which these types of companies continue to have an increasing revenues curve and continuously declining marginal costs, in effect, questioning the dogma that MC curves have to be ‘U-Shaped’.

This classical model doesn’t seem to properly fit certain data-driven companies that operate with intangible assets such as users’ personal information. This happens mainly because, for most of the online platforms provided by these firms the marginal cost of adding an additional user tends to approach zero or is extremely modest compared to the amount of revenue per user. Another reason that could explain this costs’ reduction of their inputs it’s correlated to the fact that much of their content is ‘user provided’. This allows most of these companies to sell their products and services for free or at a very low cost-of-goods sold. For these companies, the cost structure is not centred entirely on marginal costs, but on Long Run Incremental Costs (LRIC), that’s because, in practice, the true marginal investment for these firms is not for the next subscriber, but for the next ‘set’ or ‘block’ of new users, which is a number that can vary among firms.

The conflict between theory and practice in this field can find an explanation in market failures generated by the Network Effect, in this case, both on the supply side (Monopoly power- a natural result of scale and scope efficiencies) and on the demand side (Asymmetric Information- leads to a lack of true market clearing pricing, when users assume they get ‘free’ services).

From an academic point of view, this research is going to analyse and explore the gap between economic theory and pieces of evidence from the data-driven industry, in particular, taking into account in the first place the economic phenomena of the Network Effect and the Information Asymmetry between users and data-driven corporations, and then enter more specifically as regards the demand side of this market, analysing which is the economic value of users' personal data and whether the freemium and premium pricing models options might provide a fairer option for users in terms of information control and disclosure, as in the case of LinkedIn and Tinder.

Therefore, my research question is:

- “*How have data-centric companies benefited from network effect and asymmetric information?*”

The sub-questions related to it are:

- "*How can one determine the 'real' economic value of personal data?*"
- "*Can freemium and premium pricing models be applied in order to address the apparent market failures driving the success of what have become some of the largest 'winner-take-all' players in this industry?* - in essence - models that can reduce information conflicts between users and platforms.

The main research question is addressed with a study and integration between some economic notions such as Network Effects, Information asymmetry, marginal costs, and a focus on how these concepts are applied and exploited in the data-driven economy by the actors who participate in it. Concerning the information asymmetry between users and data-driven companies, some methods are listed to reduce the impact of the information asymmetry between users and online platforms. In particular, one of these methods concerns showing or making users aware of the economic value of their personal data. Consequently, this concept is then analysed in the first sub-question or "How can one determine the 'real' economic value of personal data?" The second sub-question also refers to methods for reducing online information asymmetry, but in this case through a pricing model called "Freemium".

It is the objective of this research to give an insight and a critical view of the economic theories applied to the data-driven industry and lead to real-life practical contributions in multiple sectors such as: Business Ethics (how far can those types of business models can be aligned with users wants, needs and morals), marketing for data-driven companies, Web development, policymaking on consumer's information awareness and protection, but also more humanistic field related to users' psychology and their purchasing behaviours.

The remainder of this paper is therefore structured as follows: it starts with a literature review, which begins with an introduction about data-driven companies and a brief explanation of the concept of Long Run Incremental Cost, then moves on to the analysis of the Network Effect and Information Asymmetry on companies operating in the big-data market, and finally reviews academic research conducted so far regarding concepts such as the economic value of personal data and freemium-premium pricing models in online platforms. The main and central body of the paper is structured in three chapters. The first one provides an overview of market failures and how they manifest in the data-driven economy, with special attention to asymmetric information; the second one addresses the problem of quantifying the 'real' economic value of personal data online; the third one analyses the "Freemium" pricing model for online platforms, exploring if it can be a tool that can reduce asymmetric information. The final part of the research addresses the main findings and personal interpretation of them, closing with a discussion and a conclusion.

2. Literature review

Data-driven companies have established a framework and culture for making company-wide decisions, from marketing to product development to human resource development, where data is prized and used effectively (Unscrambl, 2021). Examples of such companies and their relative platforms are Alphabet (Google), Meta (which controls Facebook, Instagram, Messenger, and WhatsApp), and further companies that offer a wide range of services, from transportation (Uber) to entertainment streaming (Twitch, acquired by Amazon in 2014). Most of the content offered by these Internet companies is provided to users for free or at a low cost and, despite that, they still make massive profit volumes. *"If you are not paying for it, if you are not the customer… you are the product!",* that's a very popular quote by Richard Serra from his short film "Television Delivers People." (1973) that explains how the television industry and the data-driven economy are comparable. Especially for data-driven companies, if they are giving something for nothing, they are making money from somewhere else, and that is far too often by selling users' information (Keller,2014).

Long Run Incremental Costs, Network Effect and Asymmetric Information

Most of the literature and economics theory regarding Long Run Incremental Cost has been focusing on the application of this forward-looking cost model in the telecommunication sector, often used to determine the price competitors pay for services provided by an incumbent operator with substantial market power (commonly a 'natural' monopolist, and normally under rate-of-return regulation).

Long Run Incremental Cost (LRIC) is the "long-run cost of providing the next *increment* of output, which should be measured on a forward-looking basis" (Den Hartog,2021). If the unit price of a product rises due to an increase in long-term incremental cost (LRIC), the company needs to raise the price of the product in order to maintain the same profit margin (Kenton,2021). When the unit price goes down, the company can lower the price of the product to maintain the same profit margin, potentially increase the demand, or operate at a higher profit margin. Adopting LRIC on pricing and costing apparently had significant impacts on performance (Jwaifel,2017). Mason (2008) stated that LRIC is typically used as a marginal cost metric for network businesses with a high percentage of fixed capital costs. According to Shirmohammadi et al., (1996), pricing methods include assessing and assigning all long-term costs (operating and reinforcement costs) associated with a new network transaction to that transaction. Therefore, incremental costs can be defined as the amount of revenue required to pay for new facilities, especially due to service customers (Kovacs & Leverett,1994). Long-term costing methods are considered to be more economically efficient as they reflect the cost of future grid enhancements as a result of increased demand/generation of nodes (Heng et al.,2019).

In economics, the term Network Effect concerns a situation where the value of a product, service, or platform depends on the number of buyers, sellers, or users that leverage it. In general, the greater the number of buyers, sellers or users is, the greater the network effect and value created by the offering are. Katz and Shapiro (1985) defined a network product as "a product in which the utility gained from a user consuming good increases with the number of other agents consuming the good", in essence, increasing return to scale. In addition, network effects can be simulated for both current and expected future sales (Farrell & Saloner, 1986). In summary, the network effect is that the benefits that consumers get from adopting or using a product depend (actively) on the current and future numbers of other consumers using the same product. The network effect seems to have a positive impact on society as it brings more benefits to consumers who consume the same goods and services. However, from a firm point of view, the "positive feedback effect" (Arthur, 1989, 1990) allows network effects to benefit larger networks than smaller ones. This term indicates that the increase in network size is much more attractive to the newer network for new buyers, indicating that it is much worth a service that enables access to this network increase. Therefore, consumers tend to obtain products or services from higher levels of network effects and higher networks, and companies with high market share may be higher profitability (Farrell & Saloner, 1985, Katz & Shapiro, 1985, 198, 194; Farrell & Klemperer, 2007). In extreme cases, network effects "Winner-Takes-All market" (Arthur, 1996) can be created. Excellent products and technologies may occur, they can act as a dominant player, and other companies can issue a market for lack of competitive advantage based on network effects (Shapiro & Varian, 1999, Lovesowitz, 2002).

Network effects are described in the literature (Katz & Shapiro, 1985; Church & Gandal, 1992) as either direct or indirect network effects. Direct and indirect network effects are related to benefit sources for network participants (Page & Lopatka, 1999). The difference between direct and indirect network effects is related to the source of benefit for network participants (Page & Lopatka, 1999). The direct network effect exists when the quality of the product is directly related to the number of other consumers of the same product. This means that the utility function of one individual is not independent of another individual's consumption decision. Instead, the benefits are greater when others are in the same network (Birke, 2009). Telecommunications networks are a prime example of the direct network effect. Purchasing communication services (telephone, fax, internet access) directly benefits existing subscribers. Existing communication subscribers are added because each subscriber eventually wants to communicate with all other subscribers (Economides, 1996).

Asymmetric information, also known as "information failure", is a market failure that occurs when one party, in an economic transaction, possesses greater material knowledge than the other party (Bloomenthal,2021). The idea of asymmetric information was first introduced by Akerlof (1973) in his work The Market of "Lemons": Quality Uncertainty and the Market Mechanism. The author explained the concept of asymmetric information in the automobile market theorizing that an unequal knowledge of information between sellers and buyers could have led to inefficient results in some markets. After that, academics and researchers agreed on the fact that information asymmetries can lead to market failures. Carwell (1996) argued that asymmetric information is responsible for obstructing purchaser rational decision-making and barring utility maximization. From another perspective, Teisl & Roe (1998) affirmed that an absence and incompleteness of information make the consumers concerned about the genuine attributes of products or services, resulting in final choices that aren't aligned with their actual preferences. What aggravates asymmetric information is the access and availability of information (Aoun Barakat & Sayegh,2021). Hobbs (2004) stated that another influencing element in this situation is the high cost of searching and seeking credible adequate and trustworthy information, which results in the customer abandoning his search and often taking an irrational decision. In this case, the customer's ability and willingness to process information are hampered by limited financial and analytical resources, leading to information asymmetry (Verbeke,2005). Finally, information asymmetry concerns appear to be particularly important if one party lacks information about the quality of another party, or if one party has concerns behavioural tendencies of other parties (Stiglitz,2000). Information asymmetry is more likely to harm users due to the fact that they do not have information about the value of their data in transactions between digital platforms and advertisers or data brokers (Economides & Lianos,2021)

The economic value of personal data

The majority of empirical studies regarding privacy have tried to point out how much individuals would value privacy in economic terms, and often focusing on people's willingness to trade private information for monetary payments. Huberman et al. (2005) studied how much money individuals would request in order to disclose some personal information (such as weight and height) by using a second-price auction. Wathieu & Friedman (2007) conducted a survey showing how participants were serene in sharing their personal information with institutions if the economic benefit of that would have been shown to them. Cvreck et al. (2006) have found how much money EU citizens would accept to disclose their mobile phone location data depending on which country they live in. According to Malgieri et al. (2017), individuals do not seem to be fully aware of the monetary value of their personal data and tend to *underestimate* their economic power within the data-driven economy and passively succumb to the monetization of their digital identity.

Many studies and empirical experiments have been conducted with the aim of quantitatively clarifying the economic value of online personal data that we provide in exchange for a service from companies operating in the data-driven sectors. Some have focused on the value of individual personal details of users, for example, Steel et al. (2013) published a series of articles for the Financial

Times about consumer data, and subsequently developed an interactive digital tool through which readers could determine a price for their personal data based on pricing benchmarks supplied by a data broker. In this context, as the authors stated: “the more specific information available about an individual, the more valuable it is to marketers”. Others focused on the relationship between the revenues and profits of companies compared with the number of users and the amount of information available to them. Beavisage and Mellet (2019) carried a research trying to provide a solid basis for analysing how personal data are constructed as economic intangible assets in industrial and market processes, in their article ‘Datassets: Assetizing and Marketizing Personal Data’, they investigate the ‘assettization’ of users data as a capital for companies operating in the data-driven economy. McDonald & Cranor (2008) tried to determine the economic value of personal data based on the time spent by users reading privacy policies.

Reducing Asymmetric Information in online platforms

When the asymmetry of information breaks down, markets operate more efficiently (Russel,2015). A strand of literature has focused on how to reduce and alleviate Information Asymmetry in the context of online platforms. Freedman & Zhe Gin (2011) have analysed the peer-to-peer (P2P) lending platform Prosper.com, showing that learning by doing can play an important role in alleviating the information asymmetry between market players. Thierer et al., (2016) considered Sharing Economy as a sector where asymmetric information could be considered not as a market failure, but as a market opportunity instead. Lewis & Gregory (2011), considering the platform eBay Motors for online used goods actions, found out that the voluntary disclosure of sellers’ private information protects the buyers from adverse selection and asymmetric information. Russel (2015) suggests that online platforms could switch to a regime of paying users for their data, which could lead to the emergence of a non-exclusive licensing market for user data when users opt-in to sharing their data with particular platforms. This will allow users to port their data to the platform to provide higher revenue and better terms for their privacy assessment. Over the last decade, "freemium" (a combination of "free" and "premium") has become the dominant business model among internet start-ups and smartphone app developers, enabling users to have free access to basic features and access to additional content for a subscription fee (Kumar,2014). In this case, this research would explore these particular pricing models (freemium/premium) for online platforms, assessing the expected reduction of information asymmetry between users (subscribers) and platforms.

Chapter 1: “Data-driven economy and market failures”

3.1 Introduction to data-driven economy and companies

The Data Economy is a global digital ecosystem where data is collected, organized, shared and valued from the information collected by a network of providers. Data entry is collected from a variety of providers, programs, applications and other sources, including search engines, social media sites, online vendors, physical store vendors, payment gateways, software as service providers, but also numerous companies installing connected devices to IoT (Internet of Things). Data-driven markets are markets where the cost of quality production is decreasing in the amount of machine-generated data about user preferences or characteristics (*user information*) which is an inseparable by-product of using services offered in such markets. Search engines, digital maps, platform markets for hotels, transportation, dating, or video-on-demand, as well as smart electricity meters comprise but some examples. (Prufer & Schottmuller, 2017)

A data-driven enterprise is an organization that has integrated data analysis into the core of its business processes. It uses the insights it derives from this data to transform its business processes. Key characteristics of data-driven enterprises include a focus on automation, continual improvement and optimization, the ability to anticipate internal and external changes, an adaptive mindset, and, most of all, a culture that fully embraces data and its potential. A data-driven company is capable of unifying the data it generates into a comprehensive model, which then serves as the basis for analysis and optimization. This isn't only a matter of IT since it's not the exclusive domain of specialists who are well-accustomed to working with numbers, in fact, every employee within the organization plays an active role in enabling the data-driven enterprise, supporting it, and driving it forward (Accenture, 2015).

Due to continued improvements within the field of data technology and therefore the rise in its usage, new markets have emerged which operate totally on the gathering, analysis, processing, and monetization of private data. While Google and Facebook are the foremost prominent operators of a data-driven business model, the combined weight of all of the smaller actors in the web advertising market shouldn't be underestimated either. Data-drivenness has become ubiquitous on the net, and it's easy to work out how it is often highly beneficial to the consumer: it allows various useful services to be provided for an occasional price and in a personalized manner. Despite such benefits, however, it's not completely without its costs or risks (Van de Waerdt,2020). For companies (and entire economies) to stay growing, the development of a "data-driven economy" is necessary: battered by global competition, commoditisation, and shrinking product/service cycles, corporations seeking to take care of or grow their profit margins will increasingly depend on creating new (and hard-to-emulate) products and services supported by insights derived from the datasets that they own or can gain access to, especially those about their customers (World Economic Forum,2014).

Digital platforms, like Google and Facebook, voraciously collect personal information from their users. This information spans many aspects of users' lives, like location, interests, activities, policy making, and social interactions. Personal information is collected without compensation to the user, aside from providing free internet search by Google or free social networking services by Facebook. Digital markets are plagued by differing types of market failure that will impact their optimal performance with reference to delivering privacy for users. These market failures may be caused by strategies employed by giant digital platforms (Economides & Lianos,2021). The exchange of data often gives rise to structural market failures, both because the investments made by companies for the collection of data on individuals, not internalizing social costs, risk leading to an over-investment in the collection of information, and because, in an environment where transaction costs and uncertainty exist regarding the assignment of ownership rights to data, market forces are likely not to be able to guarantee the achievement of an efficient situation. The possibility is realized that the interests of those who have more technical knowledge and information about the data themselves prevail. The absence of a true market mechanism can only make these relationships incomplete and inefficient. The consumer does not have a clear perception of which data are transferred, of their real value (the price) and of how they are treated, both for primary uses and, even more so, for secondary ones. This is a one-off transaction involving other assets (for instance, an APP), in the face of the dynamic use of user data. It is, therefore, the very structural configuration of the market and related transactions that are distorted and, consequently, lead to incomplete markets, which inevitably produce inefficient and unbalanced results (AGCOM,2018).

3.2 Network Effects

The term Network Effect refers to any situation during which the worth of a product, service, or platform depends on the number of buyers, sellers, or users who leverage it. Typically, the greater the amount of buyers, sellers, or users, the greater the network effect—and the greater the worth created by the offering. "The willingness to pay, for a buyer, increases as the number of buyers or sellers for the business grows" - Professor Bharat Anand, Harvard Business School. There are multiple Network

Effects, the economic literature on network effects made a distinction between direct and indirect network effects, while both arise when the utility that one consumer derives from the consumption of a service (or a good) increases with the number of other consumers purchasing the same product (Graef,2016).

• **Direct network effects** occur when the worth of a product, service, or platform increases just because the quantity of users increases, causing the network itself to grow. Social media platforms primarily have the benefit of direct network effects because the service's value grows as a consequent result of attracting more users. A commonly used example is the telephone network, since the utility of a consumer derives from purchasing a telephone may be a function of the amount of others that have already joined the phone network (Graef,2016) Apple, for example, benefited from direct network effects. The preferential treatment of messages sent from an iPhone to a different Apple device (through iMessage) has helped the corporate expand its moat within the market (Strobieski,2020).

• **Indirect network effects**, instead, occur when a platform or service depends on two or more user groups, like producers and consumers, buyers and sellers, or users and developers. As more people from one group join the platform, the others receive a greater value amount. E-commerce and ridesharing are two good business examples. (Strobieski,2020)

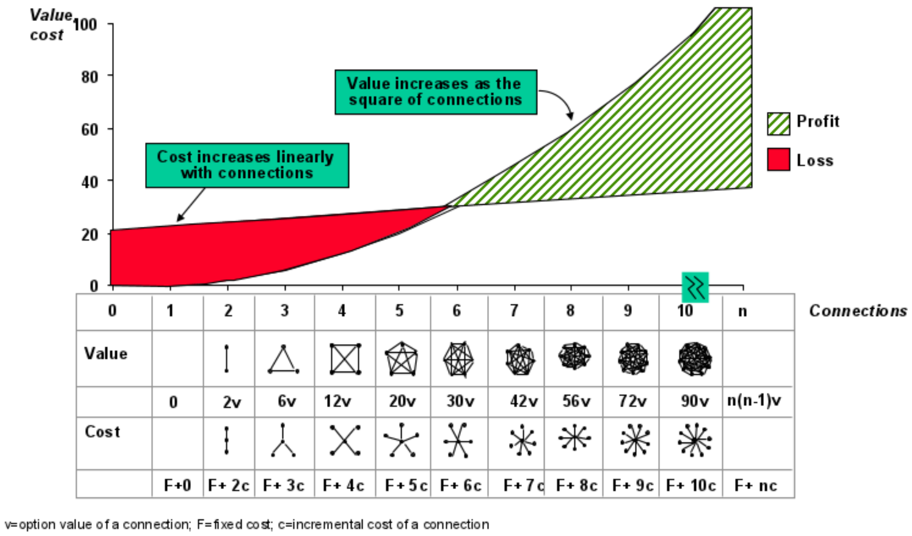

Figure 1: Network Effects. Source: Jorgenson (2020)

To give a visual representation of how network effects work, the above graph (Jorgenson,2020) explains which is the main strength of Network Effects: the fact that the *Value* (for the company) increases exponentially, while *Cost* increases linearly. We can notice that the Value increases as the square of connections, so the relationship will be n(n-1)v, while cost increases linearly with connections. Hence, the cost element will be described by F + nc, where F refers to *Fixed Cost* and c represents the *incremental cost of a connection*. From the figure, we can see that the cost of maintaining the network doesn't increase as fast as the value of the network. This implies that, in the long run, in an industry where Network Effects play a fundamental role, there will tend to be fewer actors and they will continue to expand considerably. In other words, potentially, such firms enjoy perpetual growing revenues with slowly growing, if not diminishing, marginal costs.

In the industrial era, giant monopolies were created by supply economies of scale. These have been driven by production efficiency, which reduces the unit price for creating a product or service as production increases. These supply economies of scale could give the largest company in an industrial economy a cost advantage that was extremely difficult for competitors to overcome. Nowadays, with the advent of the Internet, comparable monopolies are being formed by demand economies of scale. Unlike supply economies of scale, demand-side economies of scale take advantage of demand-side technological improvements. Demand economies of scale are driven by efficiencies in social networks, demand aggregation, app development, and other phenomena that make bigger networks more valuable to their users (Kumar,2018). Another difference between the two eras, concerning the objectives of a monopoly company, is that in the industrial era, the company would have tried to reach the maximization of the profit margin from its product, while, in the Internet era, the company aims to maximize its market share.

Many of today's most popular companies and startups are strongly influenced by network effects: eBay, Etsy, Amazon, Alibaba in the E-commerce sector. Ticket exchange: StubHub, Ticketmaster, SeatGeek; ride share: Uber, Lyft; Delivery: Grubhub, DoorDash, Uber Eats, Instacart, Postmates; social media: Facebook, Twitter, Instagram, LinkedIn, Snapchat, Pinterest. What all these companies have in common is that as they grow in size and gain more users, they are more valuable to their customers. Etsy and eBay offer significantly more value to users if the platform is used by 1 million sellers instead of 1000. Uber and Lyft offer riders more convenience and reliability as more drivers join the platform. When it comes to social media websites, users find that the channels are more interesting and diverse as more people sign up (Strobieski,2020).

While network effects are beneficial to consumers within the short term by increasing consumption utility, they also make it easier for undertakings to realize a dominant position and to bolster barriers to entry which can have negative effects on competition and innovation in future (Graef,2016). The underlying principles of network effects imply that a company, website, or platform with the highest market share will perform better and succeed in the long run. This means that its market share is expected to grow faster**.** For this reason, markets where network effects play an important role are often cited as "**Winner-takes-all markets**". In fact, companies that are ready to exploit network effects often experience fast rates of growth. For instance, when a firm "wins" in a particular country, it tends to win big and overcome the competition. Another element that boosts the growth of a company benefiting from network effects is the so-called "**First-mover advantage**", which is generally defined as the ability of a company to outperform its competitors by entering the market first for a new product category (HBR,2014). First movers into a market benefit from network effects, learning, size, and access. (Investopedia,2020)

Facebook, which has become a social media giant, provides a practical example of an effective application of the two phenomena described above. Launched in 2004, as a free social platform, it has managed to become more and more popular and conquer a higher percentage of market share, leading to the decline of its main competitor, Myspace (Graef, 2016). Ten years later, in February 2014, Facebook buys WhatsApp. The European Commission, in the context of reviewing the acquisition, stated that "the existence of network effects as such does not a priori indicate a competition problem", but, at the same time, these effects may "raise competition concerns in particular if they allow the merged entity to foreclose competitors and make more difficult for competing providers to expand their customer base "(European Commission, 2016). In this case, the "winner-takes-all market" effect was already starting with a series of acquisitions of competitors-social media such as Instagram in 2012. In 2011, Google has launched Google+ to combat threats from Facebook's social networking site, but then it was shut down in 2019. The first drawback was that Google+ was launched seven years late. By that time, Facebook not only enjoyed the "first-mover" advantage, but also took advantage of the "network effects". In that situation, Google ideally needed to provide Google+ products with features that were more advantageous than Facebook's own borderless information-sharing sales offer (Cauvery Nair, 2015).

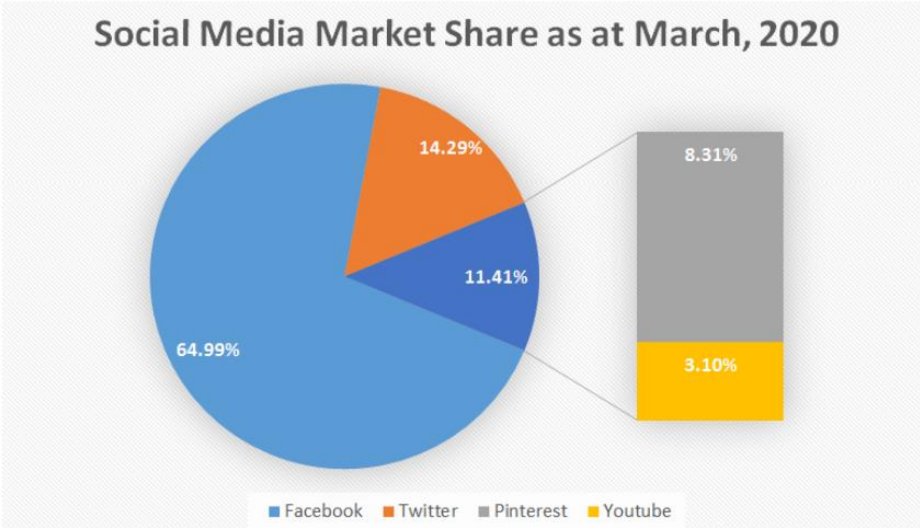

Figure 2: Social media market share 2020. Source: Vitenu-Sackey (2020)

3.3"Zero Marginal Cost" and LRIC

Marginal cost is a cost a company incurs when producing an additional unit of a good. The curve of Marginal Cost is U-shaped. This is because, initially, as a company increases its output, not only variable costs but also total costs begin to increase at a declining rate. According to current economic theory, prices should be set at marginal cost (MC). This is to maximize economic welfare without externalities. That's because such prices reflect the costs associated with providing additional units of goods. If the user evaluates an additional unit more than the cost of producing it, then producing that unit is economically efficient and vice versa. Setting the good's price equivalent to MC means that users will continue to buy additional units until it is no longer economical to produce them at that price.

But physical goods and digital goods work differently. When it comes to data-centred companies that operate with intangible assets such as users' personal data and digital goods, this classical model doesn't seem to find a complete application. The economic theory appears to be violated and it's due mainly for two reasons: the first one is that for most of the online platforms provided by these firms (e.g., Facebook, Google, Amazon) the marginal cost of adding an additional user tends to approach zero or is extremely modest compared to the amount of revenue per user. Another reason that could explain this cost reduction of their inputs it's correlated to the fact that most of their content is 'user provided' (e.g., in social networks). This allows most of these companies to sell their products and services for free or at a very low cost-of-goods sold.

Digitization in general and software development are now ensuring that the marginal costs of companies that rely on digital business models are almost zero from the very first product or service (Roland,2018). In fact, for digital products, once the original is created, the marginal cost can actually remain infinitesimal, even with vastly greater usage. It also means that retail prices are very close to zero and can still be profitable. (Kim, 2019) Companies engaged in the network economy need to make large investments to enter the market. Once the initial investment is made, the supplementary cost of creating additional units is reduced and may be negligible. As a result, there is a supply-side economy of scale and, accordingly, the average cost of delivering goods and services decreases as production increases. In economic terms, network economy industries are characterised by relatively high fixed costs and low marginal costs (Graef,2016). The graph below shows the differences in the cost structure elements of a traditional industry monopoly and a digital platform. In the traditional monopoly, marginal cost, according to classical economic theory, has a U shape because initially when a firm increases its output, total costs, as well as variable costs, start to increase at a diminishing rate. At this stage, due to economies of scale and the Law of Diminishing Returns, Marginal Cost falls till it becomes minimum. Then as output rises, the marginal cost increases. This doesn't happen for the digital platform, for the reasons explained above.

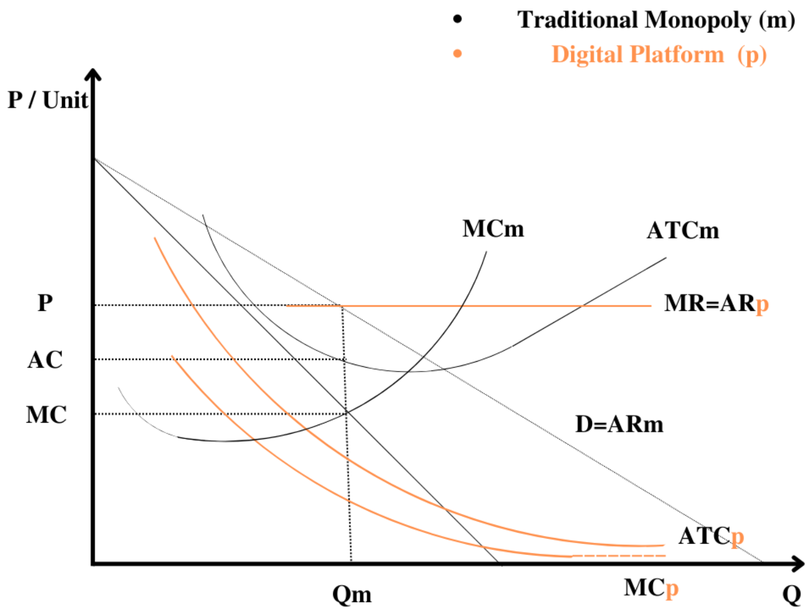

Figure 3: Declining MC for digital platforms. Own Source

In his book "Zero Marginal Cost Society", Jeremy Rifkin (2014) clearly sets out his vision of how, in some businesses, the concept of "zero marginal cost" will increasingly be present in the future. The author begins by explaining that, in a situation of "extreme productivity" and therefore "optimum general welfare", thanks also to technological innovation, the actual production cost of each additional unit (not counting fixed costs) becomes essentially zero, making the product almost free. In this case, the author says, the "lifeblood of capitalism would dry out". This "revolution" would initially affect sectors such as manufacturing with 3D printing and the educational sector with online higher education. Rifkin puts emphasis on the concept of "prosumer", referring to those individuals who are, at the same time, both consumers and producers. In the case of the data-driven economy, companies like Facebook and Google can count on a very high number of users, of which many are "prosumers", in the sense that they produce their information on mobile phones and computers and share it through video, audio and text. So, in this collaborative network world, users provide the content (therefore the product) at a marginal cost close to zero, making sure that these companies can continue to offer their services for free.

The concept of declining marginal costs for companies operating in the data-driven economy can be *quantitively* estimated through a comparison between the cost structure of a data-driven company with that of a traditional industrial company. Roland (2018) conducted a calculation to compare costs and revenues between two companies in the digital world (Google and Facebook) and a German car company (Volkswagen). Considering Google, which is a search engine, the costs it has to bear to offer its service are energy costs and those to process the search query. Transmission costs are not included as they are borne by the user. Converting the energy consumed to process a search query

(0.0003 KW / h) into energy price (0.13 € in Germany, 2018), the estimated cost for a single search query is € 0.00039. Comparing the energy cost with the potential earnings per single search query (about 2 euro cents, assuming a click rate of 1.68% and an average revenue of 1.20 € on Google Ad), the turnover of a single search query exceeds its costs by a factor of more than 500. Now considering the cost structure of a generic car manufacturer, the cost of materials (included in variable costs, together with energy costs) reaches 50% of the selling price. The ratio between the profit on a single car and its production costs is around the value of 2, which is 250 times lower than what we have seen for a search engine like Google.

Another example that gives an idea of a "zero marginal cost" business model works, is the one concerning how much an additional customer costs Facebook to process and support. As for the Google example, the answer is still almost nothing. It should be noted that the costs for the transfer and operation of servers add up over the course of a year and that heavy users with high volumes of upload data cost the company more than occasional consumers. However, the additional energy and server costs have no relation to the marketing revenue generated by an individual user on average. In order to better illustrate Facebook's revenues-costs relationship, Roland divided the advertising revenues (2017: USD 10.1 billion) by the number of regular users (2.07 billion monthly users,2017), the result is just under EUR 5 per year. Even if that doesn't sound like a lot at first, however, considering Facebook's "zero marginal cost" business model, this social network is still lucrative, in fact, in 2017 its profit reported was around USD 5 billion.

Although the concept of "Zero marginal cost" or "Decreasing Marginal costs" is an economic phenomenon in which data-driven companies provide a more truthful and likely representation, in practice, their cost structure is not centred entirely on marginal costs, but on Long Run Incremental Costs (LRIC). Long Run Incremental Cost (LRIC) is the "long-run cost of providing the next *increment* of output, which should be measured on a forward-looking basis" (Den Hartog,2021). A LRIC model is often used in telecommunications regulation to determine the price paid by competitors for services provided by an operator with significant market power, usually the incumbent (former monopoly). In theory, there is a huge number of different-sized increments to measure. However, we can group these increments into three different categories: a small change in the volume of a particular service; the addition of a whole service and finally the addition of a whole group of services.

The existing literature about the LRIC has focused mostly on the telecommunications sector but not on the new players in the digital economy. Here, the attention to the marginal incremental costs addresses the addition of storage spaces for the collection and the analysis of users' personal data and digital content (videos, photos, personal information, interests, interactions etc...) provided by the latter. In this situation, a company like Facebook will make an estimate of an average cost for welcoming and organizing a new 'block' of users, according to estimates and analysis of the average number of new users subscribed to the platform in a given period of time (for example, a semester).

3.4 Asymmetric Information

3.4.1 Explanation of this market failure and how it manifests in the big-data economy

In the economic theory, Asymmetric information, also known as *information failure*, is a market failure that occurs when one part of an economic transaction possesses greater information regarding the object of the transaction than the other party. This normally happens when the seller of a specific good or service has more knowledge about it than the buyer.

Akerlof, for his paper in 1970 *The Market for Lemons: Quality Uncertainty and the Market Mechanism*, was awarded the Nobel Prize in Economics. In his work, by analysing the used car market, he assumed that any complete and available information for a car is one-sided, suggesting that such information can only be available to the seller. Hence, the buyer wouldn't be able to identify if a car is a high-quality one ("a peach") or a defective one ("a lemon"). This phenomenon prevents certain mutually beneficial exchanges from taking place. Building on Akerlof's lemons problem, others have argued that these information asymmetries lead to distortions in people's behaviours, and to the extent that parties are misinformed or uninformed, they are less likely to be able to behave in accord with their true preferences, and hence the market fails (Thierer et al., 2016).

According to Economides & Lianos (2021), the Big Data economy stimulates the information asymmetry market failure, principally for two reasons. The first concerns the concept of "asset surveillance". For example, the numerous devices (e.g., surveillance cameras, smartphones, digital assistants, smart energy consumption monitors) that a citizen could have in his house, connected to each other through the IoT, monitor various assets such as property, cars, bodies and even relationships. Such asset monitoring creates new and valuable opportunities for the monetization of big data. It also suggests that some companies and parties can do "asset surveillance", while others cannot. The second reason concerns users of online platforms. In fact, by persuading users to give away their personal data for free, digital platforms have fostered the asymmetric information market failure, which benefits firms to the disadvantage of their users. Users are harmed by asymmetric information also because they don't know the value of their personal information for digital platforms or data brokers, which collect them as they know exactly and well which is the value of them in transactions with extern infomediaries and advertisers. In addition, users may tend to underestimate the value of their privacy. In simpler words, most users have very little knowledge or awareness of the true costs of 'free' access to many digital services, while the service providers very much do, mainly because the firms know that the actual paying customers are not the users providing the content.

From the economic point of view, the state of privacy arises with asymmetric distribution of personal data between market participants, where one side privately holds personal information. Privacy is therefore a relationship of asymmetric distribution of personal data between market players (ENISA,2012). BDA (Big Data Analytics) aggravates information asymmetry as it allows a powerful few to access and use a knowledge that the majority, namely consumers, do not really have access to. The choices of an individual regarding the transfer of their data in order to obtain a service are directed according to the balance between benefits, often immediate (e.g., access to a service) and costs (often uncertain and not known). In this context, the information asymmetry between users and operators is pervasive and structural: not only does the consumer not have all the information he would need to make an informed choice, but many of the behaviours, to be efficient, would require a degree of technical knowledge that goes far beyond the competences widespread among the population (Barakat & Sayegh,2021). A greater degree of transparency is often useless where consumers are unable, due to a structural technological knowledge gap, to understand this information. Furthermore, choices such as those relating to the transfer of one's data are made very frequently on impulse and without an assessment of the real consequences of the implicit exchange (AGCOM,2018).

3.4.2 Possible solutions to reduce information asymmetry online

In the literature regarding information asymmetry, the possible solutions for a reduction of this phenomenon report historical examples as this problem has always been present in most businesses that have ever existed. For instance, Thierer et al. (2016), explained that in the 11th century, a group of traders operating in the Mediterranean (Maghribi), tried to overcome the problem of asymmetric information with the use of trust and reputation mechanisms. Instead of travelling themselves, they used agents to achieve more efficiency. Since it was likely that sometimes some agents would abscond with capital or cheat the other merchants and escape away safely to countries with different legal systems and laws, the Maghribi built up a reputational mechanism that allows them to condition future employment based on past behaviour and banish those who have cheated by collaboratively refusing the entire dealer network to do business with them anymore**.** These reputation mechanisms worked because they help traders decide who to trade with in the future, relying on the trusted experience of others. In relation to regular transactions, those with a good reputation were rewarded with more transactions, and those with a bad reputation were not.

Nowadays, on the one hand, technology has nourished the information asymmetry in online platforms owned by data-driven companies, on the other hand, it has allowed consumers (or users) to potentially have more tools for greater awareness and knowledge about the information and quality of the product or service they purchase. Below are some examples of which are the possible (partial) solutions and mechanisms that can reduce the information asymmetry between the user and online platforms.

1) **Trust and reputational feedback mechanisms:** Market players use trust and reputational mechanisms to simplify and accelerate transactions. In transactions among impersonal agents, trust becomes a more critical component of cooperation because a buyer’s trust in a seller’s credibility decreases perceived transaction-specific risks, allowing the seller to get price premiums. Reputation facilitates collaboration, acts as an enforcement mechanism, demonstrates credibility or quality, reduces risk, encourages good behaviour, punishes bad behaviour, and helps resolve information asymmetry. Social norms also work to enrich reputation in regulating human behaviour. In combination with online review services and other information-sharing technologies enabled by the Web Sharing Economy, these reputational tools can help create more effective, and largely self-regulating markets which provide more information to a larger number of individuals than before. By opening to new innovations and opportunities while enabling greater trust, these new internet-based mechanisms are committed to revolutionizing interactions in modern markets. This will require a reassessment of traditional regulations designed to address the perceived asymmetric information market failures that are not normally improved (Thierer et al.,2016).

2) **Out-put option:** By default, on digital platforms such as Google and Facebook, users accept personal data and share it with digital platforms that impose a "take it or leave it" contract. The ability of digital platforms to allow users to accept take-it-or-leave-it-in contracts to provide personal data at zero prices is a direct result of their market advantage. There are clearly two markets here. The first one (A) concerns search services (Google) and social networking services (Facebook); the second one (B) refers to a market where it’s possible to sell personal information. These markets, in a competitive world, would operate separately from each other. Users, in a competitive world, would opt-out by default from (B). Otherwise, a user would opt-in, sell his personal information and get compensated by the platform. The amount of compensation, in this case, would depend on the type of information that the user provides to the platform. If the monetary compensation exceeds the value of the loss of privacy implied by the transaction, the user accepts participates in (B) and accepts the offer. The transaction price of the sale of personal data also fluctuates and may be negotiated individually between the platform and the user. In contrast, there is now a market failure where all trades are done at the same

zero price and some trades that were supposed to be competitive are not. The imposition of the take-it-or-leave-it contracts and the default opt-in are what make this market failure exist. The first remedy needed would be to make the “opt-out” option the one by default in the collection of personal data, and sellers would “opt-in” just if they wish so (Economides & Lianos,2021).

3) **Other solutions:** Another intuitive and helpful mechanism in order to reduce asymmetric information between consumers and competitors is to operate as monitors for each other. Underwriters Laboratory, Consumer Reports, notaries public, and online review services (e.g., Yelp) decrease bridge gaps in information. eBay and Amazon seller ratings, Uber driver reviews, and product ratings are all relevant examples of crowdsourcing reputation. Moreover, online reputation management (ORM) software solutions allow companies to track what consumers say about a brand on review sites, social media, and search engines (Investopedia, 2021).

4) **Show the monetary value of personal information:** The traditional passive approach to privacy protection only protects data based on personal/emotional (qualitative) values. To reduce the asymmetry of information in the big data era and make individuals more powerful players in this data-driven economy, it is necessary to provide additional information about the monetary (quantitative) value of their personal data (Malgieri & Custers, 2017)

The fourth mechanism, which concerns showing users of online platforms what the economic value of their personal data is, involves a turning point in the way the user interfaces with the digital platform. The user would then pass from a ‘passive’ to ‘active’ consumer behaviour, in the way of using applications and websites that require a provision of personal data to access their services. However, this requires the various data-driven companies, the laws regarding the processing of personal data, governments and users to clarify: what is personal data collected online, how it can be estimated, its quantitative value and in what how a user could bargain with the platform in the trade-off between privacy and services.

4. Chapter 2: “The economic value of personal data online”

4.1 What is personal data and how it is utilized by data-driven companies?

Personal data are defined as “any information relating to an identified or identifiable natural person” (Article 4(1) GDPR). When it comes to the digital world, personal data include information such as internet protocol (IP) addresses (unique identifiers that can be used to identify the owner of devices connected to the internet) and data from ‘smart meters’ monitoring energy usage by addresses linked to identifiable persons. Some special categories of data (*sensitive data*) are subject to stringent data-protection safeguards. They include “personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation” (Article 9(1) GDPR).

In modern information economies, the cost of storing information has declined, so that nowadays it is easier to collect, store and analyse more information about that person. Data aggregators link information from different sources into individual profiles. Some day-to-day activities can be tracked through information technology. Small pieces of personal data are placed in a database where records can be linked and tracked to create a complete document of an individual's life. For example, a company records the details of every customer transaction, and a website logs visitor behaviour. This, sometimes, may happen without the consent or knowledge of the person. In addition, hundreds of millions of people around the world may be willing to send very personal information to friends and strangers via Web 2.0 technologies (e.g., blogs and online social networks). When choosing a balance between sharing and hiding personal information (and when choosing between the exploitation and protection of personal data), individuals and organizations face complex, sometimes intangible, and sometimes ambiguous trade-offs. Individuals want to shield the safety of their data and avoid the misuse of the data they pass to other entities. However, they also like sharing, with peers and third parties, information that produces the most mutually satisfactory interactions possible (Acquisti et al.,2016).

4.2 Why it is important to understand the economic value of personal data?

Scholars and institutions have wondered why it is important to know the economic value of personal data, both from the point of view of users and from the point of view of market efficiency.

Nowadays, individuals are sharing massive amounts of personal and others' data to large data-centred companies with little or no thought to its potential monetary value and, as a result, those firms make significant profits also because their cost of "materials" is fundamentally zero. Facebook users, for instance, provide it with personal data that have the capacity to create immense long-term value for the company, in return, they are provided with a "free" service, but the transaction is absolutely asymmetrical (World Economic Forum,2014). Malgieri & Custers (2017) notes that if the "price" of personal data is shown to individuals, they may acquire higher consciousness and awareness about their power in the digital market and thus be empowered for the protection of their data privacy. Acquisti et al. (2013) observed that understanding what value individuals assign to the protection of their personal data is of huge importance to businesses, the legal community, and policymakers. It is important to businesses because, by estimating how much customers value the protection of their personal data, managers can distinguish and predict which privacy-enhancing initiatives may become sources of competitive advantage and which intrusive initiatives may trigger hostile reactions. It is important to legal scholars and practitioners because privacy is an issue that has become increasingly prominent in the law in recent years, partly due to the emergence of new technologies and online platforms, such as tracking by global positioning systems (GPS) and social networks. Finally, individual's valuations of privacy are important to policymakers because they are often required to choose between policies that trade off privacy for more necessary goals.

4.3 How to calculate the economic value of personal data?

After understanding what personal data is, the process of assessing the economic value of personal information continues with a subsequent question. *How* can we assess the economic value of personal data? Below are some mechanisms and factors to consider in answering this question.

According to Malgieri & Custers (2017), conferring monetary value to personal data requires some clarity on: 1. "How to express monetary value", 2. "Pricing factors", and 3. the actual "Pricing systems".

1. Expressing value: How to express the value of personal data. Logically, expressing the monetary value of data in a currency such as the euro or dollar makes the most sense. However, there are some issues with this approach. The primary issue is that personal information can change, become outdated and may lose some of its value. Since companies always require updated data, makes sense to state that personal data are a "dynamic product" rather than a static one. Therefore, data's value should be expressed in terms of euros or dollars *per month*, instead of just in euros or dollars. Another issue is the fact that personal data can be reused, and the number of re-using times determine its value. For a company which collects data, it may make more sense to express personal data's value in terms of euros or dollars *per person*.

2. Pricing factors: Which object is really being priced when pricing privacy or pricing personal data and related pricing factors. Single attributes with no further context don't have any price, but only if combined. Therefore, evaluating the economic value of personal data is not a matter of single attributes, but rather it concerns the pricing of combinations of single attributes or single attributes of several people. Thus, pricing personal data is about datasets and not about single data. In fact, the dimensions and the completeness of datasets are important factors in determining the personal data's price. In summary, it makes more sense to price datasets (i.e., digital identities or digital profiles) instead of pricing single individual attributes.

3. Pricing systems: OECD (2013) conducted a survey on methodologies for measuring the cost of private data, distinguishing methods that are constructed market valuation, and methods that are built on some individual's valuation. When individuals disclose their data, they suffer an objective loss of privacy in terms of greater exposure to discrimination (including price discrimination) and knowledge asymmetry, which can yield commercial vulnerability. These pricing methods can be utilized in the context of the right to acknowledge the worth of one's personal data by applying so-called "reverse liability". This suggests calculating a compensation that a possible infringer (e.g., a company or a data controller) pays ex-ante so as to be allowed to perform a probably harmful activity (e.g., processing personal data).

More specifically, the valuation methodologies presented by the OECD's study on the ways in which data might be valued in the market are: determining the market capitalizations of firms with business models predicated on personal data; ascertaining the revenues or net income per data record; establishing the market prices at which personal data are offered or sold; establishing the economic cost of a data breach; determining prices for personal data in illegal markets; reviewing economic experiments and surveys that attempt to establish the price companies would need to pay for individuals to give up some of their personal information; and ascertaining how much individuals would be willing to pay to protect their data (World Economic Forum,2014).

4.4 Quantitative calculations about the economic value of personal data

The following are studies and calculations, carried out by academics and professionals in the field, on the quantification of the economic value of the various elements that make up the personal information of individuals online.

According to a study conducted by Malgieri & Clusters (2017) general information about a person, such as age, gender and location are worth a mere 0.05 cent. People who are purchasing a car, a financial product or a vacation are more valuable to companies that want to sell those goods. Personal data from car buyers is worth about 0.21 cents per person. Personal data of people going through certain life events, such as becoming a parent, moving, being engaged or getting divorced, also prompt companies to pay more for them. For instance, personal data of a pregnant woman are worth about 11 cents. Personal data containing specific health conditions or information on taking certain prescriptions are worth about 26 cents per person. But, by adding up all these details, for most people the total is less than $ 1. The authors hypothesized that, with the exponential development of the data economy, the prices of personal data have fallen accordingly. For example, a zip code in the US cost 50 cents in 2006 and 0.05 cents in 2013. This was not only due to lower costs of data collection, but also due to a significant increase in the use of personal data for profiling and marketing. Finally, the authors remark that personal data has become ubiquitous, in particular in the U.S. where data can be freely traded, driving down prices.

In the paper "Datasset: Assettizing and Marketizing Personal Data" (Beauvisage & Mellet,2019), the authors provide a summary of the literature attempt to assess the economic value of personal data. In this case, the study analyses surveys led by private companies aimed at estimating the market worth of personal data. These economic research share the same approach: immerging Internet users in a fictitious situation and weighing the cost of disclosing personal information about themselves (age, internet history, income) against the financial benefits (discounts or monetary compensations). Therefore, the value of data is defined as the price (or equivalent in terms of service) that an individual agrees to when disclosing certain personal information. The purpose of these initial assessments was to identify the data that people place the highest priority on protection, as well as the third parties that sell it and their prices. It's also a way to estimate the cost of privacy from a user's perspective.

Table 1 : Summary of Studies on the Worth of Personal Data Item for Individuals

Source	Average valuations
Huberman et al. (2005)	Age = \$57 Weight = \$74
Danezis et al. (2005)	Location ~ £27
Beresford et al. (2012)	Favourite colour and year of birth = €1
Carrascal et al. (2013)	"Offline" information (age, address, economic status) ~ €25 Browsing history = €7 Interactions on social networks = €12 Search history = €2 Shopping = €5
Staiano et al. (2014)	Geolocation = €17 to €588 Communications = €3.40 to 51
Havas Média (2014)[1]	30% sell in the top band, i.e. "More than €500"
Orange (2014)[2]	Each piece of information ~ €15 (name, mobile number, children's ages, income, purchase history, contacts, etc.) Average total value = €170

Figure 4: Summary of previous studies on personal data value. Source: Beauvisage & Mellet (2019)

The study carried by Carrascal et al. (2013) provides a good example. The experiment was conducted on a plugin developed for Internet browsers, which, in real time, asked to a set of participants (n=168) at what price they would have sold information about their behaviour and practices online. The 'Offline' information (age, economic status, address) were estimated averagely higher than the online ones, at about €25. The average price required by participants for the transfer of their browsing history was €7, with variations depending on the type of information (Interactions on social networks=12€; visits to financial sites=15.50€; search history=2€; shopping history=5€).

In June 2013, *The Financial Times* (Steel et al. 2013) assessed the value of typical personal data (social demographics, wealth, entertainment preferences, consumption patterns) using list prices from US data brokers. Summing up the value of each piece of information, the average total value was about 20 cents. Olejnik et al. (2014) conducted an empirical study reporting purchase prices of an online advertising auction systems (known as "RTB" or real-time bidding). The author states that, to a single ad impression, browsing history is worth only about \$ 0.0005. Applying this price to the average number of websites visited and the ads displayed per website, the author estimates the generated business volume to be \$ 0.18 per user per month. The potential revenue for web users in this market is approximately \$ 0.432 per year.

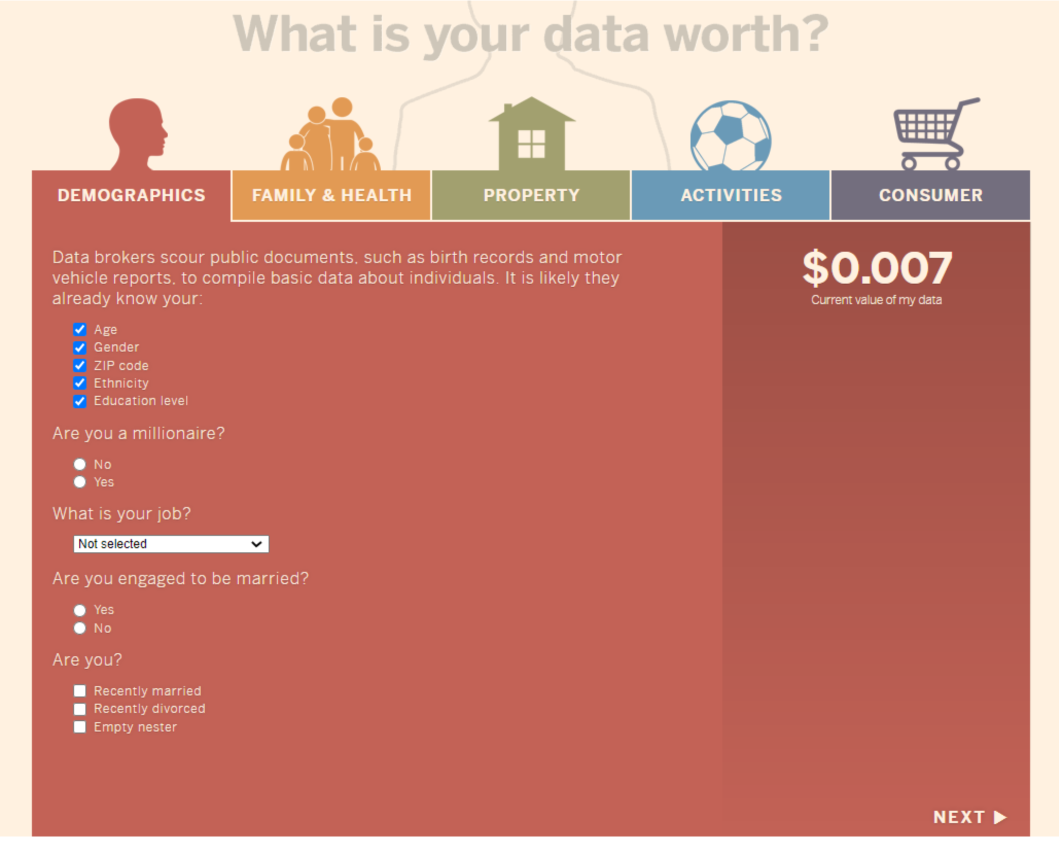

Figure 5: Financial Times's tool to estimate the value of personal data. Source: The Financial Times (2013)

The value of personal information fluctuates primarily based on supply and demand. For instance, men's data tend to be priced a bit higher than women's, as there are more women on the planet than men (Invisibly, 2021). According to survey by MacKeeper and YouGov(2020), personal data for ages 18-24 is valued significantly higher than any other demographic group, while companies spend a significant amount of money on personal data from African and Middle Eastern audiences. The same disparities exist between regions, economic backgrounds, and even education levels, depending on how a company evaluates each demographic group.

Demographic		Cost for Data Per Person	Percentage of Population
Sex Assignment	Male	$0.15	48.59%
	Female	$0.14	51.41%
Age	Age 18-24	$0.36	11.92%
	Age 55+	$0.05	32.33%
Ethnicity	Middle Eastern	$0.62	1.21%
	Hispanic	$0.01	8.09%
Family Annual Income	$40,000-$49,999	$0.02	4.94%
	$120,000-$149,999	$0.33	1.84%

Figure 6: Cost of personal data. Source: Invisibly (2021)

Another argument pointed out by this study is that how data is used and evaluated depends primarily on the collector's ambitions. Platforms (Twitter, LinkedIn, or YouTube) which provide users with a space to create their own content, use data as a bargaining chip to support their ad sales. Companies are willing to invest billions in advertising because collected data sets create a clearer picture of where to place those investments. Criminals steal personal data for different reasons (blackmail, identity theft, extortion) but the most common is to sell that information to anyone willing to pay. This group often includes other thieves, data brokers, criminal organizations, and even governments. Focusing on the evaluation of personal data by criminals, as of May 2021, hacked Facebook accounts on the dark web could be accessed for $ 65 (or equivalent in cryptocurrency); have access to an entire U.S. voters database for $ 100 and also buy a hacked and verified Coinbase account for $ 610. By comparison, brands could value someone's personal email address for $ 89 if they protect that information with security tools. Health records have proven to be of particular value to data thieves, as health records often contain a more complete collection of patient identity, background, and personally identifiable information (PII). A single Social Security number can cost $ 0.53, but a complete health record sells for an average of $ 250(see Appendix A). The more complete the dataset for the criminal, the more potential value the criminal can derive from that dataset. As a result, health care violations in 2020 increased by 55%.

5. Chapter 3: "Freemium pricing model to reduce asymmetric information in online platforms"

5.1 What is a "Freemium" pricing model in online platforms?

Freemium pricing (a hybrid of the words "free" and "premium") is a pricing strategy used by businesses when they want to offer their customers paid options in addition to free services. In general, the free option is the basic version of the service, and the paid option is the upgraded or premium version. (Patriot, 2021) Over the last decade, "freemium" has emerged as the dominant business model among Internet startups and smartphone app developers. There are several factors that make the freemium strategy attractive. Because free features are a powerful marketing tool, this model allows new companies to expand and attract their user base without spending resources on costly advertising campaigns or traditional sales force. Monthly subscription fees that are normally charged have proven to be a more sustainable source of revenue than the advertising model that was prevalent in online businesses in the early 2000s. Freemium is also more successful than 30-day free trials and other limited-time offers. This is because customers are skeptical of the tedious cancellation process and find permanent free access more attractive (Kumar, 2014).

5.2 "Freemium" and asymmetric information

In a market with information asymmetries, letting customers try a service or a product is an efficient way of reducing uncertainty regarding the quality of the product and increasing purchases (Bawa and Shoemaker, 2004; Cheng and Tang, 2010; Liu et al., 2014).

Previous literature has examined the effect of the "freemium" model from two perspectives. One stream has studied the impact of free trials on demand for premium products or services (Bawa and Shoemaker, 2004; Gallaugher and Wang, 2002; Jiang and Sarkar, 2009; Liu et al., 2014; Parker and Alstyne, 2005). Studies show that a free trial has the effect of building a brand (Bawa and Shoemaker, 2004), speeding up purchases (Jiang and Sarkar, 2009) and boosting sales of the paid product (Liu et al., 2014). Cheng and Tang (2010) examined cannibalism and network effects triggered by free trials and found that it is beneficial to offer free trials when network effects are strong. The other stream of literature deals with why consumers accept to pay for products or services after the trial period. Kim and Morris (2007) showed that consumers who experienced a trial of low-involvement products or services tend to have a better emotional and cognitive response to the product. Wagner et al. (2014) and Helm et al. (2009) empirically proved that the quality of free trials is a factor which notably affects its advertising effect, moreover that high-value free trials entail a greater evaluation of the premium product and thus motivate consumers' willingness to pay for the premium services or products. Apart from one study, almost no scholar or researcher has investigated the impact of the "Freemium" pricing model on online platforms and how it can reduce asymmetric information between the platform and the users. Hu (2019) has investigated how, in the mobile healthcare sector (healthcare apps), freemium strategies affect the demand for premium services. In the healthcare mobile sector, there is a high level of information asymmetry between the buyer(patient) and the seller(physician). The author found out that there is a positive correlation between physicians' free services provision and demand for paid services.

5.3 LinkedIn & Tinder examples

In this context, it is interesting to analyse two online platforms (LinkedIn and Tinder) that use the "Freemium" pricing model to see what information the user can access by paying, and so subscribing to the platform. Information that, otherwise, would be asymmetrically distributed between the platform and the user (to the sole benefit of the platform).

LinkedIn: LinkedIn is a social network web service, free of charge (with optional paid services), mainly employed in the development of professional contacts (through publication and dissemination of their curriculum vitae) and the dissemination of specific labour market content (e.g., job search engine, company advertising, etc.). LinkedIn users are job seekers, companies, entrepreneurs, sellers and recruiters. For a job seeker, being able to know which companies or recruiters have consulted their profile is essential information for efficient and qualitative use of the platform. This would allow him to contact the interested party directly and (likely) achieve his purpose. This information asymmetry between user and platform, whatever the type of user, can be overcome through a subscription that, among other functions, allows the user to see who has visited his profile. This is not the only element hidden by LinkedIn in its free version, others are, for example, salary insights, comparisons between applicants for the same jobs, career insight and advanced search options. The subscriptions vary on the type of user and cost from 29.99$ per month (Career), up to 79.99$(Sales).

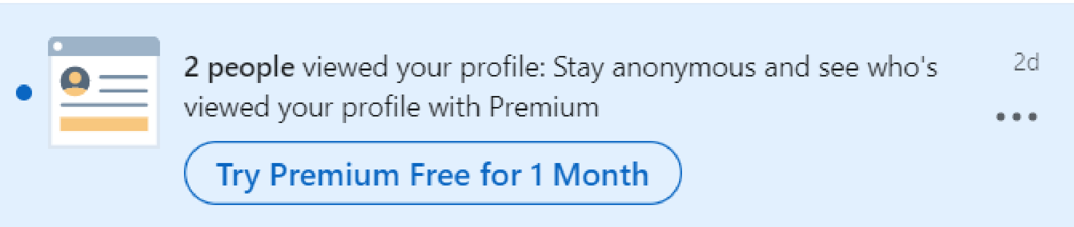

Figure 7: LinkedIn notification. Own source

Tinder: Tinder is an online (mobile)dating app. Tinder uses a "double opt-in" system, and both users need to like each other (and so "match") before exchanging messages. In Tinder, users can "swipe right" to like or "swipe left" to ignore other users' profiles, which show their photos, a list of their interests, and a short bio. In the free version, the user only knows the number of people who have shown interest in him (swiping to the right) but does not know their identity until he also swipes to the right on the same profile. Who "appreciated" you is the information that every user would like to know, because one would have more information on his side to streamline and speed up the process of matching and, hence, chatting and dating. Hence, using the platform as efficiently as possible. Also in this case, this information asymmetry can be overcome through the Gold or Platinum subscription (the two most expensive subscriptions for this app), paying € 30.99 per month for the Gold version and € 174, respectively, up to a maximum of € 624 per year. The journalist Judith Duportail, in an interview for Le Temps (2019) stated that there is asymmetry of hallucinatory information and Tinder relies on the economy of addiction, the principle of random reward, which acts as hard as cocaine and is inspired by slot machines.

6. Findings and interpretation

In this section are reported the findings of this research for every chapter, followed by a personal interpretation.

Chapter 1 findings: Regarding the network effect, we can distinguish two different nuances of the same phenomenon. One is ‘strategic’ and the other concerns ‘costs’. The ‘strategic’ one entails that network effect has two competitive advantages, often related: the first is the “Winner-takes-all market”, which implies that in the long run, in an industry where Network Effects play a fundamental role, there will tend to be fewer actors and they will continue to expand considerably, experiencing better performance and higher market share. Hence, network effects make it easier for undertakings to realize a dominant position and to bolster barriers to entry; the second advantage, the "First mover advantage", implies that the first data-centred firm that markets a new product or service in a specific sector, benefits from network effects. From the point of view of ‘costs’, the benefit is given by the digital nature of the products and services offered by data-driven companies such as online platforms and social networks. Digitization in general and software development are now ensuring that the marginal costs of companies that rely on digital business models are almost zero from the very first product or service. For most of the online platforms provided by these firms, the marginal cost of adding an additional user tends to approach zero or is extremely modest compared to the amount of revenue per user. For example, the ratio between the profit on the sale of a single car and its production costs is around the value of 2, which is 250 times lower than for a search engine like Google. In essence, for data-driven firms in a network economy, Network Effects make the Value (for the company) increase exponentially (as the square of connections), while the Costs increase linearly. In economic terms, network economy industries are characterised by relatively high fixed costs and low marginal costs. Another reason that could explain this cost reduction of their inputs it’s correlated to the fact that most of their content is ‘user provided’ (e.g., in social networks).

The information asymmetry has strengthened the power of data-driven companies more thanks to an unfair balance of information distribution between user and platform. This has happened for several reasons, for example: the concept of "asset surveillance", since asset monitoring creates new and valuable opportunities for the monetization of big data. Another reason is that by persuading users to give away their personal data for free, digital platforms have benefitted to the disadvantage of their users. Users don’t know the value of their personal information to advertisers or digital platforms, which collect them as they know exactly and well which is the value of them in digital’s platform transactions with extern infomediaries and advertisers. In addition, users may tend to underestimate the value of their privacy, not properly understanding the balance between benefits, often immediate (e.g., access to a service) and costs (often uncertain and not known). Furthermore, BDA (Big Data Analytics) aggravates information asymmetry as it allows a powerful few to access and use a knowledge that the majority, namely consumers, do not really have. Finally, the ability of digital platforms to allow users to accept take-it-or-leave-it-in contracts to provide personal data at zero prices is a direct result of their market advantage.

Personal interpretation

In a context of continuous technological innovation and digital literacy, network effects are very powerful phenomena for companies that are able to manage and control them effectively. Although the consequence is a greater market share for those companies that benefit from it, the question arises as to how and if the beneficiary company is capable of maintaining and managing this major market share in the long term. In fact, the continuous emergence of new online platforms offering digital products or services such as social networks and search engines means that maintaining the connections that trigger network effects will become increasingly complicated as, for instance, migrations of users can occur from one platform to another even just for the introduction of additional small functions. Concerning information asymmetry, nowadays, a truly competitive market that identifies the costs, value and price of a good in order to negotiate a transaction does not exist in much of the mass social media and data search segments. Hence, information asymmetry regarding the economic value of users' personal data is variable for all the players in the data-driven market. When a clear price is not established, representative parameters of supply and demand cannot be defined or observed. The value that some personal information can give to one company is not the same for another, but, at the end of the day, users are always the ones who are less aware and unaware of the prices of the transactions carried out with their personal assets (personal data).

Chapter 2 findings: In order to confer monetary value to personal data requires some clarity on various factors such as: how to express the value of personal data, because personal information can change and become outdated, so we should consider personal data as a dynamic product (value expressed in *dollars per month*); moreover, they can be reused many times, so we could value them as *dollar x person*. Another factor is the pricing factors, in fact, single attributes with no further context don't have any price, but only if combined, so pricing personal data should be about datasets and not about single data. Other ways in which data might be valued in the market are, for instance: establishing the market prices at which personal data are offered or sold; establishing the economic cost of a data breach; determining prices for personal data in illegal markets; reviewing economic experiments and surveys that attempt to establish the price companies would need to pay for individuals to give up some of their personal information; and ascertaining how much individuals would be willing to pay to protect their data.

Personal interpretation

Much of the research on how to evaluate the economic value of personal data has been entrusted directly to the opinion of the users themselves, but at the same time, many users have no idea of the value of that data in transactions between users and platform and between platform and data brokers. It would make sense, as a first step, to create awareness campaigns from neutral institutional entities, regarding the importance of reading contractual documents (regarding the use of personal data) that are accepted by default by users such as Terms & Conditions and the Privacy Policies. Thus, the user would continue to use the platform but would be able to judge which platform offers a better service (based on the ratio quality of the service/quantity and data utilization by the platform).

Chapter 3 findings: Letting customers try a service or a product is an efficient way of reducing uncertainty regarding the quality of the product and increasing purchases. For some platforms or social networks, "Freemium" pricing model allows users to access (after paying for a subscription or a premium price) sensible information that, otherwise, would be asymmetrically distributed between the platform and the user (to the sole benefit of the platform). Analysing LinkedIn and Tinder, it is noted that both platforms, in the free version, retain information that would be of fundamental importance for the user for more effective and faster use of the service. Access to this type of information takes place only after payment of the premium price, therefore a partial rebalancing of the distribution of information between the platform and the user takes place. In other words, the user "buys" part of the information asymmetry in order to take advantage of it.

Personal interpretation

The fact that a "freemium" pricing model in online platforms can reduce the information asymmetry between the user and the platform, however, raises questions about the platform's business ethics. We should ask ourselves if the platform: has decided beforehand to establish what sensible information the user is willing to pay; if it has decided to do so (and so implement the “Premium” option) after an initial period of exclusively free service; assuming an increase in the cost of the subscription, until the reduction of the information asymmetry produces a benefit to the user? How much does he value that information in using the platform, could he give it up after experimenting with both free and paid versions?

7. Discussion and conclusion

This work, in the existing literature, is proposed as the glue of several topics that have been treated here, each for a specific element or sector. So far, every subject has been dealt with individually. The contribution to the existing literature lies precisely in giving a key that connects various elements that are often complementary and logically consequential. For example, the focus on the price model "Freemium" is linked to the concept of information asymmetry, which is taken in a market failure key but contributes to the growth of data-driven companies, an asymmetry that also includes the difference in awareness by users and platforms about the 'economic' and strategic value of personal data. Lastly, this research contributes to adding information and evidence to the existing gap between the classical economic theory and what's currently happening in the data-driven industry. Given the theoretical nuance of this research, some suggestions for future studies concern the practical application of the concepts expressed in specific working contexts or in-depth studies in selected companies. In doing so, the topics covered in this research could give rise to case studies with a much greater applicative value than the descriptive and conceptual ones.

This study addresses issues that touch on various elements of the digital business of data-driven companies. These factors, such as the use and marketing of personal data, pose questions about issues such as Business Ethics and Privacy laws. From this study, policymakers could take inspiration for implementations and developments regarding the protection of personal data on the web and on digital platforms in general. In the legislative and social field, the flow of information to users could be increased, regarding the consciousness and awareness of the value of their private information and its quantitative economic valuation. In fact, many supranational institutions (e.g., UE) and governments in recent years have failed to regulate and directly control the advance and development of these Big-Data giants in their countries, could do so indirectly through campaigns to raise awareness of the user about his private digital sphere and all the elements related to it. From a managerial point of view, this study can provide useful elements for decision-makers working in digital platforms firms that intend to implement a "Freemium" price model. In fact, they should pay attention to creating a balance between the services offered through the subscription and the sensitive information that the platform discloses to the user, making him an integral part of the business model of the platform, so the user, in providing his personal data, would feel in a less passive position and powerless with regards on the utilization of the service.

This research presents limitations mainly from the practical point of view, because are analysed very wide economic phenomena that contain within them many complications related to both economic sectors and companies. A weakness of this study could therefore be the fact of covering very solid and rooted theoretical economic arguments but in a relatively new context and in a phase of continuous innovation such as that of the data-driven economy.

In conclusion, the answers to the questions elaborated for this research are presented here in summary form:

“How have data-centric companies benefited from network effect and asymmetric information?”

Data-driven companies' growth has been fostered by Network Effects mainly for two reasons. The first, is the "Winner-takes-all market" effect, which plays a fundamental role in markets where network effects operate. This means that companies, websites or platforms with the highest market share experience better performance and are expected to grow faster. The second reason is the "First-mover advantage", so that a company which is the first to enter in a new sector, fostered by network effects, will tend to outperform its competitors (in this case, the Big-Data market). Asymmetric Information between digital platforms and users gives the opportunity to data-driven companies to operate in a condition where they have exclusive access to users' private information, hence, they can exploit and sell them to advertisers and infomediaries, in the knowledge that users suffer a lacking awareness regarding the true costs of free access to many online platforms and social networks.

"*How can one determine the 'real' economic value of personal data?*"

To assess and calculate the 'true' economic value of personal data, it would be necessary first of all that all actors in the Big-data market (companies and users) have a clear definition of personal data and laws to protect them. Since there is still no institutionalized and regulated market for buying and selling data between users and platforms, a common method should be found for: how to express value (e.g., *dollar*; *dollar x month*; *dollar x person*); whether to evaluate data as individual attributes or as datasets; avoid price discrimination. Some other ways in which data might be valued are: establishing the market prices at which personal data are offered or sold; establishing the economic cost of a data breach; determining prices for personal data in illegal markets; and ascertaining how much individuals would be willing to pay to protect their data.

"*Can freemium and premium pricing models be applied in order to address the apparent market failures driving the success of what have become some of the largest 'winner-take-all' players in this industry?* - in essence - models that can reduce information conflicts between users and platforms.

Freemium or premium pricing models can overcome some information conflicts between users and online platforms. This price model is used by many platforms, in some cases (where there are no conflicts of information) does not involve a rebalancing of sensitive information for the user (e.g., Spotify, Dropbox). Instead, for other platforms and social networks (e.g., LinkedIn and Tinder), the freemium model ensures that the user, paying, can access information that, at first glance, for the platform would make sense to keep for itself because of their high value, and that in the free version were not available for the user. This means that the price paid involves a balance of benefits between platform and user, consequently, the reduction of asymmetry coincides with better and more efficient use of the social/ online platform.

References

Accenture. (2015, October). *Why you need to capitalize on the rise of the data-driven enterprise*. In *accenture.com*. Retrieved from https://www.accenture.com/nl-en/blogs/insights/data-driven-enterprise

Acquisti, Alessandro & John, Leslie & Loewenstein, George. (2013). What Is Privacy Worth?. The Journal of Legal Studies. 42. 249-274. 10.1086/671754.

Acquisti, Alessandro, Curtis Taylor, and Liad Wagman. 2016. "The Economics of Privacy." Journal of Economic Literature, 54 (2): 442-92

Adam Thierer, Christopher Koopman, Anne Hobson, and Chris Kuiper, *How the Internet, the Sharing Economy, and Reputational Feedback Mechanisms Solve the "Lemons Problem"*, 70 U. MIA L. Rev. 830 (2016)

AGCOM. (2018, June). Big data. In agcom.it. Retrieved from https://www.agcom.it/documents/10179/10875949/Studio-Ricerca+08-06-2018/c72b5230-354d-444f-9e3f-5467ca450714?version=1.0

Akerlof, G., The market for "lemons": Quality uncertainty and the market mechanism; The Quarterly Journal of Economics Vol. 84(3), pp488–500, (1970)

Aoun Barakat, Karine & Sayegh, May. (2021). Information Asymmetry in the Age of Big Data Analytics.

Arthur, W. (1989). Competing technologies, increasing returns, and lock-in by historical events, *The Economic Journal*, 99, 116 – 131

Arthur, W. (1990). Positive feedback in the economy, *Scientific American*, 262, 92 – 99

Birke, D. (2009). The economics of networks: A survey of the empirical literature. *Journal of Economic Surveys*, 23(4), 762-793

Bawa K and Shoemaker R (2004) The effects of free sample promotions on incremental brand sales. Marketing Science 23(3): 345–363.

Beauvisage T.,Mellet K.,(2019) "Datassets: Assettizing and Marketizing personal data", Orange Labs

Bloomenthal, A. (2021, January 19). Asymmetric Information. In *Investopedia*. Retrieved from https://www.investopedia.com/terms/a/asymmetricinformation.asp

Carrascal, J.P., Riederer, C., Erramilli, V., Cherubini, M. and de Oliveira, R. 2013. "Your Browsing Behavior for a Big Mac: Economics of Personal Information Online". In *Proceedings of the 22nd International Conference on World Wide Web*, 189–200. ACM

Caswell, J. A., & Mojduszka, E. M., Using informational labeling to influence the market for quality in food products. American Journal of Agricultural Economics, Vol. 78(5), pp. 1248-1253, (1996)

Cheng HK and Tang QC (2010) Free trial or no free trial: optimal software product design with network effects. European Journal of Operational Research 205(2): 437–447

Church, J., & Gandal, N. (1992). Network effects, software provision, and standardization. *The Journal of Industrial Economics*, 40(1), 85-103

Cvrcek, Dan, Marek Kumpost, Vashek Matyas, and George Danezis. 2006. A Study on the Value of Location Privacy. Pp. 109–18 in *Proceedings of the 5th ACM Workshop on Privacy in the Electronic Society (WPES '06)*

D. Shirmohammadi et al., "Some fundamental technical concepts about cost based transmission pricing", *Power Systems IEEE Transactions on*, vol. 11, no. 2, pp. 1002-1008, 1996

Den Hartog, S. (2021, July 7). LRIC 101: everything you need to know about Long-Run Incremental Costing. In *CostPerform*. Retrieved from https://www.costperform.com/lric-101-everything-you-need-to-know-about-long-run-incremental-costing/?cn-reloaded=1

Economides, N. (1996). Network externalities, complementarities, and invitations to enter. *European Journal of Political Economy*, 12(2), 211-233

Economides, Nicholas and Lianos, Ioannis, *Restrictions on Privacy and Exploitation in the Digital Economy: A Market Failure Perspective* (January 27, 2021). Journal of Competition Law and Economics, Forthcoming.

European Commission. (2016, June 12). Case M.8124 – Microsoft / LinkedIn.

European Commission (2018). GDPR, 2018. Guide to the General Data Protection Regulation.

European Network and Information Security Agency. (2012, February 2). *Study on monetising privacy: An economic model for pricing personal information*.

Farrell, J., & Klemperer, P. (2007). Co-ordination and lock-in: Competition with switching costs and network effects, in M. Armstrong and R. Porter (eds.) *Handbook of Industrial Organization*, Vol.3, Elsevier.

Farrell, J., & Saloner, G. (1985). Standardization, compatibility, and innovation, *Rand Journal of Economics*, 16(1), 70 – 83.

Farrell, J., & Saloner, G. (1986). Installed base and compatibility – innovation, product preannouncements, and predation. *American Economic Review*, 76(5), 940 – 955.

Freedman, S., & Zhe Gin, G. (2011, March). Learning by Doing with Asymmetric Information: Evidence from Prosper.com. *NBER: National bureau of economic research*. doi:10.3386/w16855

Freemium Pricing Strategy | What You Need to Know About Freemium. (2021, October 29). Patriot Software. https://www.patriotsoftware.com/blog/accounting/freemium-pricing-what-is-it/#:%7E:text=What%20is%20freemium%20pricing%3F,upgraded%2C%20or%20premium%2C%20versions.

G. Malgieri, B. Custers, "Pricing privacy – the right to know the value of your personal data" Computer Law and Security Review (2017)

Gallaugher JM and Wang YM (2002) Understanding network effects in software markets: evidence from web server pricing. MIS Quarterly 26(4): 303–327.

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

H. Y. Heng and F. Li, *"Literature review of long-run marginal cost pricing and long-run incremental cost pricing,"* 2007 42nd International Universities Power Engineering Conference, 2007, pp. 73-77, doi: 10.1109/UPEC.2007.4468923.

H. Y. Heng, F. Li and X. Wang, "Charging for Network Security Based on Long-Run Incremental Cost Pricing," in *IEEE Transactions on Power Systems*, vol. 24, no. 4, pp. 1686-1693, Nov. 2009, doi: 10.1109/TPWRS.2009.2030301.

Helm R, Mark A and Bley S (2009) The effect of free product premiums on attitudes and buying intention for durable goods: moderating effects of value and product premium fit in the dual mediation model. European Retail Research. Wiesbaden: Gabler Verlag.

Hobbs, J. E., Information asymmetry and the role of traceability systems. Agribusiness: An International Journal, Vol.20(4), 397-415. (2004)

How can the problem of asymmetric information be overcome? (2021, August 31). Investopedia. https://www.investopedia.com/ask/answers/050415/how-can-problem-asymmetric-information-be-overcome.asp

Hu, Y. (2019). How does freemium strategy affect the demand of the paid premium mobile healthcare service: From an information asymmetry perspective. Information Development, 35(1), 36–50. https://doi.org/10.1177/0266666917724496

Huberman, Bernardo A., Eytan Adar, and Leslie Fine. 2005. Valuating Privacy. *IEEE Security and Privacy* 3:22–25.

Jiang Z and Sarkar S (2009) Speed matters: the role of free software offer in software diffusion. Journal of Management Information Systems 26(3): 207–240

Jorgenson, E. (2020, May 7). *The Power of Network Effects: Why they make such Valuable Companies, and how to Harness them*. Medium. https://medium.com/evergreen-business-weekly/the-power-of-network-effects-why-they-make-such-valuable-companies-and-how-to-harness-them-5d3fbc3659f8

Jwaifel, A. (2017). *Telecommunications industry over the period 2006 - 2012: performance, efficiency and implementation issues* (Master's thesis). February , Retrieved fromhttps://repository.cardiffmet.ac.uk/bitstream/handle/10369/9935/Final%20Thesis%20Abdullah%20Jwaifel.pdf?sequence=1&isAllowed=y

Katz, M., & Shapiro, C. (1985). Network externalities, competition, and compatibility. *The American Economic Review*, 75(3), 424 – 440

Keller D.,(2014,24 May) “If it’s free, you’re probably the product”, TechTalk https://techtalk.gfi.com/if-its-free-youre-probably-the-product/ Kenton, W. (2021, July 31).

Kim, G. (2019, May 27). *Marginal Cost (Near-Zero) Pricing is a Major App Provider Advantage*. IP Carrier. https://ipcarrier.blogspot.com/2019/05/marginal-cost-near-zero-pricing-is.html

Kim J and Morris JD (2007) The power of affective response and cognitive structure in product-trial attitude formation. Journal of Advertising 36(1): 95–106

Kumar, V. (2014, May). Making “Freemium” Work. In Harvard Business Review. Retrieved from https://hbr.org/2014/05/making-freemium-work

Kumar., R. (2018, July 30). *Understanding the basics of Network Effects — The Power of the Platform. Medium*. https://medium.com/world-of-iot/understanding-the-basics-of-network-effects-the-power-of-the-platform-2cfef215fe4a

Lewis, Gregory. 2011. "Asymmetric Information, Adverse Selection and Online Disclosure: The Case of eBay Motors." *American Economic Review*, 101 (4): 1535-46

Liu CZ, Au YA and Choi HS (2014) Effects of freemium strategy in the mobile app market: an empirical study of Google Play. Journal of Management Information Systems 31(3): 326–354

M. Kajtazi, "Information asymmetry in the digital economy," 2010 International Conference on Information Society, 2010, pp. 135-142, DOI: 10.1109/i-Society16502.2010.6018811.

Meglena Kuneva, European Consumer Commissioner, March 2009 https://ec.europa.eu/commission/presscorner/detail/en/SPEECH_09_156

Nair, C. (2015, October). Understanding Facebook as a Monopoly. https://www.researchgate.net/publication/286836835_Understanding_Facebook_as_a_Monopoly

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

Olejnik, L., Tran, M-D. and Castelluccia, C. 2014. Selling off Privacy at Auction. In *Proceedings 2014 Network and Distributed System Security Symposium*. San Diego, CA: Internet Society.

Page, W.H., & Lopatka, J.E. (1999). Network externalities. Retrieved from http://encyclo.findlaw.com/0760book.pdf.

Parker GG and Alstyne MWV (2005) Two-sided network effects: a theory of information product design. Social Science Electronic Publishing 51(10): 1494–1504.

Prufer, J., & Schottmmller, C. (2017). Competing with Big Data. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2918726Nair, C. (2015, October). Understanding Facebook as a Monopoly. https://www.researchgate.net/publication/286836835_Understanding_Facebook_as_a_Monopoly

R.R. Kovacs and A.L. Leverett, "A load flow based method for calculating embedded incremental and marginal cost of transmission capacity", *Power Systems IEEE Transactions on*, vol. 9, no. 1, pp. 272-278, 1994

Rifkin, J. (2014). The zero marginal cost society: the internet of things, the collaborative commons, and the eclipse of capitalism. First edition. New York, NY: Palgrave Macmillan.

Roland F. (2018, December 20). *Zero marginal cost business models*. Arvato-System.Com. https://www.arvato-systems.com/blog/zero-marginal-cost-business-models

Russel, W. (2015, August 12). Big data, open data, and the asymmetry of information. In Big Data To Big Profits. Retrieved from http://bigdatatobigprofits.com/2015/08/12/big-data-open-data-and-the-asymmetry-of-information/

Shapiro, C., & Varian, H.R. (1999). Information rules: A strategic guide to the network economy. Massachusetts: Harvard Business School Press

Sordet, P. (2019, March 22). *«L'amour sous algorithme» dévoile le côté obscur de Tinder*. Le Temps. https://www.letemps.ch/societe/lamour-algorithme-devoile-cote-obscur-tinder

Srinuan, Pratompong. (2012). *The impacts of network effects and multi-service provision on consumer switching costs in the Swedish telecommunications market*. https://www.researchgate.net/publication/277066523_The_impacts_of_network_effects_and_multi-service_provision_on_consumer_switching_costs_in_the_Swedish_telecommunications_market/citations

Steel E.,Locke C.,Cadman E.,Freese B.,(2013,June 12) "How much is your personal data worth?" Financial Times.com https://ig.ft.com/how-much-is-your-personal-data-worth/

Stiglitz, J.E. The contributions of the economics of information to twentieth century economics. Quarterly Journal of Economics, 115(4), pp.1441–1478. (2000)

Stobierski, T. (2020, November 12). *What Are Network Effects? | HBS Online*. Business Insights Blog. https://online.hbs.edu/blog/post/what-are-network-effects

Teisl, M. F., & Roe, B., The economics of labeling: An overview of issues for health and environmental disclosure. Agricultural and Resource Economics Review, Vol.27(2), pp.140-150, (1998).

The Half-Truth of First-Mover Advantage. (2014, August 1). Harvard Business Review. https://hbr.org/2005/04/the-half-truth-of-first-mover advantage#:%7E:text=A%20first%2Dmover%20advantage%20can,in%20a%20new%20product%20category.

Unscrambl, (2021, June 15). "6 inspiring examples of data-driven companies." https://unscrambl.com/blog/data-driven-companies-examples/

Understanding First Mover Advantage. (2020, October 7). Investopedia. https://www.investopedia.com/articles/investing/111016/understanding-first-mover-advantage.asp#:%7E:text=Learning%20is%20the%20advantage%20first,a%20larger%20segment%20of%20customers.

Van de Waerdt, P. J. (2020). Information asymmetries: recognizing the limits of the GDPR on the data-driven market. *Computer Law & Security Review*, *38*, 105436. https://doi.org/10.1016/j.clsr.2020.105436

Verbeke, W. Agriculture and the food industry in the information age. European review on agricultural economics, Vol.32(3), pp.347-368. (2005).

Vitenu-Sackey, Prince. (2020). The Impact of Social Media on Economic Growth: Empirical Evidence of Facebook, YouTube, Twitter and Pinterest. International Journal of Business, Economics and Management. 7. 222-238. 10.18488/journal.62.2020.74.222.238.

Wagner TM, Benlian A and Hess T (2014) Converting freemium customers from free to premium – the role of the perceived premium fit in the case of music as a service. Electronic Markets 24(4): 259–268.

Wathieu, Luc, and Allan Friedman. 2007. An Empirical Approach to Understanding Privacy Valuation. Working Paper No. 07-075. Harvard Business School, Cambridge, Mass.

World Economic Forum. (2014). *The Global Information Technology Report 2014 Rewards and Risks of Big Data*. https://www3.weforum.org/docs/WEF_GlobalInformationTechnology_Report_2014.pdf

Appendix

Appendix A: Cost of personal data; Cost of personal data in the dark web (Invisibly,2021)

Record Type	Average Price
Health Care Record	$250.15
Payment Card Details	$5.40
Banking Records	$4.12
Access Credentials	$0.95
Social Security Number	$0.53
Credit Record	$0.31
Basic PII	$0.03

Average price for dark web data

Category	Product	Avg. dark web Price (USD)
Credit Card Data	Cloned Mastercard with PIN	$25
	Cloned American Express with PIN	$35
	Cloned VISA with PIN	$25
	Credit card details, account balance up to $1,000	$150
	Credit card details, account balance up to $5,000	$240
	Stolen online banking logins, minimum $100 on account	$40
	Stolen online banking logins, minimum $2,000 on account	$120
	Walmart account with credit card attached	$14
	Hacked (Global) credit card details with CVV	$35
	USA hacked credit card details with CVV	$17
	UK hacked credit card details with CVV	$20
	Canada hacked credit card details with CVV	$28
Payment Processing Services	Stolen PayPal account details, minimum $100	$30
	Stolen PayPal account details, minimum $1,000	$120
	Stolen PayPal account details, no balance	$14
	50 Hacked PayPal account logins	$200
	Hacked Western Union Account	$45
	Verified Stripe account with payment gateway	$1,000
Crypto Accounts	Hacked Coinbase verified account	$610
	Blockchain.com verified account	$310
Social Media	Hacked Facebook account	$65
	Hacked Instagram account	$45
	Hacked Twitter account	$35
	Hacked Gmail account	$80
Database Dumps	600k New Zealand emails	$10
	350k Czech emails	$10
	2.4 million Canada emails	$10
	4,78 million Mexico emails	$10
	380k Austria emails	$10
	USA Voter Database (various states)	$100