Online Privacy

How to Improve the Self-Regulation of Online Privacy?

Thesis presented by
Audric ENGELEN

Supervisor
Johannes JOHNEN

Reader
Eric TOULEMONDE

Academic year 2017 - 2018

En vue de l'obtention du titre académique de
Master 120 en Sciences économiques, orientation générale à finalité spécialisée

Ecole d'économie de Louvain/UCL • Place Montesquieu 3 • 1348 Louvain-la-Neuve
Département des Sciences économiques/UNamur • Rempart de la Vierge 8 • 5000 Namur

Acknowledgments

First, I would like to thank my supervisor M. Johannes Johnen from the Economics School of Louvain. He allowed me to manage this thesis in the direction I wanted while giving me valuable advices. He was always available to answer my questions whenever I needed it. He also provided me with interesting sources and tools to improve my knowledge on the subject.

I would also like to thank my family, for their support and encouragement during the writing of this thesis. I would like to particularly thank my sister and my father, with who I could discuss of my subject and get advices.

Table of Contents

1. Introduction

The world we live in is a world where more and more data are given and shared everywhere. A reason for this increase is the boom of digital instruments such as smartphones and the continuous rise of the internet. In the latest years, companies and governments have begun to realize the important use of this growing data. Users themselves begin to realize the consequences of their online data sharing. The recent Facebook drama, which revealed that 87 million users had their data unwillingly collected by a consulting firm (Badshah, 2018), has revived ongoing debates on privacy and the reactions all over the world shows that contrary to some people's beliefs, *privacy isn't dead*.

The privacy of people is at the center of many debates and one of them goes over privacy regulation. We will focus this thesis on the online privacy, meaning privacy when people use internet. We will begin by giving a general overview of the uses made with personal data. After that, we will speak about the concept of privacy itself. Particularly, we will speak over its valuation and its regulation. Finally, we will investigate privacy policies, the walls of texts users supposedly read to give their consent upon entering a website. We will show their problems and propose solutions.

After looking at the framework of online privacy, we will look at the potential solutions to improve the self-regulation of online privacy. We will combine current advices with new ideas to propose a nudge over privacy policies. Nudges have been very popular for more than a decade and have seen one of its main contributors, Richard Thaler, receives the 2017 economic Nobel Prize. In this thesis, we want to combine good ideas that are not applied yet while using new concepts in a goal of improvement of the self-regulation of online privacy.

2. Definition of the question of research

"*How to improve the self-regulation of online privacy?*"

In this thesis, we will analyze the self-regulation model of online privacy. Self-regulation means that the users need to regulate their privacy themselves. They can do it by managing their own traceability, by changing their comportment or by rejecting privacy policies. Online privacy relates to the privacy of the data of individuals when they are going from sites to sites on the internet. We will look for possibilities to improve this self-regulation of online privacy.

3. Collection of users' data

In this section, we will talk about the collection of users' data when they are surfing on the internet.

When users are interacting with a website, they are observed by first parties and third parties. First parties are the websites on which the users are directly surfing at a given moment. Third parties can be webmasters, advertisers or also website analytics engines, but they are on a different domain than the website the user is visiting (Tamura, 2017). They collect cookies and use trackers to gather information, also called personal data. A personal data is defined by the General Data Protection Regulation (GDPR) as "*any information related to a natural person or 'Data Subject', that can be used to directly or indirectly identify the person. It can be anything from a name, a photo, an email address, bank details, posts on social networking websites, medical information, or a computer IP address"* (EU GDPR Portal, 2018). The third parties web trackers can use cookies and other tracking devices to re-identify users on other sites and collect information while they are surfing on the internet. This collection of information often happens without the users acknowledging this tracking which will lead to concerns over privacy.

A paper tried to analyze the trackers' evolution on the internet from 1996 to 2016 (Englehardt and Narayanan, 2016.). They showed that the tracking became more and more widespread and complex over time. There are more and more trackers and they tend to have more functionalities than before which thus lead to more collection of information. The complexity and interconnection of trackers has also increased through time. Third parties share users' identifiers with other third parties: this practice is called *cookie syncing*.

The parties can collect different kinds of information. Some information is collected by the website owner and can take different forms such as the number of people who visited the website, the payment information to be able to buy a product or the users' login information for convenience. That kind of basic information is necessary for the good functioning of websites. Some trackers are designed to try to gather personal information about users such as their gender, age or place. A goal of those trackers is to discover and predict potential consumers' preferences. Therefore, they collect web histories, observed contents or even the time spent on places. It can detect IP addresses which will help to keep track of users on other websites and know their surfing habits.

4. Exploitation of users' data on the internet

We have briefly talked about the collection of information. Now, we will talk about the use of the personal information. There are different positive aspects coming with the growth of the data available and we will present them with a focus on the online framework.

The business of data is increasing: the market for software and hardware capable of analyzing or storing Big Data is constantly evolving (Purcell, 2013). The capability to use Big Data is a real challenge for firms. Indeed, it would give an edge to those who better use the available data because they could benefit from the positive aspects of data that we will describe here after. In the last few years, we have heard the term "Big Data" more and more thanks to an increase in the available amount of data. Big Data comes from the fact that there are way bigger datasets thus making common database software unable to work and extract value on this data (Purcell, 2013). There is no typical size which determines if a dataset is a Big Data or not. The threshold size depends on the sector, particularly on the average dataset of the sector and the software available in this sector. There are three different types of data (W3training School, n.d.).

First, there is *structured data* that can be easily exploited because the data are already stored and delimited in a way. Structured data can be used to generate information. Second, there is *unstructured data* that are raw data, unchanged, meaning no transformation was done to make it easier to exploit. Hence, it is hard to use unstructured data. They are, by instance, images, videos, PDF or e-mails. Finally, there is *semi structured* data. It ranges between the two types of data we have just seen. For example, it is a video marked with text to try to structure it. New techniques arise to facilitate the use of structured and unstructured data. Combined to the raw increase in the global amount of data, this will expand even more the amount of data used.

The data collected on users can directly serve the interest of the party gathering it or it can be sold or shared to other parties. There are two categories of uses of information: primary and secondary. Primary uses of information are the uses for which the information was collected while secondary uses of information are uses with other purposes than for what it was originally collected (Healthinfoprivacybc.ca, 2011).

Companies can use those growing data for different actions designed to improve their revenues. Here is an overview of those different actions, based on a report of Liberty Global (Liberty Global, 2012) to describe the principal uses of the data.

4.1. Improvement of user's experience

Websites can use information they collect about users to help improving users experience, mainly by offering a gain of time. When a user has an account, information such as password and language can be registered, and it becomes helpful the next time the user connects to the website. Moreover, users can look at what they previously bought or what products they were looking for the last time they visited the website. By knowing the consumers, websites can also personalize the experience by offering content or products that are more likely to interest them.

4.2. Consumer insights

With the information available on consumers, companies can adapt their process to better suit the needs. Companies can also use the information to create new products and design them appropriately. They can more effectively manage the products by getting feedbacks from forums or commentaries on commercial websites.

4.3. Product enhancements

Companies can adapt products by analyzing how they are used. They can look at when and where it is used or what part of the product is particularly appreciated. That can serve as a feedback and is important to stay ahead of the curve in a competitive world.

4.4. Targeted advertising

With the growing volume of personal data available, target advertisement is one of the most frequent ways to use data. Advertisement is one of the largest sources of revenue for internet giants and is an increasing field. By using personal data of consumers, advertisement can be targeted to consumers to suit them the most. This field has a promising future as the smartphones ad market has still to be fully exploited. The ads switch medias, spending diminish for ads on print (-7,9% in 2015) but increase on digital media (+15% in 2015) (appendix 1). In 2017, advertisers spent $185 billion on the two most used techniques of advertising (Statista, 2017), and spending equals to revenues for others. The first technique is called *display advertising*. Display advertising is advertising online by using images, audio and video to advertise (Marketing Land, 2014). The money spent in online display advertising is estimated to grow from $97 billion in 2017 to $137 billion in 2020 (Statista, 2017). The second technique is called *search advertising*. Content providers pay to be highlighted in a search engine which then brings more traffic to their content (Jansen and Mullen, 2008). The money spent on search advertising is estimated to grow from $88 billion in 2017 to $111 billion in 2020 (Statista, 2017).

In a paper of 2015, researchers tried to see the impact of sharing data on the welfare of three agents in a market (Marotta et al., 2015). There are the advertiser firms who want to show their product to the consumers. There also is an intermediate agent, the Ad Exchange, who collects the data to share it with the firms from the third agent, the online users or consumers. The model focuses on a new mean regarding targeted ad: Real-Time bidding. Real-Time bidding is an instant auction which sells the right to have an advertisement spot to the biggest bidder. The Ad Exchange collects information over users when they surf on websites. The information of users is divided in two categories: *horizontal information* which displays the consumer preference over a product and *vertical information* which displays the purchasing power of the consumer. They share all or some of that information and send user profile to the firms who want to advertise. A firm can then offer a price for the right to advertise its product to the user and the highest bidding firm is allocated the spot seen by the user at the price of the second highest bid. The model shows that different shared information and different level of display can change the welfare distribution between the three agents.

First, they analyze sharing only the consumers' preferences over products (the horizontal information) to the firms. This leads to an allocation of the benefits from this market of 30% to the firms, 30% to the consumer and 40% to the intermediate Ad Exchange. When sharing only the purchasing power of the consumer (the vertical information) the firms receive no part of the benefits generated by this market while the consumer only got 11% and the intermediate receives 89%. In the case where no

information is shared: the firms got 0%, the consumer receives 33% and the intermediate 67%. Finally, if the intermediate shares all the information of the consumer, the firms receive 43% of the benefit, the intermediary obtains 57% and the consumer receives nothing.

What we can get from this model interested in target advertising is that the type of information and the amount of information have an impact on the welfare of the different agents. Consumers are better off when only their preferences are exchanged and generally when less information is shared.

4.5. Price discrimination

Different people with different characteristics are likely to have different willingness to pay. Price discrimination happens when an equal product is sold at different prices to different people using their estimated willingness to pay. Firms usually want to be able to maximize their profit. To do so, it is optimal for them to sell at the highest price possible for each client. Price discrimination is not new and is already widely used by airlines companies. It is also commonly observed with student or senior reductions. It is often regarded as positive for the economy as it should increase competition and the economic activity.

Several reasons lead to think that the future will see more price discrimination (Odlyzko, 2003). To begin with, the amount of data collected on people is getting bigger and the ways to analyze it are better and better. That leads to an easier estimation of the willingness to pay and thus a more effective price discrimination. Moreover, the Internet makes it easier to price discriminate (Hinz et al., 2011) because changing prices becomes close to costless. Also, different users can be offered different prices at the same time. Finally, a big incentive to price discriminate is when marginal costs are low which makes large volume of transactions profitable.

Price discrimination could improve the welfare of the society. Take the example of a train where the company was able to discriminate ticket prices and in which the third class was unnecessary uncomfortable (Odlyzko, 2003). The principal reason wasn't that the bad seats were cheaper but that people who wanted to avoid the worst conditions were willing to pay for it, increasing the price of the other classes. We can see in that example that being able to collect and use information on users can improve the situation of some agents: the train could have only cheap and comfortable identical seats that people would buy at different prices depending of their estimated willingness to pay. Another example where total transparency leading to price discrimination could make people better off is the medicines market. At the moment, prices are different between countries and the US citizens are generally paying more than European citizens for the same pills. A consequence is that the poor population of elderly US citizens can't pay for needed pills. If instead of partial discrimination over general population, the discrimination was on individual information, poor people would be able to buy pills at a lower price because their willingness to pay is low. At the same time, very rich people would pay more which would diminish the social security spending. Price discrimination is particularly effective when dealing with goods which have a low marginal cost like medicines or software. They generally have big fixed or development costs but then the marginal unity is basically costless. It is profitable to sell to those with the lowest willingness to pay, even close (but superior) to marginal cost.

For effective price discrimination, arbitrage must be avoided. "*Arbitrage is the simultaneous purchase and sale of an asset to profit from a difference in prices*" (Staff, 2018). As we just saw, price discrimination is selling the same product to different people at different prices. But if people with low

willingness to pay can buy and sell to those with higher willingness to pay, it will lead to people only buying at the lowest price because someone could benefit from buying low and selling high.

Price discrimination is a good thing for standard economics as seen with the example of medicines. Nevertheless, as explained by behavioral economics, it can be rejected by public opinion. We can imagine the outrage if medias cover the news by saying "*Governments help firms to sell same medicines at a higher price to young people than to others*". In an experiment in the early 2000s, Coca Cola tried to discriminate prices in their vending machines by increasing the prices when it was hot outside. Media revealed it and negative public opinion made Coca Cola stop the experiment. The opinion could have been more favorable if it had been revealed that people would pay less when it is cold. Introducing price discrimination, even if it is proved to be positive in a certain context, could face public rejection. It should thus be introduced and presented to the population with wise terms considering the input of behavioral economics such as the importance of fairness or the importance of the reference point.

With a decreasing privacy and the increase of data available on users, Internet firms should try to use the information to carefully implement price discrimination.

5. Privacy

We covered the collection and the use of data. We talked about practices, but we will now address privacy. The collection and uses of data touch the subject of privacy and we will now define it. We will also try to get an idea of the value of privacy. Then, we will go over the concerns that privacy can raise. Finally, we will try to see the effects of different levels of privacy for the society**.**

5.1. What is privacy?

In 1890, Samuel Warren and Louis Brandeis made a law review article (Warren and Brandeis, 1890) which is considered the first publication to raise the idea of a right to privacy. According to them, privacy is first the *right to be left alone* which means not being disturbed by the press or by people having devices recording or reproducing scenes or sounds. Secondly, privacy is the *right to property* which means being recognized of its own creations and receiving all profits from wrongful use by others of said creations. It should also give the right to prevent publication as it could harm the original creator. In 1967, Alan Westin described Privacy as "*the claim of individuals…to determine for themselves when, how and to what extent information about them is communicated to others*" (Westin, 1967). In 1992, Ferdinand David Schoeman defined privacy in his book *Privacy and Social Freedom.* For him privacy is different from one person to another and is valued differently. The social norms are shaping privacy. Those social norms exist to show respect from individuals to others. Privacy is an aspect of freedom, autonomy and respect of dignity (Schoeman, 1992). In 2002, Robert Gellman says that there is a consensus for what concern some part of privacy. In the context of the collection, use, share and the processing of personal information, privacy can be described as "*fair information practices*" (Gellman, 2002).

We can see that there are different definitions of privacy and that it changed through time. We conclude that it concerns what information individuals decide to keep only for them or to share with others. Privacy is controlling what people share.

5.2. Valuation of personal information

In this part we will try to estimate the value of privacy. To do so, we will look at the value that personal information creates in the online market economy. We will also look at the value that people allocate to their own privacy.

5.2.1. Valuation by the market

The quantity of personal data is exponentially growing thanks to technological companies offering free services such as social networks, online search engine, host websites which give the opportunity to store and share videos, images… Coupled with the creation of technologies capable of treating the growing data, the value of this data will be huge.

In 2012, The Boston Consulting Group (BCG) published a report named, *The Value of Our Digital Identity* (Libertyglobal, 2012). The *Digital Identity* of individuals is all the digitally available information about those individuals. The BCG defines the three different types of value coming from the Digital Identity.

Firstly, the *consumers value* is the value that comes from the consumer surplus of Internet services. For example, the value consumers credit to Facebook usage reduced by the price they pay to access Internet at that time. It is also the value that comes from the reduction of prices coming from the use of the Digital Identity. And finally, it is the value of time saved by using digital identity applications. Secondly, *the value calculated for the businesses* is the additional revenues and the cost saved from the Digital Identity after considering reduction given to consumers. Thirdly, the *value for the government or public sector* is defined by the increase in tax revenues and the reduction of spending. The benefit of personal data application in Europe is predicted to reach €1 trillion in 2020. (estimation made without considering second-order effects of an improved economy). Governments and businesses in Europe are predicted to see an annual benefit of €330 billion. Personal data are predicted to create a €670 billion benefit for European consumers. In this report, they consider the additional value generated by 8 different sectors:

- Traditional production
- Retail
- Financial services
- Telecommunications and media
- Public services/health
- Web 2.0 communities
- E-commerce
- Online info/entertainment.

Those numbers consider all personal data collected on people and the benefit coming from it. We will here focus on the numbers coming from the personal data collected on the Internet in the three last sectors: web 2.0 communities, e-commerce and online info/entertainment.

Web 2.0 communities

Web 2.0 is the web of the user-generated contents. People aren't passively surfing on the net anymore, they are now taking an active role of interacting and sharing with others. This comportment generates a lot of personal data and the quantity is expanding with time. Social networks are making money with that data. Users share data which are then used to target advertise, which is important in this field. Indeed, the efficiency of ads will improve if advertisers can determine which type of users will see their ads. It will also increase revenues. Users can connect to more and more sites using their social networks accounts. This will increase the reach of social networks on users' information and generate even more revenue. The data collected on users can also be marketed to third parties which gives another opportunity to make money from the data. Lastly, data is used to improve services by advising users to look for content that might interest them. Thus, the user will use the platform more and share more content.

The value of digital identity in Web 2.0 communities was estimated for the EU 27, in 2011. The value for the organizations into Web 2.0 communities was estimated at €1 billion. For the consumers, the value was estimated at €8 billion. The value of the digital identity in Web 2.0 communities for 2020 is estimated to be worth €7 billion for organizations and €32 billion for users. The reason users can enjoy big platforms for free is that their revenues are generated directly from their shared contents. A BCG report of 2012 said that the consumer surplus gained from internet services which includes social Medias averages €2.900 per individual per year.

E-commerce

The e-commerce sector is an advanced user of targeted advertisement. The recommendation system, offering products that the consumer could like using its information like past purchases, is a way to generate new revenues with the data. This recommendation system is supposed to bring up to 25% of increased revenues to Amazon, the e-commerce world leader. In addition to the recommendation system, data are analyzed to derive a predicted trend which can then help sites to manage the stocks.

In this sector, reviews are a valuable source of data for companies because it can help improve products. Reviews also serve to improve the trust consumers have by comparing to other users reviews of the product. The data collected on commercial sites can be interesting for third-parties as it can help design or improve products, this data can thus be monetized.

Thanks to the e-commerce sector, consumers can identify the lowest prices. Online shopping is also more convenient. The digital identity value in the e-commerce sector for consumers was estimated at €84 billion in 2011 in the EU 27 and is estimated to grow to €181 billion in 2020 in the EU 27. For organizations it was estimated at €9 billion in 2011 and is estimated to be €45 billion in 2020, also in EU 27. One reason of this huge rise is the emergence of smartphones, which bring more data and possibilities. For example, it is possible to use location data to target advertisement of a close store.

Online info/entertainment

More and more users are watching the news, listening to music or watching videos online thanks to the importance of smartphones, tablets and connected televisions. The rise in users increases the data which increases the targeting power of this sector. In this sector as in e-commerce, recommendations are important to increase the usage and the quality of the navigation for users. Recommendations are based on previous behaviors and the behaviors of other users. Moreover, personalization can be used to increase revenues by including a fee on a personalized product or by improving the pertinence of the advertisements showed to users. As in other sectors, the collected data can then be resold to third-parties to generate additional incomes. Consumers also benefit from it, enjoying a large amount of content often for free or at a low fee because companies base their revenues on the data. Users highly value the usage of those services and have thus a consumer's surplus by using those services for free.

The digital identity value in the online info and entertainment sector was estimated for consumers at €85 billion in 2011 in the EU 27 and is estimated to grow to €126 billion in 2020 in the EU 27. For organizations it was estimated at €5 billion in 2011 and is estimated to be €20 billion in 2020, also in EU 27.

Conclusion of the valuation by the market

This report on digital identity value showed a predicted huge increase in the value of our digital information in the coming years. The value here can be considered as an increase of the welfare of the society. However, we think it only took the positive aspect of the increase of data sharing and didn't consider the potential concerns of consumers for privacy which can lead to reduction of online activity if they are not addressed.

5.2.2. Valuation by the users

After speaking about social impacts of a diminution of our privacy, we will consider the individual point of view. The valuation that users give to their privacy is hard to measure because it is context dependent. The value depends on factors such as the awareness of the people regarding the subject of privacy, that they are often not informed enough. It also depends on the way it is computed. Knowing those difficulties, we will present some studies which give an idea of the value that users give to their privacy in different contexts. In 2009, researchers experimented and discovered that people have polarized valuation of privacy (Acquisti, John and Loewenstein, 2009). That means that valuations are on the extremes and not linearly distributed. It also highlighted that people give more value to a loss of privacy than to a gain of privacy. Finally, offering a privacy friendly option before another less privacy friendly option push people towards taking the option with the most privacy. In other words, choices are context dependent (Acquisti, John and Loewenstein, 2009). The valuation of individual personal data can be as low as $0.0005 when looking at the price some companies buy and sell elements of browsing histories of users (Olejnik, Minh-Dung, and Castelluccia, 2014). In another study, people from the US give a value from $30.49 to $44.62 for the protection against errors, improper access and secondary use of personal information (Hann et al. ,2007). By analyzing the app market, a study computed how much consumers are ready to pay to avoid divulging certain information. On average, consumers would do a one-time payment of $2.28 to conceal their browser history, $4.05 to conceal their list of contacts, $1.19 to conceal their location, $1.75 to conceal their phone's identification number, $3.58 to conceal the contents of their text message and they would pay $2.12 to have an ad-free app. The study also showed that informing consumers about privacy help them to make more accurate decisions regarding their true preferences in this domain (Savage and Waldman, 2013).

Conclusion of the valuation of privacy by the users

The valuation users give to privacy depends on the way it is computed. However, it is generally low compared to the huge revenues companies make with that information. We think that comes from the unawareness of most people over the uses made with their information which leads to an underestimated importance of their data.

5.3. Privacy concerns

In this part, we will speak about the negative aspects that can be linked with privacy. Individuals are often not fully aware of the consequences of sharing personal data. They are not aware of the implications of their behaviors and it is thus quite hard for them to make an optimal decision. Moreover, the increase in data could serve malicious authorities to increase their control over the population.

Consumers' worries about data sharing reduce the potential of e-commerce. A controversial usage of personal data or an ambiguous collecting environment diminishes the likelihood of consumers sharing their data and thus using the website or buying some articles. Privacy and security concerns are the most important reason why people don't purchase online (Udo, 2001). It is important for users to be aware of the way their data are used and what data is collected. If we can develop a trustful environment regarding personal data, it would be advantageous for every party in the long term.

Moreover, there is a privacy paradox. Several studies highlighted that people tend to say they care for and are concerned about their privacy. They say it is influencing their online decision making. At the same time when they surf online, their behaviors don't match their saying. In a study, they compared users' self-reported behavior with their observed behavior in an e-commerce situation (Jensen and Potts, 2005). The study highlighted that people have false impression about their knowledge over privacy technologies and vulnerabilities. In general, people are revealing information without taking as much actions as they said they would like to protect their privacy.

Furthermore, governments could benefit a privacy reduction to increase their power. In 2020, China plan to have a national score which will award a Citizen Score to each of its citizens (USA Today, 2017). The score will be based on information collected about everyone. That could be information about having a criminal record or the good payment of bills or the purchasing habits of the citizens. It could also be based on what people say on social medias. It could also be based on people activities such as doing sports, reading books or playing video games. All those activities and data gathered by the government would then be ranked as positive or negative. A Citizen Score will then be determined for everyone in function of the positive and negative activities. What is worrying is that the score will be based on relationship with other citizens. Particularly, if a person interacts with another person with a low score, it would potentially reduce the score of the first person. A low score would also reduce the freedom of the citizen. Now, only voluntarily participants are in this project, but it will be mandatory for every Chinese citizen in 2020. This score will be available publicly so that decisions (hiring for a job, accepting children in a good school, dating someone) can be taken knowing this Citizen Score. In this case the mass of information will be used by the government in a controversial way.

There is a last point we want to talk about. On the first hand, people expect governments to protect their privacy by raising new laws, adapting the legal framework to new technologies. On the other hand, governments may want less privacy. Governments have incentives to reduce privacy in the context of tax payment. They use their power to extract information over revenues and to make people pay different amount of taxes or to detect fiscal frauds. Governments also want to reduce privacy to improve their research of criminals and frauds. That was the case when a Belgian legal obligation came into place to force everyone who owned a prepaid sim card to register their identity to their mobile operators (Alexander De Croo, 2017). The goal was to help prevent terrorism, but it is a big diminution of privacy for a very low effect on terrorism. Authorities can still find the motivation to try to stop the reduction of privacy in the private sector. That would happen if they value the protection more than the benefits coming from an increase in personal data availability.

5.4. Costs of lack of privacy regulation

We showed the huge positive impacts of the growth of data collection, sharing and usage which would give the impression that going for less privacy is obvious. Some studies tried to analyze the other side of the story and tried to estimate the cost of a lack of privacy. They say it is possible that pro data

sharing studies overestimate the negative effects of a higher personal data protection (Gellman, 2012). It comes from their supposedly false assumption that businesses would find new opportunities instead of staying in the same strategies as when privacy is low. First, we will speak about the businesses' costs of a lack of privacy protection and then we will speak about the costs for consumers when privacy is not protected. This part is based on a paper from Robert Gellman named *Privacy, Consumers and Costs* (Gellman, 2012).

5.4.1. Cost of unprotected privacy for businesses

This section will cover the problems arising for businesses in a lowly regulated privacy environment.

Sale loss

Some consumers while shopping online put merchandises in their online carts but then they don't complete their purchases. There are two principal reasons to that, both linked with privacy concerns. First, they don't want to give all the required information by the site before the transaction can be completed. Second, they don't want to enter their credit card details. Those are the principal reasons explaining why part of the consumers who try online shopping are giving up. In a report of 2000, the Federal Trade Commission (FTC) estimated that online sales were lower by $18 billion due to privacy concerns (Federal Trade Commission, 2000). We can say that a trustful environment would be beneficial for businesses and consumers.

Loss of international sharing of information

In Europe and other countries, privacy laws restrict the export of personal data to countries who don't respect a sufficient level of privacy protection. That is for example the case of the US: if a firm wants to bypass those laws and collect information from Europe, it needs to prove that it can fulfill the European requirements, or it needs to adapt to those. If it doesn't upgrade, it cannot receive the data, thus missing an opportunity to do business. The main difference is that Europe asks for more protection of the data in addition to fair information practices.

Loss of investors

There is an argument that the lack of clear privacy rules made the businesses base their profit on data while that field didn't prove to have the best profitability for investors. That could thus scare new investors away. However, that argument is made in a pro-regulation paper. We think it goes against the fact that data exploitation is a promising and growing field.

5.4.2. Cost of unprotected privacy for consumers

Here we will talk about the potential costs for consumers coming from a lack of protection of privacy. We will talk about both non-online and online arguments, with more emphasis on the second category.

Non-online specific costs

Higher prices

Some shops require registration to benefit from reductions. Those who refuse to give the necessary information to register can still buy products but at a higher price.

Junk mail

Each year a person receives on average 560 junk mail. Junk mails are different than e-mail or spam as the costs of production and delivery are significant. In 1995, a study of the US Postal Service revealed that 50% of U.S. households wished to receive less advertising mails. They have the choice to opt-out, but the proposed solutions are time consuming and thus not often used.

Identity theft

When an individual uses personal information of another person to commit a fraudulent act (such as taking a fake credit or opening a bank account with the stolen identity) it is considered as identity theft. With the rise of the Internet and with the availability of personal data it is easier to find the required personal data to steal an identity. It is also easier to use the fake identity because online commercial transactions are done without checking faces. Identity theft can lead to difficulties for the victim to clean his record and the person can have undeserved restrictions on credit, on job offers or mortgages. It is time and money consuming for victims to deal with identity theft.

Privacy laws who prevent the collection and sharing of personal data without individual knowledge would help to reduce identity theft. The cost of individuals trying to protect themselves from identity theft is also to be considered when talking about the consequences of a lack of privacy protection. For example, an annual subscription to a Credit Watch service costs about $40. The service helps to detect and to reduce the impact of identity theft. The FTC advises users to protect themselves from identity theft. People particularly should opt-out from the sharing of personal information held by third-parties. However, it is observed in practice that companies working with personal data often don't notify users when they sell it and they often don't give opt-out solutions, or only partially. Even when opting-out is available, the cost (time and expenses) to obtain it is substantial.

Online specific costs

Spam

Unprotected privacy can be linked to spam. *Spam* is described as "*unsolicited commercial email and related undesirable online communication*" (Rao and Reiley, 2012). Spam is widespread because users lack control over the way their email addresses are collected, used or shared. In 2010, valid email addresses received on average 90 billion emails daily worldwide, while 88% of that was spam. Spam is estimated to cost $20 billion annually to firms and consumers. While the opportunity cost of spam is enormous -lot of time lost-, the benefits for the spammers and products sold using spams are only $200 million which is relatively low compared to the social cost (Rao and Reiley, 2012). In addition to the financial costs, it also reduces the trust of users and their confidence to communicate their emails.

Anonymity

In some cases, users want to surf without being traceable. Being anonymous can help to reduce identity theft and spam. It can also offer the possibility to act freely. To be able to meet anonymity or their privacy needs, users need to purchase some software that makes them untraceable. For example, Anonymizer costs $50 per year and gives the ability for the buyer to surf while preventing others to trace it. Being anonymous can allow people to speak against dictatorial states or to give their true opinion. It can however also serve fraudulent activity. Protection can serve businesses by giving them

the possibility to safely share confidential information. Another possibility for people who want to be unknown is to provide false information. This behavior leads to problems for businesses to use the data and lower the value of the data.

5.5. Optimal Level of Privacy

From the definitions of privacy, we see that there are some limits linked to the concept: limits between what people consider as personal or public and limits between what stays private and what is shared. Those limits can be moved which is why we talk of privacy *trade-offs*. In economics, the interest is focused on the information aspect of privacy. We will thus focus on the trade-offs of revealing personal information. We will base ourselves on the different parts we developed in this thesis and will take arguments from the *Economics of Privacy* (Acquisti et al., 2016).

From the literatures we see that there is no simple answer to "*how much data should society optimally share?*" Indeed, the relation between welfare and privacy isn't continuous and a positive aspect of sharing more data is often counterbalanced by a negative aspect of that data being shared. We will focus on the arguments which concern the online privacy background and will thus not speak about the effect of a change in the level of privacy on the health sector or in the credit market.

From an individual point of view, sharing data often gives rewards traded against uncertain negative -and often future- events. The typical advantages of sharing data come either as an economic or a psychological benefit. Sharing data can give the user an immediate reduction or an easier use of a program which leads to economic gain. In exchange of the free product, the user could watch advertisements. Sharing data allows users to use free websites that they value such as search engine like Google or social networks. An example of psychological benefit is the happiness provided by receiving attention -*likes*- on social networks. Those benefits can be counterfeited by negative events. The sharing of data which lead to an immediate reduction on an item could lead to future price discrimination which could make the user pay more for another item. In the case of sharing a picture of himself at a party to receive immediate social acknowledgement, the user could be denied a future work because the employer looked at his profile and saw that picture.

To talk about an optimal level of privacy from an aggregate point of view, meaning for the society, we will speak about the effect of privacy on both the data subjects who share their data and the data holders who collect it. Having more privacy –that means sharing less information- has positive and negative effects on the social welfare.

People would be less afraid to download apps or visit websites if they know they can't collect all their information and that could increase the traffic on online devices. Moreover, if less information is shared, people will be less concerned about buying online. That lowered concern would increase e-commerce activity for those who were reluctant to share their information (Udo, 2001). On the other side, if less information is shared, the personalization of the offers and the knowledge on consumers' taste will drop, thus decreasing e-commerce. The possibility to price discriminate will also be diminished which can have positive or negative effects.

Individuals would benefit from a reduced misuse of their own information such as spam or identify theft. At the same time, they would lose performances coming from the improvement of the products they use. For example, if we share less information to search engines, we will have less accurate results for our searches.

Several studies highlighted that we positively value privacy. Having more privacy would then increase the welfare of individuals, due to them valuing the knowledge they aren't tracked. At the same time, it would diminish the targeting power of advertisement. That wouldn't mean fewer ads but fewer targeted ads, which would then be less valuable for companies. A decrease in ads revenues can be huge when we know that those revenues were $52.8 billion in 2015 (Lunden, 2015). That drop could be a threat for free services whose benefits come in great part from advertising.

To conclude, more privacy hasn't a clear positive or negative impact on social welfare. It even seems like the harms on profits and development would be too big to even consider a reduction in privacy. In this part, we only focused on the valuable advantages and disadvantages of having more privacy. However, it is important to realize that privacy can be undervalued by individuals. Moreover, different people have highlighted that the question of the level of privacy shouldn't consider privacy as a simple dispensable instrument (Cohen, 2012). Cohen argues that privacy shouldn't be considered as unimportant because we realize that it would undermine data analysis or limit innovation. Privacy should be a way to self-develop outside of the surveillance or of the imposed values of our society. Another article on *obscurity* (Hartzog and Selinger, 2013), which they say is a better word to use than privacy when we think about our data, goes in the same direction. They say that privacy shouldn't be reduced to the protection of data exploitation for profit by companies. Instead, it gives a "*protective state that can further a number of goals, such as autonomy, self-fullfillment, socialization, and relative freedom from the abuse of power*". We will thus take those aspects in consideration.

Here is a table summarizing the main positive and negative effects on welfare from an increase in privacy:

Positive aspects of increasing privacy	Negative aspects of increasing privacy
• Increase in e-commerce through a diminution of privacy concerns	• Decrease in e-commerce from a diminution of personalization and knowledge of consumers' tastes
• Reduction in misusages of information (spam, identity thefts)	• Reduction in products improvement
• Increase in welfare from the valuation by individuals of keeping data privately	• Reduction of advertisement revenues from a lower capacity of targeting
• Increase in freedom from reduced surveillance and by giving a "*protective state*" for individuals	• Compromisation of price discrimination implementation

5.6. Privacy regulation

In this section, we will talk about the ongoing debate on privacy regulation. As we highlighted before, the uses of data have nice effects on the economy and those effects will probably still grow in the future. Nevertheless, we cannot neglect privacy concerns and problems and we need to try to address them. There is no clear answer to which privacy level would be the best, as there are trade-offs for different levels of privacy. In a way, letting the data flow in the economy leads to positive value gained

from targeted advertising, price discrimination or product improvement. On the contrary, privacy concerns lead to loss in sales. Moreover, other negative effects of privacy such as spam or identity theft represent a high cost for the victims. It is important to try to find a regulation method which tries to keep as much as possible the good effects while trying to diminish the negative effects of a lack of privacy.

Paternalistic regulatory solutions for privacy would be easier for the users, as they don't have to deal with interacting with different subjects who have different privacy policies. However, the cost of implementing privacy protection is said to be higher than the cost of the risks of privacy violations (Rubin and Lenard, 2001). Imposing paternalistic regulation such as forbidding personal data tracking by online sites to counter the problem of online privacy would be highly inefficient. We will thus focus on the self-regulatory model, which let people manage their privacy.

5.6.1. Privacy regulation in Europe

In Europe, the General Data Protection Regulation and the ePrivacy Directive are designed to ensure the respect of the privacy of Europeans. They want to make sure Europeans benefit from more protection when sharing personal data by imposing rules to the data collector.

General Data Protection Regulation

In May 2016, the General Data Protection Regulation (GDPR) was launched and it will apply in May 2018 at the latest. This regulation goes over the protection of natural persons concerning the processing of their personal data and concerning the movement of the data (Data protection in the EU, 2016). This law begins by stating that "*The protection of natural persons in relation to the processing of personal data is a fundamental right".* One of the goals (Galdies, 2017) is to have a harmonization across Europe by applying this new regulation instead of the old directives. That will make it easier for businesses acting in Europe. Regarding the organizations outside Europe, if they take personal information (IP address and cookies are elements of the *personal information* definition used by this law) from Europeans they are subject to the European legislation. They will now only deal with one intermediary while before they needed to comply with different legislations, which will represent a gain of time and money for both parts.

This regulation deals differently between two kinds of agents: controllers and processors. Controllers are defined as "*the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law*" (Gdpreu.org, 2018). In other words, the controller determines how personal data will be processed and the purpose of this processing. The processor is defined as "*a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller*". The processor processes the data on behalf of the data controller. For the GDPR, the controller is the main agent and the one who needs to receive consent. With the regulation, data controllers must only deal with data processors who comply to the legislation by taking sufficient actions to protect the rights of the data subjects. Controllers and processors should take security actions to "*implement appropriate technical and organizational measures*" depending on the processing and use of data and the potential risks for the rights and freedoms of the individuals. The regulation shows examples of appropriate security actions regarding data such as the pseudonymization of the personal data or the ability to

recover access to the data in case of technical problems. By using pseudonymization, privacy is more protected while keeping the benefits of the data use. They also recommend testing if the appropriate technical and organizational measures are taken to provide required security of the processing

If an entity doesn't respect the new regulation, it is at risk to receive fines. For violations of record-keeping, security, breach notification and privacy impact assessment obligations, the regulators could give a fine to an entity of the maximum between €10 million or 2% of its gross revenues. For violations of obligations regarding legal justification of procession (for example the need to have the consent from the data holders), of data subjects rights and of transfers of data to different countries, the regulators could give a fine of the maximum between €20 million or 4% of the gross revenues of the entity (Galdies, 2017). We can see the importance of consent given by the data holders to the data controllers.

The regulation defines consent as "*any freely given, specific, informed and unambiguous indication of his or her wishes by which the data subject, either by a statement or by a clear affirmative action, signifies agreement to personal data relating to them being processed*". The consent cannot be obtained by force but should be freely obtained. If someone gives consent of data unnecessary for the contract because he was forced to do it by the other part, then it is not valid. Moreover, the data subject giving his consent for data collection must be aware of how it will be used. They also specify that erasing his consent should be as easy as giving it. The GDPR can thus be an opportunity for the data processors to review their uses of data and to rethink how they are processing them.

The GDPR enforces public authorities or controllers and processors dealing with large amount of data subjects to have a Data Protection officer. This person must be aware of the laws and practices of data regulation and will be helping the entity to ensure they have conformed data protection or to help for data problems.

In case of a personal data breach, meaning the destruction, loss or unwilling sharing of personal data, the controllers must notice an authority within the 72 hours following the incident. The notice is not mandatory if the breach doesn't translate to risk for individuals. Finally, the concerned individuals can make a request to receive more information about how their data is processed and the information must be given to them except if the request is unfounded. They also have the right to ask for the removal of their data when the purpose for which the data was collected is accomplished.

ePrivacy

Now, we will speak more precisely about a proposal of the European commission over ePrivacy (Galdies, 2017). This regulation is an addition to the GDPR and concerns the use of personal information on electronic devices and will give one single regulation applicable in Europe to improve the protection of personal information (ec.europa.eu, 2018). It will apply to the old electronic devices and to the new ones such as Facebook, WhatsApp or Skype. Users will have the possibility to easily say if they give consent or not to tracking cookies and identifiers. They can do so using the browser instead of needing to give or refuse consent in a load of different places. Cookies who are said to be non-intrusive and who are designed to improve users' experience on the website don't need consent. Moreover, the aim of this proposal is to ban spam form electronic devices, and particularly regulate marketing phone calls.

5.6.2. Privacy regulation in the US

We will talk here about the data protection in the United States. We will base this part on a guide to data protection in the United States (Jolly, 2017). In the US, there are multiple laws concerning the collection and use of personal data. Those laws can apply to specific data such as financial or medical data. Some laws regulate the *use* of personal data for marketing purpose and some laws are there to improve the *protection* of personal data. We will focus on the FTC which concerns privacy and data security.

Sector specific laws

Two of the biggest privacy laws are sector specific. They concern the financial information for *The Financial Services Modernization Act* and the medical information for *The Health Insurance Portability and Accountability Act*. It makes sense to have specific laws for those sectors who are particularly sensitive. It aims to improve the security of those data and to regulate the collection and use of the information.

The Federal Trade Commission

The goal of the Federal Trade Commission is "*working to protect consumers by preventing anticompetitive, deceptive, and unfair business practices, enhancing informed consumer choice and public understanding of the competitive process, and accomplishing this without unduly burdening legitimate business activity"* (Federal Trade Commission, 2018). One of its most important law regarding online privacy is the Federal Trade Commission Act. "*The Federal Trade Commission Act (FTC Act) is a federal consumer protection law that prohibits unfair or deceptive practices and has been applied to offline and online privacy and data security policies*." (Jolly, 2017) The FTC act applies to a lot of companies and individuals doing business in the US. That is one of the most important privacy laws. It doesn't concern a type of data, but it is there to prevent bad practices which would be a threat to the safety of consumers' personal information.

In a report from the FTC sent in 2000 to the congress (Federal Trade Commission, 2000), they say that they believe in the self-regulation of online privacy. Regarding privacy, they want to implement the fair information practice principles. There are four principles which are notice, choice, access and security (Privacyfirst.nl, n.d). The first principle of *Notice* means that consumers should be informed of the information practices before any collection. A good notice is essential as it will influence consumers' choice. A good disclosure of information practices online should be easily accessible, clear and understandable. It should also be unavoidable to fully inform consumers on the use of the collected personal data. The second principle is the principle of *Choice* (comparable to the principle of consent in the EU) which is that consumers must have a choice over how their data will be used regarding data secondary uses. Secondary uses mean uses not necessary to the primary action the consumers want to do. For example, if the information is used to advertise new products or to share the information to third-parties. Choice should be easy to make for consumers, they could simply click on a button online to claim their choice. The third principle is *Access*. Consumers should have the possibility to easily access their data to see if there is no problem with it. Finally, the principle of *Security* means that collectors must take reasonable actions to ensure that the data is well protected.

In 2000, the FTC realized a report which offered to legislate to push the implementation of the fair information practices online because even though it increased, lots of website were not respecting them (appendix 2). They gave recommendations, including legislative ones over what they want as notice, choice, access and security. Websites should use those when creating their online disclosures of data uses. Since then, almost all websites respect the fair information practice principles.

In 2012, in the report of the FTC (Federal Trade Commission, 2012), different propositions are made to improve consumer privacy. The first proposition is the implementation of a *Do not track* mechanism in which users can subscribe to communicate their will not to have their information tracked. It has some problems such as the difficulty to implement, to verify or to enforce it. Nevertheless, an opt-out solution needs to be created to give the user a real choice. They also discuss projects of bills to improve the transparency of information usage or to protect minors against targeted ads. The FTC continues to show its will to increase data protection and privacy choices for consumers.

5.6.3. Conclusion on privacy regulation

The Europe has focused on making laws and having a clear uniform regulation for privacy while the US have given guidelines and recommendations in addition to laws. It makes sense considering the US are generally more liberal and push towards freedom. The same general principles are applied in both places but with the GDPR, the Europe continues in its will to regulate privacy and impose US companies to adapt if they want to use European data.

5.6.4. Self-regulation model of privacy

Self-regulation means that the level of privacy will be chosen by the users themselves. Self-regulatory solutions rely on transparency and control (Acquisti and al., 2016) that we described as *notice* and *consent*. Individuals should be able to inform themselves and manage their privacy to fit their needs. However, as we will see in the next section, privacy policies aren't optimal for users. They are complex to read, have huge opportunity cost because of the time it would take to read them, and they are nudged to make people accept them (Cranor, 2012). One example of self-regulatory approach would be the opt-out list *Do-Not-Track* giving a possibility to the user to manage his privacy.

The self-regulation model is interesting because as contrarily to simple regulation, it gives the opportunity to try to keep as much as possible the positive aspects of data sharing while still giving the power to individuals who want to protect their privacy. The more a person cares about privacy, the more he should dedicate resources to protect it. People who don't care about their privacy will continue to share their personal data and keep the positive effects. At the same time, if someone is worried about his privacy he should be increasing it. That is not exactly the case as some studies highlighted. While individuals say they care about their privacy, it doesn't coincide with them taking appropriate actions to match the level of privacy they declared to want (Jensen and Potts, 2005).

If data subjects want to improve their privacy they can take several actions that we will describe (Zaharia, 2016). First, they can improve their general security by having strong passwords on their accounts which help stopping intrusions. Second, they can improve their offline privacy. To do so, they can prevent and clean malwares using specialized programs. They can also keep their operating system updated or use a guest user account on Windows to decrease software vulnerabilities. They should turn on their User Account Control, so they can actively manage what happens on their computer. Finally, when concerned for online privacy enhancement, data subjects can use a secure browser that

is less likely to be attacked. Moreover, they can manage their browser privacy settings to fit their needs. They can reveal less information by being less active on social medias. They can protect information by having a critical mind while browsing, by avoiding suspicious websites, and by being careful of the web connection they use.

Data holders, employees and customers who want to improve data security can use lots of Privacy Enhancing Technologies (PETs) (Shen and Pearson, 2011). PETs can be used by companies to meet their legal responsibilities. They can provide users with anonymization which we saw is recommended to protect information: they protect users' identity and preserve privacy. PETs can also be used to protect against network invasion. Finally, they can be used to process data or to create a privacy policy and to enforce it. We can thus see that those technologies are ubiquitous for privacy.

6. Online privacy policies

In this part, we will cover online privacy policies. "*A privacy policy is a statement or a legal document (in privacy law) that discloses some or all of the ways a party gathers, uses, discloses, and manages a customer or client's data. It fulfills a legal requirement to protect a customer or client's privacy"* (En.wikipedia.org, 2018). When a person goes on a website, he has access to the privacy policy. This text is often showed at the beginning of the session because users need to give informed consent. However, we will see there are some problems with the way consents are given towards privacy policies.

6.1. Definition

We will here describe the typical content that privacy policies have. We will give an overview of what privacy policies structure looks like and what content should figure in each part. To do so, we will base ourselves on an article named "*How to write a Privacy Policy"* (How to Write a Privacy Policy, 2017), written by a former attorney.

6.1.1. Agreement and informed consent

A privacy policy is available on websites so that users can know what type of data is collected and how this is used. "*A Privacy Policy is an agreement that explains how websites collect, use, manage and disclose user data*" (Jocelyn, 2017). Having a privacy policy is mandatory for websites and apps in most jurisdictions in the world. The primary role of privacy policies is to notice the users over the activities done with their personal data. That is the first part necessary in the model of notice and choice. By being noticed over the use of their personal data, users should be able to make an informed choice whether they want to share their data or not.

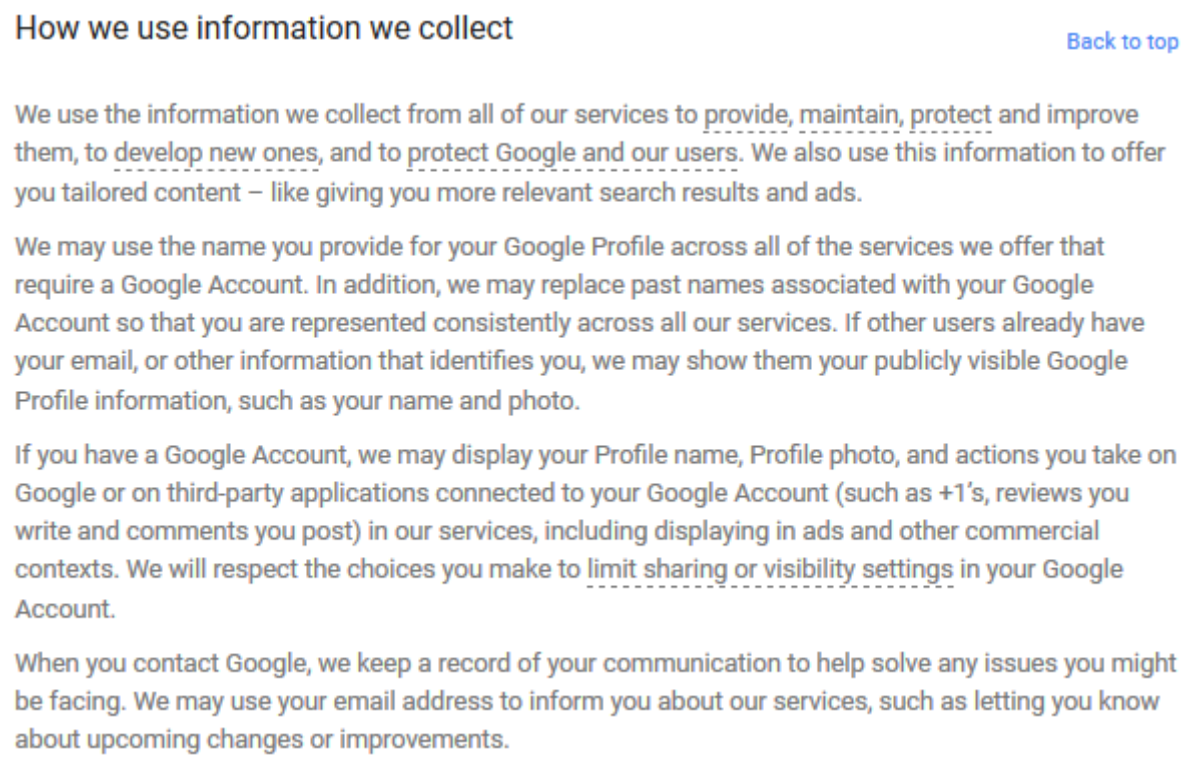
6.1.2. Information collected

In this section the website describes what it will do with the collected information. Websites often state they will use the data to improve their services or to personalize the experience of the users to their tastes. The conditional tense is used a lot and that can lead users to underestimate the perceived frequency of use.

6.1.3. Use of information

In this section the website describes what it will do with the collected information. Websites often state they will use the data to improve their services or to personalize the experience of the users to their tastes. The conditional tense is used a lot and that can underestimate the perceived frequency of use when the users read it.

Here we can see a sample of this part in Google's privacy policy[1]. It explains what they can do with what data.



How we use information we collect

Back to top

We use the information we collect from all of our services to provide, maintain, protect and improve them, to develop new ones, and to protect Google and our users. We also use this information to offer you tailored content – like giving you more relevant search results and ads.

We may use the name you provide for your Google Profile across all of the services we offer that require a Google Account. In addition, we may replace past names associated with your Google Account so that you are represented consistently across all our services. If other users already have your email, or other information that identifies you, we may show them your publicly visible Google Profile information, such as your name and photo.

If you have a Google Account, we may display your Profile name, Profile photo, and actions you take on Google or on third-party applications connected to your Google Account (such as +1's, reviews you write and comments you post) in our services, including displaying in ads and other commercial contexts. We will respect the choices you make to limit sharing or visibility settings in your Google Account.

When you contact Google, we keep a record of your communication to help solve any issues you might be facing. We may use your email address to inform you about our services, such as letting you know about upcoming changes or improvements.

We can see in the example that vocabulary and grammar are used to make practices seem less likely and more positive. For example, they use the words "*we may*" when explaining practices which makes them seem less likely. They also use terms such as "*improve*", "*help*" and "*inform*" which are all positive for the users.

6.1.4. Disclosure to Third Parties

Here, the website informs its users of the possibility to disclose their information to a third-party. The website also needs to specify how information would be shared. However, even if a website doesn't plan to share information to those companies, they still need to inform users of the possibility to share their information. It could come from legal needs if court requires the website to share information. It can also be needed to share some information to analyze the traffic of their websites with the help of another party.

6.1.5. Protection of information

In this section the website describes how they protect the information. In Europe, thanks to the GDPR, information protection will change in the coming months and anonymization and protection will be reinforced. This section of the privacy policy is generally not mandatory but is important to gain the confidence of potential users.

[1] https://policies.google.com/privacy#infouse

6.1.6. Rights of Users

Users need to be informed of their rights regarding their data. They typically are informed that they can review what data is kept about them. They are informed that they could make changes or delete the data. In some cases, it displays the right to reject cookies.

6.1.7. Notification of Changes

In this section the website declares that the privacy policy can be modified but that it would notify the users in this case.

6.1.8. Contact Information

Users who have questions can address them by contacting the website which provides contact information in this section.

6.2. Privacy policies problems

Letting the users deal with the problem of privacy lacks efficiency because of the privacy policies which are extremely complicated to understand for a great part of users. Moreover, the opportunity cost of reading online privacy policies is enormous (McDonald and Faith, 2008): it is time consuming and request an understanding of the legal and technical terms. We can assume that if the cost of reading the privacy policy is higher than the estimated value a user gives to his privacy, then this user won't even bother to read it.

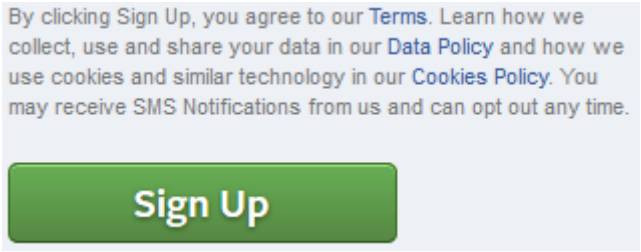
6.2.1. No good alternatives to acceptance

Alessandro Acquisti is one of the pioneers in the subject of privacy. He says that the model of self-regulation regarding privacy is not optimal (Acquisti et al, 2016). The biggest problem with online privacy policies is that their actual design doesn't give the proper choice to the users because they don't understand or read the privacy policies. We can observe two main different ways of noticing users of the use of data:

1) **Click to accept**:

It is generally expressed by a sentence such as "*By clicking Sign Up, you agree to our Terms and signify that you have read our Data Policy, including our Cookie Use*" and it only requires the user to click on a button to give consent.

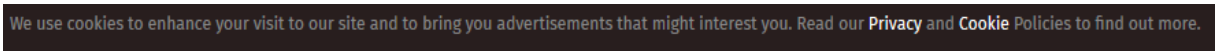
Example on *Facebook.com* sign up page.



If a new user wants to sign up on Facebook, he is forced to agree with their Data policy and Cookie use. There is no real alternative than accepting if he wants to join Facebook. We can easily see that the *Sign Up* button is more attractive than the text over privacy. That is to discourage people to read it, while still giving their consent.

2) Continue to accept:

Another way to notify users and get their consent is to say that by taking no contradictory actions, they give their consent to the privacy policy of the website. There are texts which appear while surfing on a site and give the choice to read privacy and cookies policies. "*We use cookies to enhance your visit to our site and to bring you advertisements that might interest you. Read our Privacy and Cookie Policies to find out more.*"[2]

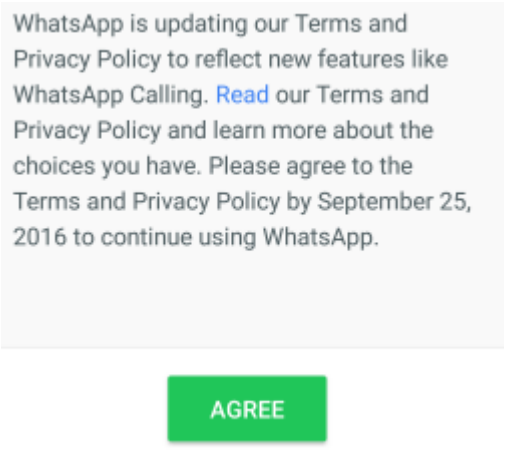


Clicking to read to cookie policy, the first sentence is: "*This page describes our cookie policy for independent.co.uk (the Website). If you do not accept this Cookie Policy please do not use this site*."

As we can see from the first line, there is no real choice for the user who wants to visit the site without agreeing to the cookie policy. Nevertheless, this site respects the EU regulation regarding notice of privacy and cookies policies. It will change with the GDPR which requires active consent that can be proven.

We can start to think that letting freedom to sites on how to notice and have consent from users will make them want to design the tools in a way that discourage users from reading them. As a reminder, The European regulation defines consent as "*any freely given, specific, informed and unambiguous indication of his or her wishes by which the data subject, either by a statement or by a clear affirmative action, signifies agreement to personal data relating to them being processed".* However, we just saw that consent is sometimes necessary for a user to keep using the website and supposed if he does. Users would need to take an active choice of leaving the websites if they disagree with the privacy policies.

For another example, here is a message that was shown to WhatsApp users in 2016 to inform them of changes in the Privacy Policy. To show disagreement with WhatsApp privacy policy changes, one needs to unsubscribe and thus lose all the advantages of the app. We can see that even if someone is worried about the use of his data by WhatsApp, it is unlikely that the worry about the data outweighs the positive effect, leading to the user trapped to accept the privacy policy changes.



Moreover, we can ask ourselves if the fact that people are needed to take actions to opt-out isn't pushing more people to stay in the default choice of acceptance. If individuals are rational, proposing them a default choice of acceptance shouldn't change the outcome. If people don't want to accept because the negative effects outweigh the positive effects for them, then they should stop using the site. We will consider this thought in the chapter 8.

[2] https://www.independent.co.uk/

6.2.2. Cost of reading privacy policies

In 2008, researchers tried to estimate the cost of reading privacy policies (McDonald and Faith, 2008). To do so, they used an opportunity-cost method. They began by estimating the reading time of all the unique privacy policies of the websites visited by a person each year. They concluded that if an American was forced to read those privacy policies, it would take him on average 201 hours every year. They then estimated the average financial value of this time –taking into consideration the different values of the time "wasted" at home and at work- multiplied by the time it would take to read them. They came to an estimated cost of $3.534 per year per American. We can think that the opportunity cost of reading the privacy policies for a person will rarely overpass the value associated to his privacy concerns. They then considered the number of Americans who surf online. If they were to all read the privacy policies, that behavior would have an opportunity cost of $781 billion per year for America.

6.2.3. Misunderstandings of privacy policies

We talked about the fact that people would rarely read the privacy policies because of the amount of time it would take them. But even if people read privacy policies, they aren't well informed about their information usage. Indeed, several researches highlighted that people face difficulties when reading the very long and technical privacy policies (Grannis, 2016). Moreover, the language used in privacy policies is misleading. Even if someone understood the complex privacy policy, it wouldn't mean that the intentions of uses of the data are well transmitted. It comes from the fact that privacy policies are often written by websites to respect the law rather than to inform the users in the best possible way (Pollach, 2007).

In 2000, the FTC highlighted 3 problems that lead to misunderstandings of privacy policies (Federal Trade Commission, 2000). First, websites tend to give general description explaining their privacy policies to make it quick to read for users. However, when reading the detailed text, the general description often goes with detailed exceptions that completely change the meaning. Secondly, ambiguous language is used to explain how users can express their choice. That can lead to users giving their consent while they wanted the opposite. Thirdly, changes in the privacy policy can lead to different uses of information that the user had refused before. In this case, they should be informed, and sometimes affirmative choice would be required.

6.2.4. Motivation of bad readability

Users are more confident to purchase in a website which has a privacy policy compared to one who doesn't have one. They are more likely to give information when they read and agreed to the privacy policies. However, the only presence of a privacy policy is said to increase the confidence of 2/3 of consumers, even if they didn't read the privacy policy (Earp and Baumer, 2003). It is thus in the interest of the sites to have a privacy policy but to make it difficult to read to avoid refusal.

Personal data is one of the key revenues for online companies, it is reasonable to think they have no real incentive to give appropriate notice of their data usage which would scare some users from agreeing to their privacy policies. It is founded to make this assumption when we can observe websites pushing users to quickly accept their privacy policies without reading them.

6.3. Privacy policies conclusions

Privacy policies are supposed to be the main element in the framework of notice and choice. The legislation is evolving, and the notice is improving. People are now offered to read privacy policies when they visit a website. However, the current privacy policies don't satisfy their primary role of informing people but are rather there to protect websites from facing lawsuits.

7. Possible solutions to improve self-regulation of online privacy

In this section, we will discuss solutions improving self-regulation of online privacy. They are designed to better inform the people by explaining how their data is generally used or by making privacy policies easier to read. Other solutions are designed to improve the choice that users face. The solutions will focus on improving information or improving the choices regarding privacy.

7.1. To inform users about the subject

We saw that people can be unaware of the collection and use of data made by companies and how important this market is. They could thus undervalue their privacy from the lack of knowledge on the subject. We could imagine some campaign of education which would inform people over privacy and personal data: how much is their data worth? What is their data used for? Then, they would have a better background to make their privacy decisions. Moreover, it could be interesting to increase the awareness of people on the existing means to protect their privacy. People who want to protect their privacy often lack the means to do so or are faced with more burden than the value they give to their privacy.

7.2. To improve the elements of online privacy policies

One of the ways to improve the current privacy framework is to simplify privacy policies. We saw that they are enormous and even if people would read them they are hard to understand. An article goes over the element of effective online notice (Grannis, 2016).

First, the *format* of effective notice implies that privacy policies should be easily readable and comprehensible. The actual format isn't attractive to read because it is a big pile of text. People don't really know where to look for information practices. Grannis proposes that privacy policies should be formatted in a list to reduce the cognitive load faced by users. In addition to the different format, it is advised to put headings or to organize key information in bullets to make it easier for consumers to locate what they are looking for. It could highlight important words. It would also be great if privacy policies were standardized so that users would be used to it and could find what they want quickly. Another idea would be to have a simple to read summary showed to consumers with a link to the full privacy policy in case they want more details.

Second, the *content* of effective notice should be more precise. Grannis says that we need accurate disclosures. It should be clearer what type of data websites collect and the way they use it. They should never contradict what they put in their privacy policy. If they say they would never share information, then they should respect it. Moreover, the language used should be precise. Generic terms should be avoided and replaced with more precision. Vague and ambiguous phrasing and grammar should be avoided. All those recommendations are interesting, and we can think that those elements will be implemented with time.

Furthermore, some projects were already created in the goal of improving privacy policies. The **P**latform for **P**rivacy **P**references **P**roject (P3P) was a project launched in 2002 designed to give websites the possibility to display their policy practices in a standard way (W3.org, n.d.). With this standardization, privacy policies could be analyzed by some programs. It could automatically deduce what the user wants for each website. This project has similarities with the element of effective notice because it proposed standardized privacy policies making them easier and faster to read and analyze.

However, this project was only applied on internet explorer and stopped working since Windows 10 doesn't support it. It was nevertheless an interesting idea and future similar projects could be applied.

7.3. Giving a real alternative

In the current privacy background, when users go on a website they are asked if they consent to the privacy policies. If they don't agree, they are asked to quit the website. That is not a great alternative, and some will say it is an unfair choice. If people don't want to accept parts of the privacy policy, they should have an easy way to apply that disagreement. For example of better opt-out solutions, the European ePrivacy directive advise to centralize into the browser either the consent or the absence of consent from users regarding the use of intrusive tracking devises or cookies. They have an easy way to notify their choice and don't need to repeat it in every websites. In the US, the FTC proposes a "*Do-Not-Track*" which would automatically signal to website if a user doesn't want to have his data tracked. Those solutions simplify the way users can protect their privacy. Moreover, websites who give opt-out solutions increase the confidence of consumers (Earp and Baumer, 2003). Having opt-out solutions could increase the sharing of reasonable information coming from an increased confidence of consumers. At the moment, a consumer is faced with a decision between accepting everything or nothing. If people could manage their data sharing and give partial agreements, some people who otherwise wouldn't give any data could give some information.

7.4. Active choice

For now, some decisions taken to regulate people's privacy make them take passive decisions. Continuing to visit a website, thus communicating consent of the privacy policies, is the main example. When faced with complex choices regarding privacy, individuals would prefer to not deal with it. Now, if they don't deal with it they accept the privacy policies. The opposite situation would be that by default everyone refuses privacy policies and needs to testify that they accept it, but that could be costly for the society as it would diminish data sharing. What we think is interesting to look for is to make people take active decisions regarding their privacy. The active decisions need to be taken in a framework where the other improvements are implemented which would permit a more informed choice.

Active decisions are optimal when people are well informed over the subject and have the capacity to deal with the decision (Carroll et al., 2009). Active decisions are good to counter procrastination. We saw that people tend to say they care more about privacy than what their online behavior suggests (Jensen and Potts, 2005). Privacy can be compared to the example of savings: people want to save more money than they do. It would be interesting to find a way to push people to take a more active and informed choice. By informing people over the value of their data, they would probably care more about their privacy. By improving notice and by making privacy policies easier to read, the complicated choice they face would be easier for them and people should thus have the capacity to deal with their privacy concerns. However, we saw before that people tend to be passive. We think that those solutions should be coupled with a nudge to make people more active regarding their self-regulation of privacy.

8. Proposition of nudge

In the recent years, the economic theory has changed its hypothesis on rationality. It tries more and more to consider irrationalities of individuals in their choices. The field of behavioral economics recognized the fact that people aren't always rational like the traditional economic theory assumed. Individuals can procrastinate, overvalue the present or tend to be passive. In 2017, Richard Thaler, pioneer in behavioral economics was awarded the Nobel Prize in Economic Sciences (Nobelprize.org, 2017). This event showed the recognition of this recent switch in the economic theory. In this part, we will talk about the passivity of the individuals and discuss how this should be addressed in the case of self-regulation of online privacy. There are two active choices that one needs to take while dealing with online privacy:

- The user needs to take time to read a hard to understand legal text.
- The user needs to either accept or opt-out of the privacy policy. That is sometimes only possible by taking undesirable actions.

There are a lot of propositions to try to improve the efficiency and readability of privacy policies. It is going in the right direction to give users the possibility to take an informed decision. However, we think it is interesting to develop a nudge with those improvements, which would render people more active with their privacy management.

8.1. What is a nudging?

"*Libertarian paternalism is a relatively weak, soft, and nonintrusive type of paternalism because choices are not blocked, fenced off, or significantly burdened. (…) [Libertarian paternalists] are self-consciously attempting to move people in directions that will make their lives better. They nudge*." (Thaler and Sunstein, 2008). Nudging is applying a nudge, trying to push towards a desired behavior for those who want it while trying to deter the least the situation of those who don't want to adopt this behavior. Nudging shouldn't reduce the number of options available but could improve people's behaviors. It must be done without coercion, by making it easy and simple to choose between the alternatives.

8.2. Irrationality

8.2.1. Default effect

Public policies often have a default choice; the choice that would apply if the individual doesn't take any action. The results coming from different default choices shouldn't be different if individuals are rational and if the cost of expressing their choice is small. That is very important to have small cost of choice, people shouldn't be trapped to stick to the default if they prefer the alternative but that can happen if the cost of changing to the alternative is higher than the gain.

A common example of effective nudging is changing the default of organ donation. What is interesting in this topic is that preferences aren't matching the actions: people say they want to be organ donor - 85% of Americans- while only 42% of those had expressed their will (Thaler and Sunstein, 2008). In 2003, in a study of Johnson and Goldstein over organ donation (Johnson and Goldstein, 2003) and the effect of the default choice, it was highlighted that by changing the default choice (being an organ donor versus not being a donor) the results were different. When the default choice was "*not being a donor*", only 42% agreed to become one while when the default was "*being a donor*" 82% stayed donor

(appendix 3). One could think of a logical explanation, that the cost to change was too big and that made people stick to the default. However, in the experiment, people were asked on a website if they wanted to be an organ donor or not, and they just had to click to express their choice. That is called the default effect.

Another example of the use of nudges is the one of retirement savings. Retirement savings are a current subject for different reasons. In most countries, pensions are paid by actual workers and given to the retirees who worked beforehand. However, because people are living longer and have on average fewer children, the ratio of the number of people needing a pension divided by the number of active workers will increase. That means that either workers will need to pay more tax or that retirement benefits will be lowered. To ensure that they can have enough money for their old age, workers will need to save during their carrier. When asked if they think they are saving enough for retirement, 68% of workers said that their savings rate is too low (Thaler and Sunstein, 2008). Saving can be compared to exercising, dieting or protecting their privacy: a lot of people say they want to do it but a fewer percentage does it. People need to have a nudge to push them towards higher savings.

8.2.2. Solutions

Fortunately, governments can push people towards what is better for them. They can make savings the default choice while still giving the possibility to opt-out of the saving program. Making savings automatic increases workers participation to savings plan (Choi et al., 2004). Another way of increasing the enrollment rate from the opt-in approach is to force people to make a choice. A company tried to go from the opt-in approach to the active decision approach for the 401(k)-saving plan. The active decision approach doesn't favor the acceptance or the refusal of the plan because neither are the default choice anymore. Instead, it forces people to take a decision about their savings. It showed that the participation rate increased by 28 percentage point (Carroll et al., 2009).

8.2.3. Possible explanation

One possible part of the explanation of the default effect, meaning that people tend to non-reasonably stick to the default, comes from loss aversion. People value an amount of loss bigger than the same amount of gain (Tversky and Kahneman, 1991). If changing from a situation to another is considered by an individual as a loss of its default situation, then that could explain the default effect. Indeed, imagine an individual is indifferent between bananas and apples; meaning he values them the same. If you give him an apple, he won't accept a trade for the banana as he would perceive the loss of the apple as bigger than the gain from the banana. That is called the *endowment effect*: the absolute gain of utility from receiving something is lower than the absolute loss of utility from losing the same thing. People give more value to things when they own them (Kahneman, Knetsch and Thaler, 1991). And from that we can deduct that the reference point matters, meaning that the same choices can have different outcomes when the status quo is different. However, not everyone agrees on this explanation and for now default effect isn't scientifically proven, but it is empirically accepted.

8.2.4. Irrationality and privacy

In the case of privacy, the default is that users accept the privacy policies by continuing to use the website after they have been noticed of them. More likely than they should, people will be passive and take no action to opt-out of the privacy policies. They will thus be giving consent more than they would than if there was no default effect.

Moreover, the weight of the endowment effect could play a role in users' decision to continue to use a site like a social network after it changed its privacy policies. We can consider someone who gives the same weights to the gain of utility from using a social network website and to the utility from keeping the privacy necessary to use the site. He is asked by the social network to reveal his place, age and activities. This person is fully aware of the privacy consequences of the subscription. If he is asked if he wants to subscribe to the social media, he will be indifferent between the two alternatives. It is because he weighs the privacy as much as the gain from the use of the social network.

We now consider that this person was already using the social network. The social network has now changed its initial privacy policy, saying he will take more data to improve its targeted advertisement. If the person didn't previously subscribe to the social network, he would choose to stay away. However, he is already using the website and sees canceling the subscription as a loss. He will thus give more weight to staying on the social network due to the endowment effect and he could accept to give more data than before.

This example could explain why procrastination when dealing with privacy policies favor the websites which have thus no incentive to push the potential users to do an informed choice before subscribing. We can suppose that people knowing there is a privacy policy will think that they will read it later and see if they prefer to stay on the site or not. In fact, that behavior would irrationally favor the use of the website.

8.3. Does online privacy choice need a nudge?

We have seen that people can be irrationally passive in certain situations and stick to the default. We will furthermore analyze the case of online privacy self-regulation decisions. In a book over nudge (Thaler and Sunstein, 2008), the authors go over "*When do we need a nudge?*" and give different characteristics which increase the need of a nudge. I will here review the different characteristics in the online privacy framework.

8.3.1. Benefit Now – Costs Later

In some situations, people receive immediate benefit from an action while the costs of this action come later. For example, they enjoy eating ice creams but will have future health problems. People are shown to be irrational in those situations and give few weights to future events. They have self-control problems. In those kinds of situations, it is interesting to structure choices to help people doing what is best for them.

For the case of online privacy self-regulation, people are asked immediately if they accept or not to share their information to the website they want to use. By giving their consent, they gain an immediate benefit of being able to use the website. That is at the price of potential future negative events such as their information being shared to third parties, their identity being hacked, etc. Because they give less credit than they should to future events, they are more likely to give consent to privacy policies. This effect is accentuated by the language used in privacy policies which are written using conditional tense and ambiguous language. It diminishes even further the perceived probability of future events.

8.3.2. Degree of Difficulty

The more difficult the choice faced by people, the more they need help to do this choice. People usually don't need help to choose if they will eat rice or potatoes, but they are more likely to face difficulties to know how to invest their money between the stock market and bank accounts. That is because investments include risk and uncertain outcomes and because people lack knowledge about it.

In the case of online privacy, people need to compare the benefits and the negative effects of sharing their data. That is a difficult choice because people lack information and need to use shortcuts and estimations. They are not educated enough about privacy. They could read privacy policies to try to have a better knowledge of the use of their data, but those are very hard to read and understand. Even if they understand the exact use of their data, it is hard for them to compare the value they give to their privacy to the benefit they receive from using the website. Indeed, as we saw before, the valuation of their data is context dependent and thus their decisions could be easily influenced. When the choice is difficult, people tend to be more passive and will thus more often stay with the default choice.

8.3.3. Frequency

Nudging is more important when a task is infrequent because people are more likely to make bad choices. They have less experience dealing with those situations and have thus spent less time thinking about it.

In the case of online privacy, the situation is frequent because people are faced with privacy policies on close to every websites they visit. However, they don't actively deal with the choice as they quickly give their consent without thinking about it. We can thus say that the situation is to some extend infrequent, but with proper adjustments it will become easier and easier for users to choose whether to give their consent.

8.3.4. Feedback

A nudge is more likely to be needed in situation where no feedback is given on the choice individuals make or when the feedback takes a long time to arrive. Imagine people simply asking someone to throw a ball. He is in fact expected to throw a ball the closest possible to a ten meters mark. He can throw the ball 1, 10 or 1000 times, if he doesn't have any feedback on his throws he won't improve.

In the case of privacy, when people give their consent to share their data on a website, they often don't receive any news from this consent. Even if they realize the advertisements are exactly the type of clothes they just looked to buy, and they feel like their privacy was violated, they would have a difficult time to remember which decisions lead to this outcome. Only very particular events give some feedback on the online privacy practices. The latest big event which could be considered as a feedback concerned Facebook. Big news outlets have highlighted the practice of Cambridge Analytica, a firm which analyzed data for Donald Trump's campaign. We already knew that they took the data of 87 million accounts (Badshah, 2018) but what is new knowledge is that Facebook was aware of those practices (Romano, 2018). When using certain apps and connecting with Facebook, users needed to give their consent to share not only their data but also data of their friends. Cambridge Analytica used the data of its users plus those of their friends to create a profiling system. They delivered it to help political campaigns that were able to target personalized political messages. Following that news, some people decided to delete their Facebook accounts. They did it either to protect their privacy or

to show their disaccord. We can see that scandals can put a light on the problem of privacy and make people realize that their choices have consequences.

8.3.5. Knowing What You Like

Nudging is more needed in situations where people have difficulties to know the outcome of their choices. For example, when travelling to an exotic country and eating at a local restaurant, people don't know what to expect from local dishes and it is thus a good idea for them to ask for advices.

In the context of privacy, people aren't sure of what they give their consent to. That comes from ambiguous privacy policies and lack of knowledge. That is why our recommended solutions include an improvement of the privacy policies and education over privacy.

8.3.6. Conclusion on privacy and nudging characteristics

We saw that the choices concerning privacy have a delayed effect and are difficult. People receive poor feedback and have trouble knowing what their choice will lead to. All those characteristics are pushing toward a need of nudging. The only characteristic that wouldn't push towards it, is that the decision is frequent and thus individuals should get better and better at it. However, people currently don't think about their choices and because of the other characteristics they don't get better at doing it.

8.4. Advantages of using a nudge

The biggest advantage of nudges is that while they have a low cost of implementation, they can have huge impacts. For example, in the UK they tested a nudge to make people pay more of their due court fines (Haynes et al., 2013). They simply sent different text messages in different groups of people and compared the payments with those of a sample of people who didn't receive any message. Sending a text has a cost close to zero and shouldn't change the decision of rational individuals. However, receiving a text significantly changed the average amount that would be paid: those who didn't receive any text paid on average £4.46 while the average of the group who was sent personalized text messages paid on average £12.87 (for the most efficient message). Such a difference multiplied by the numbers of currently unpaid fines becomes quite interesting for a low cost of implementation. Another example was when a company sent Home Energy Reports letters to US citizens, comparing their consumption to the one of their neighbors (Allcott, 2011). Those who received the letters reduced on average their energy consumption from 1.1% to 2.8%. With a very low cost of implementation (sending letters), we could potentially significantly reduce energy consumption. Those examples show the huge potential of nudges: when they are used wisely they can have huge return on investment.

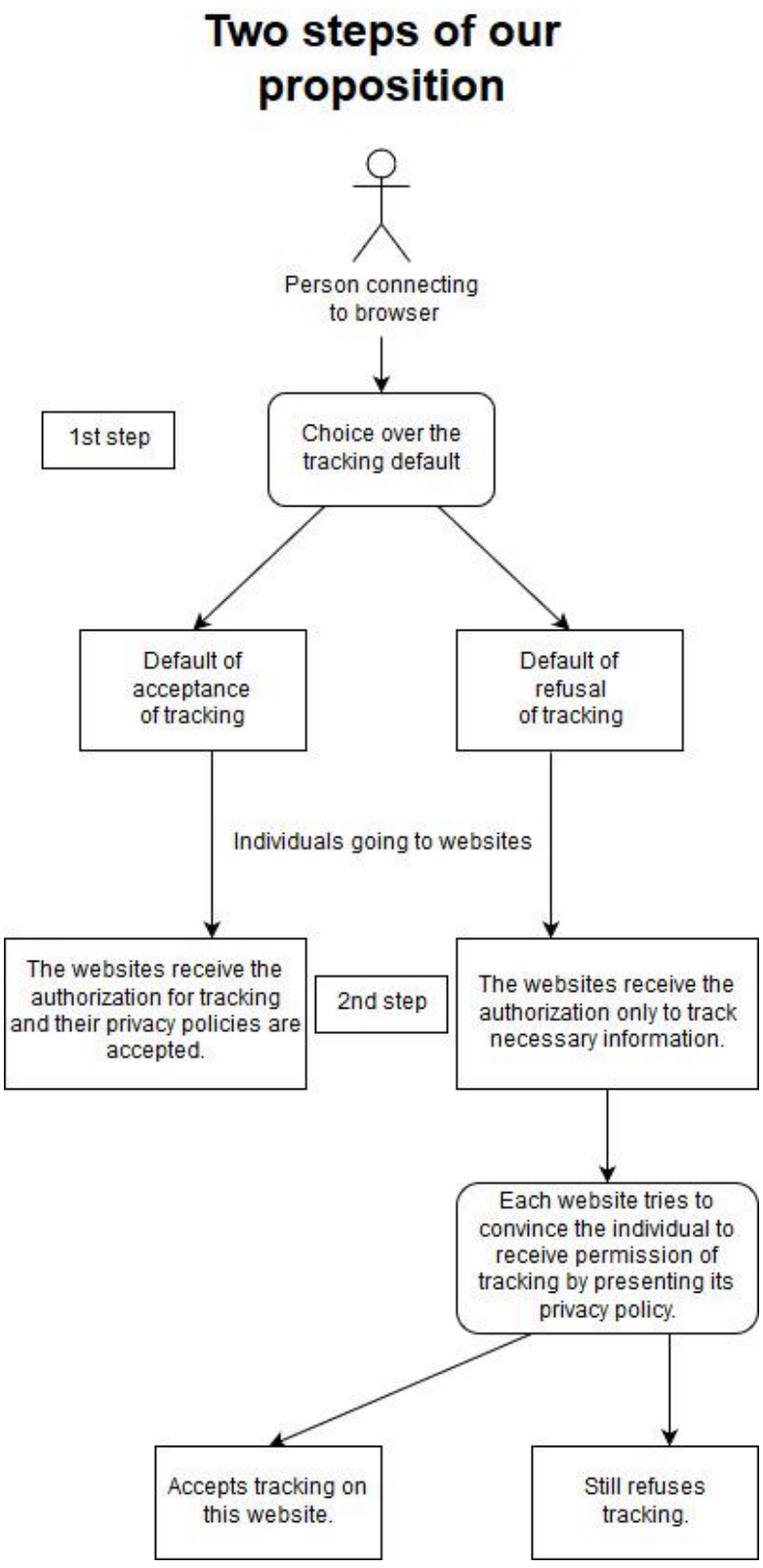
Moreover, nudges should improve the situation of a part of the population without deterring the situation of the others. Indeed, receiving letters doesn't cost much to those who really don't want to change their behavior while they have significant impact on people who simply didn't know that they had better alternatives.

8.5. Nudges for online privacy

In this section, we will propose different nudges that can be interesting to improve self-regulation of online privacy. The nudges themselves wouldn't be sufficient. Our propositions would be coupled with a better education in the domain of privacy. That is not unreasonable to imagine as it is advised to do so by number of people and the privacy debate gained exposure with the implementation of the GDPR and the Facebook drama. We should particularly try to make people realize the value of their data and the way websites and third-parties use shared data. People could make improved decisions with that knowledge. As we saw in this thesis, there is no linear relation between privacy and welfare of the society. Instead, the relation on social welfare is unclear and could go both ways. However, in addition to the trade-offs of privacy, we highlighted the importance of preserving privacy for individuals to isolate themselves from social norms and control. We also showed the potential problems such as default effect, procrastination and loss aversion that could explain an unreasonable consent rate which doesn't reflect individuals' preferences. We would like to give more power to individuals and give them another framework to either consent or refuse privacy policies.

8.5.1. Make the websites convince people

Our idea is to potentially change the default from acceptance of tracking to refusal of tracking for people who want it. It would then incentive websites to make clear privacy policies to convince people to share their data. That would only concern intrusive tracking, that is not necessary (to target ads, products and services) for the good functioning of the websites, as in the idea of the ePrivacy Directive. Our idea should make people refuse tracking more often than in the actual framework, where we saw that they tend to share more than they want. To make it clear, our idea would apply in two different steps.



On the first step, people will face a choice which will define their tracking default. In the actual framework, the default choice when people visit a website and ignore privacy policies is that they accept them and the tracking of their data. In our proposition, people will firstly face the choice of

what their default (meaning what happens when they don't do anything new regarding privacy) will be when they enter a website. They will have the choice to set their default as either acceptance or refusal of tracking. If they choose *acceptance*, whenever they enter a website they will automatically accept the privacy policy and tracking of their data. On the contrary, if they choose *refusal* they will automatically, by default, refuse tracking that are not necessary for the website to operate.

On the second step, if people have chosen *acceptance*, nothing new will happen regarding their privacy and they can continue to surf online without being bothered by privacy policies. However, if they choose *refusal* on the first step, they will still take decisions. Indeed, by default, they don't accept intrusive tracking. But our idea is that websites being visited by people who choose refusal by default will have the opportunity to convince those people to change their mind and give their acceptance to the website. To do so, the website will explain their privacy policies and the advantages of tracking to try to convince the *refusal people* to share data on their site. They would have all incentives to explain their privacy policies clearly and quickly to make it easy to read. That could solve several privacy policy problems that we described.

By applying our proposition, we want to know the difference on the level of tracking and privacy when we give the choice of the default compared to when people need to either accept or refuse privacy policies like now. Particularly, we want to know if people have a higher level of privacy with that proposition, which would better match their concerns.

8.5.2. Choice on the first step

In the first step, people won't be forced to refuse tracking of their data, but they will choose the default. They will be asked this choice on the browser such that it remembers it and signals the choice when a user visits a website. The choice could be presented as follows:

"Websites and third-parties (such as advertisers) can track your data to develop new technologies, to adapt the content you see in function of your tastes or to offer ads that are more likely to interest you. However, intrusive tracking isn't needed for websites to operate. New privacy regulation gives the possibility to choose if by default:

(1) you accept the tracking of your data which will give you a personalized online experience.

(2) you would like to only share data necessary for a well-functioning online experience.

This choice will be your default on all websites. If you want to share few data by default, you will still be offered to accept to share more data on a website when you first visit it.

- I want a personalized experience. (1)
- I want to share few data. (2)"

To help visualize what it could look like, we realized a possible design of the choice faced by people when connecting to their browser for the first time after the regulation changes (appendix 4):

Websites and third-parties (such as advertisers) can track your data to develop new technologies, to adapt the content you see in function of your tastes or to offer ads that are more likely to interest you. However, intrusive tracking isn't needed for websites to operate. New privacy regulation gives the possibility to choose if by default:

(1) you accept the tracking of your data which will give you a personalized online experience.

(2) you would like to only share data necessary for a well-functioning online experience.

This choice will be your default on all websites. If you want to share few data by default, you will still be offered to accept to share more data on a website when you first visit it.

Default: (1) I want a personalized experience (2) I want to share few data

We will now explain the reasons for the content of the proposition. We try to sincerely highlight the advantages of data sharing to keep the "data sharing friendly" users in the default of acceptance of tracking. For the introduction, we need to find a good compromise between highlighting the advantages of data sharing without pushing people who care about privacy towards acceptance of tracking. We will thus offer different propositions that will likely have different impact on the choices.

For the buttons, if we ask people to press the buttons named (1) *I want to be tracked* or (2*) I want to protect my data*, we think that even people who care few about privacy would still click on the choice to protect their data. We thus tried to choose a neutral tense. We think that testing different buttons to see the differences on the degree of acceptance is important.

As we highlighted in this thesis, information sharing has huge benefits and if people knowingly want to share their data we shouldn't stop them from doing so. However, we tried to make our message such that it doesn't push too much towards the acceptance of tracking. Indeed, those who care about privacy will accept to spend some resources (the time to accept or refuse the argumentation of the websites) by choosing the default of refusal of tracking if they perceive that the management of their privacy is more important to them than the resources spent.

The difference with the actual online privacy framework would be that people will be forced to make a choice. This choice doesn't cost them a lot of time as they only need to read a onetime message in the browser and click on their preferred solution. Moreover, people are currently passive, and websites have no incentives to render them more active which leads to the default choice of acceptance of privacy policies and tracking. With our proposition, websites will need to convince people who chose option (2) if they want to have their consent over intrusive tracking. People who accept tracking for whatever reason ("*I have nothing to hide*" or "*I don't want to lose time reviewing privacy policies*") generally wouldn't refuse any privacy policies or fight tracking in the actual framework anyway, and it would thus be a gain of time for them to not be confronted with privacy policies.

The people who will choose option (2) are those who give more value to privacy. However, we don't want to render impossible their tracking on every websites. It is reasonable to imagine that people

accept to share more information for some websites if they want to have more personalized contents. For example, they would accept to share their information to have access to offers that match their predicted preferences on some buying websites. That is why the second step of our proposition is to let websites convince people who by default refused tracking.

8.5.3. Different design of the first step

The goal of this section is to find key part of the design of the choice faced by individuals in the first step. We could change them to have different results on the degree of acceptance of tracking and thus on the degree of data sharing and privacy. We will first change the design of the **introduction**. Then we will change the design of the **buttons**.

As a reminder, our main proposition is:

> Websites and third-parties (such as advertisers) can track your data to develop new technologies, to adapt the content you see in function of your tastes or to offer ads that are more likely to interest you. However, intrusive tracking isn't needed for websites to operate. New privacy regulation gives the possibility to choose if by default:
>
> (1) you accept the tracking of your data which will give you a personalized online experience.
>
> (2) you would like to only share data necessary for a well-functioning online experience.
>
> This choice will be your default on all websites. If you want to share few data by default, you will still be offered to accept to share more data on a website when you first visit it.
>
> Default: (1) I want a personalized experience (2) I want to share few data

First, we would like to test if the introduction has a positive impact on the percentage of people accepting default of tracking.

Hypothesis 1: an introduction presenting the benefits of tracking pushes people to the default of acceptance of tracking.

To test this hypothesis, we create the exact same choice with the only difference being that the introduction will now be neutral.

Our **1st alternative proposition** to test hypothesis 1 will be the same proposition except for the part "*Websites and third-parties (such as advertisers) can track your data* ~~***to develop new technologies, to adapt the content you see in function of your tastes or to offer ads that are more likely to interest you***~~."

> Websites and third-parties (such as advertisers) can track your data. However, intrusive tracking isn't needed for websites to operate. New privacy regulation gives the possibility to choose if by default:
>
> (1) you accept the tracking of your data which will give you a personalized online experience.
>
> (2) you would like to only share data necessary for a well-functioning online experience.
>
> This choice will be your default on all websites. If you want to share few data by default, you will still be offered to accept to share more data on a website when you first visit it.
>
> Default: (1) I want a personalized experience (2) I want to share few data

By removing the positive introduction, we can compare this alternative to our first proposition and see if the difference in the percentage of acceptance is significant. It is important to notice that the explanation of the default still says that they would receive "*personalized online experience*", people are still made aware of the advantage of tracking, but we just put less emphasis on it. This alternative will serve us to test hypothesis 1, as we will describe in the section over the evaluation of the nudge.

Second, we would like to test if using concrete examples that are likely to touch people in the introduction would turn up acceptance.

Hypothesis 2: in the introduction presenting the benefits of tracking, giving known websites as examples rather than a generic positive introduction over tracking, should increase the percentage of people that take the default of acceptance.

Our **2nd alternative proposition** will be the same as our main proposition, except for the positive introduction which will be replaced by another positive introduction giving concrete examples to test hypothesis 2.

We will replace "~~***Websites and third-parties (such as advertisers) can track your data to develop new technologies, to adapt the content you see in function of your tastes or to offer ads that are more likely to interest you***~~" by "*The tracking of data renders possible, among others, the possibility for some websites such as search engines (Google, Yahoo) or social networks (Facebook, Instagram, Twitter) to make money and to stay free-to-use for users*".

> The tracking of data renders possible, among others, the possibility for some websites such as search engines (Google, Yahoo) or social networks (Facebook, Instagram, Twitter) to make money and to stay free-to-use for users. However, intrusive tracking isn't needed for websites to operate. New privacy regulation gives the possibility to choose if by default:
>
> (1) you accept the tracking of your data which will give you a personalized online experience.
>
> (2) you would like to only share data necessary for a well-functioning online experience.
>
> This choice will be your default on all websites. If you want to share few data by default, you will still be offered to accept to share more data on a website when you first visit it.
>
> Default: (1) I want a personalized experience (2) I want to share few data

In our main proposition, we talked about the advantages of data sharing. The goal is that people who weakly value privacy will take the acceptance of tracking. Here, to realize the same goal, we inform people about the fact that data sharing is the principal reason why a lot of websites are free. We highlight differently the benefits of data sharing and we want to know the impact of changing the introduction on the degree of acceptance.

Third, we would like to know if the sentences on the buttons have an impact on the individuals' choices.

Hypothesis 3: the sentence on the button has an impact on the percentage of people choosing tracking acceptance.

Our 3rd alternative proposition which will serve to test hypothesis 3 will be different of our proposition only by the sentences on the buttons. We will thus replace the sentence of the second button: ("***~~I want to share few data~~***") with a sentence reminding the word tracking: ("*I refuse to be tracked*")

> Websites and third-parties (such as advertisers) can track your data to develop new technologies, to adapt the content you see in function of your tastes or to offer ads that are more likely to interest you. However, intrusive tracking isn't needed for websites to operate. New privacy regulation gives the possibility to choose if by default:
>
> (1) you accept the tracking of your data which will give you a personalized online experience.
>
> (2) you would like to only share data necessary for a well-functioning online experience.
>
> This choice will be your default on all websites. If you want to share few data by default, you will still be offered to accept to share more data on a website when you first visit it.
>
> Default: (1) I want a personalized experience (2) I refuse to be tracked

We suppose that the use of the word "*tracked*" will increase the number of people choosing the second option, thus our 3rd alternative should have a lower percentage of acceptance than our proposition.

Fourth, we would like to know if the order of the buttons has an impact on the individuals' choices.

Hypothesis 4: the default put in the first place is more likely to be selected than if it is put in the second place.

Our 4th alternative will be the exact replica of our proposition except that the order of the button and the explaining sentences will be swapped.

> Websites and third-parties (such as advertisers) can track your data to develop new technologies, to adapt the content you see in function of your tastes or to offer ads that are more likely to interest you. However, intrusive tracking isn't needed for websites to operate. New privacy regulation gives the possibility to choose if by default:
>
> (1) you would like to only share data necessary for a well-functioning online experience.
>
> (2) you accept the tracking of your data which will give you a personalized online experience.
>
> This choice will be your default on all websites. If you want to share few data by default, you will still be offered to accept to share more data on a website when you first visit it.
>
> Default: (1) I want to share few data (2) I want a personalized experience

We think that the disposition of the two default influences the choices. Indeed, the first default to be presented will maybe be perceived as the logical choice by individuals. We would like to know if the disposition plays a role by comparing this alternative 4 and our main proposition.

We have 4 different alternatives that will likely give different level of acceptance of general consent. Before we implement our proposition for the choice of the 1st step, it is important to know the impacts of the different messages and design on the choices of the respondents. We will describe a way to test our hypothesis in the section 8.7. "*Evaluation of the nudge*".

8.5.4. Convincing message in step 2

We have proposed a design for the choice of the first step of our proposition. As explained, people choosing the default of tracking refusal will still be able to change their mind afterwards. The second step is done on each website, case by case. When going the first time to a website after having picked the refusal default, the user will be presented the website's privacy policy and an explanation of why tracking is important for them. Websites thus have a chance to change the person's choice and make him accept intrusive tracking on the website. The messages addressed to users would be regulated to be honest and with as few misleading languages as possible. We highlighted in this thesis the problems of privacy policies of voluntary making the text unclear, confusing and hard to understand. It wouldn't be in the favor of the websites anymore. Instead, they would have incentives to make privacy policies clear and readable by the users. Indeed, users will not grant permission to tracking if the privacy policy is unclear.

8.5.5. Limits of the nudge

This new framework has some limits. First, it is possible that information sharing will diminish too much from the fact that people who care about privacy will now take actions to match their concerns by refusing more tracking. That is why we propose different messages put before the choice on the first step. We want to test them to see the impact they have on the percentage of people choosing acceptance or refusal of tracking. Authorities could then choose the message accordingly with a desired level of privacy and data sharing. Moreover, websites would need to spend resources to try to convince people to accept their privacy policies and authorities would need to control if privacy policies aren't embellished to make people accept tracking.

8.6. Implementation of the nudge

Before being able to generalize our proposition, we would need to prove that it is used and appreciated by people. To do so, we would begin by contacting programmers to design an add-on considering the ideas of our end proposition. Then, people can download the add-on to apply it on their browsers. Several add-ons and apps to improve privacy already exist and some of them are popular which means that people are interested in using tools to protect their privacy. One of the most famous is "*AdBlock plus*" used on 615 million devices in the year of 2016 (Cortland, 2017). It is an add-on which blocks intrusive ads, malwares and third-party tracking cookies and script. Another famous tool is "*Disconnect Private Browsing*", a browser extension which protects against privacy harms such as tracking, malware, malvertising and third-party tracking cookies (Henry, 2015).

One of the existing privacy feature would be a very good basis to implement our proposition. In 2011, Firefox began to offer a *Do-Not-Track* mechanism. Approximately 10% of users have set their *Do Not Track* feature on (Jason, 2015). When those users are visiting a website, it signals the website, their advertisers and content providers that users don't want to be tracked (Support.mozilla.org, n.d.). The *Do Not Track* feature can be turned on in the different browsers supporting it: Firefox, Safari, Internet Explorer (activated by default on IE 10) and Opera (wikipedia.org, 2018). We could offer to implement our proposition to this feature and create our add-on from this basis. In this case, when users with the add-on go on a website, it can display its convincing privacy policy that tries to make them accept tracking.

The big flaw of this feature that would still be present in the add-on, is that it cannot force websites to respect the desire of people to not be tracked (Bradley, 2011). It is only accepted by some websites on voluntary basis. However, if the feature coupled with our proposition gains popularity and receives positive feedback, we would then offer authorities to make it mandatory for websites to respect it. To gain popularity, browsers could promote it and we could send messages to offer people to download the add-on. We could send it by mail, forums and talk about it on the internet. At the end, we could apply our proposition by having a message for every user when they connect to their browser asking if they want or not to be tracked.

8.7. Evaluation of the nudge

Evaluation is an important part of every project. It gives the possibility to rectify potential issues, to know what worked well and to improve the product. First, we want to know if our proposition help people to better match their concerns, meaning that their level of privacy will be higher. Second, we want to evaluate the impact of the different messages that are addressed to people regarding the frequency of acceptance and refusal.

The method we recommend using to test the effect of our proposition is Randomized Controlled trial (RCT). The idea of RCT is that by having a big enough sample, we can on average have similar groups. The principal advantage of RCT is that people are randomly assigned to treatment and control group and that can solve the selection bias if the numbers are big enough (Johnen, 2017). We can then compare the difference of outcomes and see if there is a link between the treatment -our nudge- and the outcome -having more privacy- (Sibbald and Roland, 1998). We can also test the different messages by sending them to different groups and analyze the differences of behaviors. We could then see the effect of the different propositions on the percentage of people who accept or refuse general acceptance of data sharing.

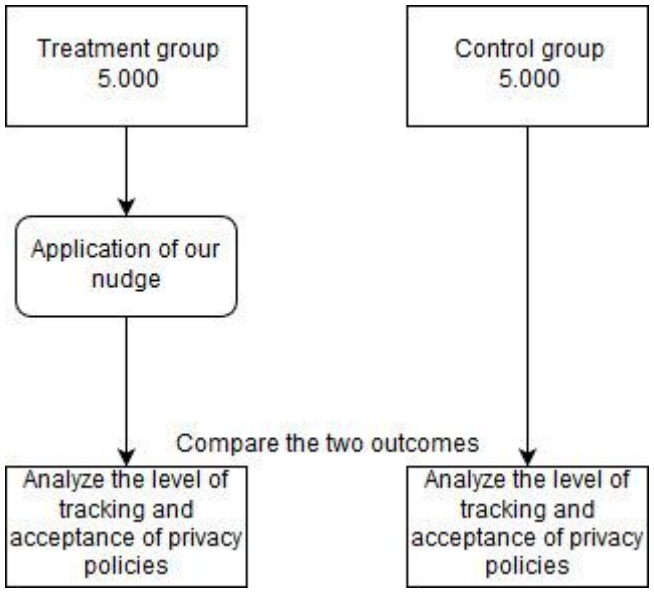
If our idea is generalized, we can first take a random sample of participants (30.000 people would be our goal) to test the different designs of our proposition and the impact of our nudge on privacy, before the actual generalization to all users. We can compare the effects of the different messages by giving each proposition and alternative to 1/6 of the sample (5.000 people each), thus using 25.000 people of the sample. There shouldn't be any bias as it would be randomized. We could thus test our hypotheses for the degree of acceptance. Then, to see the impact of the measure on tracking compared to the actual framework, we will use a control group of the remaining 5.000 people of the sample to be able to compare their behavior with the rest of the sample as we will explain in the next section.

8.7.1. Effects on tracking

We will here present the method to analyze the impact of our proposition on tracking. We will explain how we would do it before the generalization of our proposition using a sample divided in 6 groups. Then, we will explain how we could already gather data with our add-on, despite having some problems.

Test if generalization

As we described before, we can use the sample of people that received the treatment (our proposition) to see the effect on tracking. We would compare the 5 groups who choose their tracking default to the control group who stays in the actual framework and gather data about the level of tracking that individuals get and the satisfaction of users regarding this feature for their privacy. To make it clear, to test our main proposition, we would compare the behavior of the 5.000 people receiving it to the behavior of the 5.000 people in the control group.



We will here describe what we will test between our main proposition and the control group. The tests can also be realized on each of the alternative (4 x 5.000) which can serve to better understand the impacts of each change in the proposition.

First, we will analyze the level of tracking. That means that we would compare the level of tracking when applying our proposition compared to the level of tracking of a control group without our proposition. The level of tracking would depend on the percentage of people who choose the default of acceptance. It would also depend on the behavior of the people who choose default of refusal of tracking, because they will face choice to change their mind in step 2. We would compare the tracking with our proposition (or alternatives, with the same ideas) to the control group who didn't receive any choice to see the impact of choosing the default on the degree of acceptance of tracking. Once we have the numbers, we can have the samples' variances and do a t-test of equal variance to determine if the difference of the degree of acceptance of privacy policies is significant between two groups of the sample (Zaiontz, 2017). To be clear, we want to compare the actual framework with our proposition and see if there is more privacy -less tracking and less acceptance of privacy policies- for the people who received the treatment (making them choose the default).

Secondly, we will analyze the behavior of people who choose a default of refusing tracking in step 1. We want to analyze their choices when facing the convincing privacy policies of websites. Particularly, we want to know the percentage of acceptance for individuals. We want to know if there are patterns or categories between individuals, if some will have low percentage while others will have high percentage of acceptance on the websites. Differences of acceptance could be related to the websites and not the individuals, because of the convincing messages or because the type of the website (commercial, social network, search engine) influences the degree of acceptance.

Finally, we could make surveys to measure privacy concerns in the case of the proposition and compare them with the group who didn't receive the proposition. We could also see if privacy behaviors better match the privacy concerns with our proposition (i.e. people choosing refusal default if they care about privacy).

Test before generalization

To realize tests before the generalization, we could collect data with our add-on on people who are using it. However, we would have problems of adverse selection when reaching people with the add-on. On one hand, if we can contact them by mail or ads they would generally care less about privacy. On the other hand, people downloading it would be those who care about privacy (because they looked for tools, because others don't bother to download it, etc).

There is an ethical problem with those tests that need to track data over people who will show their will to stay untracked. That would need to be considered when deciding how to analyze the proposition.

8.7.2. Tests of the first step

In this part, we will explain how we could test our different messages displayed for the choice in the first step. It is important to be able to know how the design of the choice in step 1 has an impact on people's responses. In past studies, little changes in a message were shown to have impacts on the outcomes. That is for example the case in a study over tax compliance where they simply added a message to tax payment reminding letters. By including social norms or public good messages, they increased tax compliance and the difference between different messages were significant (Hallsworth et al., 2014).

Tests if generalization

We will here remind our hypotheses regarding the impact of the design of the choice in step 1 on the degree of acceptance of tracking. We will also explain how to test them with the different groups. As a reminder, the main proposition and each of the alternative messages in the first step will be allocated to 1/6 of the sample. 5.000 people will receive the main proposition and the alternatives will be tested on 5.000 people each for a total of 25.000 people.

Hypothesis 1: an introduction presenting the benefits of tracking pushes people to the default of acceptance of tracking.

To test it, we will compare the percentage of acceptance of tracking in the group which received our main proposition and in the group which received our first alternative proposition. If the difference is significant, meaning that the first group has a greater tracking acceptance percentage than the first alternative, then our first hypothesis is accepted. If not, it is rejected.

Hypothesis 2: in the introduction presenting the benefits of tracking, giving known websites as examples rather than a generic positive introduction over tracking, should increase the percentage of people taking the default of acceptance.

Here, we need to see if the difference of percentage of acceptance in the group which received the second alternative with the known websites is significantly greater than the percentage of acceptance in the group which received the main proposition. If it is the case, the hypothesis is accepted. If not, it is rejected.

Hypothesis 3: the sentence on the button has an impact on the percentage of people choosing tracking acceptance.

To test hypothesis 3, we need to see if the percentage of people choosing acceptance in the main proposition is significantly greater than the percentage of people choosing acceptance in the third alternative. If this is the case, this hypothesis is accepted. If not, we can reject it.

Hypothesis 4: the default put in the first place is more likely to be selected than if it is put in the second place.

To test our last hypothesis, we need to see if the percentage of people choosing acceptance in the main proposition is significantly greater than the percentage of people choosing acceptance in the fourth alternative. If this is the case, we can accept this hypothesis. If not, we can reject it.

With a big enough sample, results and potential differences will be analyzed and we will be able to determine which design for our proposition the authority should choose between the main proposition and the alternatives, in function of the percentage of acceptance they want and thus the level of privacy.

Tests before generalization

Before generalization, we can find alternate ways to test our messages. In the part over the implementation of the nudge, we talked about first developing an add-on of our proposition and promoting it via ads or emails. We could try to test our hypotheses by sending invitations to download the add-on. The different messages of our alternatives for our hypotheses could be used, except that instead of clicking on the proposition, users would have the possibility to download the add-on which would be equal to the default choice of "*I want to share few information*".

This is an example of how we could test hypothesis 1 with alternative 1. For testing, our first proposition would look like this:

> Websites and third-parties (such as advertisers) can track your data to develop new technologies, to adapt the content you see in function of your tastes or to offer ads that are more likely to interest you. However, intrusive tracking isn't needed for websites to operate. A new privacy add-on gives the possibility to choose if by default:
>
> (1) you accept the tracking of your data which will give you a personalized online experience.
>
> (2) you would like to only share data necessary for a well-functioning online experience.
>
> This choice will be your default on all websites. If you want to share few data by default, you will still be offered to accept to share more data on a website when you first visit it.
>
> Default: (1) I want a personalized experience (2) I want to share few data

By choosing (2), people will download our add-on and it will be as if they had chosen the default of refusal of tracking. We could then create the alternatives on the same idea, by only changing what we wanted to test, but in the case of offering the download of the add-on. So, alternative 1 would be as follow:

Websites and third-parties (such as advertisers) can track your data. However, intrusive tracking isn't needed for websites to operate. A new privacy add-on gives the possibility to choose if by default:

(1) you accept the tracking of your data which will give you a personalized online experience.

(2) you would like to only share data necessary for a well-functioning online experience.

This choice will be your default on all websites. If you want to share few data by default, you will still be offered to accept to share more data on a website when you first visit it.

Default: (1) I want a personalized experience (2) I want to share few data

We then need to compare the percentage of people downloading the add-on by clicking (2), to the number of people clicking (1) (making them leave the message, but we would keep track of the number of click). We could put the different alternatives before the download of the add-on to see the impact on the rate of downloads. We can then compare the rate of downloads and test our hypotheses like described before.

However, results will need to be taken very cautiously and the generalized tests we described before are more reliable. Indeed, there are problems going with those add-on tests. If we try to reach people by sending mails or putting ads, people who receive it will on average be less concerned about their privacy. It is more likely that someone who cares about privacy blocks ads and is more careful about giving his email address. We can thus only observe the rate of downloads of people who are on average less concerned about privacy.

Another problem with testing the different messages for the add-on is that normally we should try to promote our add-on to have the most possible downloads but putting messages that are likely to reduce the rate of downloads is detrimental to the growth of the add-on. Nevertheless, we could imagine doing it to a sample of people and then we would send others a promotional message to maximize the downloading rate. A final problem, is that it would be impossible to know if people just didn't read the message or weren't interested (and in this case, should be considered as choosing (1)) and that would give us wrong estimates on the percentages of acceptance and refusal.

9. Conclusion

To answer our research question "*how to improve the self-regulation of online privacy?*", we began by explaining the basis of the collection and the exploitation of data. We highlighted that the sector of data exploitation is increasing and a promising field. It brings advantages and will be worth more and more as the numbers of users and data shared are both growing.

Moreover, in this thesis, we put the emphasis on the concept of privacy. We described the issues of the regulation of privacy and described the way US and Europe deal with it. We then focused on the self-regulation of online privacy, meaning when individuals regulate their own privacy on the internet. We showed the potential problems of self-regulation of online privacy, particularly that individuals behave differently than what they think, that thus lead to less privacy for them than they desire. We then highlighted the main websites methods to communicate their privacy policy. They need consent from individuals, who by choosing to either accept or refuse privacy policies should be able to regulate their online privacy. Nonetheless, privacy policies include several problems of readability, understandability, loss of time to read them and lack of alternative from acceptance leading to an over acceptance of privacy policies.

We regrouped different solutions to try to solve the issues of self-regulation of online privacy. We proposed to inform users about the badly known subject of privacy, to improve the privacy policies and to render people more active to manage their privacy. The main point and innovative idea that we concluded after reviewing online self-regulation, is that we want to swap the power from the hands of the websites to the hands of the individuals. Particularly, we developed a nudge for online privacy that would give the possibility for individuals to choose the default of refusing the tracking of their data. It will thus be in the interest of the websites to develop clear and visible privacy policies to convince people who refuse to share their tracking to change their mind for the website.

We then described the way future studies or people wanting to continue the project could implement it. We described the experimentation required to test defined hypotheses over the elements influencing the percentage of people that choose the default of refusal of tracking. We also explained how to test the effect of our nudge, particularly how to check if it reduces the level of tracking and thus increases the privacy. We discussed the difficulties and problems of the testing of our proposition.

Our thesis developed a project to improve self-regulation of online privacy. With the recent change of regulation in Europe (with the GDPR), we see that this is an actual subject and that authorities try to address it. It is a step in the right direction towards more power given to the individual (that needs to be well informed of privacy policies and give formal consent), but it won't be sufficient and the debate over privacy will continue and be an important subject in the growing online world. Our proposition or other propositions should still be incorporated to answer individuals' privacy concerns.

10.References

Acquisti, A., Brandimarte, L., Loewenstein, G. (2015). Privacy and human behavior in the age of information. Review.

Acquisti, A., John, L. and Loewenstein, G. (2009). *What is privacy worth?* Workshop on Information Systems and Economics (WISE). (Acquisti, John and Loewenstein, 2009).

Acquisti, A., Taylor, C. and Wagman, L. (2016). The Economics of Privacy. Article.

Alexander De Croo (2017). *A total of 2.7 million active SIM cards registered*. [online] Available at: http://www.decroo.belgium.be/en/total-27-million-active-sim-cards-registered [Accessed 8 May 2018].

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, 95(9-10), pp.1082-1095.

Anon, (2018). *Data protection*. [online] Available at: https://ec.europa.eu/info/law/law-topic/data-protection_en [Accessed 27 Feb. 2018].

Athey, S., Catalini, C., and Tucker, C. (2017). The Digital Privacy Paradox: Small Money, Small Costs, Small Talk.

Badshah, N. (2018). *Facebook to contact 87 million users affected by data breach*. [online] the Guardian. Available at: https://www.theguardian.com/technology/2018/apr/08/facebook-to-contact-the-87-million-users-affected-by-data-breach [Accessed 2 May 2018].

Botsman, R. (2017). Big data meets Big Brother as China moves to rate its citizens, Available at: http://www.wired.co.uk/article/chinese-government-social-credit-score-privacy-invasion [accessed 4Jan. 2018].

Bradley, T. (2011). *Firefox Do-Not-Track Feature Has a Fatal Flaw*. [online] PCWorld. Available at: https://www.pcworld.com/article/217478/mozilla_do_not_track_feature_has_fatal_flaw.html [Accessed 2 May 2018].

Carroll, G., Choi, J., Laibson, D., Madrian, B. and Metrick, A. (2009). Optimal Defaults and Active Decisions. *Quarterly Journal of Economics*, 124(4), pp.1639-1674.

Choi, J., Laibson, D., Madrian, B. and Metrick, A. (2004). For Better or for Worse: Default Effects and 401(k) Savings Behavior. [online] Available at: http://www.nber.org/chapters/c10341 [Accessed 18 Mar. 2018].

Cohen, J. (2012). What Privacy Is For? *Harvard Law Review*, [online] 126. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2175406.

Cortland, M. (2017). *2017 Adblock Report*. [online] PageFair. Available at: https://pagefair.com/blog/2017/adblockreport/ [Accessed 26 Apr. 2018].

Cranor, L.F. (2012). Necessary but not sufficient: standardized mechanisms for privacy notice and choice. J. Telecommun. High Technol. Law 10, 273.

Deshpande, N., Ahmed, S., and Alok Khode, A.. (2014). Web Based Targeted Advertising: A Study Based on Patent Information.

Earp, J. and Baumer, D. (2003). Innovative web use to learn about consumer behavior and online privacy. *Communications of the ACM*, [online] 46(4), pp.81-83. Available at: https://www.researchgate.net/publication/220424005_Innovative_Web_Use_to_Learn_About_Consumer_Behavior_and_Online_Privacy [Accessed 2 Mar. 2018].

ec.europa.eu. (2018). *Proposal for an ePrivacy Regulation*. [online] Available at: https://ec.europa.eu/digital-single-market/en/proposal-eprivacy-regulation [Accessed 2 Apr. 2018].

En.wikipedia.org. (2018). *Privacy policy*. [online] Available at: https://en.wikipedia.org/wiki/Privacy_policy [Accessed 9 May 2018].

Englehardt, S. and Narayanan, A. (2016). Online Tracking: A 1-million-site Measurement and Analysis. Paper. Princeton University.

EU GDPR Portal. (2018). *Frequently Asked Questions about the GDPR*. [online] Available at: https://www.eugdpr.org/gdpr-faqs.html [Accessed 15 Mar. 2018].

Federal Trade Commission. (2000). Privacy Online: Fair Information Practices in the Electronic Marketplace [online] Available at: https://www.ftc.gov/sites/default/files/documents/reports/privacy-online-fair-information-practices-electronic-marketplace-federal-trade-commission-report/privacy2000.pdf [Accessed 14 Mar. 2018].

Federal Trade Commission (2012). *Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policy Makers*. Report. Available at: http://www.ftc.gov/os/2012/03/120326privacyreport.pdf.

Federal Trade Commission. (2018). *About the FTC*. [online] Available at: https://www.ftc.gov/about-ftc [Accessed 13 Mar. 2018].

Galdies, P. (2017). A Summary of the EU General Data Protection Regulation. [Blog] *dataIQ*. Available at: https://www.dataiq.co.uk/blog/summary-eu-general-data-protection-regulation [Accessed 27 Feb. 2018].

Gdpreu.org. (2018). *Data Controllers and Processors – GDPR EU.org*. [online] Available at: https://www.gdpreu.org/the-regulation/key-concepts/data-controllers-and-processors/ [Accessed 5 Mar. 2018].

Gellman, R. (2012). *Privacy, Consumers, and Costs : How The Lack of Privacy Costs Consumers and Why Business Studies of Privacy Costs are Biased and Incomplete*. Available at: http://www.epic.org/reports/dmfprivacy.pdf [Accessed 17 Feb. 2018].

General Data Protection Regulation [2016] (Official Journal of the European Union).

Grannis, A. (2016). You Didn't Even Notice! Elements of Effective Online Privacy Policies. *Fordham Urban Law Journal*, 42(5), pp.1109-1170.

Haddadi, H., Hui, P., Henderson, T. and Brown, I. (2011). Targeted Advertising on the Handset: Privacy and Security Challenges.

Hallsworth, M., List, J., Metcalfe, R. and Vlaev, I. (2014). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*, 148, pp.14-31.

Hann, I.-H., K.-L. Hui, S.-Y. T. Lee, and I. P. Png (2007). Overcoming online information privacy concerns: An information-processing theory approach. *Journal of Management Information. Systems 24 (2), 13–42*.

Hartzog, W. and Selinger, E. (2013). Obscurity: A Better Way to Think About Your Data Than 'Privacy. *The atlentic*. [online] Available at: https://www.theatlantic.com/technology/archive/2013/01/obscurity-a-better-way-to-think-about-your-data-than-privacy/267283/ [Accessed 6 Apr. 2018].

Haynes, L., Green, D., Gallagher, R., John, P. and Torgerson, D. (2013). Collection of Delinquent Fines: An Adaptive Randomized Trial to Assess the Effectiveness of Alternative Text Messages. *Journal of Policy Analysis and Management*, 32(4), pp.718-730.

Healthinfoprivacybc.ca. (2011). *Primary and Secondary Uses of Health Information*. [online] Available at: http://www.healthinfoprivacybc.ca/health-information-disclosure/primary-and-secondary-uses-of-health-information [Accessed 26 Feb. 2018].

Henry, A. (2015). [online] Lifehacker.com. Available at: https://lifehacker.com/the-best-browser-extensions-that-protect-your-privacy-479408034 [Accessed 26 Apr. 2018].

Hinz, O., Hann, I-H. and Spann, M. (2011). Price Discrimination in E-Commerce? An Examination of Dynamic Pricing in Name-Your-Own-Price Markets. MIS Quarterly, Vol. 35(1), forthcoming.

Jansen, B.J. and Mullen, T. (2008) 'Sponsored search: an overview of the concept, history, and technology', Int. J. Electronic Business, Vol. 6, No. 2, pp.114–131.

Jason. (2015). *How Many of Your Users Set "Do Not Track"? - Quantable*. [online] Available at: https://www.quantable.com/analytics/how-many-do-not-track/ [Accessed 15 May 2018].

Jensen, C. and Potts, C. (2005). Privacy practices of Internet users: Self-reports versus observed behavior. Department of Economics. Southern Methodist University, Dallas, USA.

Jocelyn. How to Write a Privacy Policy. (2017). [Blog] *Termsfeed*. Available at: https://termsfeed.com/blog/write-privacy-policy/ [Accessed 1 Mar. 2018].

Johnen, J. (2017). *Policy Evaluation Using Randomized Control Trials*.

Johnson, E. and Goldstein, D. (2003). Do Defaults Save Lives?. *Science*, 302(5649), pp.1338-1339.

Jolly, L. (2017). *Data protection in the United States: overview*. [online] Uk.practicallaw.thomsonreuters.com. Available at: https://uk.practicallaw.thomsonreuters.com/6-502-0467?transitionType=Default&contextData=(sc.Default)&firstPage=true&bhcp=1 [Accessed 13 Mar. 2018].

Kahneman, D., Knetsch, J. and Thaler, R. (1991). Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. *Journal of Economic Perspectives*, 5(1), pp.193-206.

Khajehzadeh, S. (2016) Big Data, Consumer Analytics, and the Transformation of Marketing. PhD. Griffith Business School.

Lerner, A., Simpson, A. K., Kohno, T., and Roesner, F. (2016). Internet jones and the raiders of the lost trackers: An archaeological study of web tracking from 1996 to 2016. In Proceedings of USENIX Security.

Libertyglobal.com. (2012). *The Value of our Digital Identity*. [online] Available at: http://www.libertyglobal.com/PDF/public-policy/The-Value-of-Our-Digital-Identity.pdf [Accessed 5 Feb. 2018].

Lunden, I. (2015). 2015 Ad Spend Rises To $187B, Digital Inches Closer To One Third Of It. *Techcrunch.*

M. McDonald, A. and Faith Cranor, L. (2008). The Cost of Reading Privacy Policies. *I/S: A Journal of Law and Policy for the Information Society*.

Marketing Land. (2014). *Display Advertising*. [online] Available at: https://marketingland.com/library/display-advertising-news [Accessed 6 Mar. 2018].

Marotta, V., Zhang, K., and Acquisti, A. (2015). Who Benefits from Targeted Advertising?

Nobelprize.org. (2017). *Richard H. Thaler - Facts*. [online] Available at: https://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2017/thaler-facts.html [Accessed 7 May 2018].

Odlyzko, A. (2003). Privacy, Economics, and Price Discrimination on the Internet. Digital Technology Center, University of Minnesota.

Olejnik, L., Minh-Dung, T. and Castelluccia, C. (2014). Selling off privacy at auction. In *ISOC Network and Distributed System Security Symposium*.

Pollach, I. (2007). *What's wrong with online privacy policies?*

Privacyfirst.nl. (n.d.). *The Fair Information Principles*. [online] Available at: https://www.privacyfirst.nl/acties-3/item/154-the-fair-information-principles-canada.html [Accessed 14 Mar. 2018].

Purcell, B. (2013). The emergence of “big data” technology and analytics. *Journal of Technology Research*.

Rao, J.M and Reiley, D.H. (2012). The Economics of Spam. *Journal of Economic Perspective*, [online] 26(3), pp.87-110. Available at: https://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.26.3.87 [Accessed 19 Feb. 2018].

Romano, A. (2018). The Facebook data breach wasn't a hack. It was a wake-up call. *vox*. [online] Available at: https://www.vox.com/2018/3/20/17138756/facebook-data-breach-cambridge-analytica-explained [Accessed 4 Apr. 2018].

Rubin, P. H. and Lenard, T.M. (2001). *Privacy and the Commercial Use of Personal Information*. Kluwer Academic Publishers.

Savage, S. and Waldman, D (2013). The value of online privacy. Working Paper.

Schoeman, F. D. (1992). *Privacy and social freedom*. Cambridge university press.

Shen, Y. and Pearson, S. (2011). Privacy Enhancing Technologies: A Review. [online] Available at: http://www.hpl.hp.com/techreports/2011/HPL-2011-113.pdf [Accessed 12 Apr. 2018].

Sibbald, B. and Roland, M. (1998). *Understanding controlled trials: Why are randomised controlled trials important?* [online] Available at: https://www.bmj.com/content/316/7126/201 [Accessed 11 May 2018].

Staff, I. (2018). *Arbitrage*. [online] Investopedia. Available at: https://www.investopedia.com/terms/a/arbitrage.asp [Accessed 11 Feb. 2018].

Statista. (2017). *Internet advertising spending worldwide from 2010 to 2020*. [online] Available at: https://www.statista.com/statistics/276671/global-internet-advertising-expenditure-by-type/ [Accessed 6 Mar. 2018].

Support.mozilla.org. (n.d.). *How do I turn on the Do Not Track feature? | Firefox Help*. [online] Available at: https://support.mozilla.org/en-US/kb/how-do-i-turn-do-not-track-feature [Accessed 2 May 2018].

Tamura, K. (2017). *What's 3rd Party Cookie, and how is it used to track users?* [online] Treasure Data Blog - Enterprise Customer Data Platform. Available at: https://blog.treasuredata.com/blog/2017/02/16/whats-3rd-party-cookie-and-how-is-it-used-to-track-users/ [Accessed 2 Apr. 2018].

Thaler, R. and Sunstein, C. (2008). *Nudge*. New Haven: Yale University Press.

Tversky, A. and Kahneman, D. (1991). Loss Aversion in Riskless Choice: A Reference-Dependent Model. *The Quarterly Journal of Economics*, 106(4), pp.1039-1061.

Udo, G. (2001). Privacy and security concerns as major barriers for e-commerce: a survey study. *Information Management & Computer Security*, 9(4), pp.165-174.

USA TODAY. (2017). *What's your citizen 'trust score'? China moves to rate its 1.3 billion citizens*. [online] Available at: https://www.usatoday.com/story/news/world/2017/11/10/whats-your-citizen-trust-score-china-moves-rate-its-1-3-billion-citizens/851365001/ [Accessed 31 Jan. 2018].

Warren, S. and Brandeis, L. (1890). The right to privacy. Harvard Law Review 4(5), 193–203.

WEforum. (2013). *Unlocking the Value of Personal Data: From Collection to Usage*. [online] Available at:

http://www3.weforum.org/docs/WEF_IT_UnlockingValuePersonalData_CollectionUsage_Report_2013.pdf [Accessed 8 Feb. 2018].

Westin, A. (1967). Privacy and Freedom. New York: Atheneum Publishers.

wikipedia.org. (2018). *Do Not Track*. [online] Available at: https://wikipedia.org/wiki/Do_Not_Track [Accessed 7 May 2018].

W3.org. (n.d.). *Platform for Privacy Preferences (P3P) Project*. [online] Available at: https://www.w3.org/P3P/ [Accessed 19 Mar. 2018].

W3training School. (n.d.). *Structured, Semi-Structured and Unstructured Data*. [online] Available at: https://www.w3trainingschool.com/structured-semi-structured-unstructured-data [Accessed 12 Dec. 2017].

Zaharia, A. (2016). Online Privacy in Under 1 Hour: Improve Your Security Fast. [Blog] *heimdalsecurity*. Available at: https://heimdalsecurity.com/blog/online-privacy-essential-guide/ [Accessed 12 Apr. 2018].

Zaiontz, C. (2017). *Two Sample t Test: equal variances*. [online] Real-statistics.com. Available at: http://www.real-statistics.com/students-t-distribution/two-sample-t-test-equal-variances/ [Accessed 24 May 2018].