THE ROLE OF CONSUMERS’ KNOWLEDGE AND ATTITUDES IN DETERMINING THE IMPORTANCE OF PRIVACY IN BIG DATA MARKETING

Master’s Thesis
Department of Marketing
5/2018
The world has changed due to the digital revolution and this means that the marketing environment has changed and that it has to keep changing. Even though the data has always existed and been a part of marketing, due to this digital revolution the marketing done based on customer data has evolved immensely as well. Big data offers marketers opportunities that seem endless, due to the exponentially growing amount of data and the new innovations that are made to harness it, but what has to be remembered is that big data marketing is not a solution without its problems.

In this research we focus on big data marketing and its implications on consumer privacy; what are the knowledge and attitudes of consumers towards privacy in big data marketing and what is their role in determining the importance of consumer privacy for the companies. The question we ask and answer is: how consumers’ knowledge of and attitudes towards privacy in big data marketing affect the reasons why companies should take privacy as a part of their strategy? The aim of the research is to explain why consumers’ knowledge and attitudes towards privacy in big data marketing are important and thus something that the companies should take into account in their strategy when looking at it from this particular point of view. These questions are examined through asking a group of informants about their knowledge and attitudes towards privacy in big data marketing and analyzing the data in its entirety before drawing conclusions from it. The research is conducted with a qualitative questionnaire sent to this group of informants and the findings are a direct result of these answers collected from them.

The findings of the research include both theoretical contributions and managerial implications. Theoretical contributions of this work presents a model that introduces a framework that explains how the knowledge and attitudes of consumers are interwoven with companies’ actions and what are the major aspects in each of the parts that have an effect on how the privacy in big data marketing is seen and this plays directly to the reasons why companies should take privacy as a part of their strategy. In the managerial implications we explain these things further and learn how either positive or negative outlook towards big data marketing from the consumers’ point of view can affect them in a larger scale and not just on how the customers’ see them. These results can be used to evaluate how the consumers see privacy in big data marketing in a specific company and what are the things that can be done to help the situation; what the companies can do to help consumers’ see privacy in big data marketing in a better light and help them understand the possible ramifications from negative outlook and benefits from a positive one.
CONTENTS

1 INTRODUCTION........................................................................................................................................ 6
  1.1 Data in the world, in the companies, and in the marketing ......................................................... 6
  1.2 Why this research? .......................................................................................................................... 8
  1.3 The research questions and the structure of this research ......................................................... 9

2 BIG DATA MARKETING ...................................................................................................................... 14
  2.1 What is big data? ............................................................................................................................ 14
  2.2 Big data management for marketing purposes .............................................................................. 16
      2.2.1 Acquiring data ......................................................................................................................... 17
      2.2.2 Organizing data ...................................................................................................................... 19
      2.2.3 Analyzing data ....................................................................................................................... 19
      2.2.4 Using data in marketing ......................................................................................................... 20
      2.2.5 Benefits of big data marketing for companies ....................................................................... 23

3 PRIVACY AND BIG DATA .................................................................................................................. 24
  3.1 What is data privacy? ...................................................................................................................... 24
  3.2 Privacy concerns for consumers and data management cycle .................................................... 26
      3.2.1 Acquiring data and privacy .................................................................................................... 27
      3.2.2 Organizing data and privacy .................................................................................................. 28
      3.2.3 Analyzing data and privacy .................................................................................................. 28
      3.2.4 Using data in marketing and privacy .................................................................................... 29
  3.3 The privacy debate and companies ............................................................................................... 30
      3.3.1 Perceived feelings of privacy and willingness to disclose information .................................. 30
      3.3.2 Customer privacy management ............................................................................................. 33
      3.3.3 Economics of data privacy .................................................................................................... 35
  3.4 Privacy as strategy .......................................................................................................................... 37
TABLE OF FIGURES

Figure 1: The relationship between sub-research questions, main research question and the research problem. 11

Figure 2: Model on how knowledge, attitudes and companies’ actions are interwoven together and what they mean. 74

TABLE OF TABLES

Table 1: Informants and their demographics. 45
1 INTRODUCTION

Technology is something that has changed our lives irrevocably. We want to live comfortably, make mundane tasks easier, and reach people no matter where they are in the world, and technology has made all of this possible. The digital revolution is both technological and economic which has benefits but also dangers embedded in it that no one yet truly understands (Kerber, 2017). Among the changes the technology has made to our lives is the way that products and services are marketed to us; the way that marketing can be done. Due to the technology, the marketing landscape has changed immensely during the previous few decades and the change is still going on today. The ways that marketers can reach consumers and the interfaces through which this can be done, have multiplied, thus making the access to consumers much easier to find. One of the things, that have had an effect on our everyday lives and that the technology can provide us, and the marketers, with, is lots and lots of data. In this part we go through few significant things that together make clear the reasons why this research is needed, what questions this work is asking, how we intend figure out the answers to these questions, and finally how this work is structured.

1.1 Data in the world, in the companies, and in the marketing

Data has always existed, and been a part of marketing, but in this day and age there is so much data that people are calling it big; it is big data. Big data is something that not only refers to the big the amount of it, it also refers to the fact that it grows in a speed that is exponentially bigger that the growth traditional data has ever been and that the traditional tools used in analyzing it are no longer enough (Zikopoulous et al., 2011). The companies come into the equation when the fact that organizations can greatly benefit from data, is examined, but the companies cannot benefit from big data if they do not know how it can be done (Jobs, Aukers & Gilfoil, 2015). As the data grows the innovations to harness it evolve and more and more of them are made; the incentive to make these innovations is that they are necessary if the companies want to benefit from it and to make big data worth anything at all (Breur, 2015). If companies are able to make big data worth something, it might lead to considerable competitive advantage for those companies capable to do so, and in this world where businesses are evolving fast competitive advantage in any field can lead be a thing
that differentiates business successes from failures (Coleman, Göb, Manco Pievatolo, Tort-Martorell, & Seabra Reis 2016). But as it is with most of the new innovations, big data is not a marketing method without any problems and data privacy is a problem that the marketers have to realize and assess if they want to use big data in a way that benefits both them and their customers.

In recent years a lack of true transparency has resulted in some hits against the trust consumers’ have on companies. Most famous of these scandals are maybe the emission scandal of Volkswagen or the slowed down iPhones of Apple, and these have resulted to companies, their practices, and the transparency, to be under the spotlight and consumers have responded in turn (Shen, 2017; Hotten, 2015). These scandals have resulted in straight costs, from fines and recalling and PR problems resulting in high ranking people from the companies’ board being fired. (Hotten, 2015). When things like this happen, it forces companies to reflect on their process and their misdoings in a way can make their business more transparent and inherently better for their customers. Privacy is one things that can become a thing that is not really protected in a way it should be inside the companies or that their consumers think it is being protected and can lead to more and more severe consequences to the consumers because the personal information that the companies collect from their customers in rising in such a rapid speed and the ramifications to people’s personal lives are so intertwined with technology and the data that is born because of it (Martin & Murphy, 2017)

The privacy issues are something that has been on the spotlight in the current events too. It has been reported that the young people between ages 16-24 spend 27 hours online every week, and an average adult spends, on average, more than 20 hours online per week, (Anderson, 2015). And because we spent more and more time online, so the data that is created by us when we are online is something that essentially contains almost all of our lives; our likes, dislikes, friends, family, our movements, our preferences, our behavior, and our important day to day interactions with institutions like banks or tax offices (Curran, 2018). The data that is collected is valuable to companies because of its penetrating nature and the coincident want of different parties to have an access to it. Another point that has to be remembered is that the data is valuable not only to the businesses and marketers that can use it in
getting new insights into the human behavior and can transform our data into something that can help them increase their bottom line, but also to governments, law enforcement, and, perhaps most worrying, to different kinds of criminals that if they gain access to this data can use it to harm individuals in ways that are later looked into in the theory part of this research.

The three biggest data breaches alone affected more than three and a half billion users of different websites and smaller breaches happen daily (Amerding, 2018). A really recent example of these data scandals is the Facebook-Cambridge Analytica scandal that has left up to 87 million individuals affected and has forced Facebook to re-evaluate their data privacy practices and has put Facebook under scrutiny in several different countries as the scandal has unfolded (BBC, 2018). This further plays into the fact that an individual person might not be able to protect their privacy in this environment, even if they would like to do so; the constant change in the said environment and even the things that are done to protect consumer privacy might not be enough for companies or governments to ensure that the data of an individual person is safe. Almost everyone has at some point had their data at risk; it might have been a request to change a Facebook or an e-mail password after someone has tried to access their account remotely or they might have had their personal data hacked as a part of a larger breach, like one of those mentioned earlier. Not all of data breaches are inherently serious in nature or cause any real harm, but the threat is there, it exists.

1.2 Why this research?

Big data is very interesting research topic; it is topical, complex, and, if you look at it from the point of view of privacy, somewhat controversial. A need for privacy is something that is present in our everyday lives, and as stated earlier the problems that arise when the privacy is at risk are common, and as the data grows so do the risks (Margulis, 2003; Martin & Murphy, 2017). So it is beneficial to understand this problem in a broader context, and realize what the actual problems that big data and, in the context of this research, the actions that marketers do to achieve their goals with the use of big data, are. After that it is possible to figure out how these problems could affect the consumers and how the companies’ actions in big data marketing
and privacy might in turn have an effect on the consumers’ reactions. There is also some personal interest involved in the selection of this topic; my interest in the subject stems from the news, from a lifelong interest in technology and simply from the desire to understand and to do research on this subject, because my personal interest on it.

There is also a clear research gap concerning these issues. There has not really been any research into the knowledge and attitudes of consumers towards big data and marketing’s role in it specifically. Most of the research into the privacy in marketing has been done either from the companies’ point of view or focusing on one specific point or on the privacy in larger scale in marketing; not on the big data (e.g. Martin & Murphy 2017, Casadesus-Mesanaell & Hervas-Drane, 2015). The privacy concerns that are affecting consumers is relatively unresearched topic and the research that has been done mostly focuses specific points and does not try to see the problem in a bigger picture (Martin & Murphy, 2017). The huge part of the research that has been done about the data and privacy is about the data related to health issues; that kind of data that is collected and used by hospitals and other health organizations but much less research is done from the marketing point of view (e.g. Mostert, Bredenoord, van der Sloot & van Delden, 2018). These kinds of aspects need to be researched to find out to what possible implications the consumers’ attitudes towards the topic could affect companies as they use data in the future and what they could achieve if they take their consumers’ worries into account and their privacy into their strategy as a part that could possibly generate trust between the companies and their customers.

1.3 The research questions and the structure of this research

The bigger research problem surrounding this work is why companies should adopt privacy as a part of their strategy. This problem could, and should, be looked at several different points of view but in this research these questions of big data and privacy are tackled from the consumers’ point of view, because as an entity the consumers form the base of companies’ operations; there is no business if the business doesn’t have any customers. Aim of this research is to explain why consumers’ knowledge and attitudes towards privacy in big data marketing are
important and thus something that the companies should take into account in their strategy when looking at it from this particular point of view. The research was done from the base of three research questions in total; two sub-research questions and one main research question. The sub-research questions, act as a defining base to the main research question and they are also linked together because they feed into each other, especially the first sub research question plays into the second. These sub-research questions are:

“What do the informants of this work know about big data marketing and privacy?”

”What are the attitudes that the informants of this work have towards big data marketing and privacy?”

Main research question is:

“How consumers' knowledge of and attitudes towards privacy in big data marketing affect the reasons why companies should take privacy as a part of their strategy?”

Research problem surrounding the phenomenon is:

“Why companies should take privacy as a part of their strategy?”

In the next figure we can see how the research questions are related to each other and to the larger research problem surrounding this phenomenon.
The methodology of the research is qualitative research utilizing open-end questionnaire as a way to acquire new primary data about the issues on hand from a selected group of informants. By applying qualitative methods to data collection and analyzing the data we can get information that is easily transformable to conclusions when one wants to describe the attitudes and knowledge of consumers have, how the process goes, and what are the links between these aspects. After this is done it is then possible to get something that might inform the companies on the decisions they make regarding the data they have on their consumers and how to best benefit from taking their customers’ privacy and the consumers’ feelings towards privacy into account, when assembling and carrying out their strategy. The data gathered from the informants is then analyzed by interpreting and analyzing the data to gather patterns and information prevalent from the data. Through all these we then find the answers to the research questions in a way that could potentially lead to beneficial information to the companies and then give them incentives to protect consumer privacy.

The structure of this work is following; first the research questions are approached from the theoretical point of view of big data and marketing and privacy issue in big data, (the structure of this part is explained further in the next paragraph). After this
is done the part four addressed the methodology of the research and this is done in three parts, the first part giving some theory on qualitative methods used in the research, second part giving the process of assembling the questionnaire, with which the new primary data is collected, and the last part of the methodology handles the process of analyzing the data collected and drawing conclusions from it. The fifth part of the research handles the analysis of data and that is done then the results of this analysis are compared to the theory and similarities and gaps are addressed. The sixth and, final main part, handles the conclusions of this work; theoretical contributions, managerial implications, and limitations and further research suggestions. After conclusions the references of this work can be found listed in an alphabetical order, and finally in the appendix you can find the questionnaire that was sent to the informants and what was not included in the main body of the research.

To understand the privacy phenomenon in marketing, you have to understand where these privacy concerns come from. This is the reason we first approach the theory from the companies’ and marketers’ point of view and then proceed to explain how those things affect the privacy from the consumers’ point of view. The theoretical framework of this work is collected from the literature surrounding big data, big data marketing and privacy collected from literature from different disciplines and not only from business literature to get a broader view into the phenomenon. The structure of the framework is built from the patterns and similarities recognized in the literature about both big data marketing and privacy and then assembled in a way that benefits the research and makes it easier to give questions that support finding the answers to the research questions. The structure of the theoretical framework is divided into two main parts, big data and marketing and big data and privacy, and then these parts are then divided again into smaller parts to get a structure with that makes sense with the later research. The first larger entity, big data and marketing, deals with the definition of big data and also the big data from the companies’ and more specifically from the marketers’ point of view. Then the second part explains the privacy, consumers’ point of view to the issue and the companies’ point of view to consumer privacy and how they react to it in more general level. The theoretical framework is structured this way because the first part creates is essential to
understanding the privacy problems that consumers face, that are explained in the second part.
2 BIG DATA MARKETING

In this chapter the big data is defined as a concept and also its main characteristics, those that separate it from the traditional data, are looked into in more detail. After this is done, in the next part the big data marketing is regarded from the big data management cycle point of view, and lastly some of the benefits of big data marketing for companies are gathered together to get a clearer picture on the issue. The data management cycle was chosen to make up the structure of this research because it is a clear and understandable way to handle these issues in a way that separates different aspects of big data easily. Doing this enables that delving deeper into the debate surrounding the phenomenon of data privacy and connecting the privacy with the big data marketing is made possible and easier to understand.

2.1 What is big data?

When we are talking about big data it is important to try and find a definition to it. Doing this, however, is a difficult task because there are so many different ways for doing just that, if the research literature is consulted. The differences of these definitions come from differences of opinions on what is included in the big data, how it is generated, and the capacity of it. In addition to the consumer big data is also in some cases defined more specifically though some specific characteristics of it, that in fact separate big data from the traditional data.

Some researchers define big data simply as all the data that there is. Data is created through human interactions with the technology (first wave of big data), the technology interacting with itself (second wave of big data) and along with the more traditional analogic data, that has been created before internet even came into the picture, they make up all the data in the world (De Prato & Simon, 2015; Breur, 2015). Even if we exclude the analogical data from the equation, the sources of data are still not homogenous; the data comes both from people and machines and sensors (De Prato & Simon, 2015). This basically means that the digital data is divided as coming from two different sources; interactions initiated by humans in the internet (audio, clicks, cookies or text) and the interactions between the machines the
consumers’ are using, but without the human initiation (user behaviors, IOT) (Weinberg, Milne, Andonova & Hajjat, 2015).

There are researches, as is explained later, that have also defined consumer big data based on its characteristics and the reason why they wanted to give big data these defining characteristics was in the need to clarify the definition. There are five different characteristics of big data: volume, velocity, variety, veracity and value. More specifically the first three are the main characteristics of big data and veracity and value were added to it later to differentiate it further.

1. Volume

Because the data comes from so many sources it means that there is a lot of it. Although big data might be referred as big data just because it is “big”, we do not actually know much data there is. The best guess we can make is that amount of data that exist in the world has grown from 3 million gigabytes in 1987 to 300 billion gigabytes to 2007 and now the amount is measured in petabytes, exabytes, or zettabytes, with one petabyte containing equivalent of 20 million filing cabinets of text and in the future we cannot even think with the measurements we are using today, because the data keeps growing and it grows fast: it has been estimated that today we create 2.5 quintillion bytes of data. (De Prato & Simon, 2015; McAfee & Brynjolfsson, 2012; Erevelles, Fukawa & Swayne, 2016; IBM, 2018). Due to this rapid growth, the market for solutions to store and analyze big data, like software, hardware, and services will double the size every two years (IDC, 2014 via Erevelles et al., 2016)

Volume is the primary characteristic of big data but not the only one, because there are large datasets that do not have the other characteristics of big data. (Erevelles et al., 2016)

2. Velocity

Velocity refers to the relentless speed of the data creation, and is the second key characteristic of Big Data (Lycett, 2013; Erevelles et al., 2016).
3. Variety

Because big data is collected from multiple sources at the same time, the richness of the data collected surpasses that of a traditional data, and the major difference is that the traditional data was structured and transactional and big data is unstructured and behavioral. (Integreon Insight, 2012).

4 & 5 Veracity & Value

The two additional V’s were added to differentiate consumer big data from large-scale sets and they are important when collecting, analyzing and extracting insights from big data (Erevelles et al., 2016; Lycett, 2013). Veracity comes to the picture because not all consumer data is accurate, and it is an issue because of the volume, velocity and variety of data are increasing (IBM, 2018, Oracle, 2013). Value on the other hand comes from obtaining insight and interpretation from the data that is deemed useful. (Lycett, 2013)

For the purposes of this research it is very important to give the definition of big data that is used in this particular research; because the researchers do not unanimously agree on the definition, or do not agree on the definition on what data is included in big data, the decision has to be made. The definition of big data that is used in this research includes all the digital data in the world, meaning that the data that is collected with traditional means is excluded, but that all the digital data, coming from human-to-machine interactions or machine-to-machine interactions (IOT), is included. This is done this way because the privacy problems, which are the other point in this research and the change in the atmosphere and environment, largely result from digital data

2.2 Big data management for marketing purposes

Big data is something that has revolutionized the marketing as we know it and it is still evolving and changing, as the technology to utilize it, and the environment which it works in, evolves and changes. Regulations that restrict the data collection or data usage are examples of the factors that are changing the environment, and one
of the reasons the data usage in marketing changes is the rabid growth of it. Although the environment is changing and how the data is used in marketing changes, it has to be remembered that using data and data analytics for marketing purposes is not a new phenomenon at all and in fact they have been used for that purpose for decades (Breur, 2015).

In order to implement the process of producing valuable information about your customers you need to have the infrastructure to gather, process and manage huge amounts of data and a one way to manage it is to do it in the data management cycle (acquisition, organization, analysis and decision) (Coleman et al., 2016). In the next part of the theory we approach big data through this data management cycle. This is done because it is easier to grasp the entity of the complex reality of big data management for marketing purposes, if there is some kind of structure in place to make this easier and through which the picture is clearer and easier to understand. At the end of this part we gather together few points about benefits and difficulties of big data marketing as a whole.

2.2.1 Acquiring data

Because of the birth of internet customers and marketers do not necessarily need to engage with one other face to face, so marketers need other means to gather information about their customers and this is where data collection tools come to picture. The defining thing in acquiring customer data is that some information can and is being asked directly from the customer and they can give their consent for doing this, while in other cases this does not happen (Martin & Murphy, 2017; Lacznia & Murphy, 2006).

First step on utilizing data for marketing purposes is of course collecting it; there is no data it is not collected first. How it is done and where it is collected are things that are not so easily understood because the process can be done in so many different ways and the sources of data are so greatly varied. Companies need to be able to collect high-velocity data not just from one source but from numerous of them, take diverse access protocols into account when doing this, and compress and filter it
without compromising data so that it is not difficult to analyze later. (Coleman et al., 2016)

Some researches like De Prato & Simon (2015) divide data collection to analogue and digital and Kumar, Chattaraman, Neghina, Skiera, Aksoy, Buoye, & Henseler (2013) divide it to traditional, digital and neurophysiological data (analogue and traditional data meaning basically the same). Digital data in its own right can be divided further on the base of where it is collected: from human to machine interactions or machine to machine interactions. (Breur, 2015; Weinberg et al., 2015). Focus groups, experiments, structured interviews, surveys, transactions, unstructured interviews are examples of how the traditional or analogue data can be collected and these are also ways in which the digital data can be collected as well; the methods might be older, but they are still used in the digital setting (Kumar et al., 2013).

Clickstream data, contents of people’s social media pages, blogs, video data and search queries are forms of data that come from online interactions that humans make when they operate internet (Weinberg et al., 2015; Davenport, Barth & Bean 2012; Kumar et al., 2013). When companies give consumers opportunity to ask about things like their products or discuss their characteristics, companies also can collect data from these interactions e.g. call center voice data or community forums (Davenport et al., 2012; Kumar et al., 2013)

There are ways of collecting data that are harder to detect and these can happen for example when a consumer tries to find information about a product they are thinking of buying and when they make the purchase (Leeflang, Verhoef, Dahlström & Freundt, 2014). These harder to detect ways are called for example cookies, tracking pixels and flash cookies, that have become popular in the late 2000’s (Couldry & Turow, 2014). These kinds of collection methods can be called involuntary methods, because the consumers might not even realize that the data is being collected (Christiansen, 2011).

The data coming from machine to machine interactions comes from different kinds of sensors, devices and radio frequency identification (Davenport et al., 2012).
Machine to machine interactions basically mean that humans do not have to be a part of the collection process at all, but it is collected all the same (Weinberg et al., 2015). Neurophysiological data is a form of data that is collected by trained professionals and examples include eye-tracking, facial electromyography etc. (Barocci, 2011; Kumar et al., 2013). Dividing data to structured and unstructured data is another way to categorize data with the way that it is collected. Structured data comes from sensors, records, scanners, files and databases and unstructured data comes from textual data (blogs and text messages) and non-textual data (videos, images and audio). Semi-structured data is unstructured data that is sorted with different types of software. (Everelles et al., 2016).

2.2.2 Organizing data

Second step on utilizing data is to find solutions for organizing and storing it. The data that is collected in the acquisition phase is usually not ready to analyze as it is. The data architecture needs to be able to organize it in such a way that it’s in formats that extract, transform and integrate the relevant information. (Coleman et al., 2016). In the age of big data storing the collected and acquired data becomes a challenge because of the staggering volume of data and the rate of speed it is being generated, thus giving companies problems with storing it, and furthermore this is something that is not currently arranged in a good way in companies and it is expensive for companies to do (Leeflang et al., 2014; Breur, 2015). Cloud databases and virtual data marts are data storages of today’s world because, compared to traditional databanks, cloud databases have more storage capacity and processing power and virtual data marts help with sharing data (Davenport et al., 2012).

2.2.3 Analyzing data

Before data can be put to the use in marketing, the marketers need to make sense of the data, which has, as explained earlier, many characteristics that make it hard to understand as it is. Breur (2015) stated that the value of big data lies in harnessing and analyzing the raw data that is collected to make it accessible and usable in creating customer value. Indeed, it is not rational to collect data if you just intend to store it. Big data analysis is also different from the traditional data analysis because
the big data is in its’ nature often noisy, dynamic, heterogeneous, complex and untrustworthy (Coleman et al., 2016). The computing infrastructure for doing this needs to be very powerful and the mining algorithms need to be scalable, making this something that only bigger companies have at this moment (Coleman et al., 2016).

Analyzing data is a complicated process because of the characteristics of it and because it needs technical knowledge from the people that do the analyzing (De Prato & Simon, 2015; Davenport et al., 2012). In addition to this analyzing data is something that needs to be done in a continuous basis because of its changing nature and by implementing a process called data cleaning needs to determine which pieces of data are useful to interpret things that are wanted (Davenport et al., 2012; Boyd & Crawford, 2012).

Statistical analysis packages and other products have been designed to help companies to handle their data, text analysis helps to utilize the text that consumers produce while on the internet, technical analysis for recording consumer journey, and predictive analysis can help with getting insights from data (Davenport et al., 2012, De Prato & Simon 2015; Leeflang et al., 2014; Allenby, Bradlow, George, Liechty & McCulloch 2014). Even if the marketers use all of these in order to analyze data, it still is not free from errors; there may be patterns that just appear to be there when they in reality do not exist (Boyd & Crawford, 2012).

2.2.4 Using data in marketing

Only after the marketers have gone through all these steps, they can begin to utilize big data in their everyday marketing practices and making decisions based on the data that is acquired, organized and analyzed can only be done if those previous steps are done well (Coleman et al., 2016). The uses of the data have a range from target marketing, personalization, consumer behavior, awareness and brands to segmentation, and strategic and tactic decisions.

Target marketing, personalization and segmentation
The personalization of advertisements and prices that is done from the base of big data is deeper in its nature that it has ever been before, and the reason is largely related to the fact that these are done by means that are unknown to customers themselves (Couldry & Turow, 2014). Facebook and Google are the most prominent examples of companies that have exchanged the data they collected to help companies deliver an ad targeted to a specific person according to their personal, but anonymous, data, and location tracking technologies can in their turn help companies target and personalize their advertising even more; based on the knowledge of the consumers in their real time locations (Couldry & Turow 2014; Kshetri, 2014).

Segmentation benefits from big data, that arise through consumers preselection of content and the data used in archetypal analysis (leans on universality of archetypes and thus give relevant observations from big data) can be also used in more specific consumer segmentation (Alemany Oliver & Vayre, 2015; Couldry & Turow, 2014) Segmentation has gone through big changes in recent years and that is linked to digitalization and thus big data (Leeflang et al., 2014)

Data coming from machine to machine interactions is an excellent tool for marketers, because it can enable the real time decision making; it is in its nature constant and thus enables real time monitoring of the data that is coming in. When companies can combine their own data with the data they can get from the other sources they can create more accurate reactions from their own customer base with personalized messages. (Breur 2015; De Prato & Simon, 2015). Online behavioural marketing is a tool that uses web tracking to create profiles about individual customers and then aim advertising and other kinds of marketing communication tailored to this one specific customer. (Nill & Aalberts 2014; Goldfarb & Tucker 2011)

**Consumer behavior**

With big data the marketers can learn to understand consumer behavior in ways that were not possible before, and making this possible is the data that comes both form human to machine and machine to machine interactions (De Prato & Simon, 2015; Weinberg et al., 2015). Digital traces for example can help to discover and understand why consumers decide to click on an ad or leave a comment on a product
(Alemany Oliver & Vayre, 2015). Click data can help companies understand the visitor traffic and the behavior of people who visit their website (page views, frequency of their visits, the time spent on their website and what they viewed during the visit) (Moe & Fader, 2004). This applies to theory as well when marketing departments can use data collected to make theories about the behaviors in a larger scale and then apply them to marketing (Jara, Concepción Parra & Skarmeta, 2014).

**Awareness and brands**

Seeding strategy for social media and advertising campaigns in there, viral marketing, brand reputation, product quality and consumer satisfaction and brand awareness are all things that can benefit from using data, and because of this big data can thus lead straight to an increase on profits (Hinz, Skiera, Barrot & Becker, 2011; Jara et al., 2014). Search engine queries and information about the times certain keywords are used can help companies to get presence on search engines and information about the keywords used (Skiera and Abou Nabout 2012 via Kumar et al., 2013; Hu et al., 2014). When companies use big data to follow consumer’s purchasing journey they can use this to build awareness, get purchases or inspire loyalty to their product (Leeflang et al., 2014).

**Strategic and tactical marketing decisions**

Bid data is also a tool from the base of which companies can get help with doing strategic and tactical marketing decisions like market transitions, segmentation decisions, understanding marketing trends, business environment, decisions on creation of new marketing products and help to do these decisions quick and give real time responses to possible changes (Davenport et al., 2012).

When the data is collected real time the companies can make decisions based on their customers location or other environmental factors, and social media analytics can detect quickly emerging and shifting trends in what their customers think and then also react quickly to these changes (Weinberg et al., 2015; Davenport et al., 2012). Consumer data that is collected from the companies’ own websites can help them change the content and advertisements on their websites so that they match their
customers’ wants and needs (Christiansen, 2011). Search trends can help with budget allocating, enhancing product design, advertising, production and inventory planning and sales forecasting and data from social media can help companies on performance measures through creating awareness of their products or services (Hu et al., 2014; Hinz et al., 2011). Because of the access that marketers have to rich, insightful and current data, they are able to make real time marketing decisions which are better than the ones they were able to make previously (Erevelles et al., 2016).

To bring some light to how big data helps with strategic decisions, the marketer has the ability to know what customer transactions are being made: what products the customers buy, what styles and colors and what their consumers think about them and their products, when they make posts on social networks and all of this can be done in real time (Erevelles et al., 2016)

2.2.5 Benefits of big data marketing for companies

Smart systems, like adaptive systems and self-governing models, are systems can go beyond what people can do by themselves with the data that comes from the machines; they use artificial intelligence to use data more efficiently and without human interference (Breur, 2015). Sales volume, innovation, customer engagement and brand loyalty are also things that can benefit from consumer data (Leeflang et al., 2014). Big data can also lead to innovations made from the discoveries made from analyzing data. These innovations can lead the considerable competitive advantage for companies when they have shortage on energy, materials, food and water. (Breur, 2015). Big data can be used to create efficiencies, enable new capabilities and open up new opportunities (Fulgoni, 2013)

Data collection is also a profitable business and not only a way for companies to understand their own customers based on the data that they have collected themselves. Companies like Facebook and Amazon collect data along their other operations and then sell it forward to other companies and there are also some companies that formed the business to the reason to collect it and then sell it (Christiansen, 2011).
3 PRIVACY AND BIG DATA

This is the second part of the theoretical framework and it delves deeper into the privacy issues and the debate surrounding big data marketing. In this part we first approach the privacy question similarly as we approached big data in marketing in the previous part. First we define what privacy means, what it means in the context of big data and what are the key questions surrounding the privacy debate. That part addresses the definition of data privacy, first from the more general point of view and then going into more detail of what privacy means in the context of data and marketing. The second part looks into the privacy debate from the consumers’ point of view and how the privacy problems that consumers face correspond to the big data management process. Then in the third part of this chapter we address some things that surround the privacy debate from the opposite side of things which is companies and things that need to be looked at when the companies make decisions about their consumers’ privacy and where these things are actually being done (companies and governments)

3.1 What is data privacy?

There is no universal concept of privacy, because of the unclear boundaries, conflicting views and because privacy in itself has many different scientific viewpoints from which it can be looked from (Margulis, 2003). Everyday definitions of privacy emphasize privacy as a control over regulation or people's unwillingness to be subjected to scrutiny, surveillance, or unwanted access (Allen, 1988 via Margulis, 2003; Margulis, 1977). Legally speaking the notion of privacy has also changed over time and there is a gap between what the privacy is legally speaking and what it is socially speaking (Kagehiro, 1990). Privacy has been previously broadly defined privacy as an construct that represents control over transactions between an individuals and other people and privacy wants to enhance autonomy and minimize vulnerability, but this definition has not taken everything into account either, such as differences in notions of control and its relationship with privacy (Margulis, 1977; Margulis, 2003)
The notion of privacy has few defining characteristics; need for it comes from changing conditions and in turn affects the conditions around us, regulating privacy can be unsuccessful, its’ at least to some extent culturally specific, it has different functions and forms that are not necessarily related to each other and that privacy contributes to self-identity and individuality. The other important aspect of privacy from the point of view of this research is what can be gained from having privacy, what can happen if you do not have it or what can happen if you lose it. The benefits that can be obtained through having privacy is having possibilities to express political views without being persecuted and discussing religion or family matters “in privacy”, supports self-definition, -assessment and experimentation and protect personal autonomy. (Margulis 2003)

If privacy is not obtained the individual in question loses the opportunities that having privacy provides. In the short term losing privacy might mean that the individual learns the limits of their autonomy and in long term it might lead for the individual to believe that they simply cannot obtain the opportunities that can be obtained through privacy (Wolfe & Laufer, 1974 via Margulis 2003). If on the other hand an individual has had privacy, but then loses it the costs of that are again different and this happens when privacy is invaded or violated (the cases of this might mean gossip and whistle blowing) (Margulis, 2003). The costs of this vary depending on the factors like content of the information released can be related to being stigmatized by your body, sexuality or race and can lead to discrimination, devaluation, stereotyping or prejudice (Crocker, Major, & Steele, 1998 via Margulis, 2003). These things in turn can lead to individual feeling violated, embarrassed and feeling unwanted or even in most serious cases to a potential loss of life, like in cases of spies in World War II (Archer 1985; Margulis, 2003)

Because of big data the personal privacy is harder to control than before (Song et. al. 2016). Despite of this the as early as in 1994 Bloom, Milne & Adler already suggested two key questions for data privacy that all the marketers should take into account, when they are using data in their marketing endeavors:

1. Should company be allowed to acquire and store information about individuals without their knowledge and consent?
2. Should company be allowed to disclose information about individuals to other parties without their knowledge and consent?

(Bloom et. al, 1994)

Even though these questions were asked so long time ago, they still remain in the center of the data privacy debate. And as Martin & Murphy (2017) concluded in their study these questions are still largely unresolved; the companies have decided that it is in their rights to do this and they do capture, store and sell this data as a part of their everyday processes (Martin & Murphy 2017; Singer 2012 via Martin & Murphy 2017). Privacy concerns are rising more rabidly amongst older generations than they are amongst younger ones, but they are rising nonetheless (Goldfarb & Tucker, 2012)

3.2 Privacy concerns for consumers and data management cycle

Customers are learning to understand the value of their information in growing numbers nowadays and they have said that they are less willing to share their information in the future and this is something that in turn can lead to more marketers trying to capture data in more covert way. (Ernst & Young 2015; Martin & Murphy, 2017). In this part we look at the consumer privacy concerns in the same data management cycle as we looked at big data marketing in the previous part. Doing it this way enables us to bring these issues in context in a way that is clear and makes the privacy concerns more manageable when we look at them from the point of view of marketing. In this part we see that consent and anonymity are two things that are certainly the key players in the consumer privacy protection environment, but they are not enough to ensure protected data by themselves (Barocas & Nissenbaum, 2014). It has to be noted that due to shortage of scholarly articles on some of the issues, in this part some of the articles used are from other sources like magazines such as Wall Street Journal to highlight some of the more practical privacy issues arising from big data marketing.
3.2.1 Acquiring data and privacy

The consent of the consumers during the data collection is what is a big issue in acquiring data, namely if people are aware of the data collection, if they understand the consequences of it and if the consent even matters (the possibility of identifying people from patterns that those who do give information away give) (Barocas & Nissenbaum, 2014). The data that is collected from the sources around the internet might turned anonymous after the collection is done, but in reality it is really easy to recognize an individual from their behavior in the websites and lead to problems on how people are treated in the marketplace. (Barocas & Nissenbaum, 2014). In addition to this the presumption that if the data is anonymous it is supposedly safe, but when it becomes less and less anonymous it becomes more and more invasive when we look at it from the point of view of privacy (Christiansen, 2011).

There are different ways that consumers disclose their information. Involuntary disclosure basically means that the consumer is not really aware that the information is being taken, like in the case of cookies and some cookies collect information even when the consumer has deleted cookies. Consumers may disclose their information voluntarily, but even this can become harmful, because the consumer might not understand that the information could be harmful, what information they are giving, or they just click through all the warnings. In this part the marketers should also take into account the age of the consumer disclosing the information, because children understand the consequences even less than the adults do. (Christiansen, 2011).

The data also is not always generated from sources that can be considered legal and reading and understanding privacy policies can be proven to be somewhat tricky, they might not be available at all or it is difficult for companies to assemble them, because no one really knows what the data can reveal. (Kshetri, 2014; Pavolotsky, 2013)

Cookies are something that people generally know about, but there are also different methods of collecting information. Information packets can track everything a person does in the internet; privacy intrusions and extreme personal profiling are things that could arise from people using these (Stecklow & Sonne, 2010). History sniffing is
something that does not rely on cookies, thus being harder for consumers to avoid, and is used to create a profile of a person by running a code based on the websites a person has visited. (Vascellaro, 2013). Lastly scraping is a method that gets the information from confidential discussions of a specific user and this can be done from several different sites by connecting pseudonyms; the data that is collected this way is said to be anonymous, but yet again it is not hard to connect people to their data (Angwin & Stecklow, 2010).

Smartphones and machine to machine interactions bring their own problems to the table; machine to machine interactions bring data that is more all-encompassing than ever and smartphones can be used to bring peoples’ behavior to marketers with their movements, moods, health, calling and spending habits etc. (Hotz, 2011; Weinberg et al., 2015)

3.2.2 Organizing data and privacy

The main reason that organizing and storing data can be dangerous from the point of view of privacy is because of hacking. If the data is hacked, it could expose information, money or identity crimes and hacking into devices that are essential to people can even lead to death. Because of the data being stored in big volumes and the peak flow and collection periods make data more vulnerable for cybercrimes. (Weinberg et al., 2015; Kshetri, 2014).

Data outsourcing is a method that many companies use, but it poses a threat because the huge concentrations of data that are stored in the same place is an easy target to hackers. Another threat is that the outsourcing companies have an easy access to this data and can use it to their own advantage and ultimately can endanger the privacy of consumers and lead to economic, emotional, psychological and social harm to the consumers whose data is stored in these companies. (Kshetri, 2014)

3.2.3 Analyzing data and privacy

By analyzing data companies can get patterns and recognize behaviors, but the problem comes when the patterns and behaviors reveal things about people that they
don’t want anyone to know. An example of this is the predictive analysis which can lead to unpleasant experiences for customers, as it did when Target predicted and revealed teenagers pregnancy before the parents knew about it. (Kshetri, 2014; Duhigg, 2012).

Data collected from the machine to machine interactions pose a new threat to privacy because of the deeper nature of the information that is collected about behavior and the person is there and because this information exist quite freely in the internet and someone may be able to hack it easily (Breur, 2015). At this point of data management cycle anonymity also becomes an issue. Is the information that is sold to third parties or kept inside the company truly anonymous and even if it is, is deanonymization truly effective when bigger datasets and better analytics can link the data to one individual quite easily (Barocas & Nissenbaum 2014; Kshetri, 2014; Pavolotsky, 2013)?

3.2.4 Using data in marketing and privacy

Data is not a stationary, it moves from one place to another and sometimes consumer really cannot know where their personal data is or who has an access to use it, (Pavolotsky, 2013). This is a problem because the data can reveal misbehaviors of consumers that they would not like to reveal and for marketers this means that amongst all this new data they have potential to exploit consumer’s secrets to get information and use it in marketing (Alemany Oliver & Vayre, 2015). This data can lead to potentially harmful outcomes such as background checks for hiring, decisions of termination and scholarships (Christiansen, 2011).

GPS data is something that is an excellent example of a kind of data that, if it would get to wrong hands, can cause problems like real time stalking, asset investigation, illegal debt collection, private investigation and kidnapping (Pavolotsky, 2013; Kshetri, 2014; Yan, 2012). Another problem that arises is when companies can use the information that an existing customer has given them to price discriminate the customer in the future. (Conitzer, Taylor & Wagman, 2012).
The customer profiles are not only derived from the information that the consumers know that they are giving to the marketers, but also from the unrelated searches they do and location based GPS on consumer’s phones. Consumers report privacy violations, feelings of vulnerability and threat of fraudulent activities when marketers use data in these ways because of their intrusiveness. (Kshetri 2014; Fong et al., 2015). Online behavioural marketing is something that has been shown to heighten privacy concerns that customers have (Nill & Aalberts 2014; Goldfarb & Tucker 2011).

### 3.3 The privacy debate and companies

In this chapter we look into the privacy debate from the companies’ point of view, their motivations for doing decisions about the privacy of their customers and how they can affect the response that their customers have for their actions in this field. First we look into the perceived feelings of privacy and customer’s willingness to disclose information and what are the factors that play into it. In the second part we look into the realities on how customer privacy is managed within the companies and lastly on the economics of internet privacy.

#### 3.3.1 Perceived feelings of privacy and willingness to disclose information.

The perceived feelings of privacy that consumers have and the willingness to disclose information are both linked to the consumer privacy debate from the other side of the coin; the companies.

The first thing that we can state from the factors that has a positive impact on consumers’ feeling of privacy is that the stronger the individual and regulatory control are; the more privacy concerns diminish (Martin, 2015; Sheehan & Hoy, 2000). All researchers do not agree on this; there is some debate on the issue whether consumer’s perceived control or perceived vulnerability is a stronger influencer on customers’ privacy concerns. However, Xu, Teo, Tang & Agarwal (2012) found that if consumers perceive that they have control on their own privacy the concern for it grows weaker but Dinev & Hart (2004) stated that the perceived control does not affect privacy concerns at all but perceived vulnerability does.
Social exchange theory and behavioral decision theory concerning data privacy and marketing, focus on the cost-benefit relationship; the consequences that consumers’ face when they disclose information to gain value from the marketers (Martin & Murphy, 2017).

Consumer information exchange considers few points; that the consumers are more likely to disclose information when they feel like they “behave” according to norms (Moon, 2000), in deeper relationships they might be willing to give information but not if they feel like it is embarrassing somehow (White, 2004), company signals (e.g. privacy seals) increase the disclosure of information by giving positive perceptions about the company giving these signals (Miyazaki & Krishnamurthy, 2002), Interface appearance, question intrusiveness and cuing privacy policy either increase or decrease the willingness to disclose information (John, Acquisti & Loewenstein 2011), “I can do it if other’s have done it first” –mentality (Acquisti et al., 2012) and the perceived control on the information they disclose and realized customization also play their part (Mothersbaugh et al., 2012). There are three points that play a huge point in these perceived feelings of control and vulnerability: trust, personalization and control.

1. Trust

If company has its consumers’ trust it might lead to willingness to disclose information, purchase intent, click through and acceptance of advertising (Martin & Murphy, 2017). What is harmful from companies’ point of view is that, if a company covertly uses some privacy-compromising technologies like cookies it compromises the trust between the company and the consumer (Miyazaki, 2008). Company can enhance trust in their customer as a promotive mechanism and they can reduce privacy concerns as a reactive mechanism. This means that trust can serve as a promoter to consumer information disclosure and as a creator of consumer-company relationship when you are talking about privacy (Wirtz & Lwin, 2009). In the context of targeted and personalized consumer content trust also is a factor in alleviating privacy concerns that consumers have (Aquirre et al., 2015; Bleier & Eisenbeiss, 2015). If a company conducts their data practices with transparency and give control
to the customer it softens the vulnerability that consumer’s feel when the practices are done (Martin et al., 2016)

2. Personalization

There is no doubt that personalization can potentially offer value to customers by e.g. product and service recommendations, there is not a consensus present in the privacy literature (Martin & Murphy, 2017). Consumers give away information and thus give away their privacy when they want to get personalization in their marketing communication and products (Martin & Murphy, 2017).

Aquirre et al. (2015) and Miyazaki (2008) for example have examined the value-information tradeoff and its relationship with the means that data is collected; it can be done openly or is it done in secret from the customer and the effect is different in both cases. Trust also plays a role in personalization; if consumers trust the company that offers them personalization, they are more willing to click through and give information (Bleier and Eisenbeiss, 2015)

3. Control

People respond better to personalized and targeted advertisements when they are able to control their personal privacy settings (Tucker, 2014). When customers’ have some control on the process of information disclosure, they are more likely to be receptive to highly customized marketing communications (Norberg and Home, 2014). Consumer’s perceived information control is in the center of mechanisms when the customers’ privacy concerns diminish along with self-protection, industry self-regulation and government mandates (Xu et al., 2012)

Although control can diminish some data privacy vulnerabilities, promote trust and reduce the emotional violation that consumer’s feel, it can also lead to increased vulnerability for consumers when they may give too much information when they feel like they are in control (Martin et al., 2016; Brandimarte, Acquisti & Lowenstein, 2012). It has to be noted also that consumers might also give false information when they choose to give it (Martin & Murphy, 2017). Other problems
marketers might face are communication avoidance, information falsification, degenerate word-of-mouth when consumers feel that their privacy is being violated (White, Zahay, Thorbjørnsen, & Shavitt, 2008) and they might feel like their privacy is being violated when the power balance between them and the company is off (Lwin, Wirtz, & Williams, 2007).

3.3.2 Customer privacy management

How the consumer privacy is managed then? In a literature review in the customer privacy in marketing Martin & Murphy (2017) say that there are not a lot of studies that are done to research this; the info is very sensitive and maybe even controversial and thus the data collection poses difficulties. Milne & Bahl (2010) stated that it is likely that organizational privacy practices are different from one another at least a little because an organization needs to take its’ customer’s wishes into account.

Rust, Kannan, Peng (2002) noted that if there is a market where privacy is not regulated at all it leads to a market for privacy, because consumer privacy almost ceases to exist. Consumer’s might be willing to pay for privacy, but as the situation gets worse and worse when the privacy is concerned, so does the quality of the value that is provided in exchange.

A company can get most profit when the customers can maintain their anonymity without obstacles and customers benefit the most when it their anonymity is costly. (Conitzer et al., 2012). Furthermore, organizations need to have some data privacy protection in order to ensure that the market system functions smoothly. (Conitzer et al., 2002 Rust et al., 2002). Casadesus-Mesannell & Hervas-Drane (2015) stated in their article and an assumption can be made, that consumers are not willing to pay much for their privacy, and companies that disclose their information can profit from so, if the competition for the data about the customers can be softened when companies can differentiate in their privacy practices (assuming that consumers are heterogeneous in this instance.)

The second point on this is well summarized in the report by Ponemon institute (2015) which states that information security failures and data breaches are becoming
more and more common and they affect multitudes of companies. But there has not been a lot of research into the effects on the performance of companies that these data breaches cause (Martin & Murphy, 2017). Kannan, Rees & Sridhar (2007) Companies might not want to, or recognize the need to, enhance their privacy protections because there is no definite information on to which extent privacy failures harm companies.

Although there is some literature on the effects of data breaches on data breaches, Martin & Murphy (2017) note that in the literature that handles privacy failures of companies, largely do not include consumer implications at all. The few points that were found was that say that an information security breach is a form of service failure, and if something happens the company transparency and control give consumer’s incentive to report vulnerabilities they can spot and soften their reaction to the breaches, thus softening the blow that a company suffers from the breach. (Malhotra and Malhotra, 2011; Martin, Borah & Palmatier, 2016)

Bowie & Jamal (2006) argue that if a company’s privacy policy is well constructed and have opt-in provisions within them, it is sufficient to protect consumer’s privacy and government intervention is not needed as such. On the other hand Romanosky, Hoffman & Acquisti (2014) suggests that because only 4% on security breaches are followed by a legal action against companies and that they are quite rare occurrence in a lifetime of a company, they will continue to favour to control their own privacy management as they please and don’t like to be regulated from the outside sources like government.

Corporate privacy policies play an important role in communication between customers, regulators and the public about how the companies collect, use and protect information about the customers and exist as an evidence about the privacy self-regulation within the companies itself (Martin & Murphy, 2017). Milne et al. (2004) states that 84% of people read companies privacy policies and report that these policies make them trust the companies more, but on the other hand a more recent study by Milne et al. (2006) found that privacy policies are longer now than ever before and they are less readable and in addition to this consumers over rely on the firms to protect their information although they probably shouldn’t.
Martin and Murphy (2017) state that marketers have a need to engage with consumers and form long time relationships with them has led to them wanting to get more and more information about the customers which in turn has led to the data collection, analytics and use in marketing in a large scale it is being done in these days. When marketers understand their customers they can give their customers better services and other benefits, so both parties do benefit from big data even though it has to be remembered that nothing comes without its problems (Martin & Murphy, 2017).

The privacy problems in the data collection phase arise mostly from this information which is collected through means that are not as direct and consumer’s do not always know what they give consent to. These means include things such as cookies that can enable tracking web movements in a large scale. (Martin & Murphy, 2017; Lacznia & Murphy, 2006)

3.3.3 Economics of data privacy

As early as in 2002 Rust et al. predicted in their study that because the cost of obtaining and processing information there is going to be a market for privacy, but it is not enough to stop the decline of privacy.

There are groups and associations on both customer and marketer side that try to preserve privacy. The groups on customer side try to educate the customers on theirs rights concerning privacy and lobby governments and regulatory agencies to take a stand for consumer privacy and on the marketer side there are associations that promote self-regulation and in organizations themselves there are people that manage the privacy protection practices (appointed because high-profile blow-ups and because of the realization that this is a one way that organization can make their customers trust them.). (Rust et al., 2002)

In addition to the people in customer and in marketer side there are also third party organizations that try to protect consumer privacy; anonymizers, infomediaries and authenticators. Anonymizers are organizations that provide customers the infrastructure to be able to be anonymous in the internet and they have an appeal to
certain type of customers and they might be ready to pay for it, however not all customer feel to need to have this kind of total privacy in the internet. In the age of big data, where costs of obtaining and processing data decline these anonymizers need to raise their prices and only those who value their privacy the most will want to buy this kind of infrastructure and this in turn has lead to some providers to let their customers choose how much privacy they need when they are surfing the net. (Rust et al., 2002)

Infomediaries are organizations that allow consumers to create their own profiles in their own terms that they store information in and then the organizations can let marketers that are interested in targeting these customers to join their network and let these marketers only target customers that have “opted-in” to wanting get targeted by those marketers. Infomediaries protect privacy by giving consumers an opportunity and a safe way to choose what information they want to provide and to whom they have given a permission to provide it to. It affects consumer’s privacy but they get something in return, they don’t need to pay but the customazitation is imperfect, thus the infomediaries appeal to those that value privacy but are willing to let go some of it to receive something in return. (Rust et al., 2002)

Authenticators are organizations that audit online marketers from the basis on their privacy practices and established privacy norms. Customers can choose only to business with these providers that have the seal of an approval by these aauthenticator and feel secure that their privacy is protected. Some customers might want this, but for some it might not be proactive enough, they don’t need to pay for the service but the services or products they are buying from these audited organizations might be more expensive. (Rust et al., 2002)

It has to be noted that if customers want privacy they might need to be ready to pay for it. There are three different categories of companies that use and sell data; some collect data and then, after making it anonymous, selling the data to other companies or just use it themselves, some companies keep their personal data within their own company, but give other companies an opportunity to target their advertisements to people according to that data and lastly some companies collect data with “the
intention of selling the information, sometimes including specific profiles or names, to third parties” (selling profiles about specific people). (Christiansen, 2011)

One of the reasons that the companies more specially SME’s are unable, or unwilling, to use big data in their marketing processes are the data security concerns, and they are the main obstacle for SME’s in regard of big data marketing according to Coleman et al. (2016). The security issue is more difficult for smaller companies because their IT expertise is lower and their database management systems are often outdated and unsupported, and they are thus more vulnerable for security breaches, intrusion and cyber-attacks. In addition to this the big data analytics is something that many of SME’s are unable to do in house so they need to outsource it and they lose even more control on their data and because of the cloud services that are being used more and more bring even more vulnerability to big data. The SME’s also have to obey legal constraints on data protection and privacy and they often cannot afford to pay for the legal expertise to understand and work in this legal environment because of its complicatedness. (Coleman et al., 2016)

3.4 Privacy as strategy

In an integrative literature review by Martin & Murphy (2017) some basic assumptions of adopting privacy as a strategy were made. By combining case studies they found that.

1. If firms are able to prioritize data privacy in a way that is authentic, and they can convince their customers of it, it has a positive effect on the firm’s market response, customer loyalty and engagement benefits.

2. The privacy dialogue with customers and regulators can have a positive effect on a company as well. The open and transparent communication with consumers is a vital condition for the firm to succeed as having the privacy as their strategy in the long run.

3. Companies should implement privacy in their promoting across every aspect of their firm, because it too affects the performance in a positive way. Consumers do
not appreciate partial or cosmetic privacy efforts so that is the reason why companies should promote privacy throughout the whole company.

4. Companies should focus on their strengths in data privacy and not the weaknesses in their practices because customers accept this better.

5. Privacy as a strategy should be a long term solution and not a short term one. In long term privacy as a strategy leads to positive performance and firms can gain from relaxed regulatory oversight, but in short term these benefits do not occur.

6. Privacy as strategy leads to high consumer trust and trust can help companies to get better performance. However it must be noted that if the company does not follow the previous points in following privacy as strategy this trust can be quickly destroyed.

(Martin & Murphy, 2017)

Casadesus-Masanell & Hervas-Drane (2015) did research on how the companies can compete with privacy: how does privacy affect competition and should firms exploit all available revenue sources, prices and disclosure or just focus on one and what is the most profitable strategy on doing so. The privacy softens the competition when firms can differentiate in their privacy policies among their customers and when they are not willing to pay a lot to get privacy (so that the parties who disclose the information further can do it profitably.). If the company is under competitive pressure they should focus on one revenue source and when consumers are not willing to pay much the companies can create revenue by disclosing consumers’ information and they should focus on price revenue if consumers are willing to pay. The writers also expect that consumers will become more aware of the disclosure practices and that transparency among companies will play an important role in consumer decisions and for companies to deliver a good service. Companies should make privacy policies more understandable in regarding disclosure but not to restrict them too much. Informed customers do value high-disclosure services.
3.5 Knowledge and attitudes

In this chapter we explain how the theoretical framework supports and is ultimately used in the rest of the research process; to put it simply the theoretical framework, introduced in the chapters two and three, provides us with the framework from where we assemble the questionnaire and the points we need to reflect on in the analysis. The theory leads us straight to these points that are explained next briefly. The theoretical framework provides the guidelines and the reference points needed when looking at the data collected from the informants, analysing it, and finally in order to answer the research questions. The knowledge and attitudes, which are things that we wanted to address in this research and are the sub-research questions, and ultimately to how these things relate to the answer to the main research question. The reason why this is and how we came into realizing this shines through from the theory.

There are a lot of aspects and technicalities related to big data marketing, which could in some cases be unknown to the general public which leads straight to the need to know what the informants know about the issue so that we could look at the different aspects of our research and see how these are linked together. The attitudes, which are the second major point that we look for in this research, come from the underlying conceptions in the theory; there are some research into this from the companies’ point of view, more specifically on how these attitudes affect some specific aspects on the privacy side of things, but not a lot into the attitudes towards big data marketing and not into the privacy and big data marketing as a whole. From the data we could however see some aspects of privacy and big data marketing that rose through and could from them then assemble the questions on relevant to the research. The third subsection in the questionnaire was collected from the privacy debate from the companies’ point of view; the specific points that were seen as important in the theoretical framework and were discussed in depth formed the points that needed more reflection from the consumers’ point of view in order to answer the research questions.

We can see clearly from looking at the literature that there is a significant research gap on the knowledge and attitudes of consumers and how they could affect both each other and companies. The research literature on privacy and big data marketing
largely focuses on how the companies see it or on few specific points on the consumer side, such as privacy policies or the personal control on data privacy. The bigger picture on these issues is still somewhat missing and this is what this research tries to answer; how do the knowledge and attitudes towards of consumers big data marketing and privacy affect each other and thus the way they see big data marketing and why companies should take these things into consideration.
4 METHODOLOGY

This part of the thesis handles the methodology of this research; what was done, why it was done, how, and with what perimeters. The first part of this chapter explains the methods used in this research, the reasons for choosing them and some general information on the how the research is conducted. After that the second and third parts explain the conducting process of this research; the second part explains the process of collecting data that we needed to answer the research questions and assembling the questionnaire, and the final part handles the process on analyzing the data, the methods used to do it, and drawing conclusions from it.

4.1 Qualitative methods of research

Qualitative methods were chosen as a way to conduct this research because it was more beneficial to do this from that point of view. In order to see how knowledge, attitudes and companies play together in the informants’ minds, there was a need for data that was richer and less rigid and would give us descriptions and examples on the knowledge and attitudes that consumers have towards big data, marketing and privacy. By doing this the results that were possible to drawn from data matched the aim of the research in better in its’ entirety and look at the process of how these aspects form and affect each other. The new primary data was collected by using a questionnaire with open ended questions that was formed from looking at the theoretical framework and looking for concepts that might relate to the research questions and the bigger research problem. The questionnaire was chosen as a way to collect primary data, because it enabled to look at the knowledge and attitudes of the informants in an environment where no one else is there to affect what they say or feel comfortable saying and using questionnaire also gave the informants a chance to really think their answers through. If the data would have been collected for example in a group interview the presence of other people in the room might have affected the answers and if it would have been a one-to-one interview with the researcher it also might have pressured the informants to answer in a way that they feel the researchers want. The questionnaire was then sent to a selected group of people, who answered to the questionnaire according to the guidelines given to them in the questionnaire and gave opinions, insights and descriptions about these questions.
4.2 Assembling the questionnaire and collecting data

In this part the process of assembling the questionnaire and the collection of data is explained. The process of assembling the questionnaire is first explained from the general point of view and then as separate entities that arose from the theory before explaining the culling process and finally explaining how the final questions were formed and why it was done that way.

From the way the theory and from the research questions it was quite easy to determine that the questionnaire should have three parts: knowledge, attitudes and companies. From there the questions relevant to each of these parts needed to be complied; this was done by going through the theory once again and collecting needlessly large and comprehensive list of every possible question that could help somehow to help give us answers to our research questions. The second part of the process was to get these questions into order and under these titles that were decided earlier. Then the process of deciding what questions should be kept and what rejected began. In some cases two or several questions could be combined in a way that they did not lose any value that they could bring and in some cases it was clear that some questions were too precise or just did not have anything to do with the research questions, despite initially it seeming like it. After the questions were picked they did, however, undergo some more editing; even if the questions were now decided on they still needed to be worded in a way that would encourage the informants to write more, so that we could get as rich data as we could get, while simultaneously keeping them easy to understand. The questions went through this editing process a couple of times and after this they were tested on several people who were not informants on this research, but could provide some insights on how the actual informants could answer them.

The first part, knowledge, is designed to help answer the first sub-research question; what do the informants of this work know about big data marketing and privacy? There were different ways of doing this, but the one that was chosen could give us enough information about the subject while simultaneously getting information that is rich enough for us to get some insights from. In this part there were three questions; first one asking about few specific, but big enough to be, at least to some
extent, universally understood, terms, the second one about ways of collecting consumer data, and the last one about how it is used in marketing.

The second part, attitudes is designed to give answers to the second sub-research question: what are the attitudes that the informants of this work have towards big data marketing and privacy? The informants were told to answer the part one first, before reading through the second and third parts, because at the start of this part some examples on both how the data is collected and how it is used in marketing are given to now explain them and refresh their memory on concepts that are part of collecting and using big data in marketing, so they can accurately answer. If it was not have been done this way, then the informants could have not known that some of their experiences are linked to big data and thus leave important information out, that could help to determine the attitudes that the consumers have towards big data. This part also had three questions which were designed to, not only get attitudes, but also try and get some knowledge of the risks of data marketing. The informants were not asked, or told, about risks directly, because it could possibly affect their attitudes at the time of them filling the questionnaire.

The third and final part of the questionnaire was about companies actions specifically. By introducing these questions we could get some information on more specific points that surround the big data and privacy phenomenon, that according to the theoretical framework are important for the issue. In addition to that, asking these questions would be helpful in making sure that these points were not overlooked by the informants: these questions were important to ask and the informants probably would not have addressed these questions in the other answers, and these would not be any data on these subjects even though and they were important if the research questions were the be answered as thoroughly as possible. At the end of the questionnaire there was also a possibility to informants to give their thoughts or opinions on big data and on privacy that they had not thought about previously or could not fit in the scope of the actual questions. This in turn helped in getting data that was not directly addressed in the questions, but that informants felt important enough for the questionnaire and about the subjects to give, giving us data we did not know to ask.
The second phase of the process was to send the questions to the informants. The questionnaire was sent to eleven people, given instructions, and few weeks to answer the questionnaire. The informants were also informed how their data was going to be handled and that it would remain anonymous, so that they could feel comfortable about answering the questions and giving their opinions and experiences on the subject. The people that were chosen were similar in age, sex and other demographics. This was because in this research it was not sensible at this point to do research on these concepts using a wider range of people because if every informant used was from largely different demographics the patterns needed might not have been there and we could not get a clear picture of differences among similar informants at all, if the was major external differences amongst the informants. The informants were from a generation (ages 20-30) that has lived through the expansion of data usage in marketing and have grown up during it, and thus they have sufficient knowledge on these concepts to give informed answers on the questions. The informants were also people that are or would be in the near future, the ones that have the most money to use and would continue to be in that group of people for next few decades, thus making them people that companies should target their products and marketing to. So while this research would not be able to give universal answers about the knowledge and attitudes on consumers and draw conclusions from them, it would target the ones that would probably be the most important consumers for the companies; whose opinions would matter the most. The number of informants was chosen because it is big enough to get enough variation in the data but small enough that the data doesn't get repetitive and new information would not emerge. After these distinctions were made the actual choosing process for informants was really straightforward; these people were chosen because they fit the demographics chosen, they could be able to give insights on the subject and were able to be in the focus group and participate. In the next table you can see some information about the informants that participated in this research:
Table 1: Informants and their demographics

<table>
<thead>
<tr>
<th>Informant</th>
<th>Age</th>
<th>Sex</th>
<th>Educational background</th>
<th>Nationality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informant 1</td>
<td>24</td>
<td>Female</td>
<td>Master’s degree</td>
<td>Finnish</td>
</tr>
<tr>
<td>Informant 2</td>
<td>24</td>
<td>Female</td>
<td>Bachelor’s degree</td>
<td>Finnish</td>
</tr>
<tr>
<td>Informant 3</td>
<td>23</td>
<td>Female</td>
<td>Master’s degree (student)</td>
<td>Finnish</td>
</tr>
<tr>
<td>Informant 4</td>
<td>24</td>
<td>Female</td>
<td>Master’s degree</td>
<td>Finnish</td>
</tr>
<tr>
<td>Informant 5</td>
<td>25</td>
<td>Female</td>
<td>Master’s degree (student)</td>
<td>Finnish</td>
</tr>
<tr>
<td>Informant 6</td>
<td>24</td>
<td>Female</td>
<td>Bachelor’s degree</td>
<td>Finnish</td>
</tr>
<tr>
<td>Informant 7</td>
<td>23</td>
<td>Female</td>
<td>Master’s degree (student)</td>
<td>Finnish</td>
</tr>
<tr>
<td>Informant 8</td>
<td>28</td>
<td>Female</td>
<td>Master’s degree (student)</td>
<td>Finnish</td>
</tr>
<tr>
<td>Informant 9</td>
<td>24</td>
<td>Female</td>
<td>Master’s degree (student)</td>
<td>Finnish</td>
</tr>
<tr>
<td>Informant 10</td>
<td>25</td>
<td>Female</td>
<td>Bachelor’s degree</td>
<td>Finnish</td>
</tr>
<tr>
<td>Informant 11</td>
<td>25</td>
<td>Female</td>
<td>Master’s degree (student)</td>
<td>Finnish</td>
</tr>
</tbody>
</table>

4.3 Analyzing data and drawing conclusions

After the data was collected the next step was naturally to analyze it and to draw conclusions from it. In this part we go through the analyzing process, how we analyzed and organized the data and used it to draw conclusions.

The data did not come in all at once but gradually over two weeks. When every informant had given their answers, all the data was collected in one document with the informants’ answers on their own, which enabled us to look at the data from every single informant as a whole and to see the underlying attitudes and opinions about the data, from their answers, from their word choices and how their answers link together. After this was done, another document was made, which had every answer to one question after one another, which gave us a whole picture of the each question and their answers. By doing this it was in turn easy to see the patterns related to each question individually and get the impression of the informants’ attitudes or the extent of the knowledge from each question.

These two documents, one with one informants’ after one another and one with each of the questions after one another, were then used together to get a clear picture of the data and simultaneously seeing the patterns in the questions of one informant and the patterns and similarities in the answers that all informants have given to each
individual question. This enables doing the work efficiently and in a way that makes
the data approachable, easy to work with and this also minimizes the possibility of
making wrong conclusions, because the data is easy to check if that is needed. This
also helps with doing confirmability audit; it easy to go back and check that the
analysis is free from personal bias of the researcher.

In the second phase, the analysis itself was done by looking at the answers
themselves, the word choices that the informants used to describe their experiences,
opinions and thoughts on the subject and the entirety of the individual informant’s
questionnaire. The way the data was first approached was color coding the different
patterns that we were trying to find. The color coding was done by focusing on one
aspect first and then moving on. It was good to start the color coding on the negative
and positive comments, because they did not need going back and forth the text that
looking for patterns and repeating words would have needed. The second phase was
to look at repeating words and expressions in the answers, not only in the color
coded positive and negative comments separately, but together and in neutral
statements too. The first thing we did was to simply look at repeating words and
color code them, but later we also color coded repeating sentiments apparent from
these word choices, for example choosing to look at words that reveal uncertainty or
certainty and color coding them differently from each other.

The next step on the process, after the color coding was done, was to start trying to
explain the patterns that we did find. We went through the data in parts first because
it was the logical way approach the data and get the most out of it relating to the
research questions we had. We went through the data part by part and question by
question and looked at the patterns relating to those questions. The thing is, because
all the questions that were posed to the informants were linked together on some
level, the answers were linked together as well. So it is not enough to just look at the
answers individually, because then one could miss something relevant. So after going
through the answers individually, we looked at these parts which correlated to other
parts of the questionnaire too; first part by part and then at the questionnaire as a
whole as well. To achieve reliability data is used in analysis comprehensively and it
is constantly checked so that there are no errors.
The next phase of the process was to start to looking at the theory and assemble the patterns we found according to it; what parts are similar to points that were in the theory, what parts are missing in their entirety or what parts of theory have not addressed specific points that we found when we were analyzing the data. Then through this found answers to our research questions; the knowledge, the attitudes and finally what are the concepts and processes behind them that could make it beneficial to companies to take their customers’ privacy as a part of their strategy when looking at the things from the privacy point of view. The sub-research questions were much more straightforward to answer and the main research question could only be answered when there was information about the knowledge and attitudes that consumers have towards privacy in big data marketing.

In the analysis there are few different parts which we use to analyse the knowledge that the informants had; the parts about big data management cycle on both of the chapters gives us the scope in which we look at the knowledge and things that are or are not mentioned; what is there to know and how this knowledge resonates within the informants and effects the other parts in the research. In the analysis we used the big data management cycle and the privacy debate from the companies’ point of view and the points that arose from there as a starting point on how we look at these aspects and then how they relate to the data that we got from the informants. The points collected from the theory about the companies’ actions were also directly utilized in the analysis; looking at the data and comparing these points to the theory gives us more insight into the knowledge and specifically to the attitudes that the informants had on the subject.
5 ANALYSIS OF THE DATA

The questionnaire that we sent to the informants was in three different parts, and this chapter progresses in the same way and is structured using those three parts as the base. This kind of structure gives us clarity and ties the data to the research questions as the first part is the part that gives us data for the first sub-research question, the second for the second sub-research question and all three, including the third one which goes into the specifics, gives us data for the main research question. Although this was the base for it, as we will discover later, all of the questions played their part in determining conclusions to each research questions, regardless in which part they were structured to be in. At the beginning of each part we first go through each question and the informants’ answers, word choices and emerging patterns and then we make conclusions from the data and compare it to theory to find answers to the research questions.

5.1 Knowledge

The knowledge part of the questionnaire had three questions. On the first question the informants were asked to describe what comes to mind when someone mentions three different words privacy, data protection, and big data. These three words were given in this order because the word privacy is the most universally understood and easiest to answer based on their own experiences, viewpoints and opinions, while the concept of data privacy is derived from the concept of privacy and is more precise, and lastly big data is a concept that requires even more specific knowledge on how data works in today’s world. There were some fluctuations on the knowledge and understanding between the words and between the knowledge among the informants, but in general it was clear that the knowledge that the informants had was uncertain, incorrect or incomplete. There were also some words that repeated through answers and some positive and negative word choices that implicated attitudes as well as gave us data on the knowledge.

The first word, privacy, was the one which the informants had the most knowledge about and understood the concept the most; it is something that is very personal for people, thus it was easiest to answer and describe. Because it was very personal there
was no need to really regard the answers as knowledge, but as something that would be something that could help the informants to answer rest of the questions and on the other hand help the analyzing process serving as a base for the rest of the answers. It was interesting to see that some informants answered this question quite clinically when some informants approached it from very personal point of view.

“[Privacy] means that my information is not being used wrong and that I know who is using them. And to what means. I can keep the information I want “myself” and no one can access it if I don’t want to. The outsiders cannot access the information or act as me” Informant 1

Some informants also stated that privacy is a basic human right and something very intangible, while some approached it along the theme of the questionnaire and focused on the privacy from the data point of view. In addition to this informant 9 was very negative on the aspect of privacy; they outright said that privacy is something that people have lost with the technology, and you could not have privacy if you are using technology in today’s world. Informants 7 and 2 on the other said that privacy is “their own home and own space” and that “privacy is a right to do things in privacy”, so it was obvious that they regarded privacy as something very personal.

“[Privacy] is own space, own home, own peace; choosing who to let in and what you share with others. Secrets and through them a kind of anxiety ... …if you can’t share things with people it can be very scary” Informant 7

The thing that was repeated in every single answer that “outsiders”, “others” and “strangers” are a threat to their personal privacy, either by spreading information about them or posing as them to some extent. This information related well to the theory of the everyday definitions of privacy; avoidance of surveillance, scrutiny or unwanted access (Allen, 1988 via Margulis, 2003; Margulis, 1977)

The second word “data protection” was something that was marginally less easy for the informants to understand and describe, because it was less personal and more defined as a concept. The informants gave many different definitions to data privacy and all of them were more or less correct. For example informant 3 said that “data protection enables privacy”, and informant 5 said that “Data protection is the same
thing as privacy”, so informants recognized that data privacy is something that is closely linked to privacy in general. Data protection was seen as something that has “something to do with computers” and something that is “there to protect personal information”. The informant 10 said that as a regular person they could not be sure that data protection is enough and that they could not be sure that their personal information is truly safe. This made the informant nervous and because they said that they work with delicate information about the customers they have. In summary this question bought out the difficulty of defining these concepts and that there is no universal understanding on what it means.

The third concept the informants were asked to define “big data”. It was clear from the get go, and even more when the data came in, that this is the concept that would be the hardest for the informants to have anything to say about. The informants had some vague understanding that the term was something to do with information and internet, but could really not define it further. Informant 11 thought big data as an entity that collects the information and informant ten said that big data might be some kind of group marketing. Only informants 2 and 8 had concrete knowledge on the term, which somewhat fits the definitions given in the theory by for example Erevelles et al. (2016), Lycett (2013) and De Prato & Simon (2015), and did not use words like “maybe” or “I assume” or something similar while they described their knowledge on the concept.

“[Big data is] information about all the movements made by individuals collected in social media and elsewhere in the internet.” Informant 2

“[Big data is] large quantities of data, that keeps on growing. It’s collected, stored, shared, analyzed and compiled into statistics.” Informant 8

Based on this data it seems like that big data is a buzz word that does not have resonance in people that have not directly been in contact with it. The information we got was also similar to that in the theory: even some of the researchers have had difficulties in defining what big data is.

In the second question the informants were asked about their knowledge on data collection methods, what they think they are and where they might have gotten their
information on them. At first glance we could see that some aspects on data collection were more widely known than others; some things were repeated through answers, while some things were mentioned once or twice or not at all. In the theory there was a lot of information on how data is collected (e.g. Weinberg et al., 2015; Davenport et al., 2012; Everelles et al., 2016); Social media behavior, search histories, internet shopping and click data were all mentioned in one or more answers and sensor data and rest of IOT data was not mentioned at all. Repeating patterns evident from the data, in addition to the ways mentioned, were that although the informants knew that their data was being collected they said that they were not sure how that happens and if they knew how their data was being collected they did not know what information the companies got from them using these methods. A few of the informants also said that they are ashamed to admit that they really do not know more on the ways that companies collect data from them and if they knew they would probably be shocked.

“I have to admit that my understanding on how the data is collected is relatively minimal. I would probably be very shocked if I knew.” Informant 4

“[Companies collect data] with cookies at least when I’m visiting their websites… …However, it feels like that companies have vast amounts of knowledge on me and in my movements so I’m not sure what other methods they have.” Informant 5

One thing that was present in over half of the answers was that they have heard that it happens, or experienced that their smart phones have listened to their conversations and they find that being the most worrying aspect of data collection. An interesting point that was also raised by informants was that the they found that they had got the information that they do have from their friends, school, the news or they have noticed themselves how their behavior in the internet has affected the advertising that they have encountered. Not one informant said that they have gotten their information from the companies themselves.

The third question was parallel the second one, but this time we asked informants to give examples of their perceptions on how the data collected is used in marketing and where they have gotten their information. It was interesting to see that in this instance the answers varied only little. Target marketing and especially advertising
was mentioned in every single answer in some form, but it was really the only thing that was mentioned in more than one answer. Only other thing that was mentioned was segmentation and it was mentioned in only one answer very briefly.

“[Data is used in marketing] for targeted advertising.” Informant 3

“Data is collected with all the methods mentioned earlier and with this data marketers do targeted marketing” Informant 1

From these we could see that only the most easily noticeable and recognizable way of utilizing data in marketing is known about or at least what they can recognize from the ways that it is actually used. But in this question, as opposed to the previous one, the informants were much surer in their answers; they did not use words that would have implicated that they were unsure in their answers. All of this probably relates to the fact that data collection is something that is done in ways that are not as visible and that it is easier to be sure about things that they can see and this could be the reason why target marketing was the only thing that was mentioned in the answers. On the other hand this also speaks about lack of transparency in companies or at least them failing to communicate their ways of collecting and using data to the consumers.

After the informants answers were analyzed, there were some clear patterns that emerged. Privacy is a personal thing and it is understood differently depending on which point of view it is regarded from, data protection is something that is needed to enable privacy in the world today, but it is not something that is generally so well understood, because it is not as personal. Big data is an unclear concept that does not resonate within our informants and while data collection is in fact better understood than data use in marketing, the informants are less sure about what is going on because they do not understand or have not seen the evidence of this happening. Data collection is seen as something that makes them unsure, while they actually know more about it than they think and data’s use in marketing is something that the consumers are surer about, but cannot in fact in most cases give more than one example of how the data is actually used. When they can see that by googling content they get advertisements of said contents and they can add it together, but when they do not realize by themselves that data collection or marketing is being
done they are either confused or unaware of it taking place. The knowledge on the concepts, the data collection methods and how the data is used in marketing fluctuated a lot, in some cases it was positive to see how much was known, but in others it was clear that in some aspects of these things the knowledge just was not there. So all in all you can say that given the missing aspects and the fact that the informants were so unsure in many of these aspects that there is certainly more to know and learn. Companies could help this by being more transparent in their practices and/or making the information more accessible.

If we compare the information that we got from the answers of the first part to the theory to the knowledge that the informants had on the subject we can see some overlaps, but also things that were not mentioned, either because the informants do not know that they are or because they do not regard as data collection or marketing methods although they are. As explained earlier there were abundance of mentions of few methods of data collection, but some things were not mentioned at all. The few more elusive data collection methods like tracking pixels or flash cookies were not known of and the informants did not discuss more traditional data collection methods or most aspects of IOT data collection either (e.g Couldry & Turow, 2014; Erevelles et al., 2016). This represents a big gap in knowledge that the informants had and how the data collection is truly made and it also poses a security threat because these methods are the ones that do the data collection in a way that the consumers are not informed that this is happening. The big data marketing, the other point we wanted to gather consumers’ knowledge on, also had some gaps on how it is really done, some things like target marketing was something that informants knew a lot about, but on the other hand, and also reflecting the results on data collection, the more covert methods and things that do not directly affect consumers as such, like using data to do strategic and tactical marketing decisions or the behavioral analysis, were not mentioned at all. All of this speaks of marketers making decisions on personal information that the informants, and perhaps consumers in general, are not privy to, and that they have not really tried to inform customers about.
5.2 Attitudes

At the beginning of this part, as explained earlier, the informants were given some information on the data collection methods and data’s use in marketing to make sure that the informants could be able to answer the questions about their attitudes with an approximate knowledge on the subjects. When the data came in it was obvious that there was clear differences of opinions and attitudes in the answers, but there was also clear recognizable patterns that emerged from the data.

In the first question of this second part the informants were asked to describe their attitudes and feelings they associate with data and data privacy and then give examples of events that have led to them feeling this way. In the answers three groups of people emerged; those who described their attitudes as neutral, those who said that they do no worry about data collection and those who were very worried about it. Also it has to be noted that even those who said that they do not worry about the data collection, still had negative comments about it. The informant 7 was the only one whose only negative comment was that if something happened they would re-evaluate their stance on these issues. The other informants that were mainly positive about data said that they do not feel that it has a great impact on their lives and do not think that the data is collected could be used against them.

“…as long as I don’t feel like that the collected data could be used against me in some way, I don’t bother myself with it.” Informant 2

“I don’t think that I’m particularly sensitive about the data that is collected about me. I don’t care, if companies get this information…. …it can make my life easier. On the other hand it is difficult to determine where the line goes… …it could be scary and feel like I was being watched.” Informant 7

The informants also talked about their feelings that they are not probably doing enough to protect their privacy in the internet, but are not that worried about it or they feel that they are just so lazy that even if they are feeling uncomfortable, they do not have the energy to care about it. The negative comments about the role of data in the privacy debate and the worries they had about it varied a lot. The informants said in multiple cases that they are afraid that the data that the companies collect about them could be used in identity thefts, that their personal information could be
revealed or that they just simply feel uncomfortable knowing that someone or something out there knows so much about them and their behavior.

“I have heard about identity and information thefts so the thought of them scares me.” Informant 1

“I’d like to protect the data collected about me more, and I don’t like targeted marketing; it makes me anxious and scared… …In the recent years I have become more and more cautious about where my data is… …I think that hackers and other risks are an everyday problem, although I’m not particularly scared that someone would hack my computer…” Informant 10

Informants also said that privacy is something that simply does not exist anymore and that it is worrying and that they are annoyed that their information is used so relentlessly in marketing.

“Having privacy in today’s world feels impossible because it is so easy to collect data nowadays… …my information is being collected all the time whether I like it or not…” Informant 8

“I don’t believe in privacy. It is too hard to influence the data collection and my data is being collected in so many levels, that controlling it seems like a futile task…” Informant 9

Informants reported that their attitudes were affected by marketing done for them, here-say and media. One informant also said that the target advertising has annoyed them so much that it has affected their decisions to buy products online negatively. They were also afraid because they do not feel that they have enough information about these issues nor they have enough know-how to protect their information adequately. Only thing that they know how to do is to avoid certain sites and look what information they are giving. The problem for them was that even if they were uncomfortable about giving the information, they felt that they needed the apps or sites they would still give the information despite being vary. They also said that even if they know how to do something to protect their privacy, they are worried about people who do not have even the information and know-how that they have. All in all the data in this question was quite negative and even most of the positive comments had caveats around them, like nothing bad has happened to me so it is not on my mind, and that they do not know how it could be used against them.
The second and third question on this part, were also integrated together. In the second question the informants were asked about instances in which they have felt that they have benefitted from marketing done with data and the third question about instances they have felt uncomfortable or unsure when their data is being collected or when they have experienced data based marketing. The instances that the informants felt like they had benefitted from data based marketing ranged from faster and better service to better prices on products. Informant 6 also reported that they have gotten inspiration to their wardrobe from the advertisements they had gotten from the data that was collected when they were shopping online.

“… From the ads for clothes I have gotten inspiration for my clothing…”
Informant 6

“I had for a long time wanted to buy a skin care product that was a little more expensive than those I usually buy, and I found that product on 35% sale in Instagram.” Informant 5

“I was buying a computer and then got some ads that might have helped me decide” Informant 10

There were also two informants who did not feel like they had benefitted from data marketing at all; they said that targeted advertisements have never made them order anything and that target marketing has not been successful because they have already decided to buy something and only after it the advertisements have surfaced and that the advertisements do not usually match what they really want.

“I have never clicked on any of the ads and they have never helped me buy anything.” Informant 6

“Usually the targeted marketing only comes when, I have already decided to buy something… …in some cases they advertise products for me that are entirely different to those I want or have been looking for.” Informant 9

It seems like that the informants were not aware how the data marketing could help them aside from the things mentioned earlier. This transcends to the information gap again; if the informants, and more importantly the consumers in general, did know how they can and do benefit from data marketing more and the companies would
really succeed in it, it could mean that their negative feelings about the issue could be tempered.

The third question related to these negative feelings again as we mentioned earlier. Two of the most prominent things that seemed to stir negative feeling on the informants were that every single website and app is asking them for information, when they do not really know how the websites could use or even need the information that they are getting and too accurate and aggressive advertising. Other things that were mentioned was fear of viruses, selling tactics that are too personalized and that they do not know where their information is going and who is getting access to it.

“[Target marketing] can sometimes get a little creepy.” Informant 3

“Advertising that is extremely personalized make me feel a little paranoid…” Informant 8

“[I feel uncomfortable] if marketing is very aggressive and they try to sell you something almost with force” Informant 10

Adjectives that were used to describe their negative feelings ranged from creepy and distressing to suspicious and unpleasant. Some of the informants said that they cannot recall any bigger instances that they have felt uncomfortable with data collection and marketing also said they have felt surprised about seeing advertisements about products that they know they need, but cannot remember where this kind of data could have been collected and while they have not been uncomfortable with it has affected their decisions about where they are going to buy their products from, because they feel that no one should know these things when they have not given this information knowingly.

In the answers the informants gave us in the first part we also got some information on the attitudes that they have on privacy and big data. Ignorance on issues also could be, and probably is, playing into the reasons why data is not seen as a threat and if it is seen as one with nothing positive about it. The fact that the informants used examples and words mentioned earlier in this part is very revealing; data
collection and data marketing is seen as something that could be, and in some cases is, beneficial, but that on the other hand the methods that marketers use to do both are vague and threatening, because they do not know what is really going on and how these things are accomplished. The thing that is potentially very beneficial for companies to know is that even when the consumers that are not otherwise bothered with the data collection and marketing, do feel strongly enough about information they do not remember giving that they could change the websites from which they buy their products from.

When we compare this to the theory we can see some similarities and some points that were left unaddressed. By asking if the informants had felt uncomfortable with data marketing or collection we tried to get information on how the data marketing is seen, but also on what they know about the risks associated with data marketing and collection. Hackers, or someone else, revealing potentially harmful information about them was something that most of the informants mentioned in their answers in some shape of form, but the concrete consequences were not discussed. In fact some informants said that they do not know how someone could use their information to harmful purposes, so they are not worried. This is telling, again, on the fact that there is a significant lack of knowledge on these issues on the consumers’ side.

If we then look more closely at the attitudes prevalent from the answers we see more interesting things. Even when there was a clear positive outlook to data marketing and collection, there still were things that made those particular informants unsure or uncomfortable, but this was not the case on those answers that had a clear negative outlook. In fact one could go as far as saying that those who had a clear negative outlook, struggled to find anything positive or even neutral on the subjects that they were asked about. The cost-benefit relationship that the social exchange theory focuses on (Martin & Murphy, 2017) was also looked in a negative light; although some informants said that they have benefitted from target marketing there were also voices that said that the cost-benefit relationship forces them to disclose information even if they do not want to do that; if they want to use an app at all, they need to give out information. Certain anxiousness was something else that came though from the answers; no one was quite sure on their reasoning on what they based their opinions on and could not in most cases recall from where they had gotten their information.
Furthermore, White et al. (2008) said that the marketers might find themselves in problems regarding communication avoidance, information falsification, and derogatory word-of-mouth when the consumers feel that their privacy is being violated and this seems like something that comes through from the answers that the informants gave. This is connected to the knowledge that the consumers have and to the attitudes, because if there was misinformation or gaps in knowledge and these along with other negative feelings lead to negative attitudes, the reactions were much stronger and as we can see in the next part of this chapter these negative attitudes lead to further problems and negative reactions to things that companies are doing about and inside these privacy issues.

5.3 Companies

The third and final part of the questionnaire focuses on companies and their efforts on informing consumers, what they could do consumers’ reactions and the informants’ trust on the companies to make decisions about their privacy. As explained earlier these questions were meant to delve into specific issues that arose from the theory as things that really have an effect on the privacy debate surrounding big data marketing. Asking about these things from the informants helped to fill in the gaps that were not addressed in previous questions.

The first question was made to ask the informants about how and if the companies have tried to inform them on how data is collected and used or if they have notices companies trying to inform about the risks in collecting data. The informants had noticed about companies trying to inform them about some data collection methods, mainly cookies, but criticized on how things were handled. They thought that companies probably only informed them about cookies and, because they are forced to do so by law and that there probably is some information somewhere, but they have felt that it is not easily accessible so they have not found it or even wanted to try. Other informants commented only that they have not seen any information and some just do not want to think about it and just click ok without even thinking about reading anything. Also, as it came up earlier in analysis, they feel that it is impossible to access sites they want to visit or apps they want to use if they do not want to give out information, so they release information even if they do not really want to.
“The websites inform about them using cookies, but it is mandatory…” Informant 6

“Cookies, but the companies are forced by law to do that…” Informant 8

“Oftentimes I don’t have a choice… …if I want to download an app” Informant 10

“I don’t usually even notice anything like that, I just click okay without reading through” Informant 5

They also commented that companies have done “very little” to inform them and on the other hand one informant mentioned Facebook and said that they have done somewhat good job on informing them about data protection when their app updates and one mentioned companies asking them if it ok to use their information in marketing. Regarding the risks all the informants were unanimous on their answers about noticing companies informing them; some informants said that they have never noticed anything and some said that they have not noticed if they have, but the answer was still no. Only thing about the risks mentioned in the theory that was mentioned in the answers that the informants gave to the questionnaires was hacking (e.g. Kshetri, 2014) and it was seen as a problem, but the other ways for using data for malicious purposes were not mentioned. Using data to reveal sensitive information (Alemany Olicer & Vayre, 2015; Margulis, 2003) was implied, but that too only in joint with hacking. Even though we gave them information on the consumer profiles, the informants did not associate them with risks in their answers.

What we learned from the straight answers and the word choices, like “not having a choice”, “forced by law” and “information is hidden so that the consumer would not notice it”, that the informants used was that they felt that companies only, again, inform them on data collection, marketing and risks because they are forced to do so by law or do the bare minimum. They do not feel that companies want to do so and in some cases force them to give information in exchange from using their sites or apps and that they do not particularly like it. There is also the unfortunate case of it seeming like that the companies are not transparent on their practices and definitely not in the case of the possible risks when doing so. This is turn can potentially harm the trust that the informants had on the companies, mentioned in the theory as being
an important point in trying to handle consumers’ perceived feeling of control and vulnerability (Martin & Murphy, 2017; Martin et al., 2017)

The second question on this part was a little different; we asked the informants put some suggestions on how companies could make them feel safe in regards of privacy in order from one to five (one was the best way and five was the worst). Then the points were added together and we looked at the answers through them and from the point of view of an individual informant as well. From the most popular to the least the suggestions were: making it possible for informants to delete their own data from the company’s database (34 points), company committing to not selling their data forward (33 points), companies making it possible to follow how their data is used (33 points), company making their privacy policies understandable and easy to access (21 points), and getting privacy certifications (15 points). The most problematic aspect for companies arising from this data is that it seems like that given the choice most of the informants would want to delete their data from companies’ databases, thus making it impossible for companies to do data marketing in a way that they have used to. This, again, speaks of the distrust between the companies and informants; only two of the informants gave this suggestion less than four points. The two suggestions that were tied in the second place were also closely linked to each other on to the one the first place, by being less radical ways of controlling their personal data than just outright deleting it. The suggestions on the third place are in turn something that the companies themselves could do to help increase trust in their consumers; the points clearly dropped between these two and the three that came before thus making it seem like control would be something that was really appealing for our informants and things that the companies could do without giving this control to them, but still trying to be more transparent on their privacy practices is not as appealing. These things corresponded well with the theory with notions of control (e.g. Tucker, 2014; Norberg & Home, 2014; Xu et al., 2012) An interesting point in this, that was somewhat contradictory but at the same time not to the theory, was the question on privacy policies; the informants did not care for them and seems like that they did not feel that they are particularly relevant, although in theory Milne et al. (2004) suggested that privacy policies make them trust companies more.
In the third question of this part the informants gave us speculation on how they personally would react if their data was not handled correctly by one company and then something would happen that would result in notable inconvenience to them. There were two major groups of answers that arose from the data; first group would contact the company, but not take any other major actions against them and the second group would directly contact some kind of authority and ask them to handle this situation for them. These authorities ranged from lawyers, to officials handling consumer matters (like consumer ombudsman in Finland) to police. Some informants even talked about suing the companies, if the breach was serious enough. It was pertinent from the answers that the informants did not have, or did not feel that they have, concrete and/or reliable information on what a consumer could and should to do in this kind of situation and it played into their reactions to the issue. Other things that the informants mentioned in their answers in regard what they would do was boycotting the company in question, demanding compensations according to the inconvenience caused and demanding answers as to why this kind of breach has happened and what the companies are going to do to compensate and to prevent these kinds of things happening again.

“I would google what to do and then contact the consumer ombudsman”
Informant 1

“I probably wouldn’t know what to do, other than contact the company.”
Informant 4

“I would contact the company and demand them to answer for what happened (possibly even take legal action)” Informant 5

“I would demand fast corrective actions and deleting the data that the company has” Informant 10

This question provoked strong opinions and feelings from the informants; although some of them had commented earlier that they are not worried about their privacy and did not feel that these threats applied to them, in this answer those informants said that if something would happen they would take strong actions towards the companies. This translates to companies, obviously, needing to make sure that these kinds of things do not happen, so that they do not lose the trust of those consumers.
that do trust them and are positive about the data collection and marketing. Some informants also said that their reaction would depend on the information that was leaked and would adjust their reaction according to this. The study by Romanosky et al. (2014) said that only 4% of security breaches are actually followed by legal action and a study by Kannan et al. (2007) said that companies do not know to what extent security breaches harm companies. The data that we got would possibly indicate that this information is no longer true, or that it could at least be moving to different direction; as the data grows so do the risks.

The fourth and the last question was made to get information about whose decisions and promises about data privacy they could trust; where they would want these decisions to be made; by governments or by companies themselves. In this case the informants were overwhelmingly unanimous; every single informant said that they would trust the decisions made by governments more than the ones made by companies. There was only one informant that said that legislation on this issue could be too complicated for it to work, but that they do not as a rule trust the companies either. The reasons why the informants felt that they would trust the governments more differentiated a lot. Some informants said that they feel that companies would only drive their own agenda and that they would not feel like they would want to protect their privacy at expense of their own profits if it came to it.

“Companies that act in the competitive markets would be biased.” Informant 1

“The companies would want to have a maximum benefit from the data collected and in this instance the consumer would always come second” Informant 2

“I don’t trust companies to handle these issues, because they have a vested interest in the issues and would probably shirk their responsibility” Informant 5

“The data collection process should be much more transparent” Informant 8

From these choices we can see, again, that the distrust between companies comes through strongly and that it could possibly harm companies. The reasons why the informants felt that the decisions made by government would be more trustworthy included; impartiality and clarity and straightforwardness of rules and regulations
that are governmentally mandated. There were also answers given by the informants that had suggestions on how they would like to the legislation around this matter to be organized; the informants wanted to governments to listen to the general public and ask them, how they would want these rules to be made and what they want from them and they would like some legislative office to really supervise these matters and that the wrongdoings would have some consequences to the companies. One informant also noted that there should be some EU-wide legislation at minimum, preferably worldwide, but said that they know that it would be really hard to actually achieve this.

“The decisions made by governments feel safer than those made according to the data protection politics inside companies. It would be even better if these was some legislation on the EU or on global level…” Informant 3

“There should be a nationwide survey on [handling the legislation]. If only the companies would be in charge, I would never feel totally safe.” Informant 10

“Same rules and same consequences for everyone.” Informant 7

The informants would also like to be able to choose what information could be collected and where it is used and they would like the companies to treat them as individuals, with individuals wants and needs for their personal data. One informant also noted that it depends on the industry in which the company operates, if they would be able to trust them; companies that specialize on the information security issues they deem trustworthy enough to make them let do the decisions for them. In the theory it was suggested that the companies do not like to be regulated from outside sources and will continue to control their own privacy practices as they please (Romanosky et al. 2014), this is something that would seem like something that the informants would not want.

In addition to the specified questions we also had had space and asked for open comments and thoughts on the data and data privacy that had arisen during the questionnaire, but had not had the space or opportunity to write about them earlier. Several informants noted that they are very unsure about the issues because they do not know enough about it to be sure. One informant also said that in their opinion the atmosphere surrounding the issue is largely negative, because media always talks
about the issue negatively; they feel that regular people would benefit from knowing more about the issue and not just about the negative aspects either but in much larger scale. The other informants who said that the consumers should know more about these issues backed their statements up with necessity to know these things because everything in this world seems to be surrounded by data, the consumers do not know how to keep themselves safe and that the consumers should be given the chance to be able to recognize when their data is being collected and when it is being used. The informants also noted that a company who would be transparent would get accreditation from their customers. Several informants also noted that they feel bad about how little they know about these issues and it would be important if there would be more discussion about these issues in the general level.

“I should probably follow more closely where my data is being used.”
Informant 3

“In a world where everything is just big data transfer, the ordinary consumers should realize how much information about them is being collected.”
Informant 6

“It is true that a consumer can benefit from targeted marketing and using data. But this is only the case when the data and the marketing is being done and used correctly and when the consumer can recognize the situations where this happens.” Informant 1

The information acquired from the informants in this part again partly fit the things that we explained in the theory, but in some parts the data supported, completed and gave different viewpoints from those offered in the theory. And again in this part the most important thing was to see how these link together with the knowledge and the attitudes of the informants.

5.4 Summary

If we look more closely on the answers we got in general we can see that, like all the questions in the questionnaire, the answers that we got from the informants are linked together. There was a lot of overlap in the answers and in the whole picture, but also differences and even if some answers overlapped a lot, the answers to different questions offered a different point of view to the issues. It was really
interesting to see that there was a lot of variety on the informants’ questionnaires as a whole; there were those who think that data collection and its’ use in marketing is bad and those who think that it is good and beneficial for them, but also everything from between.

How the informants understood privacy and responded to the first part of the first question also projected to how the informants answered to the rest of the questions; those who looked at privacy more clinically, also answered the rest of the questions in more detached manner and those who answered to it with very personal and human centered way, also reacted to rest of the question more strongly, whether their answers were positive or negative. In theory it was suggested that the consumers would benefit more if they are paying for privacy (Conitzer et al., 2012), and that there are these entities that can provide anonymity for a price (Rust et al., 2002), but the question comes from the statements that the informants gave that the privacy is a basic human right, so would it seem fair for the informants that they could be expected to pay for their personal privacy. Big data seems something that is unknown and vague to our informants; those who know what it is could more accurately answer the rest of the questions. Although data as it is seemed to be something that is understood as a concept, it was clear that the few informants who understood the scale of data, the big data, were able to give opinions that corresponded to the real state of things.

One thing that the informants did not take position on was who is collecting their information; big companies like google and Facebook? All the companies? One informant even said that they think that big data is the entity that collects data from them. This again speaks of gaps in knowledge that surround these issues. This also relates to the fact that the informants did not mention any illegal or covert data collection methods as they are; they mentioned those which they could see and then said that this might happen but don’t know how.

In the theory (e.g. Martin & Murphy, 2017; Martin, 2015; Sheehan & Hoy, 2000) there were a lot of points regarding perceived control and vulnerability, but in the case of our informants only control was something that came through directly. The information that we got seemed to imply that perceived control is not enough, but
that the customers want total control of it or they might want to delete it. Vulnerability came through as a part of the state of knowledge; when the informants did not know or were not sure about something they reacted more strongly.

Martin & Murphy (2017) also argued that both parties, companies and consumers, can benefit from big data, but in the data that the informants in this work gave it was evident that in lot of cases the consumers do not recognize the potential benefits they get from it, aside from target advertising that was not working for a lot of the informants. This is turn can again strengthen the negative reactions, because the informants implied that the companies do data marketing just for their own benefit, with an information that they provide, so in this gaze the data marketing as seen as a somewhat unfair practice.

In contradiction to the theory (Aquirre et al., 2015; Miyazaki, 2008; Bleier & Eisenbeiss, 2015) most of the informants that answered the questionnaire did not feel that the personalization has given them enough value so that they would be willing to give away their information. They felt like they were not given a choice not to do so and this plays a big part on the negative feelings that the informants had about this; there is a failure somewhere in the service process if the customers feel that a company only tells about them collecting their information because they are forced to do so by law and that they give them so choice, but to give up their personal information if they need an access to their services and that they feel like the value they get out of it is not adding up to the control they give up. In the theory trust was said to lead to willingness to disclose information, purchase intent, click through and acceptance of advertising (Martin & Murphy, 2017), but what was evident from the data that the informants gave was that because of all the things that have gone wrong, or not working the trust is not there and they feel like they are forced to do decisions with little to none information available. The companies might think that what they do is enough and they are doing that is required from them by law, but in the mind of informants this is not true and if it is so can you really make and argument that you are catering to the consumers’ needs?

It was said in the theory that organizations need to have data privacy practices in place to make sure that the market functions smoothly (Conitzer et al., 2012; Rust et
al., 2002) and that companies might not want to enhance privacy protections because they do not know if these privacy failures harm their company (Kannan et al., 2007), but from the answers that the informants gave, we can see the different ways that might lead to negative outcomes for the companies. Firstly the informants said that they would be prepared to take actions towards a company that mishandles their information, this of course is to be expected and it still might not be that harmful to the companies. But, the second point we can see is that, because a lot of the informants had so negative attitudes towards data marketing and the trust towards companies handling their data was so low, the informants said that they would like governments to take action to protect their privacy and in some cases the informants even suggested that the governments should listen to the public when they are forming the regulations. This is a problem, not only because the regulations are coming, but because based on this information the less knowledge the consumers have on these issues, the more negative their attitudes are the more strongly they would like these things to be regulated and are willing to fight for it. If the companies would like to have consumers on their side on this, and avoid stronger and stricter data regulation that the consumers might want to advocate to have, something needs to be done to each and every stage of this process.

In the theory we had a part about privacy as a strategy. From the seven points that were introduced there by Martin & Murphy (2017) we could in some part see what should be done and our research supported those points for some part, but we also saw that this is not happening as it should. The gaps that the consumers have in their knowledge about these issues strengthen their negative attitudes and reactions and these attitudes in turn reflect the uncertainty and distrust that the informants had towards big data marketing and the companies that are doing it. It is seen that the privacy dialogue with consumers can benefit companies and that this should be a long term plan. What was evident from the data that there are several different stages that need to be looked at in detail, if the companies want to get consumers abroad to make privacy as a part of strategy that really works like it is supposed to. It starts at the knowledge, weeding misinformation and giving accurate information on the data collection, marketing and privacy practices to consumers so that they can make informed decisions and changing attitudes by enhancing knowledge, changing practices that the consumers feel need to be changed and trying to communicate with
consumers, not only when something goes wrong but also about the value that the marketers can give to consumers with using data and asking consumers what they want to do with their own personal information. This in turn helps companies manage reactions towards the things they do or do not do when they are trying to manage consumer privacy and communicate with their consumers in a way that yields best results for companies and consumers alike.

When we look at the theory from the companies’ point of view the things that are kind of glossed over are things that we found out; how the process of forming attitudes and opinions towards big data marketing and privacy works and how they are linked together and with the main research question and the wider research problem. The most important thing to remember is that nothing exists in vacuum; it is easy to find that all these things are linked together in most deep way, because that is what people do. From the analysis we can see how it could be very beneficial for companies and marketers to understand that all the knowledge, attitudes, opinions and grievances that the consumers have are very closely linked together. If they manage to understand the human component in a way that takes these links into account it can help them to keep the control in their hands when it comes to consumer data and the information it gives in their hands.
6 CONCLUSION

In this part the conclusions of this work are presented, including the answers to the research questions, managerial implications of the research, limitations and further research suggestions. First we gather together the observations made during the analysis as they fit to the research questions, the theoretical contributions, then we get to the managerial implications and lastly we look at the limitations and further research venues. It became clear in the theory that companies can benefit greatly from using big data to implement their marketing processes, but that privacy is a thing that keeps fighting them in the ethical front in this pursuit of consumer information. No one knows what data can reveal from people.

6.1 Theoretical contributions

This work introduces a model that shows how the knowledge, attitudes and companies’ actions work together when talking about big data marketing and privacy and what are the specific aspects that seem to have a big effect the interplay of these three different components in the model. In addition to this we can see how these aspects can affect companies when looking how they manage consumer privacy both directly and indirectly. In some cases it was evident that the information that the informants we able to give responded really well to the theoretical framework; the informants answered the questions in such a way that made clear that if they had control on their privacy decisions they would feel more at ease with data collection and its’ use on marketing.

From the analysis of the informants knowledge, attitudes and consumers opinions about things that companies do surrounding the privacy debate the most important thing that was learned was how all of these things are connected and that there is always an unknown component, the human matter, that plays into these things and that this should be taken into account when considering data marketing and privacy. From the next figure we can see a really simplified version of the conclusions of this research, how everything is linked and seeps through to the reasons why companies should take privacy into account when they are assembling and implementing their strategy. Later in this part we go into the deeper detail about the nuances of these
aspects and the major things that affect these aspects and what was revealed about
them from the data that the informants gave.

It was clear from the get go, and more so when the questionnaires came back, that
there are no universal opinions on the matter from the consumer point of view. The
purpose of this work was not to find some, but to find patterns among the data that
relate to each other and to the theory related to these issues. There were few patterns
that arose more strongly than other and some that were there but not as strongly. In
this part these patterns are collected together and gone through individually to get the
points previously discussed to one place and getting the answers to the research
questions.

The first sub-research question was; what do informants know about privacy in the
age of big data marketing? From the data it was evident that there are significant
gaps in knowledge that the informants had about big data marketing. In the case of
data collection, the informants had some idea of how the data is collected, but they
were very unsure about their information but on the data marketing side they were
much surer on how data is used, but only really knew one way that data is utilized in
everyday marketing. An important point regarding to the knowledge was that the
informants only knew things about this on what they could see; the more covert ways
of collection and using data in marketing were on most part totally unknown to the
informants. This can potentially speak about few things; there has been no education
on these things, because they are new things, there is no true will to find out about
these things and/or that there is a transparency problem between the companies and
consumers. The question is; how long the consumers are willing to let this be and not
informing themselves, or how long the companies can do this and think that their
customers do not need to know about all of these things. Another question that needs
to be asked is that if the consumers would know about all of this, would they be
willing to accept it or would they want more restrictions on how the companies
collect and use their personal data.

The second research question was: “What are consumers’ attitudes towards privacy
in the age of big data marketing?” As explained earlier, the knowledge and attitudes
are linked together, because the state of knowledge seems to feed into the attitudes
the informants had. When they were unsure about something they reacted more strongly and more negatively to things, than to those that they were sure about. In the case where an informant did not know anything about something, this phenomenon was something that came through even more strongly. The uncertainty feeds to the negative attitudes that the informants had about this subject. The positive stuff that the informants had to say in the questionnaire was largely related to the value that they get from the target advertising and not really on anything that companies do to them to protect their privacy or inform them about it.

A few of the informants actually spoke positively about the things that the target advertising does and is helped them to find things that they need and has given them inspiration to their style or lifestyle. It has to be noted that the positive remarks were das we established earlier and that is really worrying. Even the target marketing that was in some cases seen as positive, was dividing because some informants also described it as creepy and that it has never worked for them. In summary it can be said that even though the attitudes are not fully negative towards big data collection and marketing, the problem is there. The value that the informants recognized largely did not come from marketing, but from the access to websites and apps that the and in this context this poses a problem for consumers because even if they wanted to protect their privacy they cannot really do that, without compromising their needs that has lead them to try to access these websites and apps in the first place. And this can be further applied but the things that the informants said in their answers in the third part of the questionnaire.

Knowledge and attitudes are linked together with the main research question because they speak about problems that companies have with their consumers and the questions about their privacy. The research problem surrounding our research questions was “Why it could be beneficial should companies adopt privacy as a part of their strategy?” and the main research question that plays directly into this was “How the knowledge and attitudes play into the reasons why it could be beneficial for companies to adopt privacy as the part their strategy?”

As we can see from the model all the things are linked. The specific points that rose from the analysis of the knowledge that informants had or didn’t have and affected
greatly the bigger picture of this were the role of misinformation, uncertainty and confusion. These aspects played a major part in how the attitudes then play out; misinformation, uncertainty and confusion amplified both the negative attitudes and reactions towards companies and their actions. On the attitude aspects both negative and positive attitudes are affected by the transparency in the companies and how the consumers see if the companies are transparent or not. The second specific aspect what was revealed from data as important in regards of attitudes are the personal limits that the consumers have when the attitudes shift from positive to neutral to negative; once these limits are crossed the negative attitudes feed into the problems of trust that consumers have against companies. The companies’ actions play into the consumers’ knowledge and attitudes in different ways; the companies’ actions seem to affect the attitudes more than they do knowledge. The informants reported that they have not received a lot of information from the companies directly about their data marketing and privacy policies; even if the companies put privacy policies there to read, they do not largely affect the outcome. The companies actions, such as communication or lack of it, privacy policies and data breaches, however, do affect the attitudes of consumers in a large way whether negative or positive.

The next figure summarizes and simplifies the theoretical contributions of this work and explains the interplay of knowledge, attitudes and companies’ actions and their relation to the main research question. In light blue you can see the point where the interplay is not as strong as it is in the other cases, between the companies’ actions and the knowledge of consumers.
Figure 2: Model on how knowledge, attitudes and companies’ actions are interwoven together and what they mean.
6.2 Managerial implications

The managerial implications of this research are born straight from the theoretical contributions. The knowledge and attitudes of consumers affect both directly and indirectly to the way that big data and its relationship with privacy are seen and how the consumers react to it and through that to the reasons why companies should take privacy into account in their strategy. The way that the knowledge, attitudes and companies actions affect the reasons why the companies should take privacy into account when they are assembling and implementing their strategy depends on few factors already discussed in the previous chapter.

Firstly, it seems like that out informants were agreeing on something; the big data marketing and data collection, does not work for them as it is, or at least they think that the industry is not on their side. This could pose problems for the marketers that like to have their consumers on their side, while simultaneously profiting from the data marketing. Because the data marketing as it is today, is a relatively new thing there is a lot to gain now, but if, and when, the legislation changes and catches up to the speed of innovation it could mean big problems to them, when consumers want to regain their control on their privacy. As it was stated in the theory the informants felt that if they had more control on their data and if it was more tightly controlled, they would feel more comfortable with their privacy. A significant portion of the informants wanted some kind of policy change from governments to protect consumer privacy, because they felt that they could not trust companies to not try and drive their own interest at the expense of their personal privacy. Trust was a major point that arose both from the theory and from the data and something that the companies should try to work to achieve in regards of the privacy and big data marketing.

Secondly, there seems to be a transparency problem or at least a communication problem between companies and consumers relating to this issue. If companies could eliminate the confusion by being transparent in their data collection and marketing activities it could also mean that there would not be information floating around that is in fact untrue. The gap between the benefits that informants have realized, the disbenefits and the knowledge on these issues is big. The companies could try to
eliminate disinformation and tell openly about the benefits as well as telling their consumers how their data is collected and used and their privacy protected, so their consumers’ attitudes could shift into more positive direction and that they could feel safe and informed and know how marketing benefits them. It seemed like that on the other hand because the informants were not informed on the risks, they cannot protect themselves, but on the other hand because of the lack of transparency from the companies, the informants reacted more strongly on some issues that they otherwise would have.

Thirdly it was evident from the data that the informants felt like the companies do not like to inform them of their marketing practices or the risks associated with it; the companies’ actions did not affect the knowledge of the consumers majorly. This seemed to be a thing that majorly affected the attitudes negatively and diminished the trust on the companies. This would implicate that there is a room for improvement in the communication between companies and their customers; whether it is because it would simply benefit companies that their customers know about these things in more depth or because it would create greater trust. In addition to this one thing that rises up is the question of risks and the companies’ role in protecting the data that they have; if the consumers do not know how their data can be exploited with malicious intent, how they can make informed decisions on their data privacy and trust the right companies. If there is a company that is transparent and communicates with their consumers openly it could possibly mean significant competitive advantage for the company that openly informs and assures its’ customers that their data is indeed safe. And in addition to this, maybe if the consumers would know about the data practices in the companies they would not delete their information from their databases if given the change and their reactions to misgivings on data issues and privacy would not be as strong and would not result in demanding stricter legislation on the issues.

Fourthly, what was new was that they really do not care about companies getting privacy certifications or how they present their privacy policies. This could implicate that the best thing that companies could do is to try and make ways for consumers to control their privacy in a way that would benefit them too or at least try to develop an illusion of control so that the consumers could feel comfortable with data collection
and marketing. As we can see from the theory the big data management process is very complex phenomenon and because of the speed at which new innovations are made is so fast, it could also feed into the confusion; the informants out rightly said that they feel that they do not know enough to protect their privacy and because of the lack of transparency they are possibly feeling unsure where they could get the information because they feel that the companies are just trying to protect their own interests and do not trust them enough to have the integrity to do the decisions about data privacy on their own.

And lastly, something that is worrying from the consumers’ point of view that arose from the theory was the statement that the companies might not want to take actions to protect their customers’ privacy because they do not know how harmful the data breaches actually are for a company. The informants in this research reacted kind of strongly when asked if they would take actions against a company that has mishandled their information so it could potentially be very harmful for companies to mishandle the customers’ data. Another point that has to be thought of is that, because the data is so all-encompassing and is getting more and more so all the time as the data analytics evolves, the information that could be hacked and revealed could get more and more damning for their consumers. Due to this the conclusion is that even though companies now feel that data breaches may not be so serious for them now, that does not mean that they are not in the future and that the companies should take measures to ensure their consumers’ privacy to ensure that they do not get into trouble and that their consumers’ can continue to trust them with their privacy.

Lot of the things that play into this came through in the points that we have already discussed in this part; transparency, control and the knowledge gap being a few of the most important things. The ramifications of people’s misinformation on the subject, the perceived lack of transparency and negative feelings to the subject in general can be something that makes the marketers and the companies they represent struggle in collecting and using data in their everyday practices as they are using it today. The process of how this knowledge and attitudes play towards the reasons why it would be beneficial for companies to adopt privacy as their strategy is something that the companies should take into account and recognize the need for
actions in every step in the process. The companies should also take into account the way that everything is interwoven together; if there are problems in one aspect of the process, it affects the others as well. This also means that companies need to make sure that their privacy practices are secure and up to speed for this to be beneficial; a complete transparency can be a daunting possibility, but if it is done right it might lead to enhanced trust and ultimately to better understanding of the customers and the companies themselves.

6.3 Limitations and suggestions for further research

During every step of the research it was made sure that it would be reliable and valid, but of course there are limitations to this research. It is obvious that there is a margin of error in the data that we got from the informants and that it could be beneficial to do further and more in depth research on the smaller aspects of these conclusions presented in this work so that the model could be presented in more detail. Because in this research the informants represented one group of people there are some problems in generalizing these things, so we would suggest that a comparative research should be done to ascertain that the model presented in this work works in when looking at people from other demographics as well. This work also did not handle the issues from the legal point of view; the legal issues were handled in more general way; so in order to get more specific information on this it would be good to connect legislation with the issues presented.

6.4 Concluding comments

In this research we introduced a model that could support companies in figuring out how the knowledge and attitudes of consumers might affect the way that their data marketing practices and privacy policies are seen and how they could affect them in both positive and in negative light. The results of this research have proved that if the marketers start to value results over humans, it could potentially lead to some serious ramifications to the companies that marketers represent. The true dilemma is how to balance the value that the big data marketing can yield to the marketers and consumers to the potential risks that arise from it in a way that can ultimately help
companies to enhance trust and thus bringing more value to themselves in a world that is so dependent on information.
REFERENCES


Christiansen L (2011). Personal privacy and Internet marketing: An impossible conflict or a marriage made in heaven? *Business Horizons* 54, 6, 509-514


ICD (2014). The digital universe of opportunities: Rich data & increasing value of the Internet of Things


APPENDIX

The Questionnaire

Kysely

Vastaa kysymyksiin kolmessa osassa. Lue osiot kaksi ja kolme ja vastaa niihin vasta sen jälkeen, kun olet vastannut osaan yksi.

Osio 1: Tiedot

1. Kuvaa omin sanoin mitä sinulle tulee mieleen, kun joku mainitsee sanat

a) Yksityisyys

Kirjoita tekstiä napsauttamalla tätä.

b) Tietosuoja

Kirjoita tekstiä napsauttamalla tätä.

c) Big data

Kirjoita tekstiä napsauttamalla tätä.

2. Millainen käsitys sinulla on siitä, miten yritykset kokoavat dataa sinusta? Muistatko mitään tiettyjä esimerkkejä näistä asioista, joilla dataa kerätään? Ja jos muistat, voitko kertoa, mistä olisit mahdollisesti kuullut että tätä tapahtuu?

Kirjoita tekstiä napsauttamalla tätä.
3. Millainen käsitys sinulla on siitä, miten dataa käytetään markkinoinnissa hyväksi? Muistatko mitään tiettyjä esimerkkejä tästä? Ja jos muistat, kertoisitko, mistä olet saanut tietosi?

Kirjoita tekstiä napsauttamalla tätä.

Orio 2: Asenteet

**Esimerkkejä datan keruutavoista**

Kuluttajien antama data – Data mitä kuluttajat antavat eri netissä. Esim. Facebookin tykkäykset, eri profiileihin laitettavat tiedot, tekstit

Sensoridatat – Kerätään silloin, kun käytät erilaisia Internetissä olevia koneita kuten älypuhelimia, tietokoneita, autoja yms.


**Esimerkkejä markkinointikeinoista joissa hyödynnetään kuluttajien dataa**

Kohdistettu markkinointi (Target marketing) – Kohdistetulla markkinoinnilla tarkoitetaan sitä, kun sinusta kerätyn datan perusteella yritykset näyttävät/suuntaavat sinulle sellaisia mainoksia tai tuotteita, joita luulevat sinun haluavan nähdä.

Personalisoidut mainokset ja hinnat – Datat perusteella yritykset voivat rääätäliitädä mainoksia, netissivujensa tarjontaa tai hintoja sopivaksi kuluttajille.

Real time markkinointi – Markkinointia, mitä tehdään kuluttajille heidän sen hetkisen olinpaikkaansa perusteella (gps data) tai sen perusteella, mitä kuluttaja juuri sillä hetkellä etsii (hakuhistoriasta tuleva data).
Kuluttajaprofiilit – Data eri lähteistä kootaan yhteen ja yksittäiselle kuluttajalle luodaan profiili datan perusteella, joka sitten joko myydään eteenpäin toisille yrityksille tai pidetään yrityksen sisällä markkinoinnin työkaluna.

1. Voisitko kuvailla dataan ja datan yksityisyyteen liittyviä asenteita ja tuntemuksiasi. Kerro myös, mitä on tapahtunut ja miksi sinusta tuntuu tältä? (Voit käyttää adjektiiveja kuvailemiseen ja jotkin uutiset voivat olla esimerkkejä sellaisista tapahtumista, jotka ovat vaikuttaneet käsityksiisi)

Kirjoita tekstiä napsauttamalla tätä.

2. Muistatko sellaisia tilanteita, joissa olet hyötynyt kuluttajana datan käyttämisestä markkinoinnissa?

Kirjoita tekstiä napsauttamalla tätä.

3. Muistatko sellaisia tilanteita, joissa olet tuntenut itsesi epävarmaksi, joko silloin kun dataa kerätään tai kun sinulle markkinoidaan jotain?

Kirjoita tekstiä napsauttamalla tätä.

Osio 3: Yritykset

1. Oletko kuluttajana joskus huomannut yritysten yrittävän informoida sinua siitä, miten dataasi kerätään? Käytetään? Tai näihin liittyvistä riskeistä?

Kirjoita tekstiä napsauttamalla tätä.

2. Mitä yritykset mielestäsi voisivat tehdä, että tuntisit olosi turvalliseksi liittyen omaan yksityisyyteesi? (laita järjestykseen 1-5, 1 paras tapa ja 5 huonoin)
a) Todistusten tai yksityisyyssertifikaattien hankkiminen
Kirjoita tekstiä napsauttamalla tätä.

b) Sitoutuminen siihen, että dataasi ei myydä eteenpäin
Kirjoita tekstiä napsauttamalla tätä.

c) Yksityisyyskäytäntöjen (privacy policies) ymmärrettäviksi tekeminen ja niihin helposti käsiksi pääseminen
Kirjoita tekstiä napsauttamalla tätä.

d) Oman datan käytön seuraamisen mahdolliseksi tekeminen
Kirjoita tekstiä napsauttamalla tätä.

e) Oman datan poistaminen yrityksen tietokannasta
Kirjoita tekstiä napsauttamalla tätä.

3. Mitä teksit tilanteessa, että sinun datastasi ei ole pidetty oikealla tavalla huolto yrityksen sisällä ja jotain (esimerkiksi arkaluontoisen datan vuotaminen) tapahtuisi ja siitä seuraisi sinulle huomattavasti vaivaa?
Kirjoita tekstiä napsauttamalla tätä.

4. Missä sinun mielestäsi päätökset kuluttajien yksityisyysten suojelemisesta pitäisi tehdä? Luottaisitko enemmän yritysten vai hallitusten tekemiin päätöksiin? Voisitko luottaa yrityksiin ja heidän tekemiinsä päätöksiin?
Kirjoita tekstiä napsauttamalla tätä.

Tuleeko sinulla mieleen mitään yleisiä mielipiteitä tai ajatuksia liittyen dataan ja yksityisyysteen, joita sinulla oli ennen tähän kyselyyn vastaamista tai sinulle nousi mieleen tämän kyselyn jälkeen, mitä haluaisit jakaa, joita et voinut liittää vastauksiin tai et ollut ajatellut tätä ennen?
Kirjoita tekstiä napsauttamalla tätä.
Kiitos paljon vastauksistasi ja osallistumisesta tähän kyselyyn.