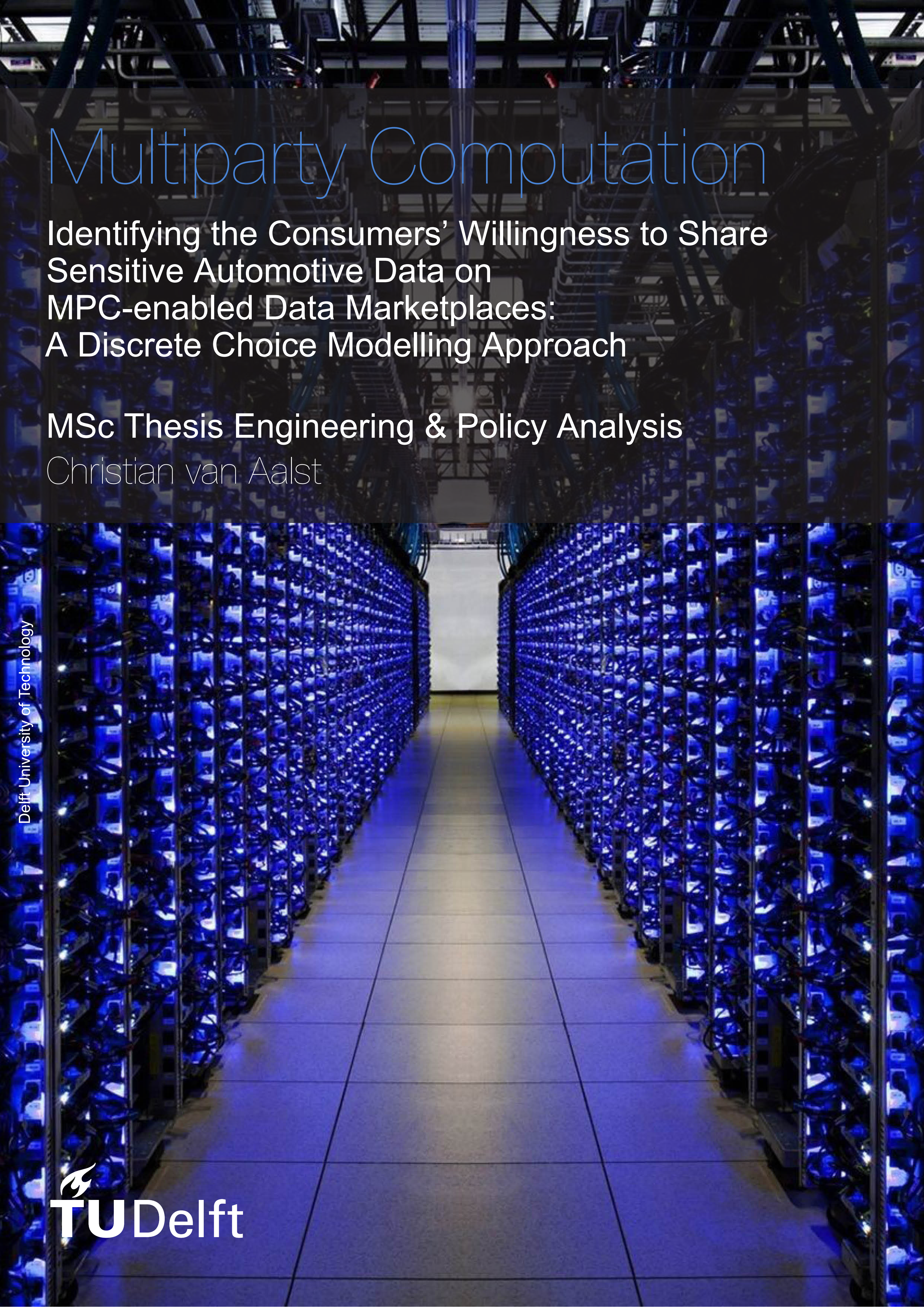
Multiparty Computation

Identifying the Consumers' Willingness to Share Sensitive Automotive Data on MPC-enabled Data Marketplaces: A Discrete Choice Modelling Approach

MSc Thesis Engineering & Policy Analysis

Christian van Aalst

Delft University of Technology



This page is intentionally left blank

Multiparty Computation: Identifying the Consumers' Willingness to Share Sensitive Automotive Data on MPC-enabled Data Marketplaces

A Discrete Choice Modelling Approach

Master thesis submitted to Delft University of Technology

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in **Engineering and Policy Analysis**

Faculty of Technology, Policy and Management

by

Christian van Aalst

Student Number: 4488571

To be defended in public on September 2, 2021

Graduation committee

Chair & First Supervisor	Dr.ir. G.A. (Mark) de Reuver	Information and Communication Technology
Second Supervisor	Dr. H.G. (Haiko) van der Voort	Organisation and Governance
Daily Supervisor	MSc (PhD) Wirawan Agahari	Information and Communication Technology

TUDelft

This page is intentionally left blank

Preface

Dear reader,

This thesis is for anyone who is is thrilled to learn about Multiparty Computation (MPC) and data marketplaces at the same time. As it is in the title, I have worked on the privacy preserving technology (PPT) MPC which situates itself in the field of cryptography. I explored what kind of effects this PPT had on consumers and tried to reveal their willingness to share automotive data on the so-called MPC-enabled data marketplaces. I aimed to do this by hands of a constructed stated choice experiment, where consumers were educated on the technology MPC and on data marketplaces and thereafter they were shown different MPC-enabled data marketplace alternatives which deviated in the factors *risk of data disclosure*, *data control*, *trust (in terms of herding effect)* and *benefit*. I estimated different types of models and ranked the factors on basis of the respondents' revealed preferences.

Personal note, in terms of process this research also requires some reflection notes. The agenda was always tight, doing research, performing experiments and writing things down. It was proven to be more difficult to launch a survey and comply with all laws and regulations set up by the ethical commission of the TU Delft. Certain phases as the start of the data collection took way more time than expected. Positive was that the amount of respondents came in very fast and the data analysis phase could start quickly. Perhaps, most of the hardships are found by problem demarcation. Where strong boundaries can lead to clarity, it can also limit the research in a way. Chosen was for strong boundaries regarding research on MPC in a user-based way, however this narrowed the research down too much in the author's experience and left out the very technical properties of MPC.

In spite of the fact that these hardships were clear lessons for future studies, being the captain of your own research and being able to execute and set sail in my own research was a very nice experience and will always be remembered.

Christian van Aalst
Rotterdam, August 2021

This page is intentionally left blank

Acknowledgements

When I started the final page of the MSc. Engineering & Policy Analysis, the thesis, I was thinking about how in earth I would perform my own research, basically most of the days from the second floor of my apartment. But, my EPA colleagues were awaiting the same routes, unfortunately these routes ran parallel. Nevertheless, Microsoft Teams became my best friend. But not the only one, you already may expect it, I have many people and friends to thank for the past six months.

Thanks goes out to my mom (Jacoline) and dad (Kees) which always supported me in every way during my time as a student in Delft. Thank you for encouraging and believing in me for the past five years and especially this last semester. Without you, things would have gone very different I guess.

I want to thank my roommates Wouter, Nick and at the end also Stan. Without the nice alternating moments between very serious discussions and laughing about everything and nothing, it would have been a lot more difficult to always focus on the thesis. You really brought joy to me.

I want to thank Nikki (girlfriend) for her tips and tricks regarding data visualisation and for always listening to my brain-dumps when we were together. Before I forget, I also want to apologize for the headache I must have given you.

Special thanks goes out to Wirawan Agahari, my daily supervisor who always guided me into the right direction and was always up to meet if there were things unclear from my side. Thank you very much for helping me the last six months and I'll wish you all the best in finalizing your work in the light of MPC-enabled data marketplaces.

In the end, I want to thank Mark de Reuver and Haiko van der Voort for the valuable feedback. Thanks for restructuring my mind each iteration, although I found it sometimes hard to admit to redirect course, you helped me push the research to a higher level. The same holds for you Wirawan.

Christian van Aalst
Rotterdam, August 2021

This page is intentionally left blank

Executive Summary

Over the last several years, the volume of data generated by the Internet of Things (IoT) has expanded at a rapid pace around the world. These often very detailed sensor data could be utilized by many different parties in order to improve services. This same tendency is seen in the automotive sector, where lots of data are gathered by newer and smarter cars. However, despite the fast data collection, car businesses seldom utilize this potential value. One way to make better use of automotive data is to share it with others via data marketplaces. Furthermore, data gathered by increasingly smarter cars is often very sensitive data. By aggregating and analyzing these car data, other parties can learn a lot about individual car users which could be experienced as unpleasant. Therefore, Multiparty Computation (MPC), a fairly new technology that facilitates encrypted and anonymized data sharing, tries to overcome this problem in order to increase worldwide data sharing.

Academics and businesses realize that these data marketplaces have huge potential but most of the research done on MPC is very technical. As Multiparty Computation is a fairly unknown and not yet massively adopted application and also not specifically on data marketplaces, it is unclear to what extent car users are willing to provide their car data. The scientific problem herein is to find the consumers' preferences in the main measurable data sharing factors which influence consumers' data sharing behavior on MPC-enabled data marketplaces. By finding and retrieving the weights of these factors, conclusions can be drawn on how to improve these data marketplaces by improving important factors. This paper aims to explore the consumers' preferences of MPC-enabled data marketplace factors to reveal the consumers' willingness to share automotive GPS data on data marketplaces in order to enhance data sharing and increase (road suggestion) products and thereby overall innovation. This knowledge can then be utilized to create a data marketplace in which data providers are more likely to join. To scope this research, the research is situated in the automotive sector which is considered advanced in terms of IoT-data generation and digitalisation.

Regarding the research gap, four research questions were revised in order to formulate a recommendation for reaching the research goal: 'Understand the factors affecting the willingness of consumers to participate and share automotive data on MPC-enabled data marketplaces by analysing the relative importance of data marketplace factors which consumers value when participating as data providers within data marketplaces'. The first research question was based on finding the most important and measurable factors which affect consumers' willingness to share automotive data based on comparisons with other data sharing sectors as most research on data sharing is in the medical-, academic and e-commerce sector. This question was developed to obtain a first qualitative set of factors which are key in data sharing and appropriate to use in future quantitative stated choice experiments. Then, the next research questions was focused on revealing the relative importance of these included data sharing factors. This question was answered in order to gain the quantitative insights in the ranking of each factor, to have a better understanding of the relative importance and to have the knowledge to find the most important bottlenecks in data sharing on MPC-enabled data marketplaces. The third research question was focused on checking whether there are differences among consumers in valuing different data sharing factors. This way, we could check whether models and consumers have differences in their 'taste' on the factors. This way, different models could be compared and statements could be drawn about the ranking of the models to have a word about which model to use in defining the conclusions. The last research question was based on attaching a general conclusion to the research and the obtained answers on the aforementioned three research questions to draw a final conclusion or recommendation.

To answer the first research question, literature research was conducted in order to find appropriate and measurable data sharing factors. To analyse the user valuation of these data sharing factors, the stated choice method was considered a appropriate method as users can hardly value a factor based on a single question. Users do not know themselves how important a factor is, however by

extracting their trade-offs in stated choice experiments, their preferences for specific factors could be extracted. During the process the literature process, 4 factors were considered appropriate in terms of importance and measurability to include in the stated choice experiment. The factors which were incorporated were *risk of data disclosure*, the probability in terms of amount of incidents per 100 incidents which were hypothetically probable when sharing automotive data on MPC-enabled data marketplaces. *Data control*, the amount of control the user had within the process of data sharing via MPC-enabled data marketplaces. This could either be in a centralised or decentralised form. The difference is about the storage and processing of the data during the MPC process, whereas a centralised setup saves and processes the data on a central server, the decentralised setup saves and processes the data in the users' own car. *Trust*, in terms of a herding effect, is about the way consumers make decisions. The herding effect shows that various consumers might imitate each other as consumers are inclined to think that other consumers are better informed and know what is right. This way, consumers herd and follow each other. Benefit, which is about the amount of money users retrieve on a monthly basis for sharing their automotive (GPS) data on monthly basis on MPC-enabled data marketplaces. An online survey was constructed to set up a stated choice experiment which was completed by 428 respondents. By processing the data with a discrete choice modeling (DCM) approach, an evaluation of the components was obtained, as well as their associated corresponding values. The analysis derived from research question two showed that risk of data disclosure was far most important to consumers, closely followed by benefit. Trust followed and data control was least important to consumers. The third research question was answered by analyzing and comparing different types of models and showed that the Mixed-Logit model (ML) was the best in explaining the obtained choice data. This model, incorporated and showed that there exists variety of taste among consumers when valuing the different factors. The fourth research question was answered by bundling the obtained information of aforementioned answers.

The results show that the feature importance of all factors is distributed as follows: Risk of data disclosure (38%), Benefit (30.3%), Trust (17.2%) and Data control (14.5%). This means that risk is good for 38% of the total importance of an alternative which is based on a combination of the four factors. What is important to see is that benefit is by far the second important factor. consumers value the amount of benefit over having more control over their automotive (GPS) data or more general trust in the MPC-enabled data marketplace. It is also shown that the benefit factor is not linear. The first increase in benefit from 0 to 10 dollar increased utility far more than an increase from 10 to 20 dollars. Furthermore, as the ML model explains the data best, it is shown that there is taste heterogeneity within the population when considering different alternatives. This means, as mentioned in the previous section, that consumers value factors differently as for example the factors risk of data disclosure or benefit. Different consumers have different preferences regarding which factors are important.

In terms of willingness to pay (WtP), in general, consumers are willing to deteriorate their low risk position (1 incident in 100 occasions) to moderate risk (5 incidents in 100 occasions) by receiving 9.50$ on a monthly basis. For an additional 20.40$, even to high risk. Consumers are very sensitive to risk of data disclosure. Furthermore, on average, consumers are willing to receive 6.90$ less on a monthly basis to have the MPC protocol installed at their own car to keep the data in their own car and have decentralised computations to share their data in a encrypted and anonymized way. Furthermore, by indexing consumers on different privacy concerned categories, it was shown that these three types of consumers (Privacy Fundamentalists, Privacy Pragmatists and Privacy unconcerned consumers) do not make significant different choices whenever it comes to choosing the most preferable MPC-enabled data marketplace to share their GPS automotive data on.

In the end, it is recommended to data marketplace platform owners to focus on the factors *risk of data disclosure* and *benefit* in order to attract more data providers and fulfill the demand of data consumers. This means, that beforehand the MPC-enabled data marketplace platform release, the techniques preserving privacy are solid. Thus by investing in data security and data security maintenance, these platform owners can mitigate the risk of data disclosure among data providers. Benefit, is like risk of data disclosure, a very cost-intensive factor. By ensuring that each data provider receives the right and fair amount of money in exchange for his or her data, consumers are more willing to share their automotive GPS data on MPC-enabled data marketplaces. These are by far the most important factors to data providers in deciding whether to use a MPC-enabled data marketplace to trade GPS data. This would mean that improvements of concerns of trust and data control would have far less effect in general. For other researchers, it is advised to focus on more complex factors within data sharing on MPC-enabled data. One may seek to dig deeper into MPC by introducing the option to be informed by a consent form, the probability of having malicious parties in the data sharing group, sharing different types of automotive data. Moreover, interviews could be valuable in asking the users themselves which factors they would find important in sharing automotive data on MPC-enabled data marketplaces. As the included factors in this thesis were based on comparisons with other data sharing sectors within literature, users could come up with very different factors. As noted before, mostly general attributes were included in this choice experiment and under limited and simple attribute levels in order to make the experiment understandable for respondents and measurable in the analysis. It is also possible to dive into the deep understanding of these general attributes, in forms of qualitative research. As mentioned before, MPC is still relatively new and behavioral user experience is lagging behind as it is not operational yet. Furthermore, as most research on MPC and MPC-enabled data marketplaces is very technical, it is important to also focus on the value proposition side of MPC-enabled data marketplaces in order to attract data providers and realise MPC-enabled data marketplaces. These types of research can yield new insights.

Contents

List of Figures

List of Tables

Part I

Demarcation of this Thesis

1 Introduction

If people were unbound from all obstacles that exist in data sharing, this would increase huge amounts of opportunities to increase innovation and wealth. Multiparty Computation (MPC) could enable the removal of these barriers and could be groundbreaking in accelerating data sharing. However, the shift to this emerging technology will impose multiple challenges which need to be faced first. Consequently, the context of exchanging data and Multiparty Computation will be presented in 1.1, and the potential and barriers of MPC are briefly stated. This thesis is situated in the automotive field, because the automotive field is full of data sharing opportunities and an important source of accelerating internet of things (IoT) data generation due to increasingly smarter cars. Per primarily research, the problem statement is identified in 1.2. Then, the goal of the research is stated in 1.3. Then first, in 1.4 the academic- and societal relevance are addressed. Thereafter, questions leading to the research goal are stated in 1.5. The research design is discussed and visualised in 1.6 and a structure and outline of this thesis in 1.8. In the end, the context wherein this project is situated is addressed in 1.7 and a comment on the suitability of this thesis within the MSc's program in addressed in 1.9.

1.1. Research Context

While worldwide data and loads of information captured by car usage increase massively, data sharing with automotive firms stagnates. M. Chen, Mao, and Liu (2014) and Kaiser et al. (2018) already stated that IoT applications generate large volumes of data everyday which could be very valuable to car manufacturers or telematics providers to make room for consumer-to-business (C2B) data sharing. IoT data are thus very valuable for improving and monitoring business processes or even for the (re-)construction of (new) business models in the automotive field (Noronha, Moriarty, O'Connell, and Villa, 2014). IoT data could thus be more and more valuable when it is shared to businesses (I. Lee and Lee, 2015).

However, while data sharing could generate more value due to aggregation, sharing sensitive data is not without risks. Due to an increase in the gathering of sensitive information by online businesses and a number of additional breaches in recent years, privacy research has gained traction in the systems engineering sector (Harborth and Pape, 2020). Even inside public organizations, there are still substantial hurdles to data exchange and reuse. The social and economic consequences of a potential disclosure of sensitive information (e.g. personal data and trade secrets) are frequently the primary reasons for people and organizations refusing to share their data (OECD., 2019). It is unclear whether individuals can effectively assess the dangers associated with data sharing, allowing them to give informed permission for using their personal information (Skatova et al., 2013). Moreover, as users are unaware and lack understanding of many possible privacy threats and consequences (Harbach, Fahl, and Smith, 2014), users make intuitive risk judgements as a result of their lack of awareness of probable outcomes (Bal, Rannenberg, and Hong, 2015). Plus it is shown that by rewarding people for sharing their data, this can increase the willingness to share (Jian and Jeffres, 2006; S. Kim and Lee, 2006), without decreasing the risks.

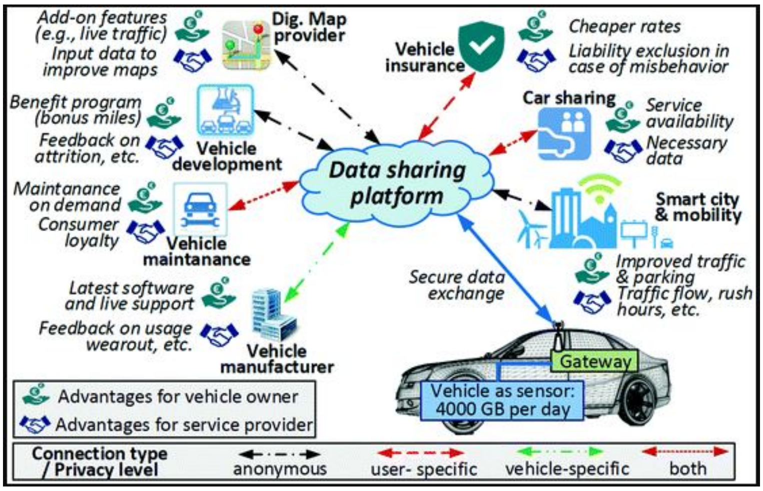
Recent years it has become clear that data sharing could accelerate innovation and welfare worldwide. This is true because it brings companies multiple benefits: as data can be exchanged and it can generate revenue (Thomas and Leiponen, 2016) or shared to other companies to improve businesses (Dwaikat, Money, Behashti, and Salehi-Sangari, 2018; Kumar, Pugazhendhi, Muralidharan, and Murali, 2018; Veselovská, Kožárová, and Zavadsky, 2018). Due to these facts, there have already multiple data marketplaces been risen up where all sorts of data is traded (Arzberger et al., 2004). However, whereas in the Dutch Horticulture industry a start is made in trading data among businesses via data marketplaces and by regulation (De Prieëlle, De Reuver, and Rezaei, 2020), the automotive sector is one of the first sectors where data marketplaces reached some form of maturity (Bergman, 2020). Nevertheless, due to a combination of regulations and safety reasons in combination with difficulties in stating data ownership, there still exists a lack of data sharing in the automotive industry (Mosterd, Sobota, van de Kaa, Ding, and de Reuver, 2021). Furthermore, based on multiple interviews done by Agahari (2020), car manufacturers often tend to have no idea about what to do with the obtained user-data. As there are thus many privacy related terms which preserve data sharing, active firms often tend to just store the data to avoid risks. Moreover, by trading these sensitive data to other companies which can in turn aggregate these data with their own data to increase its value - the manufacturers which are willing to share data actually deteriorate their own market-position vis à vis. This current business climate often leads to the decision to keep the data safe and sound on own servers which opposes data sharing. Due to this sphere, consumers are indirectly also not incentivised by car-companies to contribute their data on data marketplaces.

Because of the secretive nature of the automotive sector, which generates multiple types of sensitive data - consumers come up with well understood privacy or security concerns for rejection of data sharing. Hence, it is important for data marketplace owners to construct new secretive trading-methods on data marketplaces which strive to encourage automotive C2B data sharing to attract consumers and start accelerating tradings. These new ways of sharing IoT data from the automotive sector will make data governance an important topic as data sharing poses many legal issues, for example about data ownership and the way it may or may not be distributed (Cheong and Chang, 2007).

Nonetheless, the growth of the world's connectivity generated by the increasing use of the internet has created vast amounts of opportunities for parties to jointly calculate functions by entering their private data. However, within the pools wherein multiple participants share data, not always all participants are trusted. Generally, participants in such pool can be divided in trusted, partially trusted and even non-trusted actors. The latter two cause privacy concerns. Here, Multi-Party Computation (MPC) comes in to ensure that no private input-data is revealed to other participants (Canetti, Feige, Goldreich, and Naor, 1996). One of the first works on MPC was done by Yao (1986), where-after many more researchers started researching on MPC in the mid 80's and 90's. Noticeably, there seems to be a big gap in literature within the years 2000 to 2015, till first MPC thoughts in the automotive context about car traffic where proposed.

Today there are already multiple methods to achieve this goal of secretive sharing, but the encryption and sharing of data by Multiparty Computation on data marketplaces seems to have enormous potential. MPC is a cryptographic technique which allows to distribute data among parties without disclosing the sensitive data (Archer et al., 2018; Choi and Butler, 2019; M. de Reuver, Fiebig, Agahari, and Faujdar, 2020; Zhao et al., 2019). This thus allows for transition towards MPC-enabled data marketplaces which offers consumers the opportunity to safely share their sensitive automotive data without disclosing it. However, Kanger and Pruulmann-Vengerfeldt (2015) identified barriers which withhold MPC from being used. As MPC is still a relatively unknown and complex technology to understand, this could lead to consumers not feeling the need for MPC. Furthermore, the Usable and Efficient Secure Multiparty Computation (UaESMC), a project of the European Commission, identified barriers in terms of data visibility and transparency (Toldsepp, Pruulmann-Vengerfeldt, and Laud, 2012). However, currently the willingness of consumers to share data with automotive companies MPC-enabled data marketplaces is still unknown. Since it is often not explicitly said why consumers refuse or fail to involve in exchanging automotive data on MPC-enabled data marketplaces(Gubbi, Buyya, Marusic, and Palaniswami, 2013), it is valuable for MPC software developers which try to increase the speed of this move to a exchanging-culture, to look into consumer-behaviour on data marketplaces to determine their preferences and trade-offs in automotive data sharing by MPC. The scientific contribution herein lies, to distinguish and show the importance of these barriers in a well set up experiment in order to improve that perceived level of adaption of MPC.

Figure 1.1: Vehicle usage data used by various actors (Kaiser et al., 2018)


Add-on features (e.g., live traffic)
Input data to improve maps
Dig. Map provider
Vehicle insurance
Cheaper rates
Liability exclusion in case of misbehavior
Benefit program (bonus miles)
Feedback on attrition, etc.
Vehicle development
Car sharing
Service availability
Necessary data
Data sharing platform
Maintanance on demand
Consumer loyalty
Vehicle maintanance
Smart city & mobility
Secure data exchange
Improved traffic & parking
Traffic flow, rush hours, etc.
Latest software and live support
Feedback on usage wearout, etc.
Vehicle manufacturer
Gateway
Vehicle as sensor: 4000 GB per day
Advantages for vehicle owner
Advantages for service provider
Connection type / Privacy level
anonymous
user- specific
vehicle-specific
both

The research is positioned within the automotive sector, as cars are currently getting more and more advanced electronic add-on services (IoT) that generate data, allowing them to serve more as a central platform, driving a change in the value proposition (Athanasopoulou, Bouwman, Nikayin, and de Reuver, 2016). This way, everyday, more and more data is gathered which is key for data sharing on MPC-enabled data marketplaces and therefore increases innovation and welfare. Figure 1.1 clearly illustrates the current situation and potentials in the automotive sector when data sharing occurs. Hence, by gathering these loads of data, security and privacy are the core topics to talk about in the willingness of people to share data (Viereckl, Ahlemann, Koster, and Jursch, 2015).

People's inclination to exchange information in data marketplaces is influenced by a variety of known and unknown factors that vary by person and industry. Literature describes a bundle of common factors in data sharing within the C2B automotive-sector: trust (Kanger and Pruulmann-Vengerfeldt, 2015; M. Spiekermann, 2019), the amount of data control (Koutroumpis, Leiponen, and Thomas, 2017, 2020, the risk of data disclosure (Koch, Krenn, Pellegrino, and Ramacher, 2021) and current benefits (Derikx, De Reuver, and Kroesen, 2016). In addition, based on conjoint analysis of Derikx, De Reuver, and Kroesen (2016), person-specific characteristics such as age, gender and educational level are showing to have significant influence on in the willingness to exchange automotive data of customers. In this research, we are going to compare these very concrete and measurable factors with other sectors and perform the stated choice experiment in the context of MPC-enabled data marketplaces.

1.2. Problem Description

The data created by IoT-equipment is primarily owned by the manufacturers of these devices but is frequently private consumer data (Javaid, Zahid, Ali, Khan, Noshad, and Javaid, 2019). External parties on the other hand, may benefit from access to such data; the difficulty is granting them permission on terms that data owners deem appropriate. There's a chance to create a data marketplace where IoT-users can sell personal automotive data and data buyers can buy it. Academics and businesses realize that these data marketplaces have huge potential (Mišura and Žagar, 2016). However, much research on data marketplaces is too technical or about; the pricing of data (Fricker and Maksimov, 2017) and value proposition (M. Spiekermann, 2019), or marketing data (Leon et al., 2013), but hardly any research is done on the willingness of people to share on these data marketplaces. To reach a strong platform, both data suppliers and customers need to adapt to the platform. This makes it important to study on why customers do or do not adapt to these platforms in order to contribute to literature on C2B data sharing via central platforms.

Academics are realizing that MPC could create value in various sectors and on multiple societal fronts (Bestavros, Lapets, and Varia, 2017; Lapets et al., 2018; Lapets, Volgushev, Bestavros, Jansen, and Varia, 2016). Furthermore, most of the past research on MPC is technical, but little research is done on the user research of MPC. Also, hardly any research is done on the effect of MPC-technology on data marketplaces yet (Roman and Vu, 2018). There are enormous amounts of studies performed on revealing the consumer's willingness to share data in general, however most of these studies are about health data (K. Kim, Sankar, Wilson, and Haynes, 2017), scientific data (Ghosh, 2018) or e-commerce data (Sarkar, 2015), and very limited in the automotive setting. The automotive setting could be seen different due to its commercial aspects and its sensitive data whereas health and scientific data often is situated in a non-commercial setting and e-commerce data not involves sensitive data. Therefore, as the automotive sector reaches out to commercially-tinted sensitive data, this sector differentiates among the sectors where consumers behavior in data sharing is studied extensively. Specifically, it is not yet studied which factors within automotive data sharing in a MPC-enabled data marketplaces are most explaining consumer willingness to share data.

Therefore, it is interesting to investigate which of these factors are the most important for people in order to help the companies which are developing data marketplaces to focus on the most important aspects. Perhaps software developers are focusing mostly on reducing of risk of data disclosure whereas trust is far more important. Or the focus is mostly on the amount of control people have over their data whereas the amount of benefit people get is more important. And so on, business developers could be mostly focusing on the business model and the benefits whereas people might be less interested in benefits and are more aware and worrying about the security of their data. The problem is thus, that it is yet unknown what data sharing factors on MPC-enabled data marketplaces are most important to consumers. A combination of all these aspects will translate into ”MPC-configurations”, which will consist of variations of all these factors together.

The research problem in this context is thus visioned from perspective of the data marketplace developers. These are seen as the problem-owner in this research as they want to clarify which factors are important to people in order to improve these and thus attract more people in automotive data sharing to enable car manufacturers in aggregating these data to improve their services. Within the context of the knowledge gap relating the unknown willingness of people to share data in C2B automotive data marketplaces that is described earlier - this paper aims to explore the effects of data marketplace factors on the consumers' willingness to share automotive data on data marketplaces in order to enhance data sharing.

To achieve this, understanding which factors influence the willingness to share data via MPC-enabled data marketplaces is vital in order to improve people's interest in sharing automotive data and participate in data marketplaces.

1.3. Goal of this study

The study's goal will be to address the stated research gaps and provide a specific solution that will help solve the given scenario. Below, the research goal is stated:

Understand the factors affecting the willingness of people to participate and share automotive data on MPC-enabled data marketplaces by analysing the relative importance of data marketplace factors which people value when participating as data providers within data marketplaces.

To put it another way, the research will investigate the factors that influence citizens' interest in engaging in a data marketplace for automotive data, as well as the respective value of these factors. By comparing the factors, this will be accomplished. Gaining an understanding of the important aspects and their relative importance will help figure out what needs to be done to change the current data-exchange environment. As a result, in order to fulfill the study's goals, this study will be problem-solving and practice-oriented.

The goal of this exploratory study is to collect and analyze quantitative data in order to identify trends. (Goundar, 2012). To explore the potential and people's willingness to share data via MPC-enabled data marketplaces, discrete choice-experiments are envisioned to be the right tool to accomplish this. This method of examining actors' behavior using Discrete Choice Modeling (DCM) could be regarded as exploratory in nature. (Salampessy et al., 2015). Also, the appeal of data marketplaces and the illustration of causal links can be shown between factors in addition to identifying patterns. As there is a lack of understanding of the current functioning of MPC-enabled data marketplaces, this is consistent with the overarching purpose of this study, which is to visualize the influence of MPC in specific data marketplaces and contribute to global data sharing acceleration.

1.4. Practical and Academical relevance

The practical relevance of this study is to get more insight into the relative importance of factors that influence the participation of automotive data providers on MPC-enabled data marketplaces. These digital platforms allow automotive firms and telematics providers to collectively create value with these data. This data marketplace could be seen as a multi-sided platform which enables the interaction between participants as data providers and automotive firms (S. U. Lee, Zhu, and Jeffery, 2017). To have a successful platform, the amount of data providers is very essential. These platforms are typified by the network effects, which indicate the correlation between providers and end-users (M. de Reuver, Sørensen, and Basole, 2018). When data markets reach a critical mass of demand, additional end-users are likely to enter as the service will become more attractive. Thus, for sustainability reasons, this study could bridge the gap to increase the supply of data which is highly relevant and an important research aspect regarding online platforms (M. de Reuver, Sørensen, and Basole, 2018). The participation of both sides is needed but it is yet unclear how this participation is guaranteed. Furthermore, value development is a driver to keep ahead of the competition, so not utilizing a significant part of future value which can be utilized is undesirable. Awareness of what data providers consider to be essential in joining a data network and providing data to share with the community will help to ensure data provider participation. Platforms typically seek 'generativity', which refers to the platform being a self-contained system capable of producing new content, structure, or behavior without the need for feedback from the system's creator (Tilson, Lyytinen, and Sørensen, 2010).

Whenever it comes to the academical relevance of this thesis, because little has been known about why consumers falter or refuse to share the automotive data (Gubbi, Buyya, Marusic, and Palaniswami, 2013), knowing which considerations are relevant would contribute literature in the data sharing environment. As a consequence, when academics know which factors are important to consumer in data sharing on data marketplaces, this could be generalized and further used in technical research on data marketplace architectures or the types of usage of privacy preserving technologies on these platforms. In the end, this research in consumer valuation of various factors in automotive data sharing on data marketplaces, contributes with a proposed ranking of factors, to literature by giving an overview of the important data sharing factors to focus on in further research. This will also assist researchers in analysing the value from the rapidly increasing amount of produced IoT data as sharing IoT data with other organizations is closely correlated with value creation from IoT data (Jernigan, Ransbotham, and Kiron, 2016).

1.5. Research Questions

Solving the study questions would give the necessary competence to accomplish the above-mentioned research goal. The MPC technology is expected to have multiple direct and indirect (co-variables) impacts on peoples' perspective of data sharing, such as impacts on trust, on perceived data control, or on preserved privacy levels. At this step, there is not yet considered which attributes to include in the choice experiments. To answer this question, we need to consult the literature and reports on MPC and consider ourselves which of the dozen of attributes seem most likely to have impact on peoples' choice behaviour. This way, the choice experiment does not get too big for respondents to argue their choices. The first sub-question will gather:

1: What factors could drive consumers' choices regarding sharing automotive data on MPC-enabled data marketplaces?

After the retrieval of sufficient responses, it is important to analyse which attributes thus play a significant role in explaining directions of choices to generalize findings for the population. This way, the results can be used in further research in conceptualising MPC-enabled data marketplaces. By obtaining the choice data, this will provide a list of the respective values within the field of sharing data via these types of data marketplaces:

2: What is the relative importance of factors that influence the willingness of consumers to join and share data in MPC-enabled data marketplaces?

It is important when comparing different configurations of MPC, that the implications and limitations in the relative importance are clearly addressed. This way the over- and underestimations of the retrieved results are mentioned and the validity and reliability of the performed research can be discussed. Therefore, different models are estimated and socio-demographic effects on the main factors will be tested.

3: How does the respective value of factors differ across different people?

The last research question is aimed at analysing the results of all prior research questions in order to translate the findings into a clear story and to have a word to improve the willingness to share automotive data by data providers which are in fact car users or people in general. These revelations may assist people in getting a greater understanding of the conditions that led to the knowledge gap that this study is looking into.

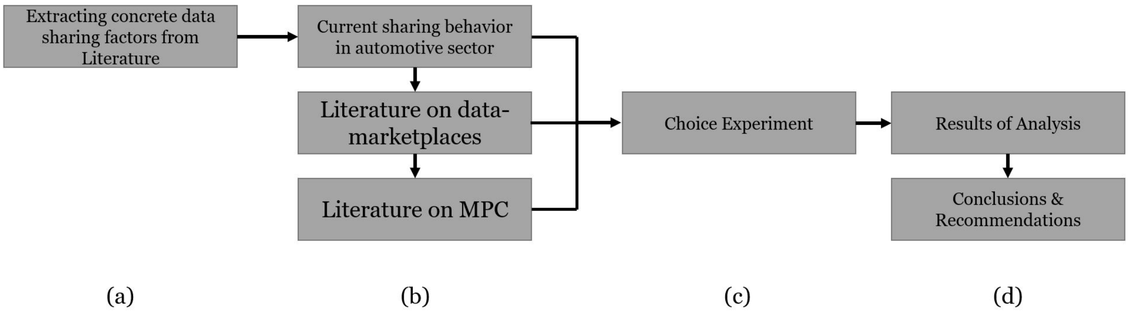
4: What are the policy inferences of the factors and their respective value?

1.6. Research Design

After identification of the data sharing factors within the automotive sector by extracting literature, question 1 will be answered qualitatively. Essentially, this research is a quantitative research as multiple explorative methods are used to answer the other research questions. As explorative research is about collecting and analyzing numerical data which can further be used to find certain patterns (Goundar, 2012), discrete choice-experiments are set up in order to obtain stated preferences of consumers in data sharing on MPC-enabled data marketplaces. This Discrete Choice Modeling (DCM) method for analyzing consumers' behavior, can be described as explorative in nature (Salampessy et al., 2015). Besides finding patterns, attractiveness for specific combinations of factors can be estimated and causal relationships can be shown. This fits with the overall goal of this research which consists of visualising the effect of combinations of data sharing factors on peoples' willingness to share automotive data. Furthermore, sampling will be done via a respondent pooling platform where respondents are incentivised by monetary benefits to collect the data efficiently. The respondents group will be broadly based on all adults without other requirements.

To structure this research, a research framework is set up to visualize the process and make it more clear. Figure 1.2 shows the research framework which is a schematic representation of the research and shows the steps to reach the research objective (Verschuren, Doorewaard, and Mellion, 2010). In the first step (a), literature review is performed on the comparable data sharing factors in different sectors in order to use these in the automotive sector. In step b, literature is searched on the three core concepts as sharing behavior in automotive sector, data marketplaces and MPC. Step a and b form the fundament of step c, here the choice experiment is constructed and released. Then, in the final step (d), data analysis will be performed and will be used to estimate the logit models and reveal consumer preferences for certain data sharing factors on MPC-enabled data marketplaces. Finally, in step d, the conclusion and recommendations will be drawn and these will be compared to the obtained literature.

Figure 1.2: Research Framework


Extracting concrete data sharing factors from Literature
Current sharing behavior in automotive sector
Literature on data-marketplaces
Literature on MPC
Choice Experiment
Results of Analysis
Conclusions & Recommendations
(a)
(b)
(c)
(d)

1.7. Context of this Research Project

As the automotive sector seems a potential field for first implementation of data sharing platforms, providers of these platforms or providers of certain data sharing methods (as Safe-DEED) are very interested in this innovation already. This research paper, wherein automotive C2B willingness to share in data marketplaces by MPC-technology is analysed, is part of a larger PhD Research Project which in terms is embedded in the larger European Safe-DEED (Safe Data-Enabled Economic Development) project.

The Safe-DEED ambition within the field of data sharing is to accelerate the data sharing process, together with knowledge institutions as car manufacturers, automotive suppliers or telematics service providers. The Safe-DEED project brings together experts from the fields of cryptography, data analytics, entrepreneurial ventures, and law to work on strengthening security systems, increasing trust, and spreading privacy-enhancing solutions. Their goal is to do this via MPC.

1.8. Outline of this thesis

To begin with chapter 2, this part is all about determining the factors related to the consumers' willingness to share sensitive data in various sectors. This way, comparisons can be made with other sectors and factors can be obtained which can eventually be of value in data sharing in the automotive sector. Here, the focus lies in retrieving factors, selecting measurable factors and defining these measurable data sharing factors. It is about fining literature which underline the importance of these factors and aggregating these findings to current research. This subsection is the essence of this research and needed to construct the choice experiment and the answer research question 1.

In chapter 3, the first part is about the core concepts as the background and domain. Here data sharing in the automotive sector, data marketplaces and the MPC technology are described. These three parts form the theoretical background and domain wherein consumer data sharing is analysed. As data marketplaces and MPC are relatively new concepts, this is argued necessary.

In chapter 4 and 5, the methodology and the experimental design are proposed. Choices for multiple models and techniques will be well argued. The experimental design will be shown which respondents will fill in later on. In this chapter, the core of the experiment will be constructed in order to perform the analysis part in the next chapter.

In chapter 6, the results of the survey and are presented including the analyses. Different models will be estimated and their model fits will be drawn. furthermore, estimations on the factors and significance levels will be stated. Research questions 2 and 3 are answered in this chapter.

In chapter 7, the results will be attached to multiple conclusions plus limitations and these will be discussed. Recommendations on future research will also be given. In this chapter, an answer will be given on the fourth research question.

1.9. Suitability with MSc Program

Because of three factors, this thesis project is exceptional for the MSc Engineering & Policy Analysis degree.

Firstly, this research has a clear technical component embedded: investigating consumers' preferences regarding the data sharing factors on MPC-enabled data marketplaces for identifying the willingness to share data through discrete choice modeling (DCM). These results can be further used to focus on the most important aspects of data marketplaces by data marketplace owners or developers of MPC.

Secondly, the main flaw that is causing the situation is Europe's innovation shortfall of low growth, insufficient innovation and environmental and societal grand challenges (Commission et al., 2012). It is vital to develop breakthrough technologies and transform them into innovations (new products, processes, and services) that are adopted by the rest of the economy in order to enhance future production efficiency. Therefore, this research has a strong societal aspect to it: contributing to the 'Decent Work and Economic Growth' and the "Industry, Innovation and Infrastructure" Sustainable Development Goals of the United Nations. This way, it is possible for creation of decent jobs and improved living standards. Furthermore, enhanced data trading allows for more efficient resource utilization and stimulates research and development, both of which can help to unleash vibrant and challenging economic forces that produce revenue and jobs.

Lastly, this research covers multi-actor involvement. Consumers provide data which is the fundament for all actors in 1.1 to start data sharing. Data marketplace owners have to be involved in supplying the platform, and software developers are needed to provide the technology and knowledge to help consumers and private automotive firms in providing the correct environment for these innovations to comply to the governmental laws and ethical values.

This demonstrates that this is a study of a socio-technical process that requires both technical and institutional expertise, making it appropriate for a thesis project in the MSc Engineering Policy Analysis.

Part II

Literature Research

2

Factors relevant to willingness to share on MPC-enabled data marketplaces

In this chapter, first the literature review strategy is described in 2.1. Then, the literature review is performed and focused on people's incentives and subsequent factors which affect their willingness to share data in other sectors in 2.2. These factors will be mapped and chosen based on generalizability, measurablility, and perceived importance. Thereafter, in in 2.3, the factors which are deemed relevant in this stated-choice research are addressed and more in-depth literature is sought on these factors. The same holds for the demographic variables which will are potentially affecting the willingness of people to share data on MPC-enabled marketplaces. Thereafter, a conclusion is drawn on the factors in the experiment in 2.4.

2.1. Literature review strategy

This study can be defined as a method of collecting and summarizing prior findings in a methodical manner (Baumeister and Leary, 1997; Tranfield, Denyer, and Smart, 2003). Reviewing literature by a decent methodology is very important and critical to successive findings in any educational way of research (Webster and Watson, 2002). By integration of findings, literature reviews can address powerful research questions like no other study does (Snyder, 2019). Structuring a literature review is deemed necessary to discover clear knowledge gaps and to determine the need for further investigation (Levy and Ellis, 2006). However, the traditional ways of describing the current literature often lack systematical approaches or transparency on all research done in the field (Tranfield, Denyer, and Smart, 2003). It frequently indicates a lack of understanding or the correct information needed to make decisions about the integration of various studies. Therefore, in this study, it is important to look at both sides of the evidence instead of selective cherry-picking, otherwise problems within the results and conclusions sections can arise (Snyder, 2019).

In this literature review, the proposed approach by Levy and Ellis (2006) is taken as guiding strategy to create structure. Three steps are distinguished. In this systematic way, this research will be conducted. Step 1 is conducted in order to explicate the reasons of individuals to share personal data and connect the most relevant data sharing-factors to this concept (current Chapter 2). Step 2 will be used to explicate everything about data marketplaces and sharing data in the automotive sector (Chapter 3). Then, step 3, shows all about the MPC-technology and why the included data sharing-factors are influenced by MPC (Chapter 3).

Furthermore, as key terms to search for will probably yield many articles on the web, the right approach is required to check which sources are valuable to include in a evaluation (Snyder, 2019). First, criteria will be set up (i.e. Year of publication, Language of the article, Method). In this literature review, scientific papers, conference papers and journals beyond the year 2010 are mainly included in order to build further on the most current information. The repositories which will be used to retrieve the relevant studies are Google Scholar and the TU Delft library database. Only Dutch and English written papers are considered in order to speed up the process by avoiding translation issues. Sources are mainly retrieved by hits on keywords and forward- and backwards snowballing to stay in the line of research.

In table 2.1, an overview of all used keywords is stated where the review is based on. Some of the sources where already obtained by reviews for the research proposal, these might deviate from the systemic approach, but this seems not a big problem. The systematic literature review was aimed to provide theoretical foundation for the main study, the stated choice experiment. This literature review was divided in three different parts as data sharing, automotive data on data marketplaces and MPC. To compose a broad literature of instant sources, plenty of search terms were used.

For the topic ”data sharing”, searched was at how individuals experience this and what factors they find important when sharing data. After obtaining a first glimpse of the factors and different fields, a choice was made to narrow the scope to a small group of factors to do a more detailed literature study on these afterwards. These factors were narrowed down on the criteria: generalisability, measurability, and perceived importance. Thereafter, more literature was searched on these factors by specifically stating these factors in the search terms (”trust”, ”risk and benefits”, ”data control”). Thereafter, for finding interaction effects, basic demographics as ”Age”, ”Gender” and ”Education” were added to explore if these demographic factors affected people’s willingness to share in other fields (as medical-, academical and e-commerce sectors) or preferably in data sharing in the automotive sector. To gain knowledge about how data marketplaces work, different synonyms were used for data marketplaces: ”e-marketplace”, ”data platform” and ”data market” delivered relevant sources. Moreover, these terms were linked to the terms ”automotive”, ”car” or ”vehicle” in order to find information on specific automotive data marketplaces. The same held for the literature search on MPC, Multiparty Computation has many different forms: ”Secure Multiparty Computation”, ”Multi-party Computation”, ”MPC”, ”SMPC” etc. For every category, each iteration, the complete first page of hits was scanned for relevant sources. Relevant sources were researched till these yielded no relevant additional information anymore. The following table shows the search terms which were used during the literature research process.

Thus, by splitting the material into three parts (Section 2.2-2.4, Section 3.1-3.2.4, and Section 3.3-3.3.3) the major purpose of this study is to make a clear perspective of the underlying base of information. Data sharing-factors, automotive data on data marketplaces and the MPC technology section. In all sections, a conclusion section will be included to show why the read information is important to start each next chapter. To summarize, all deemed relevant data sharing-factors will be addressed, automotive data marketplaces will be explained, and the MPC technology will be introduced and elaborated more closely. In the end, all these obtained information will be used as fundament for the next phase: the experimental design.

Table 2.1: Literature Review Search Terms

Topic	Search Terms
data sharing	data sharing AND (individuals OR consumers) ”data sharing” AND online data sharing AND ”risks and benefits” ”data sharing” AND trust ”data sharing” AND ”herding effect” ”data sharing” AND ”data control” ”data sharing” AND e-commerce ”data sharing” AND ”social media” ”data sharing” AND ”academic sector” OR ”research” ”Privacy calculus” AND ”benefit structure” AND ”gender differences” ”Age” AND ”willingness to share data” ”Educational level” AND ”willingness to share” ”Educational level” AND ”willingness to share” AND dataplatform AND (vehicle OR car) ”familiarity with technology” AND share AND data
data marketplaces	”Internet of Things” AND (”willingness to share” OR adoption OR acceptance) AND (”data platform” OR e-marketplace OR data marketplace) AND privacy ”willingness to share” AND ”data marketplace” AND (vehicle OR car) e-marketplace AND adoption OR acceptance (”privacy preserving technologies” OR ”privacy enhancing technologies”) AND (”data marketplaces” OR ”data markets” OR ”data-marketplaces”)
MPC	”Multiparty Computation” OR ”Multi-party Computation” OR ”Secure Multiparty Computation” (SMPC OR MPC) AND data AND ”automotive” MPC AND (applications OR ”Use-cases”) MPC AND ”anonymization technologies” MPC AND privacy AND data ”multiparty computation” AND ”secret sharing” AND ”datamarketplace”

2.2. General data sharing factors by individuals over other sectors

Why do people share data? One might assume that people have certain reasons to share personal or sensitive data. The way one investigates these reasons for sharing data has a lot to do with the way of interpreting peoples’ concerns which are related to the willingness to share data and the things they gain by sharing data. The willingness to share data is influenced by various types of factors which we will try to explore in the this section.

In this research, to keeps things clear, we will follow an utilitarian view on privacy which implies that privacy can be given up again other forms of utility (i.e. money). This is true as long as the benefits exceed the sacrifices of losing privacy. This is important to mention, because our approach to the willingness to share is based on utilitarian models. Privacy concerns can be defined as a deviation of the required privacy interest and privacy interest which are satisfactory. So to summarize, in this research, we assume that people strife to obtain happiness and this is reached by an obtaining things with the highest utility. So, it is assumed that sharing data brings people thus something that increases their utility and this increases happiness which is the main goal in life. Based on this assumption, research for reasons and factors for data sharing will be conducted in order to use these in the stated choice experiment to deviate utilities and find patterns.

There are multiple different fields where individuals are already willing to share data. The medical sector, the academical sector or think of social media platforms or e-commerce. It is interesting to dig deeper in the motives for data sharing in these sectors and try to connect these to data sharing in the automotive sector as in this sector the research is shallow. Are there factors that could be similar in the automotive sector? Let us split up each field and search for individual’s factors which affect data sharing in these fields and try to connect them with data sharing in the automotive sector.

2.2.1. Medical Sector

In a systematic literature review by Nanibaa'A et al. (2016) of individual's perspectives on data sharing in the US, 4659 adult participants were asked to share their information with other academic institutions. 92% of the participants were willing to share data with academical institutions. The majority was positive in donating the data because this could lead to **advancements in future research**. Around 40% recognized the **benefits** of sharing but wanted to share as long as the potential **risks** were disclosed or the data to be restricted (Haga and O'Daniel, 2011). Furthermore, the study was not able to perform analyses on interaction effects of demographic factors due to most of the respondents did not consent for those analyses. A quite similar study in the US, performed by Sanderson et al. (2017), showed that there was little to no evidence that suggested that differences in socio-demographic factors lead to interaction effects in people's willingness to share sensitive medical information. Another study, of the Vanderbilt University showed that 18.5% of the 4050 staff respondents were more likely to share data with the university's biobank than with the national database (Brothers, Morrison, and Clayton, 2011). Other studies showed that most people are concerned about the fact that the government could access the database, 39% would not consider would not provide their biomedical data if governmental parties could access these (Beskow and Dean, 2008). Most of these respondents mentioned to **mistrust** the government and police having access to these types of data (Hiratsuka, Brown, and Dillard, 2012). They pledged for more **transparency** from the researchers-side about how the data is handled and stored.

Other study by Bell, Ohno-Machado, and Grando (2014), asked 70 healthy respondents if they were willing to share their health records for **research purposes**. This research showed that the volunteers were more comfortable when they were given the options to choose about which portions of their healthcare data would be shared and with which institutions or persons. 83% showed had a strong preference on having **control** over their own specific data and 68% of the respondents mentioned that they were having serious privacy concerns about the situation that their data would be used for **unknown profit purposes** (Bell, Ohno-Machado, and Grando, 2014). However, 80% of the respondents were still in favor of data sharing to improve quality in the care for everyone while there is general distrust in the project (Dixon et al., 2014). The indications for control on their own medical data affected their behavior in a positive way towards the willingness to share. Again transparency and data control were important factors for the decision to share medical data for research. Other factors were the type of health-information (Whiddett, Hunter, Engelbrecht, and Handy, 2006) or **own health condition** (Willison et al., 2009), the **level of anonymity** (Mamo, Browe, Logan, and Kim, 2013), and the **type of funding of the research** and which type of actors they were sharing their data to (Willison et al., 2009).

Although sharing participating member clinical trial data offers potential benefits, several pharmaceutical sponsors and researchers have advised caution due to worries about possible danger to research participants. The opinions of clinical trial participants about the hazards of data sharing are mostly unknown (Mello, Lieou, and Goodman, 2018). Mello, Lieou, and Goodman (2018) conducted a survey on the risks of sharing medical information where 771 participants from three medical centres were attracted. This was done by mail and by waiting room hardcopied paper-surveys. Their study showed that merely 8% of the respondents were feeling that potential negative consequences outweighed the benefits of sharing data. This induces that there people do not feel that there are many risks involved. The greatest concerns were that the sharing of data would make other less willing to join clinical trials (37%), 34% thought that these data could be used for marketing purposes, 30% saw the danger that the data could be stolen. Far less concerns were in the range of discrimination (22%) or the exploitation for profitable purposes (20%). The potential to undermine patient identity and privacy is perhaps the most serious danger in clinical data exchange, and one that has been underplayed to some extent. When patients engage in clinical studies, researchers make a promise to protect their privacy (Rosenblatt, Jain, and Cahill, 2015). This could be one of the reasons that few people are having strong concerns about the risks of data sharing in the medical field.

A European based study on motivations for data sharing for research in diabetes by Shah et al. (2019) showed that attitudes towards data sharing diverges between countries on protecting privacy, beliefs about risks and corresponding benefits and advancing research. Around 50% showed that data control is important, and control over with whom the data is shared is more important than which types of data are shared. Danish people found control on data types more important and Dutch respondents more on with whom the data was shared. The findings state than even with anonymized data, people prefer privacy above all other things. Resulting conclusion was that a movement from consent forms about data use and re-use could be better adapted to participant-specific choices about with whom to share with and which types of data (Riordan, Papoutsi, Reed, Marston, Bell, and Majeed, 2015; Shah et al., 2019).

2.2.2. E-commerce sector

Within the e-commerce sector, tons of data are also shared every day. The Internet has grown in importance as a marketing tool as a result of fast technological advancements, causing privacy issues (L. Chen and Liu, 2015). Individuals and companies benefit from the acquisition, use, and exchange of private information online (Sánchez and Viejo, 2017). It is shown that the reasons for sharing data in the e-commerce sector are dependent on the attitude the customer has towards the way the data is or will be handled (Anic, Škare, and Milaković, 2019). This also allows for the limits of people's autonomy and the misuse of their private details (identity fraud, spying, scamming, prejudice, and manipulation), all of which have a detrimental impact on internet companies as well as the implementation of new information communication technologies (ICT) (Acquisti, Brandimarte, and Loewenstein, 2015; Choi and Butler, 2019). Recent events have sparked heated debate over what should be done to safeguard people's privacy and how much **data control** customers should have over internet purchases (Sánchez and Viejo, 2017).

Not only data control is important to the consumer, their beliefs about the possible risks and danger of **data disclosure** also play a role according to Anic, Škare, and Milaković (2019). It is the company's reputation which displays the trust of consumers in the way they handle their customers' personal data. Therefore, the perceived risk people feel by sharing e-commerce data to specific companies, plays a particular role in the decision whether to share their data. D. J. Kim, Ferrin, and Rao (2009) also highlighted the supreme role of trust and risk in the context of decisions in online shopping. The perceived risk relates to **trust** in a sense, whether these companies handle the sensitive data properly is based on trust as consumers cannot check in detail where or how these data is stored (Y. Chang, Wong, Libaque-Saenz, and Lee, 2018). Also, on the internet, sensible governmental management enhances perspectives and reduces both perceived danger and increases peoples' willingness to share data (Choi and Butler, 2019; Dienlin and Metzger, 2016.

2.2.3. Research data sector

The research data sector is about researchers and academics sharing their research papers and making these publicly available on the internet. Data is an intellectual asset that helps to scientific development, according to the findings. Data generated by academics or research groups can be re-analyzed and processed by other scientists using various methodologies or techniques to uncover new information or discoveries in this setting. By literature search, several factors were found which influence data sharing behaviour with regard to research data. Chawinga and Zinn (2019) comes up with the following factors; **absence of compensation**, **control over the data**, **fears of data misuse** and **seniority** and **age**.

Following Chawinga and Zinn (2019), missing of benefits is the most important factor for researchers to not share research data. It's not odd that most studies mentioned compensation as a barrier to data sharing since it appears that data sharing provides researchers with few or no intrinsic advantages; they can only be acknowledged or credited by re-users. This way, besides being responsible in advancements in science and minimizing total research costs, the individual researcher has no additional reason to share his or her results with other researchers.

Then, as also in the e-commerce, the amount of control which persons or institutions have over their own research papers, are found important in researchers' data sharing behaviour. Scientists seeking control over their data do not imply they are afraid to disclose it; they simply want to know how the information is stored, how it's used, and for what reasons it is being used, as well as be credited or credited by others who are utilizing it. Likewise, in Germany, 80 percent of academics said that maintaining control over shared data was critical (Fecher, Friesike, and Hebing, 2015). As there are certain ambiguities around sharing data in popular data sharing platforms, academics may want control over their data. Intellectual property rights are one of the unsolved issues (Milia et al., 2012). Because giving academics more control over their data motivates them to disclose more, the study indicates that academics should have sufficient control over the data they disclose when formulating institutional and national research policy.

Another factor which tends to influences the willingness of researchers to share data is the fear of data misuse. Misuse incidents decreased the willingness of academics to share research data (Cragin, Palmer, Carlson, and Witt, 2010). Data can be used for intentional misuse as ”falsification, commercial misuse, competitive misuse or flowed interpretation” (Fecher, Friesike, and Hebing, 2015, p. 16).

In the end, it is shown that older and more senior researchers were more willing to share research data with other researchers as they value data sharing more than young academics (Milia et al., 2012). Young researchers are often less skilled and their research data is of low quality and therefore they afraid of posting their research online (Chawinga and Zinn, 2019). Whereas, older and senior academics are more mature and see the bigger picture of advancements in research and thereby are more willing to share their research data (Tenopir et al., 2011).

2.3. Selection of general data sharing factors

The previous section lighted a glimpse on comparable data sharing sectors and the subsequent factors found in literature research. The factors which will be used in this research will be based on the provided information of the previous section. These factors are chosen based on the aforementioned criteria as generalisability, measurability and perceived importance. The chosen factors are likely to affect peoples' willingness to share automotive data on MPC-enabled data marketplaces. In this section, a selection is made to include the most important factors which will be included in the choice experiment.

After the first literature scanning on factors in different sectors than the automotive sector, the following factors were found by comparing and looking to the aforementioned different sectors. The following table 2.2 shows which factors were commonly mentioned and whether these are chosen to be in line with this study's purpose.

Not all factors are included for further investigation, a selection is made. To speak of the factors *Risk of Disclosure* and *Data Control*, these are seen as key-factors which influence the willingness to share data when compared to Wang, Duong, and Chen (2016). It is always important to be aware of the risks within data sharing and it is important to know what your capabilities are in measurements of the amount of control as mitigation measure. These two factors are expected to be important to people in the automotive field of data sharing (Treiblmaier and Chong, 2011).

Furthermore, the factor *Trust* is also included. More investigation is needed about why trust influences data sharing behaviour. This factor is assumed to be relatively complex to design.

The factor *Benefit* is expected to be important (Chawinga and Zinn, 2019). As it is known that people are sensitive to monetary or non-monetary benefits (, people could be able to throw privacy concerns overboard in exchange for benefits. This is interesting and insightful relative to the other factors. This could give insights to what extend risk, data control or trust are accepted in exchange for some sort of value.

In the end, the factors Age, Gender, Educational level and familiarity with the technology, are included and these are measured as side-effects. These are assumed to be the co-variables which could give insight in the differences between respondents. All factors and co-variables will now be further investigated in detail in each subsection, to gain more knowledge and definitely decide whether to include them in the experiment or not.

Table 2.2: Factors found by comparing sectors

Factor	**Sector(s)**	**Generalisable to automotive sector?**	**Measuarable to concrete levels?**	**Perceived relevant?**	**Chosen for this study?**
Risk of data disclosure	Medical, E-commerce, Research	Yes	Yes	Yes	Yes
Data Control	Medical, E-commerce, Research	Yes	Yes	Yes	Yes
Trust (herding effect)	Medical, E-commerce, Research	Yes	Yes	Yes	Yes
Institution to share with	Medical	No	No	No	No
Purpose of sharing	Medical	Yes	Yes	No	No
Type of data	Medical	Yes	No	No	No
Benefit	Medical, E-commerce, Research	Yes	Yes	Yes	Yes
Gender	Medical	Yes	Yes	Yes	Yes
Educational level	Medical	Yes	Yes	Yes	Yes
Advancement in future research	Medical, Research	Yes	No	No	No
Reputation sharing partner	E-commerce	No	No	No	No
Governmental laws and regulations	E-commerce	Yes	No	No	No
Data misuse	Research	Yes	No	No	No
Seniority or Age	Research	Yes	Yes	Yes	Yes

2.3.1. Risk of Data Disclosure

While technology advances, different definitions of sensitive data rise on the web. In this paper, we see sensitive data as information that people want to keep classified, unless specific parties are granted permission by the owner of these sensitive data (Pretty, 2020). Moreover, we follow OECD. (2019) and define the factor risk of data disclosure as; the probability of a data breach which can be determined as a loss due to unautherized parties having access to to personal data which is due to failure of the organisation to completely safegaurd these data via privacy preserving techniques. In this study only automotive data is targeted. Due to the high confidentiality of these data, these data are often paired with strong laws and regulations to discourage cybercriminals (Skatova et al., 2013). Sensitive data, as defined by the courts, is information that must be secured from unlawful disclosure and named personally identifiable information (PII). If these data falls in the wrong hands, this could turn out very badly for the consumer and the platform owner. Important to note, in this study we simplify the factor risk of data disclosure by leaving out the risk of data disclosure which go beyond data security. OECD. (2019) state that these risks are based on contractual agreements and consent forms between data providers and consumers in data sharing and are about sharing data to third parties. This is also important, but left out in our scope on MPC-enabled data marketplaces.

An important factor in data-security of online data sharing is the amount of risk which is involved. This risk is mostly based on the level of encryption of the data on the centralised or decentralised servers (Rao and Selvamani, 2015). This is about the risk of decryption of the sensitive automotive data which is shared in good trust. It is rather difficult to address the amount of risk which is involved in a single process as this is also dependent on the parties involved. Risk of data disclosure can be addressed by probabilities in terms of percentages. Often, risk is not concretely addressed and defined by ordinal values as ”High”, ”Medium” or ”Low” risk (Koch, Krenn, Pellegrino, and Ramacher, 2021). These values are often based on the amount of occasions in the past. In Koch, Krenn, Pellegrino, and Ramacher (2021), the risks of information disclosure are defined on the likelihood and the impact in the LINDUNN methodology which is based on Linkability, Identifiability, Non-repudiation, Disclosure of information, Unawareness and Non-compliance. Priorities are attached to these different cases which clearly outlines where to stress on in data sharing applications in order to keep this safe. Trabelsi, Salzgeber, Bezzi, and Montagnon (2009) performed a study in data disclosure risk evaluation and came up with statistical tools to evaluate data disclosure risk. This is a more technical approach and was linked to combinations of attributes within a dataset and their combined probability of information disclosure was calculated. This approach is more solid, however still based on many probability assumptions and very difficult to assess in this type of research. MPC is not yet implemented which denies the possibility to retrieve a basis of disclosure incidents to use as starting point.

A risk assessment is thus often based on estimations on the possible impact and the likelihood a malicious attack may have. This impact assessment is further highly based on the type of data which is shared (Krasnova, Spiekermann, Koroleva, and Hildebrand, 2010; Skatova et al., 2013). Some types of automotive data have higher value or contains more sensitive information than others. This is something which varies and needs clarification within the experiment. In the experiment this type of data will be fixed on automotive GPS data, as it is expected that most respondents are familiar with this type of data is every respondent makes the same assumption on the type of data which is shared. This is important as this highly affects the way people perceive the level of risk of data disclosure (Xie, Teo, and Wan, 2006).

2.3.2. Data Control

This factor is based on the control which users have to mitigate data disclosure and to increase the data-security. The opportunity to define which data will be shared and to decide how the data is shared to preserve data-security. This is the only lever consumers have to mitigate data-breaches. As full prevention of security breaches is impossible, it is therefore valuable for users to have knowledge about which steps to perform in mitigating the damage (Dhillon, 2015). This control can translate itself in measures of instructions, feedback and levers to cancel the computation-process. Data Control is very dependent and related to the platform architecture. Low data control comes with a centralised architecture, whereas high data control relates to a more decentralised architecture where the data stays at your own car during the computation process.

M. Spiekermann (2019) constructed a taxonomy for data marketplaces in order to give an overview and to categorize different types of data marketplaces. One of these factors is the factor platform architecture, which is about the platform design. A distinction can be made between centralised, decentralised and hybrid platform architectures (M. Spiekermann, 2019). For the centralised approach this means that data is provided by different suppliers on a central server which enables better control on the data. This overcomes all types of technological difficulties for data providers and consumers. On a decentralised architecture, the data is kept at the supplier’s which raises technical difficulties regarding processing and storage of data but increases the data-sovereignty. Furthermore, a decentralised architecture enables direct trades, which is possible by distributed ledger technologies (DLT) such as block-chain where trades are verified by market participants (Koutroumpis, Leiponen, and Thomas, 2017, 2020). Hybrid architectures bundle these two architectures by decentralised trading and supplementary technical support in forms of infrastructure from the centralised platform architecture. It may be interesting to include the factor *Platform Architecture* as consumers can be affected in datasharing behaviour by this factor.

S. Spiekermann (2007) worked on the first researches regarding data control in ubiquitous computing environments using privacy enhancing technologies. She assumed that when users experience a certain sense of data control, they indirectly experience themselves exercising the right of privacy. She defined 6 different catagories within control; Power, Contingency, Helplessness, Choice, Information and ease-of-use. The choice to choose between architecture types and thereby the amount of control one has, is an important aspect within this study. By letting people choose between different types, makes people feel they have certain power to steer the outcome to the desired outcome (Langer and Abelson, 1983). In this case about sharing automotive data via data marketplaces, this would imply that respondents can easily state where there data is stored during the computation process. Furthermore, a study performed to discover the factors of self-disclosure on social media by Xu, Dinev, Smith, and Hart (2008), showed that control perceptions affect peoples' formation of privacy concerns. Krasnova, Spiekermann, Koroleva, and Hildebrand (2010) stated that literature back then already provided useful insights into the factors on data sharing and information disclosure. They found that benefit, control, and beliefs relating to risk and trust are the main factors whether to share data or not. They all have a certain amount of overlap and affect each other. Trust, data control and benefit can be risk-reducing factors for platform participants (Krasnova, Spiekermann, Koroleva, and Hildebrand, 2010). Malhotra, Kim, and Agarwal (2004) concluded that among these factors, data control could be viewed as the most active component in preserving privacy.

Wang, Duong, and Chen (2016) also did research on peoples' intention to disclose personal information via mobile apps. This factor-analysis research was also based on privacy calculus perspectives. They built up on Culnan and Armstrong (1999) who found that consumers have a reduced sense of being invaded and become less concerned about disclosure risk if they feel they have certain control over their information and how it will be processed. So by offering a high level data control, this mitigates the concerns about the involved risks in their personal data being processed (Malhotra, Kim, and Agarwal, 2004). See figure 2.1 how the factors are assumed to interact in a data sharing environment.

2.3.3. Trust (Herd Effect)

Trust is often described in a way of a person or institution's confidence on the fact that the other party will behave integer and responsibly conform the rules. As there is trust between two parties, this means that parties are convinced that they will meet the rules and expectations of agreements and do not exploit each others' weaknesses (Pavlou, 2002). These weaknesses are dependent on the context in which trust is needed. In the scope of current research, these weaknesses lie in the sensitive automotive data of consumers which is used by other parties in the MPC process to compute functions (see 3.3.1).

Before engaging in a data marketplace, a crucial threshold of trust needs to be overcome (H. H. Chang and Wong, 2010). Trust becomes an issue in situations were risks are involved, people's behaviors in these situations are dependent of trust (Luo, 2002; Sirdeshmukh, Singh, and Sabol, 2002). Kramer (1999) defined trust as ”the reliance on the integrity, ability, or character of a person or thing”. Trust itself is dependent on multiple factors, in a consultancy report of Otonomo (2020), trust in datasharing is branched into different aspects as: transparency, safety, previous experiences, incentives (in terms of return) and unknown factors why people are reluctant in datasharing and have a certain level of trust. This section relates to how the MPC-technology is viewed or regarded as being more user-friendly. The people's comprehensibility with the computation technology and if this meets the requirements of the quality standards. (Chiregi and Navimipour, 2016). Furthermore, a good reputation of a company is also one of the key enablers of increasing consumers' trust in the company and to start sharing their (sensitive) data (Chiregi and Navimipour, 2016). This factor will be included as in the form of a *herding effect*, as trust is in essence a complex factor to design and by defining trust in a way of herding behavior, we can make the factor more measurable in terms of levels.

Mattke, Maier, Reis, and Weitzel (2020) did research in herding effects regarding advertisement clicks on social media platforms. The herd theory itself is a commonly described theory which offers a useful theoretical foundation on further insights regarding the way people make decisions (Banerjee, 1992; Bikhchandani, Hirshleifer, and Welch, 1992). The herd theory explains how behavior of other persons or close relatives affects one's own current behavior (Mattke, Maier, Reis, and Weitzel, 2020). The herd theory states that a user who is observing others, only looks at their behavior and not at the reasons behind that behavior. H. Sun (2013) states that the decision to decide to do the same as the predecessors is based on two aspects: discounting own initial formed information and imitating others. As mentioned before, the herd can consist of only unknown persons or can also include known relatives as a friend or experienced person (Bikhchandani, Hirshleifer, and Welch, 1992; H. Sun, 2013; Tucker and Zhang, 2011). The first aspect shows a preference by assuming that others are better informed and make a better decision by higher cognitive analyses.

This theory on herd behavior can be linked to the concept decision making to share data on MPC-enabled data marketplaces. The observation of others already sharing automotive data on MPC-enabled data marketplaces could trigger one's decision to also start sharing automotive data. For the same exact aspects as in the herd theory, people may discount their own information and imitate more experienced persons or friends or even unknown persons in the decision to share automotive data. This way, trust in the form of a herding effect will be used in the experiment.

2.3.4. Benefit

Within datasharing, privacy interest can be affected by multiple activities. Solove, Rotenberg, and Schwartz (2006) differentiated four activities which affect people's privacy: (1) Collection of information, (2) Altering of information, (3) Spreading of Information (4) Contravention. Finn, Wright, and Friedewald (2013) updated Clarke's (1999) types of privacy, as they were outdated due to recent technological advancements in scanning devices. The following seven privacy types are considered: personal privacy, behavioural privacy, communicational privacy, data privacy, privacy of senses, spatial privacy, and privacy of association. Moreover, spatial privacy and data privacy are often affected by the way mobile devices or cars are used or insurance companies act.

As disclosure of sensitive information often leads to increasing privacy concerns (Bansal, Gefen, et al., 2010), these negatively affect people's willingness to utilize online applications (Malhotra, Kim, and Agarwal, 2004; Miyazaki and Fernandez, 2001). People perform ”calculus of behavior” to map the outcomes of uploading sensitive data (Lwin and Williams, 2003). As a result, people are mapping trade-offs between potential pros and cons for releasing sensitive information. Thus, in order to make people willing to use applications which require uploads of personal information, people have to be compensated for it. Hann, Hui, Lee, and Png (2007) proposed two methods to mitigate the cons of data disclosure on the intention to use: (1) by offering benefits in the manner of monetary rewards, or (2) by offering privacy policies regarding management of usage of personal data. Jen, LU, Wang, and Chang (2013) also showed that monetary benefits (i.e. direct payouts or discounts) are positively affecting people's intention to use digital services.

There is much research done in the field of rewarding people for certain actions, also for data- or information sharing. Jian and Jeffres (2006) and Willem and Buelens (2007) state that by direct and indirect incentives, the intention of datasharing can be increased. And, as performance based reward-systems are also able to encourage people to start datasharing (S. Kim and Lee, 2006), rewarding people with benefits on data marketplaces may increase their willingness to share sensitive data. Also Derikx, De Reuver, and Kroesen (2016) showed that people were willing to share personal automotive data in exchange for 9.54 euro per month. People are basically compensated and exchanging privacy concerns for monetary value. As these exchanges are possible and currently a big part of data marketplaces, we include the option of having certain *Benefit* in exchange of data input.

2.3.5. Gender

In the following subsections, apart from the MPC, the personal demographic dimension is branched into measurable factors which might affect people's behavior in datasharing. These are the demographic-factors. These will be explained factor by factor as across people, many significant variations exist in terms of intention to exchange data. It is interesting to map these in order to have a clear understanding of the reasons why adoption levels can deviate between people.

Sex differences in the use of mobile products and social networking sites exist (Kennedy, Wellman, and Klement, 2003; D.-Y. Kim, Lehto, and Morrison, 2007). The foundation of choosing the demographic attribute *gender* lies within the social role theory of Eagly (1987), which tries to explain gender differences in behaviour. Multiple studies show that within datasharing, differences arise between genders. This is mostly because of the difference between men and women in the processing of information which happens different patterns (Bem, 1981; Venkatesh and Morris, 2000). In Ziefle, Halbey, and Kowalewski (2016), significant differences in the shared type of data were found between males and females (a factor findable in the next subsection). Where females were more reluctant to share locational data, males were more reluctant to share lifestyle habits. Furthermore, Y. Sun, Wang, Shen, and Zhang (2015) and Hui, Teo, and Lee (2007) showed that males are more sensitive and willing to interchange privacy for a certain form of benefit than females.

2.3.6. Age

As for gender, also the demographic variable *Age* could be a predictive variable in people's willingness to share sensitive data on data marketplaces. For instance, Leon et al. (2013) showed in a study regarding users' willingness to share to online advertisers, that senior citizens were less eager to disclose demographic data but more inclined to give locational data than younger citizens. However, a study in researching which factors affect the willingness to share in electronic healthcare data showed no effect in age but in other factors as race, educational level, benefit and control (K. Kim, Sankar, Wilson, and Haynes, 2017). Contradictory, research showed that younger scientists were more willing to share their findings, but were disclosing less information to the world relatively to more senior scientists (Tenopir et al., 2015). Nevertheless, as there is rarely any evidence on correlation of the factor *Age* on the willingness to share in data marketplaces of automotive data - *Age* will be included.

2.3.7. Educational Level

As for educational level, it might have a certain correlation between the willingness to share data on data marketplaces. Higher educated people often have a better understanding of threats within digital services (Kowalewski, Ziefle, Ziegeldorf, and Wehrle, 2015). Unless there is little knowledge available about the relation between educational levels and to which extent this leads to more datasharing on data marketplaces, Haeusermann, Greshake, Blasimme, Irdam, Richards, and Vayena (2017) showed that higher educated people are more eager to exchange genetic information for medical research. Because this is yet unclear in the field of data marketplaces, *Educational level* will be included.

2.3.8. Familiarity with technology

As the digital world and thereby digital threats are evolving, people increasingly get more privacy concerns. Karpati (2011) described digital literacy as the set of basic abilities which are required to protect and retrieve digital information. There are levels within digital literacy, which define the control and understanding of people in certain digital environments, as also on MPC-enabled data marketplaces. Therefore, we follow Harborth and Pape (2020) which state that digital literacy affects the way people understand and value privacy preserving technologies as for instance MPC. As MPC-enabled data marketplaces are a relatively new concept and expected is that few people know about the MPC technology or data marketplaces, digital familiarity with privacy preserving technologies (PPT) or data marketplaces is used. This way an indication can be retrieved and correlations can be analysed.

Familiarity with technology could be an important factor in one's eagerness to exchange sensitive information on web applications. The concept of *familiarity with the technology* relates to people's individual familiarity in computer-related functions and technologies (Bunz, 2003; Jenkins, 2006). As Hargittai (2007) described that certain levels of familiarity can either promote or discourage people to get involved in parts of the internet which involve the contribution of personal data and the control over it. In this sense, *familiarity with the technology* could serve as a form of motivation to empower users to undertake control of their sensitive data online. Park (2013) divided this factor in Knowledge, Information control behavior, Internet experience and sociodemographic characteristics (as age etc.) and came up with very confirming results that more digital familiar people were more in control and willing to share data online.

2.3.9. Westin's (1968) privacy index

Westin did research in the way consumers valued businesses and laws & regulations regarding data privacy. He provided statements that respondents may use to express their degree of agreement, ranging between strong disagreement (1) and strong agreement (5). Based on these combinations of values, people were categorized in different groups: Fundamentalists, Pragmatists and Unconcerned. We're interested if these different categories of people make different choices in the stated choice experiment regarding alternatives of MPC-enabled data marketplaces where we thus vary the factors *Risk of data disclosure*, *Data control*, *Trust (Herding effect)* and *Benefit*. These statements will be included and asked in the survey, after the choice-experiment. These will be further explained in chapter 5.

This first research-background part of the thesis discussed the literature on the workings and factors of MPC and the demographic factors which may affect the willingness to share automotive data on MPC-enabled data marketplaces. To have a more clear understanding of these links between factors, figure 2.1 depicts a theoretical model, including direction on causalities between entities. In the next chapter, these polarities can be useful in designing priors within an efficient experimental design by Ngene software to have an indication and to decrease the number of required shown options to reach significant estimates. Thus, this will ensure that less respondents are required in the experiment while obtaining same valid and reliable results.

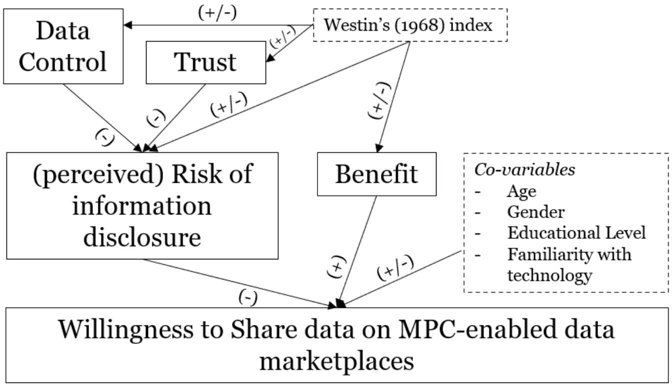
2.4. Conclusion

As all factors which will be included in the choice experiment are clear now, the first question of current research is answered:

1: What factors drive consumers' choices regarding sharing automotive data on MPC enabled data marketplaces?

The final factors which will be considered in the experiment are thus the following: Risk of Disclosure, Data Control, Trust (in terms of a herd effect), Benefit, Age, Gender, Familiarity with technology, Educational level and Westin's (1968) privacy index. These main factors are list in table 2.3 and will be further branched into various levels in the next sections, in order to construct the choice experiment and let respondents make trade-offs. In figure 2.1, a simple conceptual model is constructed where all factors are decomposed with their expected polarity on the willingness of people to share automotive data on MPC-enabled data marketplaces. This + or - sign indicates the expectation of which the factor contributes to the utility of an alternative. Higher Data control, Trust and Benefit are predicted to influence people's willingness to share in a positive way whereas Risk will probably lead to a negative influence on people's willingness to share data. For the co-variables, it is yet hard to have an indication on the influence on people's willingness to share on MPC-enabled data marketplaces.

Figure 2.1: Conceptual model of datasharing factors and co-variables



(+/-)
Westin's (1968) index
Data Control
Trust
(+/-)
(-)
(-)
(+|-)
(+/-)
(perceived) Risk of information disclosure
Benefit
Co-variables
- Age
- Gender
- Educational Level
- Familiarity with technology
(+)
(+|-)
(-)
Willingness to Share data on MPC-enabled data marketplaces

Table 2.3: Factors and co-variables influencing willingness to share data

Factor	**Definition**	**Reference**
Risk of data disclosure	refers to “the probability of a loss, unauthorised access to or disclosure of personal data as a result of a failure of the organisation to effectively safeguard the data”.	Definition of(OECD., 2019 cited.)
Data Control	refers to the degree in which parties have power and authority over their automotive data in terms of data storage and processing on MPC-enabled data marketplaces.	Definition of (Dhillon, 2015) and (M. Spiekermann, 2019) modified to current scope.
Trust (Herd Effect)	refers to the way the herd effect affects consumers’ decision making process of sharing automotive data on MPC-enabled data marketplaces.	Definition of (Bikhchandani, Hirshleifer, and Welch, 1992; Chiregi and Navimipour, 2016)
Benefit	refers to the degree in which benefits for trading automotive sensitive data affects people’s willingness to start data automotive sharing on MPC-enabled data marketplaces.	Definition derived from results of (Derikx, De Reuver, and Kroesen, 2016) and (Agahari, 2020).
Co-variable	**Definition**	**Reference**
Gender	refers to the degree in which males and females, or different genders categorized by ”other”, value the main factors which are assumed to be affecting the willingness to share on MPC-enabled data marketplaces, differently.	Definition of (Bem, 1981; Kennedy, Wellman, and Klement, 2003; D.-Y. Kim, Lehto, and Morrison, 2007; Venkatesh and Morris, 2000).
Age	refers to the degree in which different age groups value the main factors which are assumed to be affecting the willingness to share on MPC-enabled data marketplaces, differently.	Definition of (Leon et al., 2013) modified to data marketplaces.
Educational level	refers to the degree in which different groups of educational level value the main factors which are assumed to be affecting the willingness to share on MPC-enabled data marketplaces, differently.	Definition of (Haeusermann, Greshake, Blasimme, Irdam, Richards, and Vayena, 2017; Kowalewski, Ziefle, Ziegeldorf, and Wehrle, 2015) modified to data marketplaces.
Familiarity with technology	refers to the degree in which different groups of familiarity with the technology value the main factors which are assumed to be affecting the willingness to share on MPC-enabled data marketplaces, differently.	Definition of (Bunz, 2003; Jenkins, 2006) modified to data marketplaces.
Westin’s (1968) privacy index	refers to the degree in which different groups of privacy concerned groups value the main factors which are assumed to be affecting the willingness to share on MPC-enabled data marketplaces, differently.	Definition of (A. F. Westin, 1968)

3

The Core Concepts (Background & Domain)

This chapter is all about datasharing of automotive data in the automotive sector, the concept of data marketplaces, and the privacy preserving technology MPC. First, in section 3.1, a concise but clear introduction will be given on datasharing in the automotive sector to gain specific knowledge about the domain where the MPC-enabled data marketplaces could be of added value. Then in section 3.2, data marketplaces are introduced wherein each subsection will give a detailed explication on the architecture of data marketplaces and why these could be valuable in the automotive sector to use for sharing automotive data. The last subsection (3.2.4) of the data marketplace section, introduces privacy preserving technologies (PPT) in order to understand the next section (3.3) better, which is all about the Multiparty Computation technology. This privacy preserving technology (PPT) will be explained in detail in order to have an understanding of how this PPT can be of any use on data marketplaces where automotive data is traded.

3.1. Data sharing in the automotive sector

Automobile manufacturers are beginners in the field of the data sharing. According to KPMG's (2020) analysis on vehicle data sharing, the digital services and shared mobility markets were practically non-existent. However, as we speak, this is becoming a competitive industry in which European car manufacturers are competing not only with one another, but also with an increasing number of non-European car manufacturers. Data has become increasingly crucial in the worldwide automotive business. Despite the fact that manufacturers have been collecting data from their connected vehicles for many years, they have only just begun to look at the possibility of sharing these data(Mosterd, Sobota, van de Kaa, Ding, and de Reuver, 2021). Because the economic potential of their data is unpredictable, manufacturers have generally been hesitant to do so last years (KPMG, 2020). In 2019 and 2020, however, the manufacturers gradually increased the sharing of automotive data with external parties, signing agreements with both data aggregators and data marketplaces. Manufacturers are gaining additional revenue sources as a result of these agreements.

The global automotive environment is complicated, and alliances change frequently. Here Technologies, which is owned by Audi, BMW, Daimler, Mitsubishi, Bosch, and Continental, is a self-deployed datamarketplace to share and monetize automotive data (Hansen, 2020). But there are many more marketplaces in which automotive data is shared. Otonomo, Caruso, VINChain, and AMO labs are all very popular automotive data platforms where data is traded and supplied by manufacturers[1]. Huge volumes of data are being generated as the number of connected cars on European roads continues to grow. This data is important to a variety of parties, and OEMs (original equipment manufacturers) are increasingly looking into methods to monetize it while maintaining control over the data and dealing with consumers' data privacy concerns (McKinsey, 2016). On an European level, there have been debates

[1] https://datarade.ai/platform-categories/automotive-data-platforms

on market trends, but due to a conflict of stakeholder interests, no resolution has yet been identified that is fair to all parties concerned.

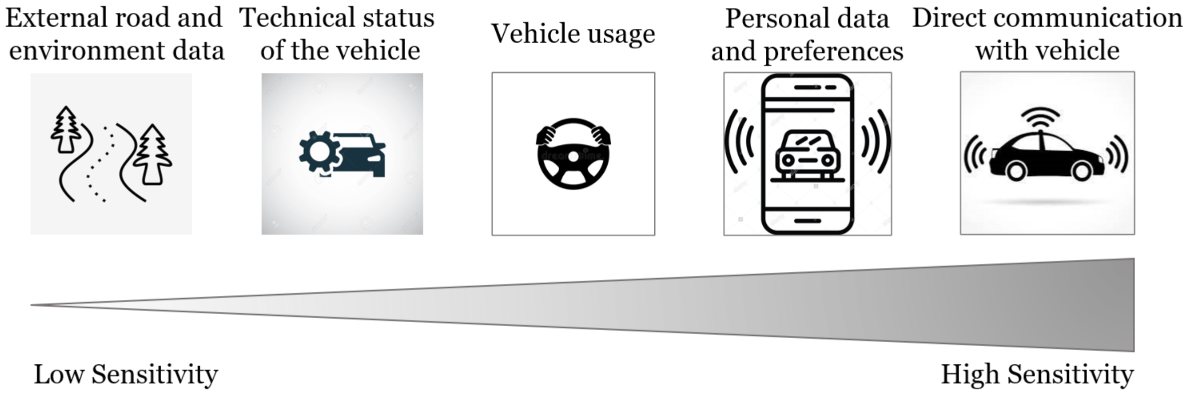
As the automotive data sharing ecosystem is emerging in tremendous speed, a variety of actors is already active in the sharing environment. To give a certain idea, the following actors are making their mark currently:

Automotive data analytics is already being used by *car repair companies* to effectively gain knowledge about the usage of cars, to better structure repairing choices plus to enhance dealer-customer interaction (Delgado, 2021). Furthermore, *car manufacturers* are working on the mechanics that will make up the environment for detecting, processing, and leveraging automotive data. Despite the large number of apps available, car manufacturers are now faced with the task of figuring out how to use data to 1) directly reach end customers; 2) strengthen their own product and service portfolio; and 3) help improve their target audience (Delgado, 2021; Otonomo, 2020). Then there are the *insurers*, the insurers are using automotive data by offering utilisation contracts, investigating occurrence policies (e.g., short-term, area vehicle insurance), and are increasing their awareness of consumers' behavior beyond the yearly main event pinch point (Keller, Eling, Schmeiser, Christen, and Loi, 2018). *Road-side assistance providers* are gathering and analysing real-time radio signals from vehicle sensors and automated warnings, they are also optimizing rescue vehicle routing, and are evaluating incident data to deliver helpful information to car manufacturers and infrastructural operators. *Startups* are designing new apps, manufacturing innovative tools (e.g., retro-fittable motion controls), and are delivering services via novel monetization strategies as Spotify via carplay. *Government institutions* are defining the parameters for the gathering and distribution of vehicle data. They also have the power to impose public-beneficial car data-enabled services, such as emergency call features, and to control contentious issues like connected car technological qualification, data property rights, and intellectual property rights over shared technologies and services (McKinsey, 2016). Regulators are using road traffic data to alleviate congestion and car incidents by collaborating with infrastructure operators using big data.

3.1.1. Data viewed in Automotive Sector

Cars generate a variety of data output kinds. As the disclosure of private data results in general in increasing privacy concerns (Bansal, Gefen, et al., 2010), each type of data brings its specific risks when it is shared to others. These categories imply thus different perceived customer sensitivities which are shown in figure 3.1:

Figure 3.1: Car data categories based on data sensitivity


External road and environment data
Technical status of the vehicle
Vehicle usage
Personal data and preferences
Direct communication with vehicle
Low Sensitivity
High Sensitivity

These classifications denote varying levels of customer sensitivity when it comes to data sharing. According to McKinsey (2016), consumers are most willing to submit data in the low-sensitivity data types as external exogenous conditions, car technical states, and driver activities. The technical status of the vehicles which include the motor temperatures, technical malfunctions and airbag status is also on the low sensitivity end. The vehicle usage data, which is about speed, GPS, load weight is perceived as medium sensitivity. Then, personal data as radio channel, driver's identity are on the more high sensitivity end. The most sensitive data are the data which are directly communicated from the vehicle to the telephone or vice versa. Think of calendar, Spotify, telephone or e-mail data (McKinsey, 2016). These data types are thought to be more ”objective” and therefore less critical (McKinsey, 2016). Nonetheless, as with vehicle location and usage, effective management of this data is necessary. Consumers, on the other hand, consider data categories to be riskier when it is about sharing personal data and interests, as well as the substance of personal conversations sent from within the car. Customers are more hesitant to disclose these types of information since it is considered personal (KPMG, 2020). Consumers expect a decent amount of benefit in return for sharing data in the ”nonpersonal” category where they are most inclined to do so. Once it concerns to what people could get in exchange for the highly sensitive data they give, their demands are even higher, and being aware of data management faults is critical for the image of the sector players. Consumers' tolerances for sharing automotive data are constantly shifting (Otonomo, 2020). Although a large portion of customers is initially opposed to data sharing, over half of those who voice reservations say that their apprehensions will fade if specific assurances are provided according to the research reports of McKinsey and KPMG.

In the end, automotive data is being used to create value in one of three areas (or a mix of them) by players in the expanding automotive market. First, players make money by selling products/services to clients, customizing advertising, and selling data to third parties. Second, they're leveraging automotive data to cut expenses by improving RD efficiency or reducing necessity repairs, for instance. Third, businesses are improving security and safety by leveraging the potential of automotive data to speed up safety interventions that protect drivers from actual injury or the theft of their personal property or documents.

3.2. Introduction to data marketplaces

To elaborate more on how automotive actors can share data in the automotive environment, an introduction on data marketplaces is useful. In this subsection, data marketplaces are elaborated by different aspects and taxonomies within literature.

In this research we follow Abbas, Agahari, van de Ven, Zuiderwijk, and de Reuver (2021) and define a data marketplace as a multisided platform where data providers can sell data to data consumers in exchange for financial transactions. People have the possibility to engage in various sorts of data marketplaces nowadays, ranging from one-to-one bargaining, to many-to-many market-platforms (Koutroumpis, Leiponen, and Thomas, 2020). As Roth (2009) explains, these markets must meet certain standards of perceived safety-levels, low transaction costs and mass of demand and supply in order to be attractive to consumers. These four aspects are the fundament of data marketplaces. M. Spiekermann (2019) already came up with a clear set of multiple dimensions and factors of data marketplaces. Let us follow his taxonomy to decompose data marketplaces.

Each data marketplace has a certain *value proposition*, this value proposition is the core idea of the data marketplace and provides thus value for the users of the platform. In M. Spiekermann (2019), two types of value propositions are differentiated. Transaction-centred and data-centred data marketplaces. The difference herein is that the former also provides tools for data-analysis and visualisation to gain direct insights in the data, besides providing the required infrastructure to bring data buyers and sellers together.

The *market positioning* is about the independence of the platform. This is about whether the platform owner is also a data seller or if it is merely the neutral marketplace operator which is neither seller nor buyer (Richter and Slowinski, 2019).

The *degree of openness* on a data marketplace is based on the degree of accessibility. A closed data marketplace is only specified for specific actors whereas an open data marketplace is open to anyone or any company that wants to trade data. The former opens a lot of cooperation between actors, however this reduces the control over the quality of the data on the data marketplace (S. Spiekermann and Korunovska, 2017). In an closed data marketplace, the opposite on hand.

The *degree of integration* is about the types of data which are traded (Lange, Stahl, and Vossen, 2018). A data marketplace can have a broad domain and general data offerings across various sectors. There are also data marketplaces which are specified on data within one specific domain. The aspect *data transformation* makes a distinction between data marketplaces where the data is syntactically checked and prepared (S. Spiekermann and Korunovska, 2017). It is about whether on the data marketplace only raw data is traded and forwarded or that the data is also normalised and and aggregated or converted in an uniform format. With aggregation, the data is organised in report-based packages which makes the data ready for instant analyses.

The characteristic *platform architecture* differentiates centralised, decentralised and hybrid platform architectures (M. Spiekermann, 2019). For the centralised approach this means that data is provided by different suppliers on a central server which enables better control on the data. This overcomes all types of technological difficulties for data providers and consumers. On a decentralised architecture, the data is kept at the supplier's which raises technical difficulties regarding processing and storage of data but increases the data-sovereignty. Furthermore, a decentralised architecture enables direct trades, which is possible by distributed ledger technologies (DLT) such as blockchain where trades are verified by market participants (Koutroumpis, Leiponen, and Thomas, 2017, 2020). Hybrid architectures bundle these two architectures by decentralised trading and supplementary technical support in forms of infrastructure from the centralised platform architecture.

The *price model* is all about the way the price is determined which the data buyer pays to the data seller, 6 basic forms are distinguished (Stahl, Löser, and Vossen, 2015). Public institutions or non-profit institutions offer their data often for free. This helps the data marketplaces to gain the mass of demand of users. Fixed-price is a specified price, and the subscription price model gives parties access to the data over a certain period against a price. The pay-per-use model calculates which data is used and measures the final price on a summation of all used datasets. The progressive model is specified on the popularity or demand for the datasets. Thus, the price of an dataset increases when more consumers buy that particular dataset. This is used when data spreading is to limited.

The *revenue model* explains the profit generation of data marketplaces. This is mostly done by commission on each data purchase. It is also possible that fees are calculated on memberships of data marketplaces, the use of storage space or use of data service. On Freemium data marketplaces, basic functions are free of charge and for more functions the user has to pay a fee. There are also flat rate tariff data marketplaces where users pay a lump sump for an amount of time to use the complete platform with all its services. At last, there are also completely free of charge data marketplaces, often constructed by non-profit organisations as governments.

As this study is scoped at automotive data, it may be interesting to have an overview about the kind of automotive data marketplaces there are currently running online. The following table 3.1 contains an classification of B2B and B2C data marketplaces where automotive data is traded and is based on categories of M. Spiekermann (2019). In the automotive field of datasharing, different types of data can be traded from consumers (i.e. car users) to different business actors as shown in figure 1.1 by Kaiser et al. (2018). Most of these are already traded on these data marketplaces.

Table 3.1: Classification overview data marketplaces which trade automotive data

Data marketplace	**Value proposition**	**Data transformation**	**Platform architecture**	**Revenue model**	**Type of data**	**Founded**
Caruso	Data	Aggregation	Centralised	Membership Fee	Vehicle information, in-vehicle data, process data	2017
Otonomo	Data	Aggregation	Centralised	Transaction Fee	Traffic data, BMW Car data, Mercedes-Benz data, Avis budget group data, road sign data	2015
Here	Data	Aggregation	Centralised	Freemium	Location-based data	2018
IOTA	Transaction	Raw Data	Decentralised	Transaction Fee	Sensor data, but unclear for the mobility	2017
MDM	Data	Raw Data	Hybrid	Transaction Free	Traffic data, Parking information, Information on road works, Incident alerts, Petrol station prices, Static road network data, Environment data	2010
VINChain	Data	Aggregation	Decentralised	Progressive	Historical vehicle data	2017
AMO	Transaction	Raw and Aggregation	Decentralised	Progressive	V2X data, In-car data, Environment data	2017

The data marketplaces above, are all commercial and still online. However, various early founded data marketplaces already closed due to multiple challenges. This has to do with the obstacles in datasharing which are the lack of **trust** and lack of data security (S. Spiekermann and Korunovska, 2017) which induces the **risk of data disclosure**. Data owners are fearing that they lose **data control** if these data is used by other parties (Miller, 2014). This has a lot to do with the vulnerability of data marketplaces to the risk of strategic behavior by users, as the opportunity exists that their data can be easily and illegally transferred to third-parties after being sold to a party on the marketplace (Koutroumpis, Leiponen, and Thomas, 2020). This in term raises doubts on the pricing mechanism of the data and increases the intention to construct small one-to-one, bilateral data marketplaces where high-confidential and high-value data is traded. This type of data marketplace has strong rules and invasive monitoring but provides little space for economies of scale unfortunately (Koutroumpis, Leiponen, and Thomas, 2017). Another challenge lies in the pricing mechanism as many customers find the prices for the data too high. This is partly because consumers are not qualified to weigh the prices and on the other hand do not know that ensuring data quality is can also raise the costs (Miller, 2014; Stahl, Löser, and Vossen, 2015. In the end, this situation urgently asks for clear valuation procedures, however these are not yet available because valuing data is far harder than valuing material goods (Moody and Walsh, 1999).

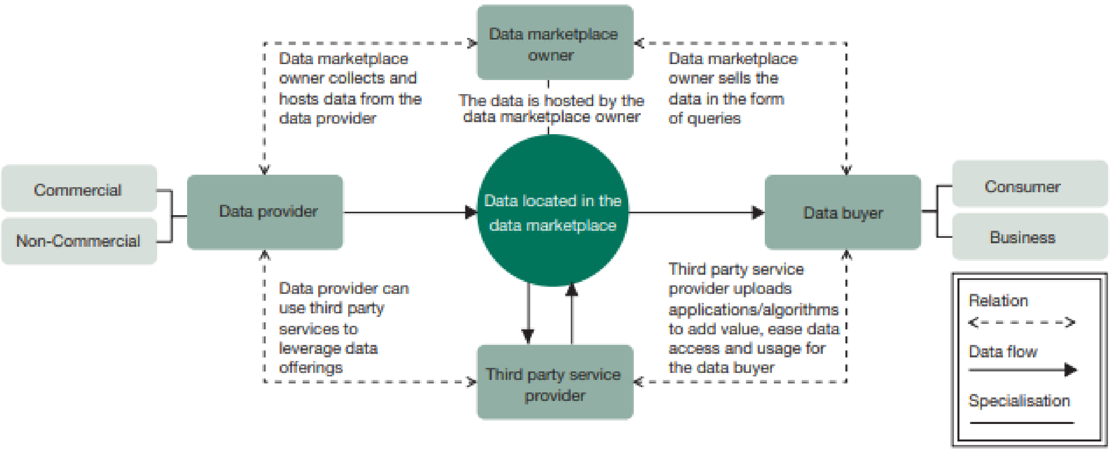
Besides Spiekermann's taxonomy, Abbas, Agahari, van de Ven, Zuiderwijk, and de Reuver (2021) conducted a study where they categorized all relevant literature on data marketplaces by the STOF model (Bouwman, Faber, Haaker, Kijl, and De Reuver, 2008) to structure the articles and investigate the state of the art of data marketplace research. This STOF model (Service, Technology, Organisation and Finance) is perfectly suited for this approach as it is designed for ICT services as data marketplaces (Abbas, Agahari, van de Ven, Zuiderwijk, and de Reuver, 2021). Currently, most research is dominated by literature scoped on the technical aspects of data marketplaces. Abbas, Agahari, van de Ven, Zuiderwijk, and de Reuver (2021) argue that the reason behind this trend may be due to the funding programs of the EU which are mostly focused on technological development. Furthermore, most platforms are still in the platform design processes, which are all about technical aspects instead of organisational or financial or service aspects. This might be one of the reasons that these data marketplaces are not yet commercially active, knowledge and research about non-technical topics are lacking (Abbas, Agahari, van de Ven, Zuiderwijk, and de Reuver, 2021). Consider the service category, within this category, the basis is laid to propose business models and increase commercial exploitation. However, few studies are found which also discuss non-technical literature; M. Spiekermann (2019), S. Spiekermann and Korunovska (2017) and Schomakers, Lidynia, and Ziefle (2020).

Nevertheless, there are certain directions towards many-to-many data marketplaces by means of distributed ledger technologies (DLT) such as blockchain, to reach a more securely monitored and decentralized data marketplace which could enable trading of high-value and confidential data on large scale. Instead of a central mediator, these transactions would be conducted and validated immediately by market players (Koutroumpis, Leiponen, and Thomas, 2020). The negative externalities of users' strategic behavior to take data of ledger and the possibility of data breaches could still remain. Nonetheless, it could potentially be a viable start for multilateral data market designs (Koutroumpis, Leiponen, and Thomas, 2020). Strangely, most research is done on these multilateral data markets taxonomies by M. Spiekermann (2019) (as in table 3.1) and Fruhwirth, Rachinger, and Prlja (2020), whereas most data marketplaces based on bilaterally negotiated contracts (Koutroumpis, Leiponen, and Thomas, 2017). For that reason, van de Ven, Abbas, Kwee, and de Reuver (2021) conducted a study to gain more knowledge on both types of contracting in order to set up a broad taxonomy which also includes business model dimensions. This better knowledge of data marketplace business models might aid in commercialization, making data more available and usable to consumers, corporations, and governments (van de Ven, Abbas, Kwee, and de Reuver, 2021).

3.2.1. Platform actors and interactions

In data marketplaces there are mainly three or four actors involved. Two of them are the well known buyers and sellers. In our view, we follow M. Spiekermann (2019) where in addition, the infrastructural manager (data marketplace owner) is added as third main actor. It may happen that the third party service provider is also the data marketplace owner, then the fourth actor is not present. Figure 3.2 below shows how the four different actors interact with each other.

Figure 3.2: Schematic representation of data marketplace actors (M. Spiekermann, 2019).


Data marketplace owner
Data marketplace owner collects and hosts data from the data provider
The data is hosted by the data marketplace owner
Data marketplace owner sells the data in the form of queries
Commercial
Non-Commercial
Data provider
Data located in the data marketplace
Data buyer
Consumer
Business
Data provider can use third party services to leverage data offerings
Third party service provider uploads applications/algorithms to add value, ease data access and usage for the data buyer
Third party service provider
Relation
Data flow
Specialisation

The *data providers* can be seen as the data sellers of the system. These are individuals which sell their automotive data on the data marketplace via applications or devices. These sellers frequently have minimal awareness of the platform's internal dynamics, requiring simple and succinct user interfaces. Furthermore, the amount of sellers can be very large but on the same time providing limited amount of data. In the end, they are often providing very sensitive user-data which raises tracking and profiling issues which urge for strong data security. These data providers can be commercial or non-commercial as mentioned in the previous section. The *data buyers* are the consumers in this case. These are often organisations who are aiming on aggregation of users' data to perform Machine Learning or profiling tasks (Giaretta et al., 2021). The sellers and buyers are connected by a certain infrastructure (data marketplace) which is in need for governance. Often, additional actors provide the computational power to the platform in order to enable the datasharing and computation processes, these are the *third party service providers*. They facilitate the services which consumers can use to make better use of the bought data. As mentioned in the previous section these could be aggregation of data or the preparation. Then, in the end, there is the *data marketplace owner* which is responsible for the platform and the main facilitator of storage of the data.

3.2.2. Platform Incentives and Actor requirements

Before sellers connect to the platform, there have to be certain incentives to attract them. Benefits and requirements are the constraints which need to be fulfilled in order to bring up the supply side of these data marketplaces. The following requirements are needed to attract sellers (Giaretta et al., 2021):

1. **Data Control**, the ownership of the data must be maintained at the seller-side. Full control over storage and distribution must be preserved.
2. **Data Privacy**, without authorization from the seller, no entity must be able to access the sellers' data. This aims at minimizing the risk of data disclosure.
3. **Benefit**, this ensures that value generated by the sellers will largely be given back to the sellers on the marketplace.

From the consumers' side, these other requirements are envisioned (Giaretta et al., 2021):

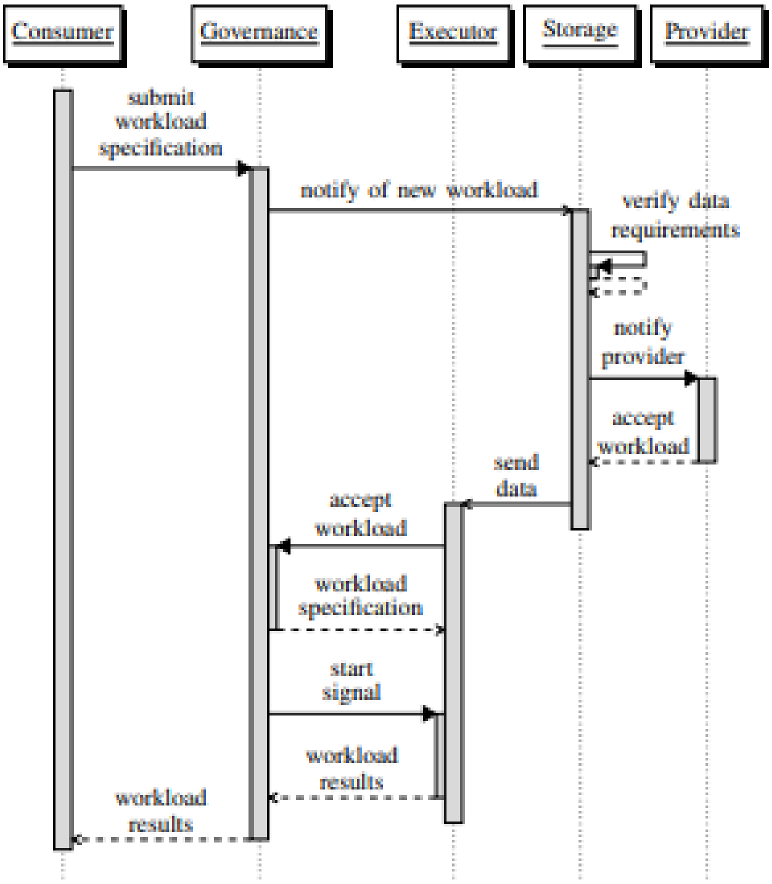
1. **Workload Confidentiality**, this is focused on the situation that whenever a buyer buys certain data - this must not be available to any other consumer somewhere for free. Even the trade and transaction information of could represent certain information.

2. **Data Authenticity**, which is focused on identification and rejection of data that is intentionally wrong and affecting cleaned and honest information.

3.2.3. Platform Roles

To have a certain idea of the different roles and functions within a datamarketplace, figure 3.3 shows a sequence diagram of an high-level trade on a data marketplace. First, data consumers submit their specifications in which they specify the preconditions where the data must fulfill. These conditions can be specified on validity or reliability, the minimum amount of data or the type of data provider. The providers continuously produce (automotive) data which are stored on the storage subsystem of the data marketplace. They must determine whether or not they want to contribute their data to this effort. As a result, the database is in charge of permanently keeping these data and connecting buyers to sellers. The executors (third party service providers) are purely providing the computational power to execute the workloads. The governance actor (data marketplace owner) is mainly auditing the platform and keeping track of all transactions of workloads between data providers, executors and consumers. This layer also distributes the rewards (benefits) and verifies that no actor is behaving maliciously (Giaretta et al., 2021).

Figure 3.3: Sequence diagram of trade on datamarketplace (Giaretta et al., 2021)



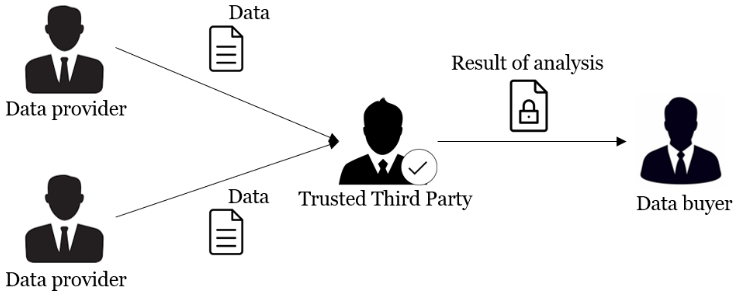
3.2.4. Privacy Preserving Technologies (PPT's)

As it is in the name, privacy preserving technologies (PPT's) are designed to ensure privacy on (personal) data. Privacy preservation is an essential concept in big data since it is required to give security to data once it is moved or communicated between multiple parties so that other organizations do not learn what data is shared between the original parties (Aldeen, Salleh, and Razzaque, 2015). As a result, privacy preservation differs from traditional data security. We follow Yu (2009) and define PPT's as encryption technologies that regulate access to data, aiming to prevent information disclosure by unauthorized parties. Thus, when the output data is important and private, privacy preservation techniques in data sharing entails concealing the information. This way, PPT's help to keep the data classified which enables people to share (personal) data which would otherwise not been possible. Within literature, there are various sources explicating the different types of PPT's. It may be useful to explain multiple PPT's to show the difference between MPC and other PPT's and why specifically MPC, could be useful on data marketplaces.

The GDPR accepts various encryption methods as long as the keys to decrypt are only available to the ones who are entitled to have them. However, most data analyses are not compatible with encryption methods, often the clear raw data is needed (Christen, Gordijn, and Loi, 2020, p. 293). We follow (Christen, Gordijn, and Loi, 2020) which describe two different encryption technologies which are compatible for privacy preserving computations. Thus, these technologies preserve privacy and are simultaneously able to directly perform computations on the encrypted data. Thus, these technologies calculate a function based on encrypted data, and the ones who are able to decrypt it get the same answer as if the function was calculated on the real raw data. First another current, less technical way of data sharing will be introduced.

Currently, confidential information from different parties can be shared and analyzed through the intervention of a *Trusted Third Party (TTP)* (Sousa, Antunes, and Martins, 2018). As the name suggests, a TTP is a party that is trusted by the parties that want to combine their confidential data sets in order to arrive at answers (TNO, 2020). Figure 3.4 illustrates in a simplified way the process of analyzing data with a TTP. The data purchaser can be an external party interested in the aggregated output of the analysis, but can also be the data provider. To guarantee that the third party handles the data confidentially, a confidentiality agreement is usually signed (Al-Sharidah, Syed, Alsannat, and Gaddourah, 2020). The disadvantage of this method is that it is often expensive, it requires a great deal of trust in the third party and that there is a security risk if the data is insufficiently protected (Lapets, Volgushev, Bestavros, Jansen, and Varia, 2016). The data providers have the option of being selective in which confidential information they disclose to the TTP (Lapets, Volgushev, Bestavros, Jansen, and Varia, 2016). Nevertheless, due to a lack of technology that can guarantee the privacy and security of the data, there are risks associated with sharing data with a trusted third party (Roman and Vu, 2018).

Figure 3.4: Analysis of data via a Trusted Third Party (TTP)



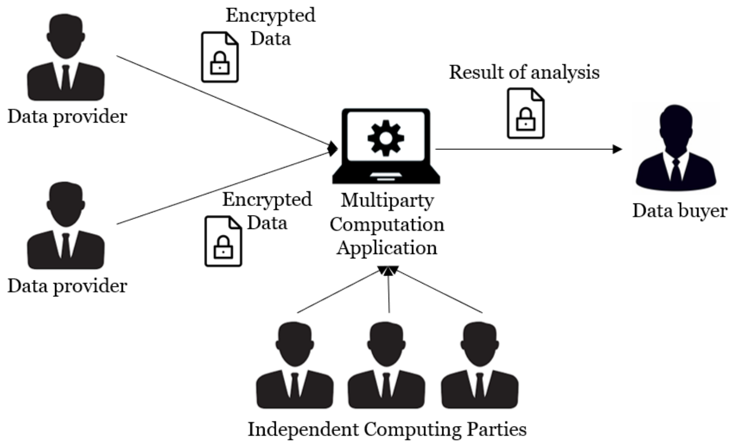
Besides trusted third parties, within the privacy preserving technologies there exists *homomorphic encryption* (Gentry, 2010). Homomorphic encryption could serve as an anonymization method. Homorphic encryption, is a combination of multiple encryption techniques which allow for computation processes on encrypted data (Gentry, 2010). These provide confidentiality while preserving data encryption which makes them very reliable. However, this leads to many additional operations and an increase in time-duration during these computation processes. Therefore, on huge data-sets it is currently a big task to apply this technology in a efficient manner. Alternative strategy is to make the data no longer linkable to the individuals which is called *differential privacy* (Dwork, 2006; Dwork, Roth, et al., 2014). The data is no longer personal which lifts the legal restrictions that apply. There are differences between these techniques while they look that similar (Apfelbeck, 2018). For instance, in MPC, multiple data providers are required to start computations. Furthermore, differential privacy makes use of the introduction of random noise in the data analysis without encrypting data. In the end, one has to look at the ways how these technologies can improve or complement each other to reach robust data security as in these various use cases (e.g. Alter, Falk, Lu, and Ostrovsky, 2018; Pettai and Laud, 2015) (Agahari, Dolci, and de Reuver, 2021). As MPC makes it possible to simultaneously share data in a safe and privacy preserving manner, MPC is expected to have relatively more potential to be of value to data marketplaces than the other privacy preserving technologies. This is because multiple data providers can simultaneously provide automotive data which could enable more efficiency in the data sharing process and multi-sided results. Therefore MPC is further used in this thesis and will be explained in the next section.

3.3. Multiparty Computation (MPC)

This section is written in order to broaden the reader's understanding on MPC, as the experiment will be based on MPC-enabled data marketplaces. This section is the last missing part besides having elaborated the data sharing factors, the automotive data sharing context, data marketplaces. Now, this privacy preserving technology which aims to facilitate secure data sharing of sensitive automotive data on data marketplaces. As MPC can be perceived as a complex privacy preserving technology (PPT), the main components of the MPC technology are first elaborated, and two examples are given in this section. Second, a glimpse of the general MPC-architecture is given in 3.3.1 in order to understand the different steps within a MPC-process. Then, various current MPC applications are stated in order to broaden the knowledge in 3.3.2. In section 3.3.3, the conclusion is given on MPC-enabled data marketplaces and why these should influence the selected data sharing factors in a positive way. Then, the next chapter will address the method and experimental design wherein the chosen data sharing factors of the previous chapter will be used to deviate in MPC-enabled data marketplaces.

Multiparty Computation (MPC) is a technique where the use of Third Trusted Party (TTP) could be neglected as shown in fig 3.5. MPC requires the presence of multiple parties which together input the data and compute these functions to gain knowledge. However, due to these active participation of multiple parties, increases the delays in communication between the parties which makes it more demanding to employ MPC on platforms which perform many operations (Giaretta et al., 2021). Nevertheless, MPC is besides homomorphic encryption the only currently possible technique where multiple parties can calculate a function based on the encrypted data (Apfelbeck, 2018). MPC could therefore potentially be suited as technique for data sharing on data marketplaces where the raw data is aggregated and needed for complex calculations. In the next section 3.3, MPC will be extensively explicated.

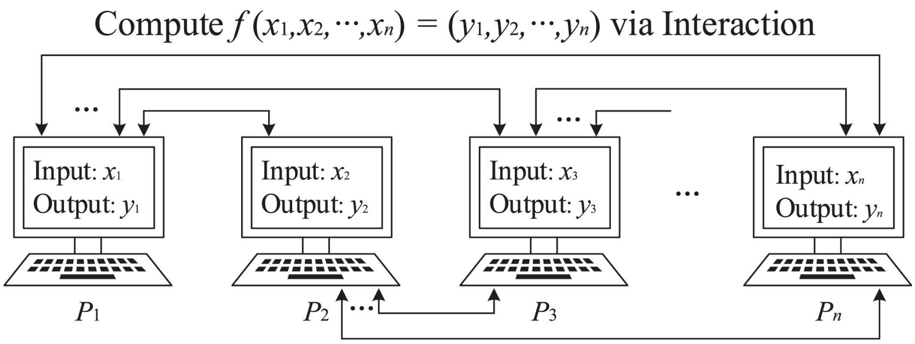
Figure 3.5: Analysis of datasharing via Multiparty Computation (MPC)



The MPC-technology can be a possible solution for quite diverse applications. Either within organisation layers or between different parties on data marketplaces. Between unknown or known, trusted and non-trusted parties, or just to secure a connection to enter a certain profile. Before explaining the MPC-technology further in detail, we explicate why MPC could affect the way people look at our four different identified data sharing factors. As MPC is a technology meant to preserve privacy, this may affect the way risks of data disclosure, data control and trust (in the form of a herding effect) are perceived by consumers. Furthermore, even benefit could be valued differently than in data marketplaces where no privacy preserving technologies are applicable. For instance, it may happen that people perceive data marketplaces more safe and more trustworthy, and feeling themselves more in control of their own data and therefore do demand different benefits than in old fashioned data marketplaces. We expect people to behave differently and value the four proposed factors differently in MPC-enabled data marketplaces. The MPC technology could be a valuable tool in enabling 'Computation on Encrypted Data' (CoED) (Archer et al., 2018), thus to extend data marketplaces with data analytic-properties. This will be discussed later on in section 3.3.3. Now, let us therefore dive deeper in the inner-workings of MPC.

In MPC, multiple IP's (actors), possessing disclosed datasets - can interactively and jointly compute a objective function by inputting their datasets (Zhao et al., 2019). As in the following figure 3.6, a schematic visualisation is shown of multiple actors jointly interacting in order to compute a function by their contributed input data.

Figure 3.6: Visualisation of Multiparty Computation (MPC) (Zhao et al., 2019)


Compute $f(x_1,x_2,\cdots,x_n) = (y_1,y_2,\cdots,y_n)$ via Interaction
Input: x_1 Output: y_1
Input: x_2 Output: y_2
Input: x_3 Output: y_3
Input: x_n Output: y_n
P_1
P_2
P_3
P_n

This figure could for example visualize a computation, determining which company yields the highest yearly revenue. For example, suppose that four companies (i.e. Adidas, Bol.com, Carglass and Decathlon) want to compute a function based on private inputs a, b, c, and d respectively. They agree to compute the following function:

$$F(a,b,c,d) = max(a,b,c,d) \tag{3.1}$$

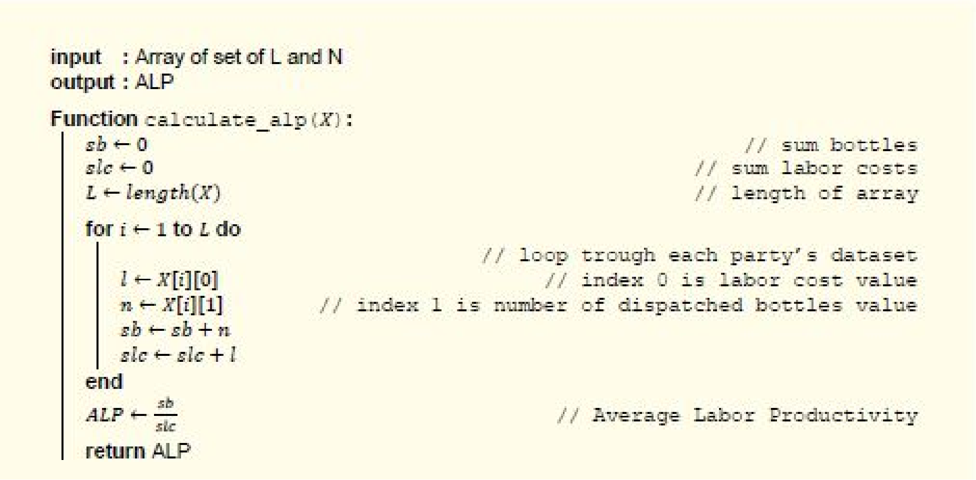
while keeping their private inputs disclosed.

Before engaging in the computation of this function, all parties engage in a predefined protocol. We follow Archer et al. (2018) which interprets a protocol "as a set of instructions in a distributed computer program". That application consists of a set of interactive fixed steps which are made explicit to all parties involved in advance. Each party provides a confidential bit of data and receives a final output. In general, the protocol can thus be seen as an functionality that deterministically maps inputs to outputs without randomness.

For the above example, if for instance Carglass' input c, is the general output - this company will know they have the highest input. Whereas Adidas, Bol.com and Decathlon will only know that their input is not the highest. This basic example can easily be generalised to more complicated functions combined with multiple inputs and outputs which subsequently are further used as inputs. One can imagine that it can be rather difficult to decrypt one's input values.

Another example could be an benchmark application between for instance different wine-companies on their Average Labor Productivity (ALP). Therefore, the different companies have to deliver their labor costs in euros (L) and the amount of dispatched bottles (N). Due to MPC, this benchmark can be computed without disclosing their sensitive L and N. In figure 3.7, an example function is written down.

Figure 3.7: Visualisation of Benchmark Code (Petronia, 2020)



```
input  : Array of set of L and N
output : ALP
Function calculate_alp(X):
    sb ← 0                                                // sum bottles
    slc ← 0                                               // sum labor costs
    L ← length(X)                                         // length of array
    for i ← 1 to L do
                                              // loop trough each party's dataset
        l ← X[i][0]                                // index 0 is labor cost value
        n ← X[i][1]             // index 1 is number of dispatched bottles value
        sb ← sb + n
        slc ← slc + l
    end
    ALP ← $\frac{sb}{slc}$                           // Average Labor Productivity
    return ALP
```

In this example, an arbitrary protocol is executed where companies collude to extract information from eachother in order to compute a function. The envisioned properties which MPC protocols aim to ensure are thus:

1. *Input Privacy*: ”The information derived from the execution of the protocol should not allow any inference of the private data held by the parties, bar what is inherent from the output of the function” (Archer et al., 2018, p. 2).
2. *Robustness*: ”Any proper subset of adversarial colluding parties willing to share information or deviate from the instructions during the protocol execution should not be able to force honest parties to output an incorrect result” (Archer et al., 2018, p. 2).

However, while these companies from the example do not want to reveal their data, it keeps possible, even when the program is secure - to extract some information from other companies. Maurer (2006) calls this the vulnerability which could be explained as the threshold for opponent to inflict corruption to the collaboration. Of course, this example is not really complex and benchmarks often are based on more data which induces that this example could lead to misleading conclusions (Petronia, 2020). As Zare-Garizy, Fridgen, and Wederhake (2018) accurately describes, more data is needed to draw correct conclusions - but this increases the traceability of actors' input values. Extra information as devaluation of goods, transportation costs are also needed in order to fully grasp the correct benchmark. But, this could increase the possibilities to link all results to specific IP's or rather companies. As Petronia (2020) clearly describes, even though a protocol is safe, the calculation itself may leak classified info about the inputs.

The issues generate challenges in the MPC protocol requirements which need to be fixed. There is thus still room for attacks of an opponent which affect the privacy, correctness, fairness and the output (Zhao et al., 2019). Based on the perceived level of crime within a group of data-contributors, different models are set up to categorize for the required level of security of the MPC application. These situations may also exist on data marketplaces so the following situations are defined to give an certain idea of the scenarios:

- *Semi-honest adversary model*, where the users will probably execute the protocol as delivered, but might try to derive knowledge from the solution.
- *Malicious adversary model*, where corrupted parties might deviate from the instructions of the protocol based on enemy's orders (Bestavros, Lapets, Jansen, Varia, Volgushev, and Schwarzkopf, 2017; Catrina and Kerschbaum, 2008).
- *Covert adversary model*, where parties try to cheat if they know that they will not get caught or if they expect the loot to be bigger than the damage of getting caught (Zhao et al., 2019).
- *Rational adversary model*, where participants will deviate from the protocol if it will benefit their utility function. (Miltersen, Nielsen, and Triandopoulos, 2009).

In the end, there is a constant trade-off between advancement of security and practicality. While in more advanced secure MPC models the imposters among the participants have less room for strategic behavior, the entire service will be computationally more expensive and therefor less practical. Nevertheless, Zhao et al. (2019) mention that a MPC-protocol is only secure if it can handle any malicious attacks in the current state of security. As MPC is fairly new (Choi and Butler, 2019), not all technological concerns have been resolved by today. (Zhao et al., 2019). As, within the Usable and Efficient Secure Multiparty Computation (UaESMC) research[2], various implementation challenges already have been risen up (Kerik, Laud, and Randmets, 2016). Simultaneously, individuals are more ready to accept models that aren't as good as they could be. Bestavros, Lapets, Jansen, Varia, Volgushev, and Schwarzkopf (2017) address that weak models which are more advanced and efficient in a technical sense, can prove to be valuable in computations without collaboration. Other ways to enhance the security level of the protocol is by accompanying risk profiles or reputation-based systems (Bestavros, Lapets, Jansen, Varia, Volgushev, and Schwarzkopf, 2017).

Furthermore, complex variables as **trust** and **risk of data disclosure** may affect the way people are willing to share data on MPC-enabled data marketplaces. Recently, M. de Reuver, Fiebig, Agahari, and Faujdar (2020) examined and demonstrated that the visualisation of Multiparty Computation has effect on how these two factors are perceived. It may be useful to estimate how these two factors influence peoples' behaviour regarding automotive datasharing on MPC-enabled data marketplaces, so that we can check whether there exists a direct connection between increased security and trust (in terms of a herding effect) in this technology. In the end, in order to gain from this type of datasharing and information winning, it depends completely on the party's next step which is to analyze and utilize the gained information.

3.3.1. MPC Architecture

In figure 3.8, an example architecture of MPC is visualized. Often, the distributed computation by multiple servers is used as shown in the middle *Computing parties (CP)*. A common MPC system is defined by three fundamental roles whereby each person or institution which is involved in a MPC computation may hold one or more of the following roles (Archer et al., 2018);

- *Input Parties (IPs)*, ”delivering sensitive data to the confidential computation” (Archer et al., 2018).
- *Result Parties (RPs)*, ”retrieving results or partly results from the confidential computation” (Archer et al., 2018) .
- *Compute Parties (CPs)*, ”jointly computing the confidential computation by using the confidential input-data” (Archer et al., 2018).

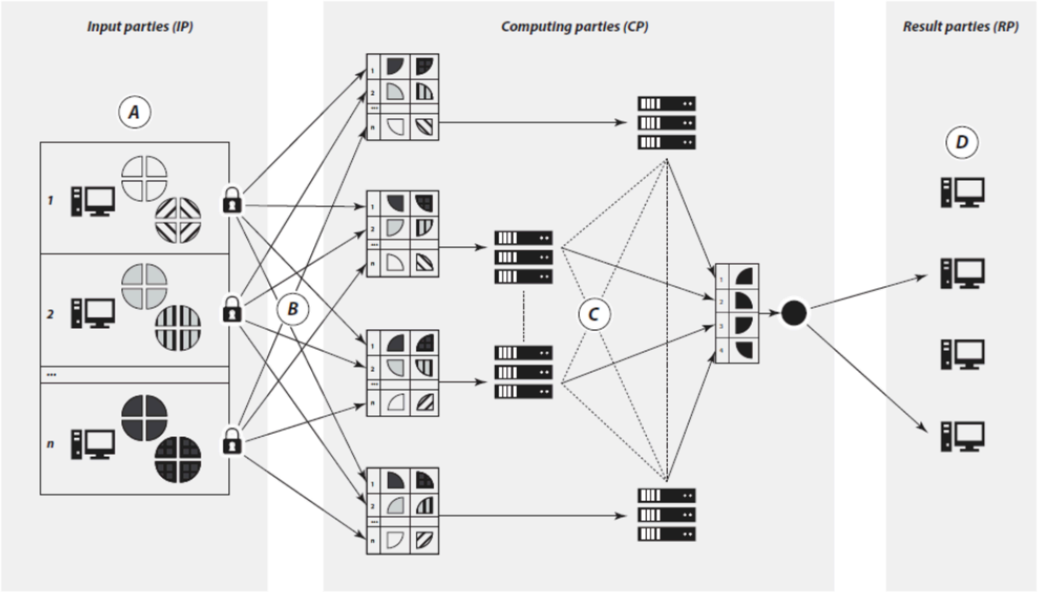
But, these roles often diverge in practice, as often very heavy computations need to be executed which require many parties to take the role of CP in the computation-process. As this can be heavily demanding on companies' total computation power, this could lead a burden on resource requirement which hinders MPC adoption.

The process of MPC is divided over 4 steps in the figure. The first step (A) is the submission of the data, this data is then encrypted and collected through applications or web-based forms or different plug-ins. The second step (B) can differ between architectures. This step B is the distribution of the IP encrypted input data to different servers, this is secured by a HTTPs channel[3] in Bogdanov, Talviste, and Willemson (2012) - which serves as secure transportation of data. While in Bogetoft et al. (2008), in a small application, each share of input data is encrypted by various public keys and sent to storage servers. In this example, step C is performed on a decentralised architecture where all parties have the data on their own servers (Keep in mind that this can also be saved and computed centrally, as in our experiment later on).

[2] see the project at: https://cordis.europa.eu/project/id/284731

[3] see here for elaboration on HTTPs: https://www.cloudflare.com/learning/ssl/what-is-https/

Figure 3.8: Example MPC architecture with data flow (Bestavros, Lapets, Jansen, Varia, Volgushev, and Schwarzkopf, 2017; Bogdanov, Talviste, and Willemson, 2012; Bogetoft et al., 2008)



In step C of figure 3.8, the multiparty computation process starts. Each party performs the same instructions given by the MPC protocol on the data-shares they possess. In the last step D, the output of all functions is distributed over all *Result Parties (RPs)*. It is important that the environment architecture among the *Compute Parties (CPs)* safeguards against recreation of data-shares to the original input value (i.e. by private or public keys). Therefore, it is also important that *Compute Parties (CPs)* (and the *Input Parties (IPs)*) are independent and are not inclined to collaborate strategically.

3.3.2. Various current MPC Applications

As MPC is a co-educational technique that does not rely on trusted individuals or institutions, Archer et al. (2018) conclude that MPC can be used (in a variety of different forms) to create more trustworthy institutions than those we have today. To get a certain sense about the use cases of MPC, Archer et al. (2018) already proposed multiple use cases which can be divided in two sets: small encryption of keys and encryption of entire databases. Here follow a couple of use cases to broaden the understanding:

- The Jana system, an MPC-enabled secure database, offering PDaaS (Private Data as a Service). This is unique among encrypted databases due to the ability to offer also encryption during the processing of data instead of the storage only.
- Cybernetica, similar MPC-enabled database. Unique in the fact that it is not an central-based encrypted database but decentralized by distributed ledger technology.
- Off-exchange matching, Alternative Trading System (ATS) which specializes matching services to match and facilitate buyers and traders in securities. These retrieve lot of data about buyers' willingness to pay (WtP) and sellers' willingness to accept which is encrypted by MPC.
- Privacy preserving statistics and passwords, several applications on mobile devices make use of user profiles including passwords. These passwords are very vulnerable if stored in plain-text. Therefore, by MPC between the server and mobile devices, long hash values can be computed in order to secure the users' credentials.
- Cloud key management, TLS key protection which is heavily used in the banking sector to secure keys in ATM's or to secure data in payment card systems.
- Voting and benchmarking mechanisms for elections, where all people vote but individual votes can hardly be disclosed (Sousa, Antunes, and Martins, 2018).

3.3.3. Conclusion: Potential of MPC-enabled data marketplaces

This conclusion explicates why MPC would help data marketplaces to overcome the barriers as in the lack of trust, security (risk of data disclosure), privacy and transparency (Agahari, Dolci, and de Reuver, 2021). As explicated in the the data marketplace-section (3.2), data marketplaces suffer from technologies that enable data control on the platform in order to keep data ownership to the data providers owning the data. As MPC is currently not yet applied on data marketplaces, we follow Agahari, Dolci, and de Reuver (2021) who within an European founded organisation Safe-DEED are investigating on adoption of MPC on data marketplaces to ensure data control to data providers. They found out that MPC could overcome the pitfalls of lacking security and privacy (1), the provider having control over their own data (2) and ensuring the correct computations (3). MPC could even generate value to data marketplaces.

Three aspects of MPC could overcome these barriers, as MPC is based on (1) computational security, (2) based on protocols for computation and the correct control mechanisms by suspicious actions, and (3) therefore correct executions of computations (Agahari, Dolci, and de Reuver, 2021). These three aspects could make it possible to preserve the provided data at the data provider and enable data ownership and the preservation of of results by the multiparty computation. These three values could thus have a positive impact on the factors **risk of data disclosure**, **data control**, **trust** and **benefit** and therefore this user-based study is interesting to perform to check whether this effect is true. These values are solutions to the data marketplace objectives with the help of MPC.

However, beware, Agahari, Dolci, and de Reuver (2021) also recognised several factors which may hamper the adoption of MPC on data marketplaces. The technology is relatively complex to understand this negatively affects the perception of MPC (1), due to this it may that data marketplace owners or users may not see the need for this technology (2). Even if they see the need for this technology, there is no recourse availability yet as the maturity is not adequate to fulfill the desires (3). The radical change to adopt MPC could also be one of the factors which affects the adoption (4). Furthermore, some implications arise by implementing MPC on data marketplaces. First of all, by implementing MPC on data marketplaces, this changes the platform architecture of the data marketplace. As MPC works in a decentralised way, centralised data marketplaces need to change to decentralised platform architectures (Agahari, Dolci, and de Reuver, 2021). Also, data marketplaces which only provide their service in a way of a broker, have the opportunity to offer new types of services as data analytics. And, constructing MPC-enabled data marketplaces can result in additional costs.

Nevertheless, by promoting MPC on data marketplaces, this could generate awareness of the technology which may further incentivizes people to dive deeper in MPC. The only studies which are currently focussing on MPC-enabled data marketplaces are Koch, Krenn, Pellegrino, and Ramacher (2021) and Roman and Vu (2018). The former couple initiated a type of architecture for data marketplaces where MPC is attached to blockchain based smart-contracts (Agahari, Dolci, and de Reuver, 2021). Koch, Krenn, Pellegrino, and Ramacher (2021) tried to already provide privacy-preserving data analytics with the use of MPC.

To conclude, besides that there are still many barriers to overcome when conceptualizing MPC-enabled data marketplaces, MPC could bring value to data marketplaces which may affect the four selected user-based data sharing factors in this study: risk of data disclosure, data control and benefit. In the next chapter, the methodology behind the stated choice experiments will be explicated in order to move towards the experimental design where the experiment on these MPC-enabled data marketplaces will be realized and released. Then, answers will be given on how important these factors are relatively to each other.

Part III

Experiment

4 Methodology

The methodology is the building block and crucial step before entering the experimental phase. In this chapter, first an overview is given on the general approach of this study in 4.1. Then, the method which is used to gain and analyse the data will be explained in 4.2. This will serve as the framework and fundament of the next part; the experimentation phase. This chapter is thus the fundament for the experimental design and the further input to the stated choice experiments and analysis. In the sections 4.3, 4.4 and 4.5, different models are explained which are further used to compare among each other to have a certain sense of which model to use and which model performs best. Several models are thus used in order to address the research questions and sub-questions stated in section 1.5. The model and the reason why this model should be used will be described there. The relative weights of the components will be calculated using survey data acquired objectively. The selected research approach will now be explained in greater detail.

4.1. Research Approach: Modeling

As explorative research is about collecting and analyzing numerical data which can further be used to find certain patterns (Goundar, 2012), discrete choice-experiments are set up in order to obtain stated preferences of consumers for exploring how consumers value the four data sharing factors on MPC-enabled data-marketplaces. This Discrete Choice Modeling (DCM) method for analyzing actors' behavior, can be described as explorative in nature (Salampessy et al., 2015). Besides finding patterns, attractiveness for data-marketplaces by MPC can be estimated and causal relationships can be shown. As there is a lack of understanding of the functioning of MPC data-marketplaces, this fits with the overall goal of this research which consists of visualising the impact on automotive businesses in certain (hypothetical) data-marketplaces and contributes to the Safe-DEED project which is focused on worldwide acceleration of data-sharing.

The approach to come up in this thesis is the modeling approach. Discrete choice sets about MPC-enabled data marketplaces will be constructed for car owners and the requirement of sufficient respondents needs to be fulfilled. After receiving the choice-modelling data, a R-model will be used in which (mixed) Logit-Models are applied to the retrieved data. This will determine the weights of particular factors which make the data most likely. The model's log-likelihood (LL) and the significance of attributes can be evaluated. Moreover, this model could iteratively be improved to identify if different additional attributes (which were not involved in the initial model) make the estimation of the model's likelihoods even better.

Advantages of this stated choice modeling approach is that the approach is very flexible in nature, it is not bounded by existing data-marketplaces due to its experimental design (Mandeville, Lagarde, and Hanson, 2014). Furthermore, DCM is a fairly mathematically elegant way to determine market shares of certain MPC data-marketplaces, because it shows people's preferences of attributes without asking them these difficult questions directly. Apart from these advantages, one has to be very aware of the hypothetical bias in this type of modeling and of the respondents (Rakotonarivo, Schaafsma, and Hockley, 2016). Because of the experimental nature of this approach, people's choices might not reflect what people would choose in reality. One should also be aware of constant trade-off between reliability and validity of DCM which is reflected in the results (Carson et al., 1994).

4.2. Research Method: Discrete Choice Experiment

In general, consumers currently have a lack of understanding in MPC (Bogetoft et al., 2008), and data marketplaces are not yet commonly known, the consumers' willingness to share automotive data on MPC-enabled data marketplaces is yet unknown. It is therefore hard to ask consumers directly how they value their own willingness to share automotive data on a platform which is unknown to them with a technology which is perceived as relatively complex. Thus, people do not know exactly how important they value different data sharing factors in this new context. Therefore, Stated Choice Experiments (SCE's) will be constructed in order to let people make choices between MPC-enabled data marketplace options, instead of asking the willingness directly. This way, by Discrete Choice Modeling (DCM), the way in which consumers value the data sharing factors can be retrieved. Thus, by letting people make these choices between different options, the evaluation and estimation of the data will obtain the preferences of people. It is therefore very important that the choices display clear and real trade-offs between different factors and are correctly formulated according (technical) rules to minimize standard errors and maximize validity and reliability. Also the requirement on sufficient respondents needs to be fulfilled. The choice data of the people are the crucial data which needs to be obtained in order to analyse the perceived factor importance. This data is thus key for the next step, in which our constructed software programs will help analyze the important elements within the data.

Advantages of this stated choice modeling approach is that the approach is very flexible in nature, it is not bounded by existing data marketplaces due to its experimental design (Mandeville, Lagarde, and Hanson, 2014). Furthermore, Discrete Choice Modeling (DCM) is a fairly mathematically elegant way to determine market shares of certain MPC-enabled data marketplaces, because it shows people's preferences of attributes without asking them these difficult questions directly. Apart from these advantages, one has to be very aware of the hypothetical bias in this type of modeling and of the respondents (Rakotonarivo, Schaafsma, and Hockley, 2016). Due to the experimental nature of this approach, people's choices might not reflect what people would choose in reality. One should also be aware of constant trade-off between reliability and validity of DCM which is reflected in the results (Carson et al., 1994).

There are various Discrete Choice Models that can be used in estimating the factors which are varied in the SCE's. In this research, the Multinomial Logit Model (MNL), the Mixed-Logit (ML) and the Random-Regret-Model (RRM) are used to analyse the choicedata. Each different model has its own technological niceties and advantages and disadvantages. Nevertheless, by having a broad analysis on the obtained choicedata by different models, a more explicit analysis can be drawn which enables the room for discussion and conclusion on the different models which enlarges the utility of this research. In the following subsections, each model is explained in a succinctly manner. But first, let us compare this study with similar studies which were also based on DCM. This way, we can better argue why we used DCM instead of basic surveys or interviews.

The tools are of great importance. Two different software-tools (Ngene and R) are needed in order to fully grasp the desired information. Ngene is a software tool, capable of generating survey structures for a broad spectrum of discrete choice experiments. Ngene assists in optimizing a survey by maximising the information and making choice experiments more realistic and familiar to respondents. Afterwards, the filled in choice experiments data is used as input for the constructed R-model wherein the weights will be estimated which make the data most likely. The weights display the importance of a factor in a numerical way. Besides this, significance levels, standard-errors, Log-Likelihoods (LL) and data marketplace market shares can be obtained which are needed to evaluate the decision making process whether MPC is valuable within data marketplaces. The mentioned terms are all important in concluding if the results are suited to be generalized to the whole population instead of just on the group of respondents.

4.3. Multinomial-Logit model (MNL)

In the Multinomial-Logit model (MNL), the estimation of each factor, called the beta, is based on the Maximum Likelihood-principle. This principle finds the set of parameters that makes the filled-in choice data most likely. Instead of the likelihood, the logarithm is used which becomes very large and negative. This is done as this gives easier numbers to calculate with and to report. MNL models are also based on random utility maximization (RUM). This holds that the alternative with the highest utility, has the best possibility of being selected by people. Utility of an alternative is based on the summation of each factor plus a error term. Let us first handle all key elements in a RUM MNL choice model, for each individual n:

1. Alternative-subscripts (e.g. MPC-enabled marketplace 1, MPC-enabled marketplace 2) i, j
2. Factor-subscripts (e.g. risk of data disclosure, benefit) m
3. Factor-values (e.g. low risk, 20 dollars) x
4. Weights (e.g. for risk of data disclosure, benefit) (**to be estimated**) β
5. Randomness (error term) ϵ
6. Decision rule (for the moment: RUM)

n's utility of alternative i:

$$U_{in} = V_i + \epsilon_{in} = \sum_m \beta_m * x_{im} + \epsilon_{in} \tag{4.1}$$

i is chosen by n if:

$$\sum_m \beta_m * x_{im} + \epsilon_{in} > \sum_m \beta_m * x_{jm} + \epsilon_{jn}, \forall j \neq i \tag{4.2}$$

Note that weights are estimated at the sample level, not for each individual. For simplicity, it is assumed that factor-values are the same for each individual

Thus, a random sampled individual chooses the alternative whose total utility is the highest among all alternatives in the choice set.

$$TotalUtility = SystematicUtility + errorterm = \sum_m \beta_m * x_{im} + \epsilon_{in} \tag{4.3}$$

Systematic utility: sample-specific summary of all that can be related to the observed factors (e.g. risk of data disclosure, data control, trust, benefit)

Error term: everything else that governs the individual's choice (unobserved factors, heterogeneity in tastes, randomness in choices)

So: even when the systematic utility is highest, that alternative may still not be chosen by a particular individual in a particular choice situation. In other words, a choice can only be predicted by a probability: a higher systematic utility leads to a higher choice probability. The probability of i getting chosen is then:

$$P(i) = P(V_i + \epsilon_i > V_j + \epsilon_j, \forall j \neq i) = \frac{exp(V_i)}{\sum_{j=1..J} exp(V_j)} = \frac{exp(\sum_m \beta_m * x_{im})}{\sum_{j=1..J} exp(\beta_m * x_{jm})} \tag{4.4}$$

(where i is included in J)

The Newton-Raphson-method is often used for the estimation of the parameters (the weights of the factors)(C. Chorus, 2017). A log-likelihood (LL) is thus the logarithm of the likelihood that the found betas make the obtained choice data most likely.

The MNL-model gives the following outcomes:

- The parameter estimates (betas)

- The final log-likelihood (LL)
- The standard errors associated with the parameter estimates (se's)

The LL can be used to determine a model's fit, by using the McFaddens' rho-squared (McFadden et al., 1973):

$$p^2 = \frac{LL_\beta}{LL_0} \tag{4.5}$$

Where LL_β is the log-likelihood of the estimated model, and the LL_0 is the log-likelihood when all betas were zero (which would mean random model). If p^2 is equal to zero, the model does not estimate better than throwing a dice. If it is 1 (which is almost impossible), the model has a perfect fit.

However, we need to adapt a statistical perspective in order to check whether this model has a good fit in relation to other estimated models and not merely due to coincidence. To check this, there are two methods:

- For nested models (where one model has more parameters than the other), a Likelihood-Ratio Test is performed (C. Chorus, 2017).
- Non-nested models: Ben-Akiva & Swait test (Ben-Akiva and Swait, 1986)

Standard errors can be used to compute the 95%-confidence intervals (CI), this interval has the 95% probability of containing the true population beta:

$$P(-1.96 \leq \frac{\hat{\beta} - \beta}{SE} \leq 1.96) = 0.95 \tag{4.6}$$

4.4. Mixed-Logit model (ML)

The Multinominal logit model is a relative easy and nice way to estimate choice behavior, however it has some disadvantages when taking it to reality. The MNL-model ignores the correlation within 'nests of alternatives' which are similar in unobserved and/or observed factors. The random errors of the MNL are i.i.d. (independent and identically distributed, they are all drawn and assigned independently), which may be unrealistic. This leads to biased estimation outcomes. Solution: Mixed Logit Model (ML). As a result:

$$cov(\epsilon_{n,MPC-enabledmarketplace1}, \epsilon_{n,MPC-enabledmarketplace2}) \neq 0 \tag{4.7}$$

The i.i.d. error term assumption is the source of the IIA-property (Independence from Irrelevant Alternatives): the relative popularity of two alternatives does not depend on a third one. This doesn't hold for some situations, thus the use of it can lead to flawed results. Adding a constant to the utilities or enriching the specifications of systematic utilities will not help to overcome this problem, as there still would be variation across individuals in terms of unobserved utility.

To sum up, the i.i.d. assumption of logit models (MNL) is invalid if:

- One or more subsets of alternatives share common factors (e.g. lack of privacy)
- Utility associated with these factors varies across individuals (some like, others don't like data sharing due to the risks involved)
- This variation is not fully captured in the systematic utility (correlation between alternatives in terms of unobserved utility)

Solution: an extra error term is added that represents variation of the utility of the common unobserved factors. (note: "DC" stands for Data Control)

$$\begin{aligned}
U_{n,MPC1} &= \beta_{Risk} * Risk_{MPC1} + \beta_{DC} * DC_{MPC1} &&+ \epsilon_{n,MPC1}\\
U_{n,MPC2} &= \beta_{Risk} * Risk_{MPC2} + \beta_{DC} * DC_{MPC2} + \upsilon_{n,decentralised} &&+ \epsilon_{n,MPC2}\\
U_{n,MPC3} &= \beta_{Risk} * Risk_{MPC3} + \beta_{DC} * DC_{MPC3} + \upsilon_{n,decentralised} &&+ \epsilon_{n,MPC3}
\end{aligned}$$

where $V_{n,decentralised}$ ~N$(0$,$\sigma_v)$

The size of sigma reflects the degree of correlation between (unobserved) utilities. It is estimated from the data (it's an extra parameter). If it is zero, then ML -> MNL. It is preferable to always estimate a ML when in doubt about the aforementioned utilities.

Another biased assumption of MNL is that it assumes that tastes are the same within the population, assumed is that there is exactly one taste parameter (beta). However, these tastes will vary. This again ignores correlations between unobserved utilities of alternatives with similar factors. Solution: specify a probability density function for one or more of the betas in addition to the additional error term only:

$$\begin{aligned}
U_{n,MPC1} &= \beta_{Risk} * Risk_{MPC1} + \beta_{DC} * DC_{MPC1} &&+ \epsilon_{n,MPC1} \\
U_{n,MPC2} &= \beta_{Risk} * Risk_{MPC2} + \beta_{DC} * DC_{MPC2} + v_{n,decentralised} &&+ \epsilon_{n,MPC2} \\
U_{n,MPC3} &= \beta_{Risk} * Risk_{MPC3} + \beta_{DC} * DC_{MPC3} + v_{n,decentralised} &&+ \epsilon_{n,MPC3}
\end{aligned}$$

where $V_{n,decentralised}$ ~N$(0$,$\sigma_v)$ and $\beta_{n,Risk}$ ~N$(\beta_{DC}$,$\sigma_\beta)$

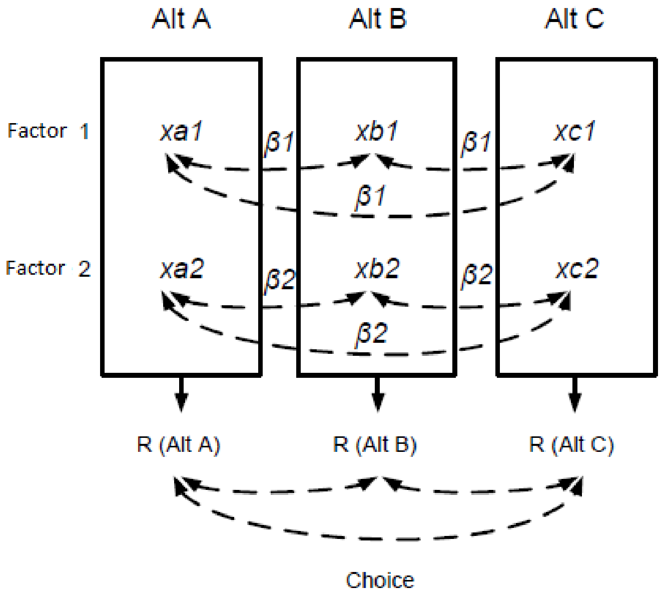
At last, MNL assumes that choices made by the same individual are uncorrelated. However, if a traveller makes T consecutive choices, those choices are correlated as well. If he chooses car at trip t=1, it is likely that will choose car at t=2 as well. What goes wrong is: if every choice is independent of all the others, every case provides an equal amount of information. But the choice at t=2 provides less information than the choice at t=1. So, the model is expected to have more information than it has in reality. The model will assign too much certainty to the estimated parameters so it will underestimate the SE's of parameters which leads to overestimated t-values which can lead to a misconception that parameters might look significant, while they are not. Solution: use ML and thus make the preferences individual-specific. To conclude, MNL is more user-friendly and has a way faster runtime whereas ML is closer to reality but has a way slower runtime. Even billion-dollar investments are often still estimated by MNL due to these reasons.

4.5. Random-Regret-Minimization model (RRM)

One of the most interesting findings from behavioral economics: people care less about absolute factor levels than about relative factors levels. Reference points can be status quo, norms, expectations, peers. Losses loom larger than gains of equal magnitude (C. Chorus, 2017). This model is different than the MNL and ML as it is based on regret minimization (RRM) instead of utility maximization (RUM). The core assumptions:

- People choose the alternative with least regret (C. Chorus, 2012)
- Regret = the sum of regrets associated with binary comparisons with all other alternatives (C. Chorus, 2012)
- Attribute-regret is a convex function of factor difference

Figure 4.1: Comparisons factors in RRM



Achieving regret is assigned more weight than attaining rejoice. Summation over all attributes generates regret. The minimum regret alternative is chosen. Here the generic RRM model is formulated:

$$R_i = \sum_{j \neq i} \sum_{m} \mu * [ln(1 + exp[\frac{\beta_m}{\mu} * (x_{jm} - x_{im})]) - ln(2)] \tag{4.8}$$

Where R_i is the regret of alternative i.
The sum is taken over all competing alternatives within j and over all factors m.
A weight β_m is taken according to the importance of a factor m.
Each performance of factor m is compared to other factors within j.

Different uses of μ:

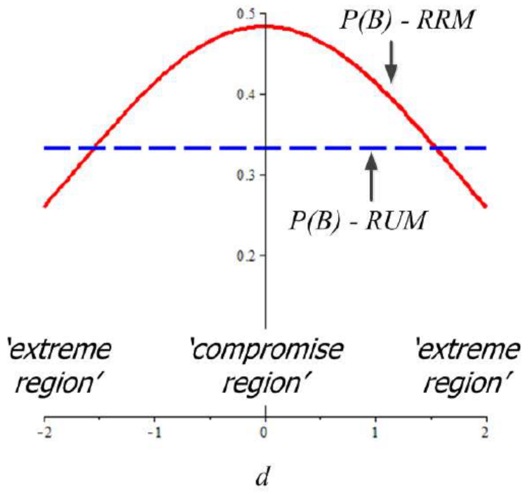
- μ = 0.01 (towards 0), this means no care for rejoice, only for regret (pure-RRM)
- μ = 1 , this is an conventional Random-Regret Model (C-RRM)
- μ = 100 (towards infinity), rejoice and regret are equally important in this case (MNL)

In the end, the probability that an alternative i is chosen over j in a RRM model is:

$$P(i) = P(RR_i < RR_j, \forall j \neq i) = \frac{exp(-R_i)}{\sum_{j=1..J} exp(-R_j)} \tag{4.9}$$

Note: RRM predicts a preference for compromise alternatives as seen in this plot (choice probability of B):

Figure 4.2: Compromise Alternative RRM



This is because being an extreme alternative (very strong performance on some attributes, very poor on others) is inefficient in terms of regret.

When testing if a RRM model explains the obtained data statistically better than a RUM model, we can't use the LRS of the RUM, as this assumes nested models (difference in parameters). Now the goal is to compare non-nested models. Thus, the Ben-Akiva & Swait test is performed:

$$p = NormSDistr(-\sqrt{2 * N * ln(J) * \frac{(LL(RUM) - LL(RRM))}{LL(0)}} \tag{4.10}$$

Where, NormSDist(x) = the probability that draw from standard normal < x
N = Number of observations
J = number of alternatives in the choice set

When interpreting the RRM-parameters: the absolute value of beta gives the maximum increase in regret associated with the comparison of two alternatives, caused by one unit deterioration in the considered alternative's factor. The differences between C-RRM and RUM are often significant but small (C. Chorus, 2012). Much bigger differences become apparent when deviating the μ. To check whether the chosen μ is statistically significant, the difference with 1 is used from the C-RRM model:

$$tratio(\mu \neq 1) = \frac{\hat{\mu} - 1}{SE(\hat{\mu})} \tag{4.11}$$

5 Experimental Design

This chapter will outline the complete experiment design which will be based on a methodology where the goal is on aiming to explore which factors affect the willingness to share on MPC-enabled data marketplaces. Therefore, a high validity and reliability is aimed for in order to lay confidence in these findings. We first start with explication of the participant selection criteria in 5.1. Thereafter, all general platforms used for the data collection (5.2) and the experimental setup (5.3) are defined. The factors will be branched into levels in the stated-choice section in 5.4. This chapter concludes in section 5.5 with an example choice set. The design choices will be shown which are the foundation for the next chapter, namely the data gathering and data analysis.

5.1. Participant Selection

In order to see if perceptions are conform the population, a large group of individuals is asked to participate in the experiment. This way, in the data analysis part of this study, these obtained data can be used to say meaningful things about the population. A large enough sample group of respondents is required to approach the population and improve the validity of the results. The targeted group of respondents in this study consists of adults (18+) whose (driving) data will be hypothetically processed. As the targeted population is very big and contains all kinds of people, incentives (voluntary) response sampling will be used to obtain the sufficient amount of participants. This voluntary response sampling method could lead to bias, as respondents who know more about a subject are more likely to participate in the experiment (Nield and Nordstrom, 2016). However, by rewarding the volunteers, it is aimed to gather a appropriate representation of the population to increase the validity.

The survey development tool Qualtrics will be used to construct the stated choice experiment. This offers a number of advantages. Qualtrics hosts its experiment on own servers and stores these data there. Each noted response by Qualtrics will be able to be placed on own devices for data analysis. Qualtrics satisfies all GDPR privacy-related rules which are set up by the European Union. This strategy is only based on respondents from Qualtrics and Prolific. So no social media platforms will be considered in this strategy as data-heterogeneity is no issue in this study as the target group is not very specific. For the minimum sample size requirements per factor, we follow C. Chorus and Molin (2017) and Bliemer and Rose (2005, p. 11) with the following equation in order to reach a 95% certainty that it is statistically significant. That is:

$$N \geq (\frac{1.96 * se_f(\beta_f^*)}{\beta_f^*})^2 \tag{5.1}$$

$Where:$
$N\ is\ the\ sample\ size$
$se_f\ is\ the\ standard\ error\ for\ factor\ f$
$\beta_f^*\ is\ the\ estimate\ of\ factor\ f$

For a general rule of thumb according to the amount of respondents which are needed for the entire choice experiment, we follow Rose and Bliemer (2013, p. 1024):

$$N \geq 500 * (\frac{L^{max}}{J * S}) \tag{5.2}$$

$Where:$
$N\ is\ the\ sample\ size$
$L^{max}\ is\ the\ largest\ number\ of\ levels\ for\ any\ factor$
$J\ is\ the\ number\ of\ alternatives$
$S\ is\ the\ number\ of\ choicetasks$

In our case (L=3, J=3, S=9) this would results in a minimum of 56 respondents within the sample size to reach significance. The recommended amount of respondents from Qualtrics is 120. This is not expected to be an issue to reach by Prolific.

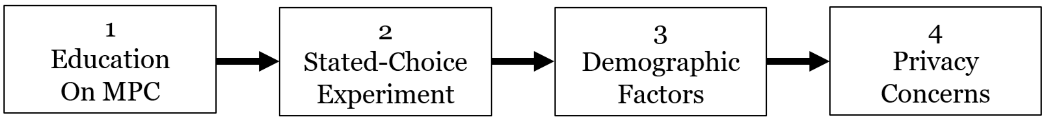
5.2. Gathering of the data

The data collection will be organized by a self constructed survey and choice-experiment by the survey-software Qualtrics[1]. This software tool is supported by the TU Delft (TU Delft, n.d.). To analyze the data, an constructed R-Script will be used to gain information from the raw data. The analysis is done by the software programm RStudio[2]. All respondents will be reached by the platform Prolific[3], which is an platform for participant recruitment. Prolific provides the option to set up filters to have customer pre-screening, an age requirement of eighteen years old will be set. Within this platform, each respondent will be compensated for filling in the survey or experiment truthfully. This platform is expected to be suitable for this study as it offers several advantages. Firstly, the main advantage is the speed in which participants are gathered when the eligibility criteria are broad, the median amount of time needed to reach a N = 100 is around 1.5 to 2 hours (Prolific, 2020). Furthermore, this platform offers a wide reach. Moreover, despite that the conditions of lab and online testing are very different, evidence is growing towards that findings are being comparable. See Crump, McDonnell, and Gureckis (2013) which replicated many experiments on MTurk or Peer, Brandimarte, Samat, and Acquisti (2017) which did similar research on the platforms Prolific and Crowdflower. At last, whereas most university studies are based on samples which are often highly concentrated in the age range 18-23 and often highly educated, Prolific has a older participant pool with a range of education- and employment levels.

5.3. Design

This section elaborates the way the research is set up. Figure 5.1 shows an schematic representation of the process. By collecting data about the backgrounds of respondents and their digital literacy and experience with datasharing, it is optional to analyze whether there exists correlation between people's background and their prioritization in the stated choice experiments. First the education and stated choice experiment is executed in order to have the best attention of the respondents. Thereafter, the demographics and privacy concerns will be asked. It is expected that respondents will take approximately 20 minutes to complete the survey. This section further explicates the factors plus levels. The relevance of the factors is already discussed in section 2.

Figure 5.1: Experiment Process



The education on MPC consists partly of a 3-minute video (initial is 6 minutes) in which the MPC

[1] https://www.qualtrics.com/
[2] https://www.rstudio.com/
[3] https://www.prolific.co/

technology is explained. This video[4] is developed as part of the Safe-DEED deliverable (D2.6) and part of Petronia's (2020) MSc thesis. The video contains a short introduction on the goal of MPC plus a small MPC example to make people more familiar with the technology. The example is about a group of colleagues which want to know who of the group earns the highest salary, without disclosing their own salary. In the example, the respondents are shown how MPC is used as a solution in order to obtain the name of the person earning the highest salary. The representation of MPC in the video is a simplified display of the technique which gives the advantage that people are expected to understand the concept easily. However, as it might be better to educate respondents more extensively about MPC, this could also confuse them. The respondents can watch the video as many times as they desire, this way they can fully understand the concept. After watching the video, the respondents are asked to check the box if they watched the video and are ready to start the choice experiment. This is required to proceed the experiment and this way there is a certain check about whether the respondents understood the concept.

After the video, an textual announcement is made on the stated-choice experiment to introduce the method to the respondents. This announcement also includes the scenario in which the respondents need to make a choice between three different alternatives each time. This first block is estimated to take 7 minutes. The scenario in which people make choices between the alternatives is the following; people are asked to exchange personal car data to help navigation companies improve their services (i.e. road suggestions). By improving their services, consumers will have a better product in the future. The context of helping navigation companies is chosen as this situation is expected to be most familiar to the respondents group, this way we expect the situation to be clear to the respondents to increase validity. Furthermore, by ensuring that the services will be better in the future due to this data sharing, the respondents have an initial incentive to share their GPS data. Shown is, that sharing their car data will be done on MPC-enabled data marketplaces. Beware, in the scenario they share sensitive GPS data and disclosure of these data by malicious parties could have negative effects on the consumers. This is the situation where the respondents are in. This scenario is held constant in each choice situation because if this is not the case, people can make different assumptions and this affects the validity of the experiment in a negative way. See appendix I for the full detailed survey.

n the stated choice experiment, each time, one option with three alternatives is given to the respondents. Three alternatives are chosen in order to gain more information per choice set than in a choice set with two alternatives (C. Chorus, 2017) (see for the justification the example in Appendix C). Each respondent retrieves the same 9 options. Each alternative is a combination of the four factors; risk of data disclosure, data control, trust and benefit. See for example figure 5.3. Every time, respondents need to make a choice on which alternative they would prefer in the scenario of sharing their GPS data to navigation companies. By showing the respondents multiple combinations, the average importance per factor can be pinpointed. This is important to base future decisions on. This part of the survey is estimated to take 5 minutes.

The demographic factors are asked after the experiment. This way the respondents have the best focus on the experiment. The demographic questions are straightforward questions on their gender, age, nationality, employment. Furthermore, people's familiarity with privacy preserving technologies and data marketplaces is asked. These questions are found relevant in explaining relations between respondents and the choices they make in the stated choice experiment of section 5.4. This block is expected to take around 5 minutes.

At the end of the experiment, it is interesting to get a certain sense of the privacy concerns of the respondents in order to link these to respondent-choices in the analysis phase. A. Westin and Interactive has committed themselves over the years 1993 to 2003 on showing people's privacy segmentation and on constructing the core privacy orientation index in order to compare groups of people in their feelings about privacy today. They also constructed an index on online consumer privacy, in which we follow A. Westin and Interactive (1999) in their criteria to derive the privacy index. By doing this, the following statements in table 5.1 will be asked to respondents to anwers on a (1) "strongly disagree" to (5) "strongly agree" scale:

[4] https://www.youtube.com/watch?v=90jcXCHsBF0&ab_channel=Safe-DEED

Table 5.1: Aspects of A. Westin and Interactive's (1999) privacy index.

Year of Study	Criteria used for deriving privacy index
1995-2003 Privacy Segmentation & Core Privacy Orientation Index	(1) Consumers have lost all control over how personal information is collected and used by companies. (2) Most businesses handle the personal information they collect about consumers in a proper and confidential way. (3) Existing laws and organizational practices provide a reasonable level of protection for consumer privacy today.

These three statements in which people need to fill in to what extend they agree or disagree with these statements. The values for fundamentalists, unconcerned and pragmatists were found to be:

- ”Privacy Fundamentalist: At the maximum extreme of privacy concern, Privacy Fundamentalists are the most protective of their privacy. These consumers feel companies should not be able to acquire personal information for their organizational needs and think that individuals should be proactive in refusing to provide information. Privacy Fundamentalists also support stronger laws to safeguard an individual's privacy” (A. Westin and Interactive, 1999).
- ”Privacy Unconcerned: These consumers are the least protective of their privacy - they feel that the benefits they may receive from companies after providing information far outweigh the potential abuses of this information. Further, they do not favor expanded regulation to protect privacy” (A. Westin and Interactive, 1999) .
- ”Privacy Pragmatists: Privacy Pragmatists weigh the potential pros and cons of sharing information; evaluate the protections that are in place and their trust in the company or organization. After this, they decide whether it makes sense for them to share their personal information” (A. Westin and Interactive, 1999).

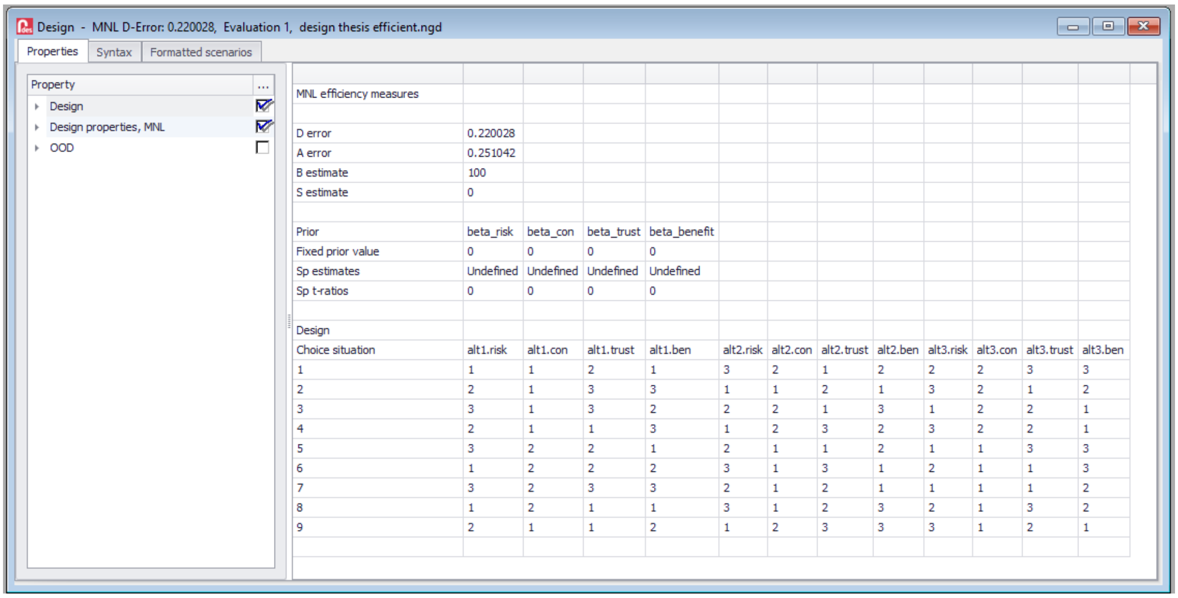
5.4. Stated Choice Experiment

To retrieve which factors are most important to respondents, multiple sets of stated-choice questions are presented in the experiment. Each factor influencing the willingness to share on MPC-enabled data marketplaces has certain levels which are altered between different questions between MPC-enabled data marketplace. The respondents are each question asked, to choose their most preferable MPC-enabled data marketplace on which they would like to share their sensitive automotive GPS data on. This section holds the levels of the factors and the design by Ngene where the choice sets are based on.

For constructing a experimental design, two different methods are found appropriate. Either efficient or orthogonal-designs based on basic-plans. Basic plans are template designs which are set up by mathematicians in order to reduce standard errors and thus have as reliable parameters as possible (Molin, 2017). These basic plans hold likewise properties as efficient designs and are focused on reducing the total number of questions for the respondent. However, these orthogonal designs do not allow for prior values on factors (Molin, 2017). These efficient designs are in technical terms focused on minimizing the D-error which is a measure of co-variance between the factor-levels which are shown to the respondents (Bliemer and Rose, 2005). D-error is a metric that evaluates how effective a design is in retrieving knowledge from experiment participants. The smaller the D-error, the better the design. Hence, by minimizing the D-error, the coherence between choice-tasks will be reduced. In the experiment design phase in Ngene, efficient designs were able in 9 choice-tasks whereas orthogonal designs are possible from 18 choice-tasks per respondent. Less choice tasks per respondent are preferred in this case.

As full-factorial designs, which entail all possible combinations of factor-levels, are way too big for respondents to answer and would lead to respondent-fatigues and eventually to unreliable or invalid estimates - we choose to employ an efficient design to construct various of different scenarios of resulting in different choices. See figure 5.2. Based on mathematical optimization and the possibility for prior estimates on factors, these efficient designs try to make the standard errors minimum (Kuhfeld, 2006). Priors are predefined expected parameters which are based on the expectation of the researcher, in order to ”help” the models estimate the parameters already around that prior. However, as data sharing in this type of situations and by this relatively new technology is quite new, there are no specific priors for the constructed data sharing factors found in literature. Also, we do not give our priors a predefined polarity (+0.01 or -0.01) as these are also not based on literature, and whenever priors are chosen wrongly, one this could affect the factors, the choice probabilities and the model fit in a wrong way (Kuhfeld, 2006).

Figure 5.2: Estimate Efficient Design by software package Ngene



MNL efficiency measures												
D error	0.220028											
A error	0.251042											
B estimate	100											
S estimate	0											
Prior	beta_risk	beta_con	beta_trust	beta_benefit								
Fixed prior value	0	0	0	0								
Sp estimates	Undefined	Undefined	Undefined	Undefined								
Sp t-ratios	0	0	0	0								
Design												
Choice situation	alt1.risk	alt1.con	alt1.trust	alt1.ben	alt2.risk	alt2.con	alt2.trust	alt2.ben	alt3.risk	alt3.con	alt3.trust	alt3.ben
1	1	1	2	1	3	2	1	2	2	2	3	3
2	2	1	3	3	1	1	2	1	3	2	1	2
3	3	1	3	2	2	2	1	3	1	2	2	1
4	2	1	1	3	1	2	3	2	3	2	2	1
5	3	2	2	1	2	1	1	2	1	1	3	3
6	1	2	2	2	3	1	3	1	2	1	1	3
7	3	2	3	3	2	1	2	1	1	1	1	2
8	1	2	1	1	3	1	2	3	2	1	3	2
9	2	1	1	2	1	2	3	3	3	1	2	1

Table 5.2 contains four factors which are deemed relevant for the stated-choice questions regarding data sharing on MPC-enabled data marketplaces, plus their levels. These data sharing factors are found through literature review study. This table corresponds to table 2.3 in section 2.3 where the definition and references are stated. The most appropriate and distinctive factors are included. This is because various factors had overlap or were hard to materialize in the form of levels.

Table 5.2: Factors including levels influencing MPC

Factor	**Levels**
Risk of data disclosure	1. Low: the consumers are exposed to 1 incident in 100 occasions (sales of data). 2. Moderate: the consumers are exposed to 5 incidents in 100 occasions (sales of data). 3. High: the consumers are exposed to 10 incidents in 100 occasions (sales of data).
Data Control	1. The MPC protocol is installed centrally at data marketplaces. Your car data is transferred to the central MPC computation server hosted by data marketplace operator. The computation is performed centrally. 2. The MPC protocol is installed at your car. Your car data stays with you. The computation is performed in your car.
Trust	1. Hardly anyone you know uses this technology when sharing data on data marketplaces. 2. About half of the people you know use this technology when sharing data on data marketplaces. 3. Almost all the people you know use this technology when sharing data on data marketplaces.
Benefit	1. Participants receive no benefit for inputting their automotive data. 2. Participants receive 10 dollar per month for inputting their automotive data. 3. Participants receive 20 dollar per month for inputting their automotive data.

As the application of stated choice experiments in the field of MPC on data marketplaces is very new, the levels for each factor are chosen based on comparisons of course material. For the design of factor levels of *Risk of Data Disclosure*, levels (in terms of percentages) are based on Travisi and Nijkamp (2008) which operationalised environmental accompanied with health risk-levels in the Italian agriculture sector. Hauber et al. (2013) did materialize risk in the same way but in the field of health-care, specifically risk in medicines. As MPC is not yet in operation, there are no relevant sources which state the amount of data disclosures. Therefore, we follow Koch, Krenn, Pellegrino, and Ramacher (2021) who initialized a table which gives an overview about the threats in data sharing (see Appendix **??**). Own assumptions are made about numerating low, medium or high likelihoods of a threat. Data disclosure due to sharing car data by MPC on data marketplaces is assumed to be low if it happens in 1% of the times people share. Medium is assumed to be 5% and high 10% of the time.

For the factor *Data Control*, 2 levels are initialized. This data sharing factors is most related to MPC specifically. Following Archer et al. (2018), MPC is decentralised by design, which means that this relates to a decentralised architecture where the automotive GPS data stays at your own car during the computation process. Nevertheless, a distinction is made between a centralised and decentralised architecture which means that choices have to be made on whether the data is stored and processed on a central server or in their car. This is done to retrieve the consumers' preferences regarding the architecture of MPC. Furthermore, images are included to show the difference between centralized and decentralized data control to respondents. This is done as it is expected to be a complex and hard to comprehend factor to the respondents.

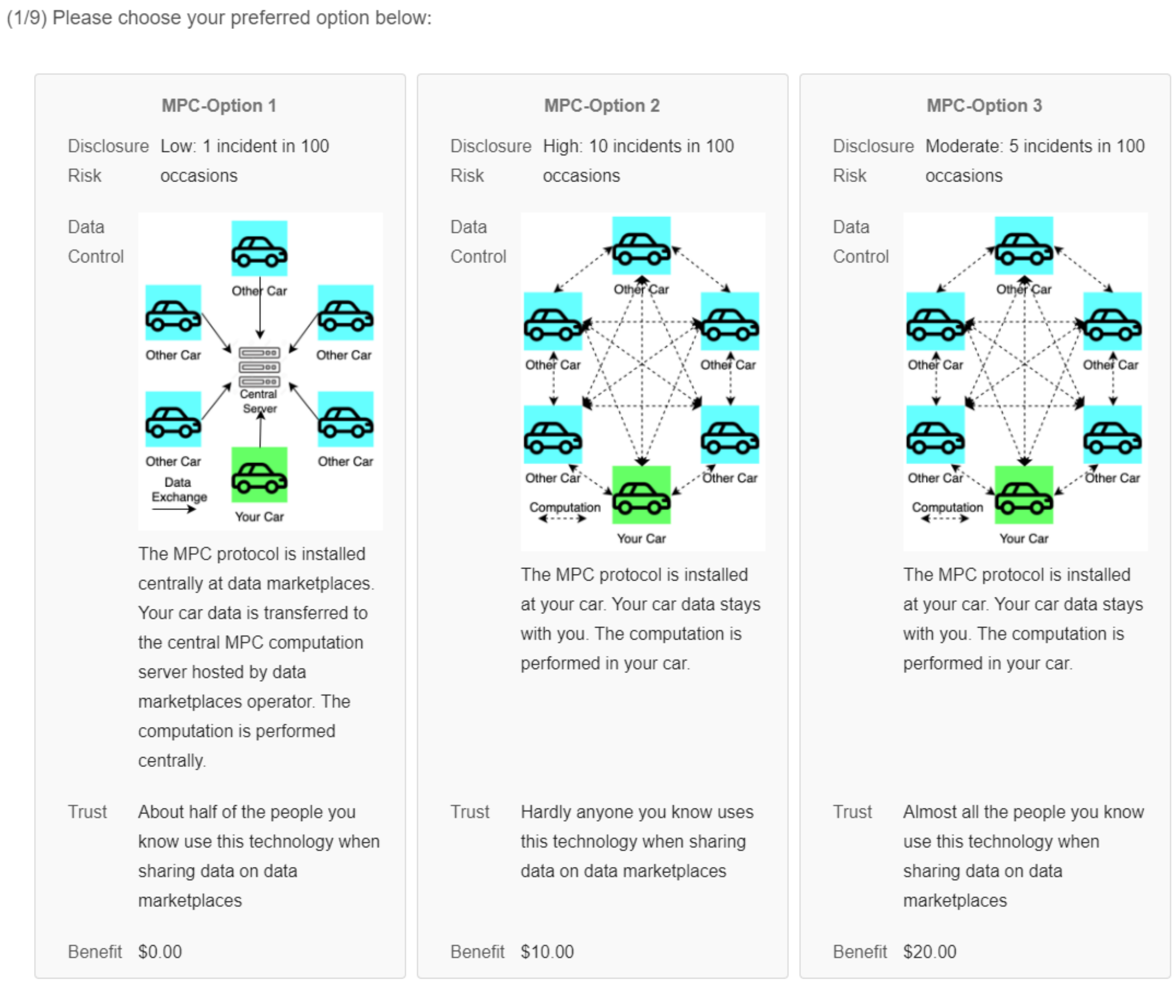
The factor *Trust*. In economic and financial decision-making, herding and imitating may reflect a social learning experience, but this will be regulated by feelings and socio-psychological factors that determine sensitivity to social influence (Baddeley, 2010). The herding effect is based on the assumption that people follow each other when the consequences are preferable and safe. The levels are based on own initialisation as their is little stated choice research done on herding effects in data sharing. Only in stock market decisions.

In the end, the factor *Benefit* is a common factor within many choice experiments and lends itself for measuring people's Willingness to Pay (WtP) for certain increases on other factors. The other way around however, we can show in which degree people are willing to accept an devaluation of a factor in exchange for a (monetary) compensation. In Derikx, De Reuver, and Kroesen (2016), people were willing to shoot privacy concerns on automotive data for an average 9.54€ per month to insurance companies. This is in the middle of our range between zero and twenty dollars, which is preferable as it adds reliability in the estimation of this parameter.

5.5. Conclusion

Current chapter shapes the the fundament of the proposed experiment. The experimental design was build on demographic factors, questions regarding privacy concerns and the stated choice experiment with its efficient design. For the demographics: age, gender, educational level and familiarity of technology are concerned. For the factors: risk of data disclosure, data control, trust (in terms of a herding effect) and benefit are concerned. These are all based on the literature research input of chapter previous chapters regarding data sharing, data marketplaces and MPC. A mathematical efficient design is applied to reduce the number of choices a respondent has to make to 9 choices. This reduces respondent fatigues and aims for more valid results. As mentioned earlier on, Ngene is used to generate the efficient design where standard errors and the required number of choice tasks are minimized. This complete design code is visible in Appendix A and the colorized design in figure 6.7 in the next chapter. Figure 5.3 shows a final example choice task which respondents will fill in. As each single choiceset is different because the alternative-factors deviate in levels, respondents make trade-offs each time between factors and its levels which the model will pick up and will estimate preferences for certain alternatives. The option which is generating the highest utility is chosen which tells our model more and more each choice about the importance of factors. See the following figure for an indication of a choice task. In the end, this completed design allows us to move to the next phase: the data collection phase. The following chapter and sections will outline the data collection and subsequent data analysis.

Figure 5.3: Example Choice task



(1/9) Please choose your preferred option below:

	MPC-Option 1	MPC-Option 2	MPC-Option 3
Disclosure Risk	Low: 1 incident in 100 occasions	High: 10 incidents in 100 occasions	Moderate: 5 incidents in 100 occasions
Data Control	The MPC protocol is installed centrally at data marketplaces. Your car data is transferred to the central MPC computation server hosted by data marketplaces operator. The computation is performed centrally.	The MPC protocol is installed at your car. Your car data stays with you. The computation is performed in your car.	The MPC protocol is installed at your car. Your car data stays with you. The computation is performed in your car.
Trust	About half of the people you know use this technology when sharing data on data marketplaces	Hardly anyone you know uses this technology when sharing data on data marketplaces	Almost all the people you know use this technology when sharing data on data marketplaces
Benefit	$0.00	$10.00	$20.00

Part IV

Completion

6

Results & Data Analysis

The results of the choice experiment are addressed and analysed in this chapter. Beginning with the data-cleaning in the next section 6.1 to ensure reliable and valid data. Thereafter, the demographics are discussed in section 6.2 and the stated choice experiment in 6.3 where all factors are discussed. In the end, multiple conclusions will be drawn in the conclusions section 6.4 where after in the next chapter 7 the overall conclusions will be drawn and the reflection will be held. This chapter, will focus and will answer the second research question.

2: What is the relative importance of factors that influence the willingness of consumers to join and share data in MPC-enabled data marketplaces?

6.1. Data Collection

Via Prolific, the data gathering phase was divided in two stages. The pre-test accompanied by the main test. The pre-test was set up in order to check whether the factor estimates where in the right direction and a check if the estimated time was rightly chosen. Especially, the attribute benefit was checked if it was not disturbing the results. Furthermore, it was a check on whether the Qualtrics conjoint design was applicable as input in the estimation of the logit models. In the following table the two different tests are described.

Table 6.1: Collected data sets

ID	Source	N	Motive	Collection date	Remarks
Pre-test	Prolific	74	Monetary incentive	18 June, 2021	Stopped early at 74/120
Main test	Prolific	428	Monetary incentive	21 June, 2021	Completed

The data collection of the pre-test commenced at 16:32 local time on June 18, 2021, using Prolific. The average monetary compensation per hour was £16.31/hr with 147,008 of 147,944 allowed respondents and 46 open places. We watched the procedure unfold in real time once it was released. During that process, we came to the conclusion that the conjoint design was randomized for each respondent by default. This means, that each respondent saw different combinations of alternatives per choiceset. This could possibly lead to data formatting and model estimation difficulties. Therefore, the process was stopped early to check whether this was fixable. After inspection, it was concluded that the benefit attribute was not disturbing the results and all attributes were of the right direction. However, the estimation was somehow skewed due to the conjoint design and respondents did not need the full twenty minutes to finish the survey on average. In the end the respondent pooling was stopped and 5 responses were destroyed based on time criteria. The time barrier was set on 7 minutes as many respondents filled in the survey within 10 minutes. Only time criteria and bot criteria were considered as there were no right or wrong answers in the stated choice experiment.

After importing the efficient Ngene design, the data collection via Prolific of the main test commenced at 11:53 local time on June 21, 2021. The mean reward per hour was £17.65/hr with 147,731 of 147,942 allowed respondents. Within 2 hours, the 428 responses were reached. This time, different from the pre-test, the same choices were shown to every respondent. This omitted the data formatting and estimation barriers and increased data validity. 21 responses were destroyed based on the time criteria or incomplete response criteria. Therefore, 21 new responses were retrieved which added up to the wanted 428. This response rate is 373% of the minimum response rate required by Qualtrics. This should be a sufficient sample to base the model estimations on and state conclusions about the population. Both collected data sets were merged later on, unfortunately due to the different designs the model fit decreased with 19.2%. Therefore, the main test will be used as primary data to base the results and conclusions on.

6.2. Demographics & Privacy Concerns

After prompting of the respondents for their preferences regarding data sharing by MPC, various demographic questions were asked. Additionally, respondents were questioned to state personal privacy concerns regarding data sharing based on A. F. Westin's (1968) privacy statements. In the next figures, various demographics of the respondents are analysed. Also, in tables 6.2-6.4, all demographics are summarized.

Table 6.2: Demographic characteristics of the respondents (Part A)

Characteristic	N	Percent	Mean	SD	Remarks
Gender (N=428)			1.5	0.510	
Male	216	50.5			
Female	210	49.1			
Other	2	0.5			
Age (N=428)			2.72	0.878	
18 - 24	211	49.3			
25 - 34	147	34.3			
35 - 44	53	12.4			
45 - 54	11	2.6			
55 - 64	6	1.4			
Degree (N=428)			3.85	1.644	
less than High school diploma	14	3.3			
High school diploma	114	26.6			
Some college, no degree	87	20.3			
Associate degree	13	3.0			
Bachelor's degree	120	28.0			
Master's degree	69	16.1			
Doctorate or professional degree	11	2.6			

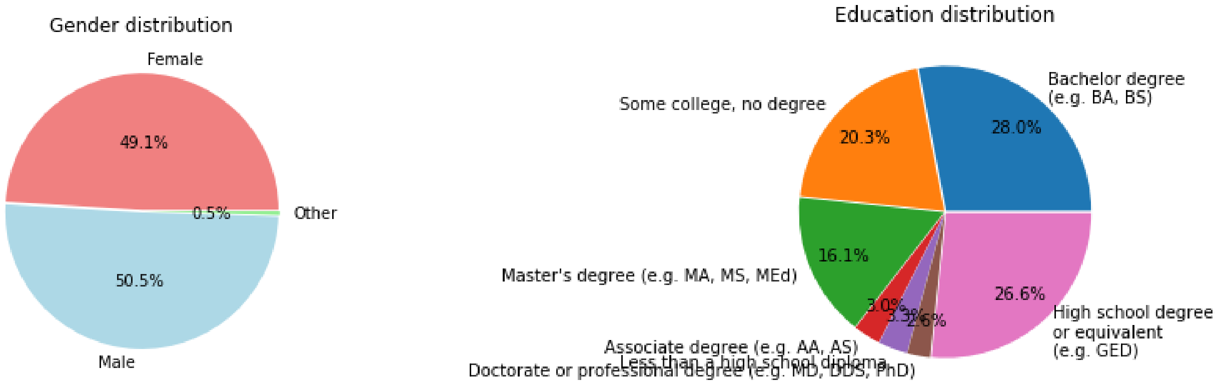
Table 6.3: Demographic characteristics of the respondents (Part B)

Characteristic	N	Percent	Mean	SD	Remarks
Employment (N=428)			3.73	2.31	
Full time (40 hours+)	122	28.5			
Part time	55	12.9			
Other	18	18.2			
Student	184	43.0			
Retired	1	0.2			
Homemaker	6	1.4			
Self-employed	35	8.2			
Unable to work	7	1.6			
Industries (N=428)			12.47	5.12	
Forestry, fishing, hunting	5	1.2			
Real estate	2	0.5			
Mining	4	0.9			
Professional, scientific services	54	12.6			
Utilities	1	0.2			
Management of companies	5	1.2			
Construction	12	2.8			
Admin, support, waste management	11	2.6			
Manufacturing	17	4.0			
Education	50	11.7			
Wholesale trade	2	0.5			
Health care	33	7.7			
Retail trade	28	6.5			
Entertainment and recreation	31	7.2			
Transportation	19	4.4			
Food services	21	4.9			
Information	31	7.2			
Other services	69	16.1			
Finance and insurance	26	6.1			
Unclassified establishments	7	1.6			
Car ownership (N=428)			2.54	1.76	
Yes	196	45.8			
No	84	19.6			
No (but lease/ rental)	15	3.5			
No (but family member)	133	31.1			

Table 6.4: Demographic characteristics of the respondents (Part C)

Characteristic	N	Percent	Mean	SD	Remarks
Experience data marketplaces (N=428)			2.92	1.07	
Shared data on data market for multiple times	75	17.5			
Shared data on data market once	38	8.9			
I know what a data marketplace is, but never shared on it	162	37.9			
Before this survey, I had never heard od data markets	153	35.7			
Familiar with PPT's (N=428)			2.20	1.51	
Yes, I knew about PPT's before this survey	261	61.0			
Due to this survey, I have now an idea of what a PPT is	155	36.2			
No, I still do not quite have an idea of what a PPT is	12	2.8			

On gender, the group is fairly balanced as shown in figure 6.1a. 49% females and 51% males and 0.5% other. The age distribution is skewed towards the younger people, this can be expected as the survey is launched via Prolific which is based on monetary incentives. It can be expected that younger people, especially students are more familiar with these types platforms as older people probably already have a job and do not want to use their spare time to get relatively low extra income. This distribution is not normal. As shown in figure 6.3, most of the respondents are students. Furthermore, in figure 6.4 the educational distribution of the respondents seems normal as the higher degrees (i.e. PhD and MA) are in the minority, high school and bachelor degrees form the middle and around 25% of the respondents has only finished high school or did not aim for a degree. Figure 6.2a shows that this dataset consist of respondents mostly originating in the UK, South Africa, Poland, Portugal, USA, Spain, Italy and Greece.

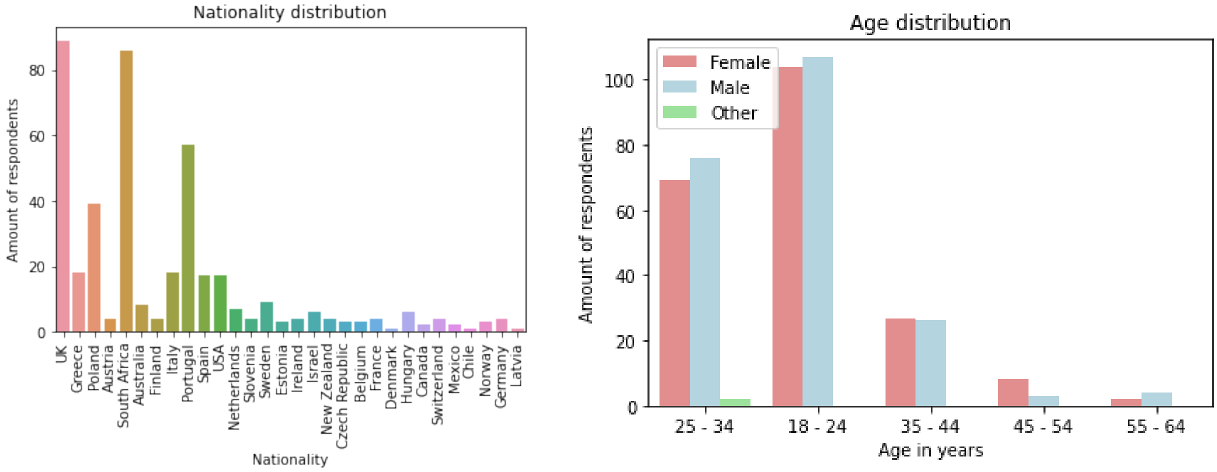


(a) Gender distribution

(b) Education distribution

Figure 6.1: Gender and Education (N=428)

The following figures 6.3 and 6.4 show the employment distribution and the subsequent employment sectors. As mentioned, mostly students and full-timers are among the respondent group. Far most re-



(a) Nationality distribution

(b) Age/Gender distribution

Figure 6.2: Nationality and Age and Gender (N=428)

spondents work in other services. Most respondents work in healthcare, in educational services or are working as professional in scientific or technical services. Thereafter, entertainment, finance, information and retail trade follow. The sector distribution seems skewed to the more technical employment sectors.

Figure 6.3: Employment Distribution (N=428)

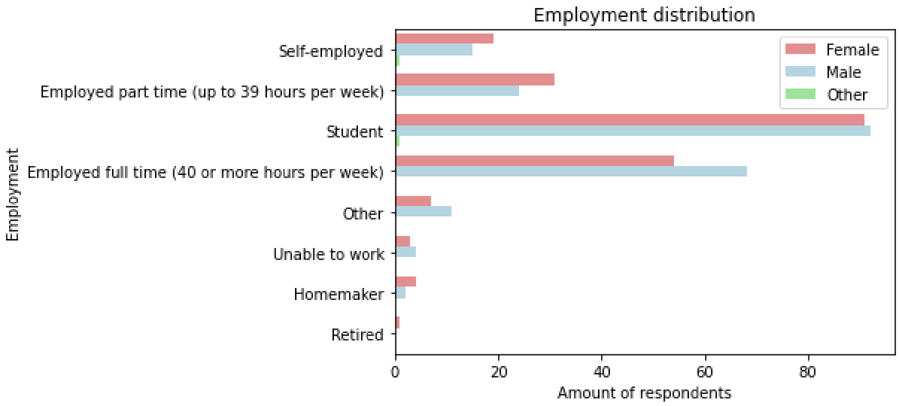
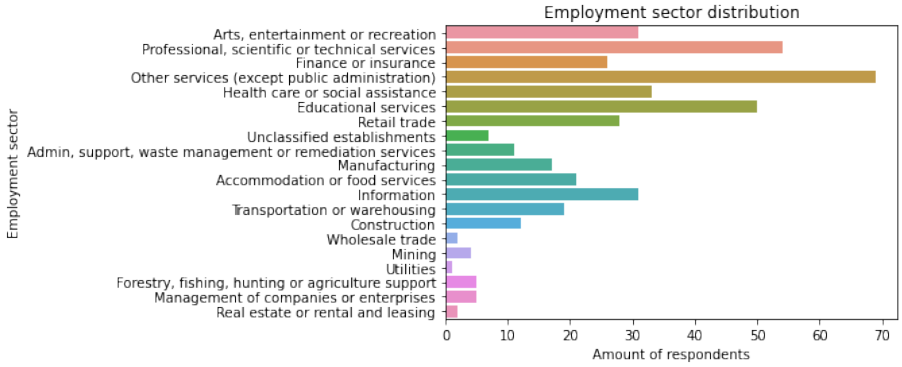
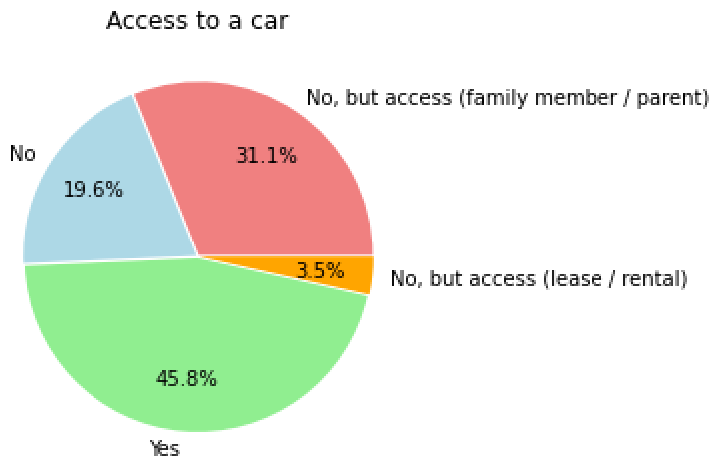


Figure 6.4: Sector Distribution (N=428)

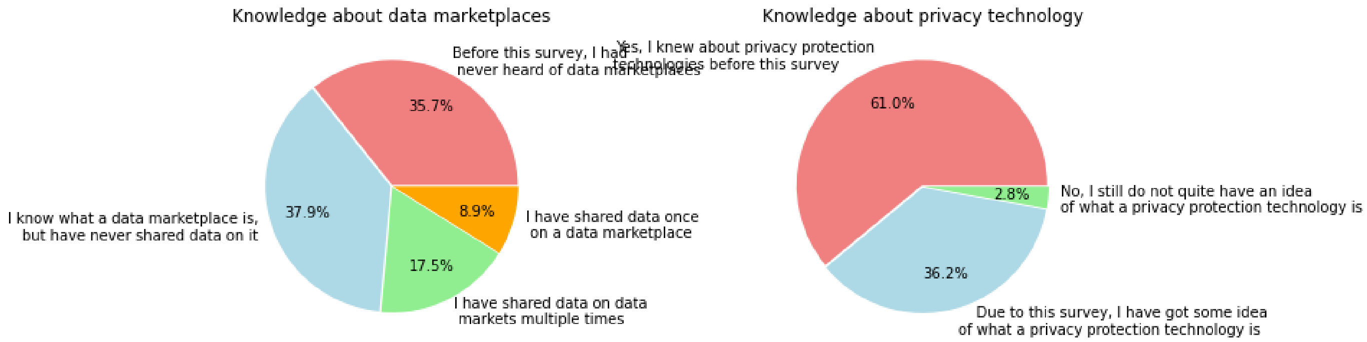


As co-variables for the stated choice experiment, peoples' accessibility to a car and their knowledge about data marketplaces and privacy preserving technologies (PPT) could be explanatory. As shown in figure 6.5, almost half of the respondents have access to a car and around 35% of the respondents have access to a car via family members or via lease. Around 20% has no access to a car at all. This seems normal.

Figure 6.5: Car Access (N=428)



Then, in figure 6.6a it is shown that around a third of the respondents did not know what a data marketplace was before attending this survey. Around 38% did know what a data marketplace was before this survey. In the end, more than 25% says that they've even shared data on data marketplaces before. This combination of people seems far from normally distributed. Perhaps, not all of these respondents did have the same idea of what data marketplaces are. This would mean that in total, around two-third of the respondents already did know what a data marketplace was before attending this survey. This is unreal.



(a) Knowledge about data marketplaces

(b) Knowledge about privacy preserving technologies

Figure 6.6: Knowledge about data marketplaces and PPT (N=428)

Also the distribution of figure 6.6b seems skewed. More than 60% of the people claims to know what privacy protection technologies were before attending this survey. On the other hand, this could be true when looking at the distribution of the working sectors of the respondents.

At the end of the survey, the respondents were shown three different privacy statements where they had to fill in to what extent they agreed or disagreed with these statements. The following figures show the distributions of the three different statements, divided over each gender. In table 6.5, the distribution of agreement is shown.

Table 6.5: Privacy Statements

Statement	**Strongly Disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**	**Mean**	**Standard deviation**
Consumers have lost all control over how personal information is collected and used by companies	1.6%	12.4%	14.0%	52.3%	19.6%	3.76	0.961
Most businesses handle the personal information they collect about consumers in a proper and confidential way	7.7%	34.6%	31.1%	20.6%	6.10%	2.83	1.04
Existing laws and organizational practices provide a reasonable level of protection for consumer privacy today	4.9%	18.9%	36.4%	29.7%	10.0%	3.21	1.02

Multiple one-way analysis of variance (ANOVA) tests were conducted in order to determine whether there were any statistically significant differences between the means of three or more independent (unrelated) groups on the privacy statements (see Appendix B for explicit analyses). The following results were found (p<0.05):

- Older people agree statistically more on privacy statement 1, that consumers lost all control over their data (p=0.048).
- Higher educated people disagree statistically more on privacy statement 2, that businesses handle the consumer personal data in a proper and confidential way (p=0.049).
- People with more experience on data marketplaces agree statistically more on privacy statements 2 and 3. That businesses handle consumer personal data in a proper and confidential way and that the existing laws and organizational practices provide a reasonable level of protection for consumer privacy today (p=0.014 and p=0.032).
- No significant differences were found between genders, working sectors, people with different PPT knowledge or between different types of car ownership.

Furthermore, we followed Kumaraguru and Cranor (2005) who showed that Westin (1999) re-coded the respondents and their privacy statements into three different groups in order to analyze whether these groups make different decisions regarding the conjoint analysis when it comes to risk, data control, trust and benefit. The following groups were found within the dataset:

- **Privacy Fundamentalist** (15.7% of N=428): respondents who agreed (strongly or somewhat) with the first statement and disagreed (strongly or somewhat) with the second and third statements.
- **Privacy Unconcerned** (6.5% of N=428): respondents who disagreed with the first statement and agreed with the second and third statements.
- **Privacy Pragmatists** (77.8% of N=428): all other respondents.

In section 6.3.2, the interaction effects of these groups with the four different data sharing factors will be addressed on the basis of MNL estimation.

6.3. Stated Choice Experiment

This section presents the analysis of the acquired dataset through the stated choice experiment. First, the chosen analytical model will be described, along with a few observations on data manipulation. The findings of the experiment for data sharing via MPC will next be analyzed.

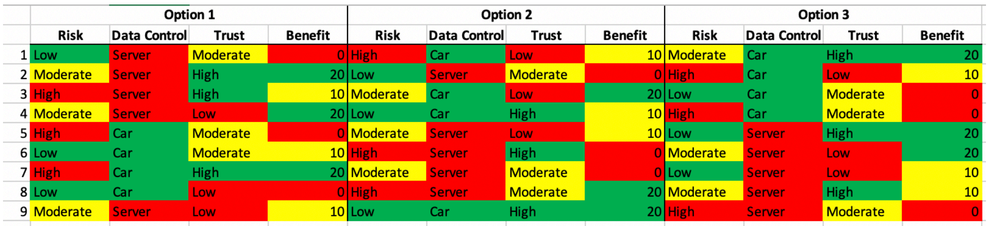
There are several points to be made about the data handling and the used analyses models. The dataset was analyzed using an MNL-style model, which is already explicated in 4.3. As Train (2009) shows, this model enables the researcher to extract trade offs and evaluate the relevance of each factor in the dataset. Thereafter, relative values can be determined from utility-variations resulting from altering attribute levels. These advantages flow through to the study's objectives, allowing researchers' preferences to be pinpointed. While logit models like MNL cannot reflect taste variation, they can't manage scenarios when unobserved factors are correlated across time (Train, 2009). Mixed Logit (ML) models however, which hold a variable capturing taste heterogeneity, can help to reduce correlation. However, due to the absence of nests of (labeled) alternatives, and the need to generate a multitude of discipline-specific examples, ML could not be used to its full potential. This is shown in 4.4. In the end, these comparisons between the models are performed to the rank models after data analysis and to conclude about validity and reliability measures which will show which model to use in the conclusion section. The most significant and best performing model will thus be used.

Table 6.6 offers a summary of the given answers of the processed choice experiment. Each choiceset was based on an initial design, which showed that each option differed. This means, that each choiceset, each alternative was different. Therefore, the choice probabilities of table 6.6 without the initial design gives little information because it is an unlabeled experiment. Each alternative has the same name and just a descriptive number to diversify the different alternatives. Therefore, the design from section 5.4 is presented in a colorized (green, yellow, red) manner in figure 6.7 beneath, to give some meaningful insights. Here, green is preferable and red is not preferable. By quick inspection, it is seen that alternative 2 in choice set 9 is very dominant and chosen by 84% of the respondents due to its good attribute levels. Unfortunately, this choice set contributes little information due to its dominance.

Table 6.6: Distribution of given DCE choices (N=428)

Choiceset	Alternative 1	Alternative 2	Alternative 3
1	19%	14%	**67%**
2	**57**%	23%	20%
3	19%	**47%**	34%
4	25%	**70%**	5%
5	11%	13%	**76%**
6	**67%**	9%	24%
7	29%	18%	**53%**
8	35%	14%	**51%**
9	11%	**84%**	5%

Figure 6.7: Colorized Design



	Option 1				Option 2				Option 3			
	Risk	Data Control	Trust	Benefit	Risk	Data Control	Trust	Benefit	Risk	Data Control	Trust	Benefit
1	Low	Server	Moderate	0	High	Car	Low	10	Moderate	Car	High	20
2	Moderate	Server	High	20	Low	Server	Moderate	0	High	Car	Low	10
3	High	Server	High	10	Moderate	Car	Low	20	Low	Car	Moderate	0
4	Moderate	Server	Low	20	Low	Car	High	10	High	Car	Moderate	0
5	High	Car	Moderate	0	Moderate	Server	Low	10	Low	Server	High	20
6	Low	Car	Moderate	10	High	Server	High	0	Moderate	Server	Low	20
7	High	Car	High	20	Moderate	Server	Moderate	0	Low	Server	Low	10
8	Low	Car	Low	0	High	Server	Moderate	20	Moderate	Server	High	10
9	Moderate	Server	Low	10	Low	Car	High	20	High	Server	Moderate	0

6.3.1. Multinominal logit model (MNL)

The first model is the basic MNL model. Only those attributes are shown that were used in the choice experiment. This MNL model, is the most basic function for expressing a decision model's usefulness. In first place, only linear attributes were considered. This model can be compared based on the adjusted rho-square value with different models as ML or RRM. The base MNL-function which represents the total utility per alternative used in the remainder of this research is as follows:

$$Alternative_1 = \beta_{RISK} * RISK_1 + \beta_{DC} * CONT_1 + \beta_{TRUST} * TRUST_1 + \beta_{BENEFIT} * BEN_1$$
$$Alternative_2 = \beta_{RISK} * RISK_2 + \beta_{DC} * CONT_2 + \beta_{TRUST} * TRUST_2 + \beta_{BENEFIT} * BEN_2$$
$$Alternative_3 = \beta_{RISK} * RISK_3 + \beta_{DC} * CONT_3 + \beta_{TRUST} * TRUST_3 + \beta_{BENEFIT} * BEN_3$$

The full Ngene syntax can be found in Appendix A. Now the basic model characteristics and model fit will be shown in table 6.7. This also includes the parameter estimates (betas) and subsequent significance levels. Furthermore, the utility contribution graphs and the relative importance of factors will be shown in Appendix H. Also, results on the willingness to pay (WtP) will be discussed.

Initially, an effort has been made to fit a basic statistical model in order to obtain understanding about respondents' attitudes for sharing data by MPC-enabled data marketplaces. That is, each and every response, as well as each attribute, was analyzed. Given the huge number of replies received, it is regarded a necessary step in the investigation. A variety of analytical indicators are used to determine whether model outputs are significant and the model as a whole is doing well. Factor and model importance are measured using a variety of metrics. As mentioned in the method section, significance indicators (p and t) indicate the level of significance per factor. In this research, by significant, meant is 5%. This means that t>1.96 and p<0.05. As all factors are generalized and do not differ per alternative, 4 factors are estimated. For the model fit, the LRS and Mc Fadden's rho squared are used which are also both explicated in the method chapter. The full model is shown in Appendix D.

Table 6.7: Multinominal Logit Model (MNL)

	Beta Estimate	Standard Error	t-value	p-value
β_{RISK}	-0.69036	0.025414	-27.164	0.000
β_{DC}	0.34476	0.039552	8.717	0.000
β_{TRUST}	0.25173	0.022341	11.268	0.000
$\beta_{BENEFIT}$	0.05092	0.002440	20.869	0.000
Number of observations	3852			
0-Loglik	-4231.855			
Final-Loglik	-3351.745			
LRS /w 0-Loglik	1760.22			
Mc Fadden's rho-squared	0.208			
AIC	6711.49			
BIC	6736.52			

Table 6.7 shows the results of the MNL model. Here are the results:

1. The model's estimations do not exhibit significance issues. All attributes are found to be highly significant as all p-values < 0.05 and all t-ratios are > 1.96. To put it another way, it's highly implausible that the population's attribute effects are equal to zero. To clarify, the dependent variable in the choice model is the observed respondent's choice behavior. This model exceeds the null-hypothesis with a significance of LRS = 1760.22 and p = 0.00. Rho-squared value is 0.208. Values between 0.2 and 0.4 indicate a good model fit (McFadden et al., 1973). The AIC and BIC values indicate the model's focus on mitigating information loss (Hauber et al., 2013).

For AIC and BIC, low values are aimed for. However, AIC is useful for comparing models, but it does not tell anything about the goodness of fit of a single, isolated model. By estimating the next ML model, this AIC and BIC value will tell us something.

2. It is shown that the risk of data disclosure attribute is most important to people in general when making decisions. Benefit is the second most important attribute, then trust and at last data control. The importance is related to the utility difference of the lowest and highest attribute-values, called the range. Therefore, not to confuse benefit with trust. Benefit's lowest level (0$) has 0 utility, and the highest (20$) has 1.0184. Trust can have an higher estimate, however the utility difference is 0.50346. Furthermore, all factor-polarities do agree with the expectations. Only Risk is affecting people's utility negatively, even more when it increases. See Appendix H for detailed utility graphs and scores.
3. The optimal package, which is the highest preferred bundle across respondents and maximizes their preference and utility, is a combination of low risk of data disclosure, data control in own car, everybody has trust in the application and for uploading the car data they aim to receive 10$ per month.
4. Respondents are willing to pay 9.50$ or to earn less) on a monthly basis to reduce their level of risk of data disclosure from moderate to low. Or 29.90$ to reduce their risk of data disclosure from high to low. This relation is thus not linear.
5. Respondents are willing to pay (or to earn less) 6.90$ on a monthly basis to keep their data for MPC processes in their own car.

To summarize, this model has a good fit which means that based on these estimated parameters the model makes the data most likely to be true. The parameters are in the right directions and are all significant which is good. This means that the model explains the data significantly better than a model which is based on random choices.

6.3.2. Multinominal logit model with privacy groups (MNL + interaction)

The following model is based on the previous basic MNL model, but now with interaction effects included. The interaction effects are based on attribute estimate differences between privacy fundamentalists (coded ”3”), privacy pragmatists (coded ”2”) and the privacy unconcerned people (coded ”1”) which are coming from . The full model syntax is shown in Appendix E.

Table 6.8: Multinominal Logit + Westin interaction Model (MNL + interaction)

	Beta Estimate	**Standard Error**	**t-value**	**p-value**
β_{RISK}	-0.766889	0.118045	-4.2252	1.193e-05
β_{DC}	0.489044	0.183633	2.0016	0.022663
β_{TRUST}	0.442638	0.103825	3.3120	4.6314e-04
$\beta_{BENEFIT}$	0.048018	0.011333	2.7786	0.002730
int$_{RISK}$	0.036342	0.054887	0.4240	0.335781
int$_{DC}$	-0.068668	0.085416	-0.6045	0.272751
int$_{TRUST}$	-0.090892	0.048206	-1.5112	0.065371
int$_{BENEFIT}$	0.001379	0.005269	0.1747	0.430663
Number of observations	3852			
0-Loglik	-4231.855			
Final-Loglik	-3349.011			
LRS /w 0-Loglik	1762.22			
Mc Fadden’s rho-squared	0.2086			
AIC	6714.02			
BIC	6764.07			

Table 6.8 shows the results of the interaction-MNL-model. Here are the remarks:

1. All main parameters are still significant, however due to the introduction of the interaction effects they are less significant than in the basic model.
2. It is shown that all standard factor estimates are somewhat smaller than the estimates of the standard MNL model, this can be explained due to the introduction of the interaction effects which pick up a little of the systemic utility of the parameters estimates of the basic effect (C. Chorus, 2017).
3. The model fit is not significantly increased. In other words, by addition of the interaction effects, they do not explain the choices better which are made by respondents. Which means, as they are insignificant, that there does not exists a significant difference between these groups in valuing different data sharing factors.
4. The interaction effects are all not significant. Int_{TRUST} is close to significance (p=0.065). It can be stated that with 90% certainty, the westin grouping variable affects people’s decision making regarding the trust variable within alternatives. Fundamentalists (3) have the lowest utility in the trust variable. In addition, even while int_{RISK} and $int_{BENEFIT}$ are not significant, fundamentalists attach the highest utility on risk and benefit.
5. It is conform the expectations that the sign for int_{DC} and Int_{TRUST} are negative. This means that decentralised data control and high trust are least important to privacy unconcerned people and privacy pragmatists and most important to respondents seen as privacy fundamentalists.

To summarize, as all interaction effects are not significant and the main parameters are less significant as before, it seems better to base the decisions still on the basic MNL model in explaining the data. Let us move to the ML model.

6.3.3. Mixed-Logit Model (ML)

The previous MNL builds on the assumption that there is no taste-heterogeneity between people, which means that people all have the same taste regarding attributes of alternatives. In each, that all people like monetary benefits with the same magnitude and nobody dislikes it. This is unreal, therefore, the second estimated model is the ML model where attributes are based on distributions. Only those attributes are given that were used in the choice experiment. This model can be compared based on the adjusted rho-square value with different models as MNL or RRM. The full model syntax is shown in Appendix F. Table 6.9 shows the ML-estimates which are used in the remainder of this research:

Table 6.9: ML model estimates

	Beta Estimate	Standard Error	t-value	p-value
β_{RISK}	-1.26489	0.069933	-18.087	0.000
β_{DC}	0.44651	0.090536	4.932	6.722e-07
β_{TRUST}	0.47216	0.050470	9.355	0.000
$\beta_{BENEFIT}$	0.09124	0.006165	14.800	0.000
σ_{RISK}	0.96936	0.064806	14.958	0.000
σ_{DC}	1.51249	0.100607	15.034	0.000
σ_{TRUST}	0.76970	0.051689	14.891	0.000
$\sigma_{BENEFIT}$	0.09459	0.006265	15.099	0.000
Number of observations	3852			
0-Loglik	-4231.855			
Final-Loglik	-2912.14			
LRS /w 0-Loglik	2639.43			
LRS /w MNL	879.21			
Mc Fadden's rho-squared	0.3119			
AIC	5840.28			
BIC	5890.33			

Table 6.9 shows the results of the ML model:

1. This model does also not exhibit significance issues. All factors are found again to be highly significant as all p-values < 0.05 and all t-ratios are > 1.96. This model outperforms the null-model and the MNL model with a significance (LRS = 2639.43, p = 0.00) and (LRS = 879.21, p = 0.00) respectively. Rho-squared value is 0.3119 which is a huge increase in model fit in perspective to the basic MNL model. Now that the second model is estimated, MNL and ML can be compared based on the AIC and BIC values which indicate the model's focus on mitigating information loss (Hauber et al., 2013). These metrics also hugely decreased. To conclude, there exists heterogeneity in attribute taste among people because this estimated model is significantly stronger than the null-model and the MNL model.
2. First of all, it is shown that all parameters are bigger compared to MNL. This comes due to allocation of estimates from error term to sigma of the estimate. The ranking of the estimates is still the same. The importance is related to the utility difference of the lowest and highest attribute-values. This means that Risk of data disclosure stays the most important factor, followed by benefit, trust (in terms of a herding effect) and data control.
3. The optimal package, which is the highest preferred bundle across respondents and maximizes their preference and utility, is a combination of low risk, data control in own car, everybody has trust and for uploading the car data they aim to receive 10$ per month.

This ML model explains the data best which shows that there is taste heterogeneity among the respondents and in this case thus within the population. This model outperforms both previous MNL models. We will use this model in the conclusions section to base our statements on.

6.3.4. Random Regret Minimization Model (RRM)

As last, we test the RRM model which is based on regret minimization instead of utility maximization to check whether this model explains the data better than the ML model. The full model syntax is shown in Appendix G. This model has thus a preference for compromise alternatives, thus it is interesting to see if this characteristic better explains the choices made by the respondents compared to the ML model.

Table 6.10: RRM model estimates

	Beta Estimate	**Standard Error**	**t-value**	**p-value**
β_{RISK}	-0.47181	0.017488	-26.978	0.000
β_{DC}	0.24887	0.026872	9.261	4.862e-11
β_{TRUST}	0.16771	0.014684	11.421	0.000
$\beta_{BENEFIT}$	0.03426	0.001644	20.832	0.000
Number of observations	3852			
0-Loglik	-4231.855			
Final-Loglik	-3340.851			
LRS /w 0-Loglik	1782.008			
LRS /w MNL	21.788			
LRS /w ML	857.422			
Mc Fadden's rho-squared	0.2105			
AIC	6689.7			
BIC	6714.73			

Table 6.10 shows the output of the RRM model:

1. No significance issues are found. All factors are found again to be highly significant as all p-values < 0.05 and all t-ratios are > 1.96. This model exceeds the null-model with a significance LRS = 1782.008, p = 0.00. The MNL model with a significance of LRS = 21.788, p = 1.5225E-06 (Ben-Akiva and Swait, 1986). Rho-squared value is 0.2105 which is a relative small improvement in model fit compared to the basic MNL model and this is a huge decrease in model fit compared to the ML model.
2. First of all, it is shown that all parameters are smaller than MNL. This comes due to allocation of estimates from error term to sigma of the estimate. According to C. G. Chorus (2010), RRM parameters are often twice as close to zero as their utilitarian counterpart model (MNL). The ranking of these estimates is still the same. The importance is related to the utility difference of the lowest and highest attribute-values which again means that risk of data disclosure is most important, then benefit, the trust followed by data control.
3. Whereas Zeelenberg and Pieters (2007) state that regret is experienced in situations where difficult and important decision are made, this survey could be not of this league. Moreover, whereas Simonson (1989) identified that regret occurs when there is the existence of compromise-effects in choice situations and where decision makers need to explain their choice to others, respondents where not obliged to justify their choices in this experiment. Therefore, these choices could not have led to regret which makes this RRM model not the best model to base the ranking of the data sharing factors on and their relative importance when compared to the better explaining ML model.

6.4. Conclusions

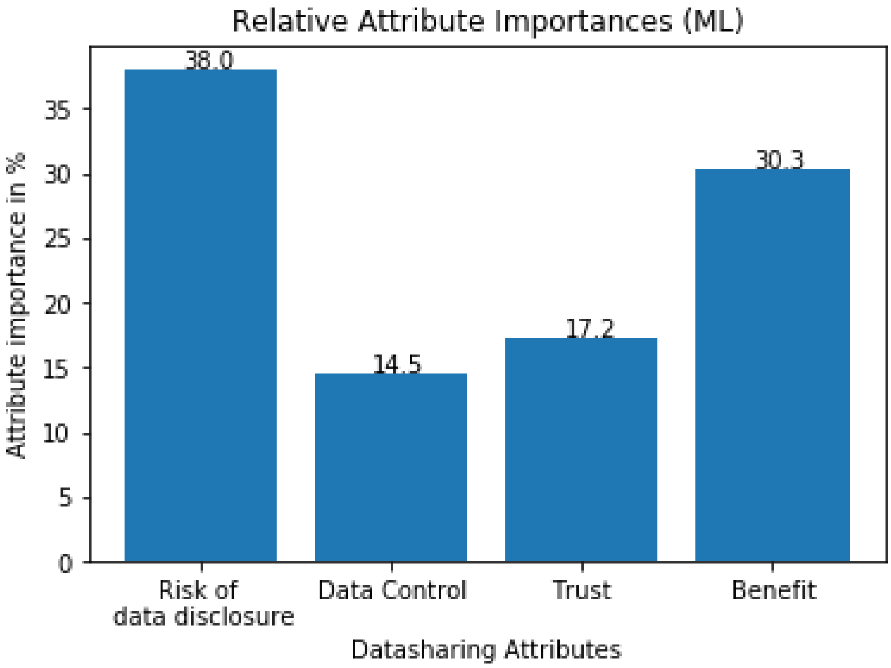
The results of the mixed logit model imply that the ML model performs better than the null-, MNL-, MNL with interaction and RRM model. Therefore, the outcomes of the ML model are applied to interpret and predict respondents' willingness to share data via MPC-enabled data marketplaces. Important to conclude, this respondent group over represents younger people which study so the conclusions cannot be generalized to the whole population in general. Furthermore, the sample seems over represented in the professionals in scientific or technical services, the educational services, and information technology - which may be the reason that most respondents are familiar with data marketplaces and know about privacy preserving technologies. Moreover, the respondent group is quite similar distributed like A. F. Westin's (1968) study on different privacy concerned groups but this study showed that their decisions regarding automotive data sharing on MPC-enabled data marketplaces are not different. In the following table 6.11, all models are summarized to have an overview about the model fits.

Table 6.11: Model Fit of different models

Model	**Number of parameters**	**Mc Fadden's Rho-square**	**Final log-likelihood**	**LRS /w Null model**
Null model	0	0.00	-4231.855	
MNL	4	0.208	-3351.745	1760.22
MNL + interaction	8	0.2086	-3349.011	1764.45
ML	8	0.3119	-2912.14	2639.43
RRM	4	0.2105	-3340.851	1782.008

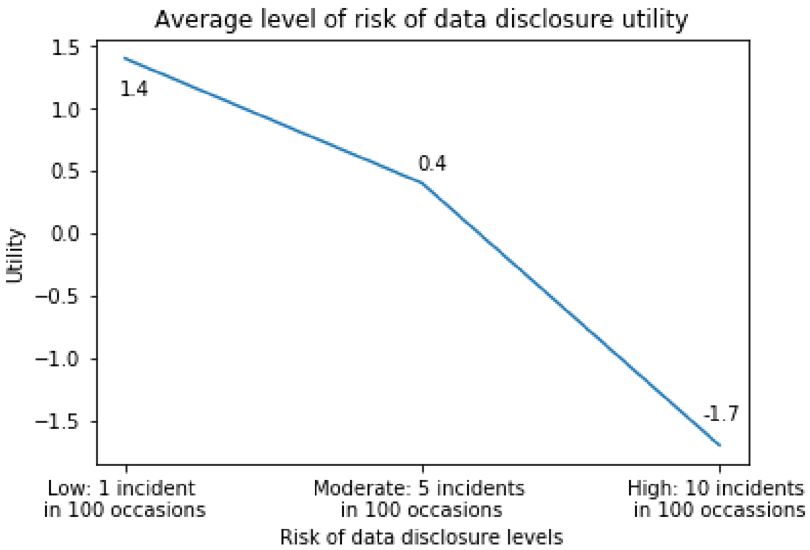
The following figure 6.8 shows the relative importance of the attributes as percentages. In other words, the average measurement of influence an attribute had when the respondents were choosing their preferred alternative. The higher the score, the more weight it carried in the decision-making process (the scores add up to 100%). Each attribute will be discussed, beginning from most important to least important.

Figure 6.8: Relative Attribute Importance (ML)



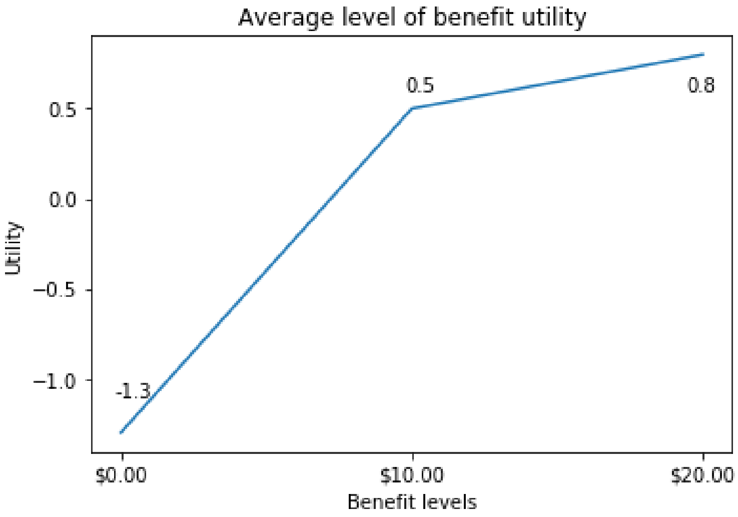
The risks estimate (-1.2649) a consumer is exposed to by sharing data on data marketplaces will influence the decision to share on a data marketplace the most. This attribute has been found as being the most crucial based on the maximum likelihood principle on the filled-in choice data. Relatively, risk is good for 38.0% of the total importance. Important to say, every model showed that risk is the most important factor. Too high risks of data disclosure will eventually lead to discontinuation of sharing. It is therefore very important to have the level of security within MPC on-point.

Figure 6.9: Risk utility (ML)



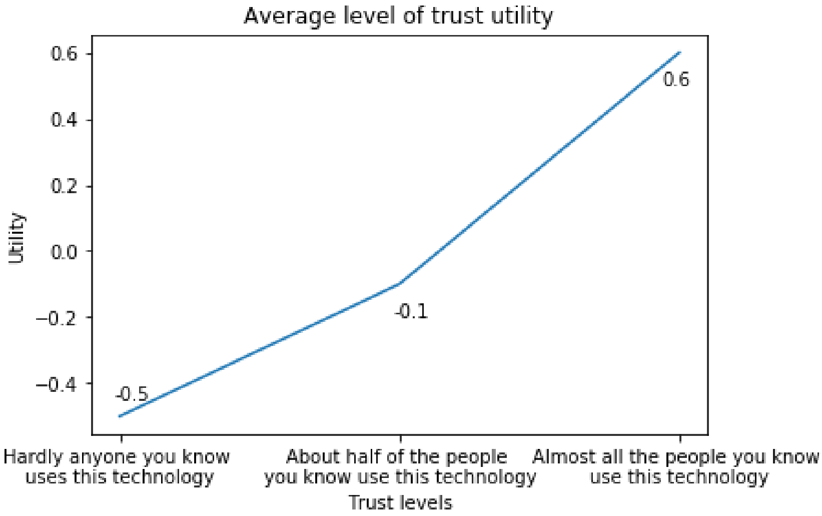
The benefits (0.0969) that the consumer perceives to sharing automotive car data influences the decision to participate as second most important. Almost a third (30.3%) of the choice for a specific platform is based on the amount of benefit people receive. However, as shown in figure 6.10, after reaching 10 dollars of benefit per month, the curve flattens and people experience not as much additional utility for additional monthly benefit. As this attribute is perceived as very important to respondents, it is important to have a certain benefit factor incorporated in MPC-enabled data marketplaces.

Figure 6.10: Benefit utility (ML)



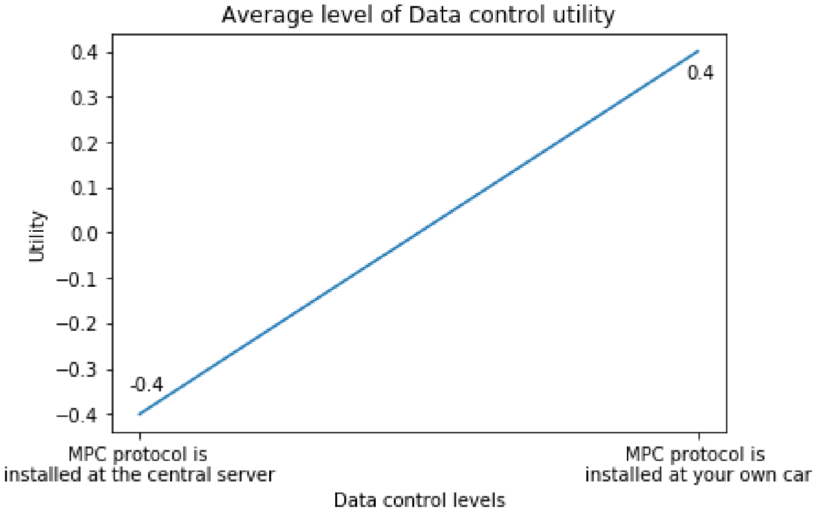
General trust (0.4722) in the MPC technology and MPC enabled data marketplace is weighted third most important. The choices are for 17.2% based on trust. Trust in this research is mostly based on other people having a certain feeling about MPC or do already use it. A herding effect. This has causality with the certain mass of demand a platform needs in order to attract more people. From this experiment it is shown that it is still relatively important that people attract other people in the technology. Moreover, the effect is stronger whenever more people are sharing their automotive data using MPC-enabled data marketplaces as the average level utility is not fully linear as shown in figure 6.11.

Figure 6.11: Trust utility (ML)



At last, data control (0.4465) is perceived least important to people. 14.5% of a choice is based on the amount of data control people can still have, while inputting their data by MPC on data marketplaces. It was not expected to be the least important factor, even less important than the herding effect. People like to have their data stored in their own car, but it is not that important relative to the other factors as risk or benefit. Furthermore, as this variable only has two levels, linearity cannot be tested as seen in figure 6.12.

Figure 6.12: Data control utility (ML)



7 Conclusion

The final reflections on both the study process and the important findings are found in this chapter. In this chapter, the significant findings are condensed into a set of concluding remarks (in 7.1) that provide a good overview of the answers to the main study questions. This study's societal and theoretical consequences are also examined in 7.2 and 7.3. Here, one may see how the main results add to the relevant literature and how they might affect society. Finally, the limits of this study are highlighted in the discussion section in 7.4, followed by the future research options in 7.5.

7.1. Answers to the main research questions

The goal of this study was to learn more about people's willingness to contribute automotive data on MPC-enabled data marketplaces. Before diving into the research's major conclusions, it's important to clear up any confusion about what MPC factors mean in the context of this study.

> *1: What factors drive consumers' choices regarding sharing automotive data on MPC-enabled data marketplaces?*

During the literature reviews, it was discovered that a wide range of factors were thought to influence information exchange. However, most of these factors were not concrete enough or had measurement issues in generalizing these among different MPC-enabled data marketplaces. Furthermore, because of the contextual nature of the this research question, the factors that could be tested had to be reduced down to those that had to do with decision-making. This method revealed the necessity to develop a set of data-sharing incentives and barriers on MPC-enabled data marketplaces. In this research in the automotive sector, it was chosen to include the factors 1) risk of data disclosure, 2) data control, 3) trust (in terms of a herding effect) and 4) benefit. These factors were based on literature research and the following the criteria: relevancy, measurability, and generalisability, because aimed was on uses experience regarding data sharing on MPC-enabled data marketplaces and the discrete choice modeling approach based on standard consumer choices asks for an understandable and clear design.

> *2: What is the relative importance of factors that influence the willingness of consumers to join and share data in MPC-enabled data marketplaces?*

he stated choice experiment showed that sharing of automotive car data on MPC enabled data marketplaces is prone to a variety of internal and external factors. The model with a rho-square of 0.3119 did explain a decent amount of choices. However, as the model did not get close to 1, this means that there still exists unexplained variance which can be explained by additional or different factors. These factors are not concrete and were hard to include in the choice experiment. Either due to complexity or due to being too ambiguous. Nevertheless, for the included factors in all models, all factors appeared to be relevant in decision making between MPC-enabled data marketplaces. As expected, risk of data disclosure affected the utility negatively and all other factors affected the utility in a positive way. Risk of data disclosure was found to be the most important factor among respondents, followed narrowly by benefit. Trust in terms of a herding effect was thirdly important and data control is found to be least important. Furthermore, it is shown that there exists taste heterogeneity between people regarding the four included factors as shown by the ML model by the estimated factor sigmas. In the end, it is remarkable that the demographic factors had no significant effect on people's choices in this stated choice experiment. Neither did Westin's (1968) privacy index where privacy fundamentalists, privacy pragmatists and privacy unconcerned people were compared by the estimation of interaction effects on the data sharing factors.

3: How does the respective value of factors differ across different people?

As highlighted above, it is been proven that the relative importance of attributes varies across people as the ML model was the best approximation of the ”true” model. It is proven that different consumers have different preferences regarding data sharing factors when sharing automotive GPS data on MPC-enabled data marketplaces. However, no significant interaction effects were found between different demographics and choices in the stated choice experiment. Furthermore, in this research we did found significant differences between consumers based on their age, educational level and experience with data marketplaces. Older participants feel that consumers lost all control over their personal data statistically more than younger participants and the higher educated people more often disagree that businesses handle consumer data in a confidential and proper way than less educated people. Contradictionary, the more experienced people with data marketplaces, agree more on the fact that businesses handle the personal consumer information in a proper way and that the existing laws and organizational practices provide enough protection for consumers today.

4: What are the policy inferences of the factors and their respective value?

The key conclusions have a variety of policy consequences for how MPC-enabled data marketplaces are set up. As MPC is not yet massively adopted, modifications can be more easily adapted. As it is shown, the choice for sharing data via MPC-enabled data marketplaces is mostly dependent on the amount of risk of data disclosure which is involved and the opportunity to have a certain benefit. The concept of MPC is exactly invented in order to increase safety during sharing of sensitive data. However, it seems important to inform people more about the technology. These educational functions should be applied to governmental authorities such as local authorities and the EU, as well as financial institutions and academic institutions. As was shown in the data, most people do not really know what happens with their personal data and do not know where it is stored or if it is even sold. So raising awareness regarding the types of risks would be a first task in combination with explaining what MPC is. Furthermore, many comments by respondents pointed out to the fact that they would like, just as in this survey, to choose themselves how their data is stored and distributed. In other words, have a say in the way the data is handled. Benefit, which is not an MPC factor in the end, showed that by giving people a certain incentive to share their data, people are more willing to share their automotive GPS data on MPC-enabled data marketplaces. People are constantly making trade offs between the amount of risk and the amount of benefit they receive. This incentive helps to gain the appropriate mass of demand which creates publicity for MPC in general for people to start making use of this method of hidden sharing. In the end, trust in terms of a herding effect in this research, was one of least important. This contradicts literature on trust in data sharing in the e-commerce and research sector (Chawinga and Zinn, 2019; D. J. Kim, Ferrin, and Rao, 2009) in general as people base their decision to use specific products or share valuable data on trust in general. However, this research gave insights about the way people herd in terms of online automotive data sharing.

7.2. Importance of the research in societal view

Finally, decision-makers which desire more data sharing activities on MPC-enabled data marketplaces, can take the political consequences as a guideline. The answers to the accompanying questions can be used by research bodies active in data marketplaces. As a result, relevance is determined by the conversion of outcomes into consequences and potentials for all types data sharing platforms. It should also be noted that understanding the dynamics of information sharing technology acceptance by users is essential in order to respond to the main research problem. The journey toward universally available and safe automotive data sharing is will hopefully accelerate as a consequence of the behavioral findings of this research. Simply put, this research contributes to the fundamental movement of aggregation of consumer data to improve automotive operations for all, by examining how those at the forefront of new knowledge creation behave through sharing automotive data on MPC-enabled data marketplaces.

7.3. Importance of the research in theoretical view

This study's additions to the research are shown in three ways. Firstly, where most of the research is on the adoption of privacy preserving technologies (PPT's), this research focuses on the consumer preferences in these PPT's on data marketplaces. This new stated choice approach in the field of PPT's showed that the difference consumers have different preferences and make different decisions regarding risk, data control, trust and benefit. Secondly, this study showed that the technology MPC makes consumers make data sharing trade-offs differently than in normal sharing situations. Thirdly, this study showed regarding the specifics of automotive GPS data the consumer preferences where strongly based on risk of data disclosure and benefit.

A contribution has been made in performing a different method in the field of automotive data sharing on MPC data marketplaces, namely the stated choice experiment. Finally, user-experience research on MPC is performed instead of too technical research on MPC. This study ventures into the field of quantitative research, allowing for a comparison of the drivers and inhibitors under consideration. By drawing upon comparative literature on data sharing in the academic-, e-commerce- and medical fields, measurable and generalizable data sharing factors have been found which are deemed relevant on automotive data sharing. Although this set of factors is rather small, the start is made in the new field of MPC-enabled data marketplaces and how people rank these factors differently. These factors can be used by researchers in future similar studies on automotive data sharing on data marketplaces.

In terms of knowledge gaps, multiple contributions are made in the sense of perceived user experience in terms of data sharing factors on MPC-enabled data marketplaces. First, the behavior in data sharing automotive sector. In perspective of quantification, the effect of the benefits noticed with Derikx, De Reuver, and Kroesen (2016) in relationship to car data sharing is supported. In their research people were willing to share their car data to insurances for an average amount of 9.54 dollars. In current research it is shown that people are willing to deteriorate their risk of data disclosure from low to moderate if they receive an amount of 9.50 dollars. Therefore, Lwin and Williams (2003) research on calculus of behavior which was assumed beforehand, is supported. Furthermore, the way in which Koch, Krenn, Pellegrino, and Ramacher (2021) defined risk of data disclosure, in probabilities in terms of percentages, showed that people weigh this attribute the heaviest. Risk of data disclosure which was expected to being supported, is supported. However, respondents rank the way data control within MPC-enabled data marketplaces is specified, as lowest. Whereas Choi and Butler (2019) foresaw issues regarding cloud-based computations and consumers losing control over their data, controversially, people value this issue as lowest in situations where they share GPS data on MPC-enabled data marketplaces. Nevertheless, there is certainly an effect, however it is almost three times less important compared to the risk of data disclosure. In the end, the herding effect which is derived from Chiregi and Navimipour (2016) is also shown to be significant, people do thus base their decision on other people regarding sharing automotive GPS data on MPC-enabled data marketplaces. We reject the effect of Westin's (1968) privacy index interaction effect on the way choices data sharing choices are made.

In the end, it is advised to data marketplace platform owners to improve on the factors benefit, trust and risk of data disclosure. Raising awareness about the way in which consumers' data is handled on data marketplaces could improve the situation. The majority of consumers not really knows in much depth how their data is being used, stored or sold. Consumers are aware of it but don not really understand at a low level how data control could help mitigate the risk of data disclosure. There is room for more awareness and education so that consumers can protect themselves. Furthermore, meeting the requirements of consumers to decrease risks and improve the benefits, data marketplaces would gain more participants which utilize and trust the platform may be the trigger to speed up worldwide automotive data sharing to gain more value from it and innovate in better automotive products and services.

7.4. Discussion and limitations

Plenty of remarks can be made about the findings of the foregoing analytical efforts. They should be formulated according to overall circumstances, as stated choice experiments for data sharing by MPC-enabled data marketplaces produced novel insights.

The model is not prone to significance issues. The sample size of the data set can be considered large enough and thus provides more accurate mean values. Also, the outliers are better identified that could skew the data, and smaller error margins are found. As a result, fluctuations are avoided, and outliers have a less impact on model stability (Rose and Bliemer, 2013). It seems that this dataset mitigates these barriers which serves the model fit. However, the sample size is skewed younger (technical) people which makes it hard to generalize all results to the population. Furthermore, it is shown that heterogeneity is existing in this context of data sharing, this is proven by the ML parameter estimations. However, the model fit of 0.3119 is still far from perfect. With the promising concept Multiparty Computation, one seeks to find useful, insightful results. Here it was initialized to explore user decisions regarding data sharing on MPC-enabled data marketplaces. During literature search, it became quickly clear that it was hard to find tangible attributes which could be varied in a stated choice experiment as MPC is perceived as unknown, in essence very technical and even not yet massively adopted. It was prioritized to include the most tangible attributes in the choice experiment. This resulted in a small amount of 4 parameters. These parameters were prone to be very basic in order to make it simple for respondents to think about trade offs. Where widespread believe that the number of attributes in studies should be limited to 6 or 7 (Conjointly, 2020; Molin, 2017), this good but far from perfect model fit is likely the effect of having too few attributes which need to explain the choices of respondents. By introducing more attributes, certainly more information would be picked up which served the model in terms of model fit.

Thus, as this research was based on primary general data sharing factors, factors as impact of data disclosure, a factor representing the laws and regulations on MPC-enabled data marketplaces, factors specifying the actors to share with or a factor representing the type of automotive data to share on MPC-enabled data marketplaces could have been important to gain deeper understanding in current context. Also, it can be questioned if for the factor risk of data disclosure, the levels were a reflection of reality. As MPC is not yet massively adopted, it was unclear to which extent risks still existed. Furthermore, by making these parameters rather basic, one can hesitate to which extend current research is still closely connected to the principle of MPC. Rather than focusing on the relative relevance of specific MPC components, the emphasis of this research as well as its main research question is on a behavioral examination of the motivations and barriers of data sharing on MPC-enabled data marketplaces. Moreover, the relation between the inner-workings of MPC and the implemented factor data control is this research could be questioned. As MPC is initially based on a decentralised architecture (Archer et al., 2018), the data sharing factor data control deviated between centralised and decentralised ways of data processing and storage. A combination of these facts question the academic contribution to the user experience of MPC in the research.

Regarding the data sharing factor comparisons with findings in literature, many similarities were found. By Wang, Duong, and Chen (2016) who studied similar intentions to disclose personal information but then via mobile applications, same rankings were found regarding the factors risk of data disclosure, benefits and data control. Also H.-S. Kim (2016) found that also as expected here of consumers in this research, young Facebook users were lacking awareness and privacy concerns when sharing GPS data via online applications. And also Han, Min, and Lee (2015) showed that benefits are more important to consumers than risks, because in general people perform calculus of behavior and weighing potential benefits almost as heavy as potential risks. Therefore, as (monetary) benefits increase, people are finding benefits more important than potential risks of data disclosure.

Also, because sharing data on data marketplaces is not familiar to people and can be stated as quite new, this type of stated choice experiment could introduce certain bias. That is, as people are not experienced with it, their choices might not reflect what they should do in reality. As Gerber, Berens, and Volkamer (2019) already researched in the risk awareness of people in data sharing on online social networks, most people tend to be not aware of all the potential risks and subsequent severe consequences of these risks. This way, therefore, it could have been good to include a ”no-share option” among the choices for types of MPC-enabled data marketplaces. However, when this option was proven to be dominant by being chosen for 90% of the time, this would have led to huge information losses (Molin, 2017). As this experiment was quite straightforward, respondents did grasp the idea better and people found it a really clear and interesting topic. However, perhaps the trade-offs were too clear and to basic that significantly all types of respondents valued the attributes the same on average. This could be one of the reasons that the interaction effects had little influence on the model’s estimations. In the end, many things could have been set up differently, the question is if this would have improved the experiment results. This topic is quite new for most of the respondents and they based their decisions quite similar regarding literature on data privacy and potential benefits.

Although these researches were in line with ours, there are also several studies which found several demographic interaction effects in online data sharing. So did Wang, Duong, and Chen (2016), which found differences between males and females in the willingness to share data on mobile applications. Where we also expected differences between genders, age groups and car users in valuing data sharing factors as benefits, risk of data disclosure, trust (in terms of a herding effect) and data control, no interaction effects were found in this study. It was also not expected that different privacy groups, based on Westin’s (1968) categories, did not statistically weigh the data sharing factors differently. Furthermore, speaking of generalizability of this research, this research is very specifically scoped on user research on MPC-enabled data marketplaces where automotive GPS data is traded. It is hard to state to which extent these results or this research can be used in data sharing situations different than MPC-enabled data marketplaces.

In the end, the used methodology poses numerous implications to the outcomes of the study. The stated choice experiments stay hypothetical. Different than revealed choice experiments, people have to imagine into the situation and then make the decisions they would make in reality. It is always questionable if these choices would have been the same in real situations. Perhaps in real situations, many different or additional factors play a role in one’s decision to choose for certain alternatives. This generates noise to the validity of the results. Whenever revealed choice experiments were possible, this would have been more valid. People then choose what they actually chose when they were in the situation.

7.5. Further Research & Recommendations

The current study is the first step in a preliminary investigation into the behavioral area of consumer automotive data sharing via MPC-enabled data marketplaces. While this study provides a solid foundation for ranking MPC-related factors in data sharing, there are numerous avenues for future research. By introducing complicated variables, this restricted focus on data sharing could be widened. One may seek to dig deeper into MPC by introducing for example malicious parties as attribute. Furthermore, for example, Faujdar (2019) found that different types of MPC user interfaces influenced the way people were willing to upload sensitive data on digital platforms. This could be taken into account in similar stated choice experiments regarding MPC-enabled data marketplaces. As noted before, mostly general attributes were included in the choice experiment under limited attribute levels. It is also possible to dive into the deep understanding of these general attributes, in forms of qualitative research, and introduce additional levels to these attributes. For example, the herding effect which was measured in three levels of other people being active in automotive data sharing. This could be far more extensively experimented by following Baddeley (Baddeley) and Milne, Rohm, and Bahl (2004) to combine herding effect with research in consumers and the option to read the privacy policies on data marketplaces to investigate to what extend consumers take their own responsibility in retrieving knowledge about the sharing platforms.

Also in terms of risk of data disclosure, the accompanied data sharing factor as the impact of data disclosures could be implemented in similar stated choice experiments. Therefore, first extensive impact assessments are required in order to map the possible threats in data sharing as shown in Koch, Krenn, Pellegrino, and Ramacher (2021). By including this type of factor, one can investigate whether different people value the factor risk of data disclosure differently when the accompanied impacts are known. This in the end is also expected likely to depend on the type of automotive data (KPMG, 2020; McKinsey, 2016; Whiddett, Hunter, Engelbrecht, and Handy, 2006) one shares on MPC-enabled data marketplaces.

The privacy calculus model is used in this experimental study to evaluate consumers' calculations regarding different data sharing factors in sharing automotive GPS data on MPC-enabled data marketplaces. The findings show that perceived benefits have a higher effect on automotive customers' decisions to share automotive information via MPC-enabled data markets than perceived trust and the data control as also shown in Wang, Duong, and Chen (2016). Data marketplace owners should highlight the availability of privacy-preserving technology and assist users control their automotive data to improve perceived benefits. Furthermore, data marketplace owners should decrease the perceived severity of the threat to attract customers to provide sensitive automotive data. Thus, as MPC is still relatively new and behavioral user experience is lagging behind as it is not yet massively adopted, solving the privacy dilemma on the consumers' side by explaining more about MPC may help data marketplace owners attract more aggregated and privacy-preserved automotive data. As a result, both consumers and data marketplace providers might benefit from a win–win scenario.

References

Abbas, A., Agahari, W., van de Ven, M., Zuiderwijk, A., & de Reuver, M. (2021). Business data sharing through data marketplaces: A systematic literature review. https://doi.org/10.18690/978-961-286-385-9.6

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, *347*(6221), 509–514.

Agahari, W. (2020). Platformization of data sharing: Multi-party computation (mpc) as control mechanism and its effect on firms' participation in data sharing via data marketplaces. *Proceedings of the 33nd Bled eConference*.

Agahari, W., Dolci, R., & de Reuver, M. (2021). Business model implications of privacy-preserving technologies in data marketplaces: The case of multi-party computation.

Aldeen, Y. A. A. S., Salleh, M., & Razzaque, M. A. (2015). A comprehensive review on privacy preserving data mining. *SpringerPlus*, *4*(1), 1–36.

Al-Sharidah, A., Syed, A., Alsannat, E., & Gaddourah, A. (2020). How cybersecurity policies enable ir 4.0 emerging technologies. *International Petroleum Technology Conference*.

Alter, G., Falk, B. H., Lu, S., & Ostrovsky, R. (2018). Computing statistics from private data. *Data Science Journal*, *17*.

Anic, I.-D., Škare, V., & Milaković, I. K. (2019). The determinants and effects of online privacy concerns in the context of e-commerce. *Electronic Commerce Research and Applications*, *36*, 100868.

Apfelbeck, F. (2018). Evaluation of privacy-preserving technologies for machine learning. https://medium.com/outlier-ventures-io/evaluation-of-privacy-preserving-technologies-for-machine-learning-8d2e3c87828c

Archer, D. W., Bogdanov, D., Lindell, Y., Kamm, L., Nielsen, K., Pagter, J. I., Smart, N. P., & Wright, R. N. (2018). From keys to databases—real-world applications of secure multi-party computation. *The Computer Journal*, *61*(12), 1749–1771.

Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., Moorman, D., Uhlir, P., & Wouters, P. (2004). Promoting access to public research data for scientific, economic, and social development. *Data Science Journal*, *3*, 135–152.

Athanasopoulou, A., Bouwman, W., Nikayin, F., & de Reuver, G. (2016). The disruptive impact of digitalization on the automotive ecosystem: A research agenda on business models, platforms and consumer issues. *Proceedings of the 29th Bled eConference*.

Baddeley, M. (2010). Herding, social influence and economic decision-making: Socio-psychological and neuroscientific analyses. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *365*, 281–90. https://doi.org/10.1098/rstb.2009.0169

Bal, G., Rannenberg, K., & Hong, J. I. (2015). Styx: Privacy risk communication for the android smartphone platform based on apps' data-access behavior patterns. *Computers & Security*, *53*, 187–202.

Banerjee, A. V. (1992). A simple model of herd behavior. *The quarterly journal of economics*, *107*(3), 797–817.

Bansal, G., Gefen, D. et al. (2010). The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. *Decision support systems*, *49*(2), 138–150.

Baumeister, R. F., & Leary, M. R. (1997). Writing narrative literature reviews. *Review of general psychology*, *1*(3), 311–320.

Bell, E. A., Ohno-Machado, L., & Grando, M. A. (2014). Sharing my health data: A survey of data sharing preferences of healthy individuals. *AMIA Annual Symposium Proceedings*, *2014*, 1699.

Bem, S. L. (1981). The bsri and gender schema theory: A reply to spence and helmreich.

Ben-Akiva, M., & Swait, J. (1986). The akaike likelihood ratio index. *Transportation Science*, *20*(2), 133–136.

Bergman, R. (2020). A business model taxonomy for data marketplaces: Data trade in various trading structures.

Beskow, L. M., & Dean, E. (2008). Informed consent for biorepositories: Assessing prospective participants' understanding and opinions. *Cancer Epidemiology and Prevention Biomarkers*, *17*(6), 1440–1451.

Bestavros, A., Lapets, A., Jansen, F., Varia, M., Volgushev, N., & Schwarzkopf, M. (2017). Design and deployment of usable, scalable mpc. *Theory and Practice of Multi-Party Computation Workshop*.

Bestavros, A., Lapets, A., & Varia, M. (2017). User-centric distributed solutions for privacy-preserving analytics. *Communications of the ACM*, *60*(2), 37–39.

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy*, *100*(5), 992–1026.

Bliemer, M. C., & Rose, J. M. (2005). Efficiency and sample size requirements for stated choice studies.

Bogdanov, D., Talviste, R., & Willemson, J. (2012). Deploying secure multi-party computation for financial data analysis. *International Conference on Financial Cryptography and Data Security*, 57–64.

Bogetoft, P., Christensen, D. L., Damgård, I., Geisler, M., Jakobsen, T. P., Krøigaard, M., Nielsen, J. D., Nielsen, J. B., Nielsen, K., Pagter, J., et al. (2008). Multiparty computation goes live. *IACR Cryptol. ePrint Arch.*, *2008*, 68.

Bouwman, H., Faber, E., Haaker, T., Kijl, B., & De Reuver, M. (2008). Conceptualizing the stof model. *Mobile service innovation and business models* (pp. 31–70). Springer.

Brothers, K. B., Morrison, D. R., & Clayton, E. W. (2011). Two large-scale surveys on community attitudes toward an opt-out biobank. *American journal of medical genetics Part A*, *155*(12), 2982–2990.

Bunz, U. (2003). Growing from computer literacy towards computer-mediated communication competence: Evolution of a field and evaluation of a new measurement instrument. *INFORMATION TECHNOLOGY EDUCATION AND SOCIETY-ALBERT PARK-*, *4*(2), 53–84.

Canetti, R., Feige, U., Goldreich, O., & Naor, M. (1996). Adaptively secure multi-party computation. *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, 639–648.

Carson, R. T., Louviere, J. J., Anderson, D. A., Arabie, P., Bunch, D. S., Hensher, D. A., Johnson, R. M., Kuhfeld, W. F., Steinberg, D., Swait, J., et al. (1994). Experimental analysis of choice. *Marketing letters*, *5*(4), 351–367.

Catrina, O., & Kerschbaum, F. (2008). Fostering the uptake of secure multiparty computation in e-commerce. *2008 Third International Conference on Availability, Reliability and Security*, 693–700.

Chang, H. H., & Wong, K. H. (2010). Adoption of e-procurement and participation of e-marketplace on firm performance: Trust as a moderator. *Information & management*, *47*(5-6), 262–270.

Chang, Y., Wong, S. F., Libaque-Saenz, C. F., & Lee, H. (2018). The role of privacy policy on consumers' perceived privacy. *Government Information Quarterly*, *35*(3), 445–459.

Chawinga, W. D., & Zinn, S. (2019). Global perspectives of research data sharing: A systematic literature review. *Library & Information Science Research*, *41*(2), 109–122.

Chen, L., & Liu, H. (2015). A review of privacy protection in e-commerce. *Journal of*.

Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile networks and applications*, *19*(2), 171–209.

Cheong, L. K., & Chang, V. (2007). The need for data governance: A case study. *ACIS 2007 Proceedings*, 100.

Chiregi, M., & Navimipour, N. J. (2016). A new method for trust and reputation evaluation in the cloud environments using the recommendations of opinion leaders' entities and removing the effect of troll entities. *Computers in Human Behavior*, *60*, 280–292.

Choi, J. I., & Butler, K. R. (2019). Secure multiparty computation and trusted hardware: Examining adoption challenges and opportunities. *Security and Communication Networks*, *2019*.

Chorus, C. (2012). Random regret minimization: An overview of model properties and empirical evidence. *Transport Reviews*, *32*(1), 75–92. https://doi.org/10.1080/01441647.2011.609947

Chorus, C. (2017). Lecture notes on analysis by discrete choice models [powerpoint slides].

Chorus, C. G. (2010). A new model of random regret minimization. *European Journal of Transport and Infrastructure Research*, *10*(2).

Christen, M., Gordijn, B., & Loi, M. (2020). *The ethics of cybersecurity*. Springer Nature.

Clarke, R. (1999). Internet privacy concerns confirm the case for intervention. *Communications of the ACM*, *42*(2), 60–67.

Commission, E.-E. et al. (2012). The grand challenge-the design and societal impact of horizon 2020.

Conjointly. (2020). Avoid common mistakes with practical tips for setting up conjoint studies. https://conjointly.com/guides/tips-for-setting-up-conjoint-studies/

Cragin, M. H., Palmer, C. L., Carlson, J. R., & Witt, M. (2010). Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *368*(1926), 4023–4038.

Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating amazon's mechanical turk as a tool for experimental behavioral research. *PloS one*, *8*(3), e57410.

Culnan, M. J., & Armstrong, P. K. (1999). Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation. *Organization science*, *10*(1), 104–115.

De Prieëlle, F., De Reuver, M., & Rezaei, J. (2020). The role of ecosystem data governance in adoption of data platforms by internet-of-things data providers: Case of dutch horticulture industry. *IEEE Transactions on Engineering Management*.

Delgado, R. (2021). How car manufacturers are using big data. https://datafloq.com/read/car-manufacturers-are-using-big-data/1204

de Reuver, M., Fiebig, T., Agahari, W., & Faujdar, V. (2020). D2. 4 user experiment report.

de Reuver, M., Sørensen, C., & Basole, R. C. (2018). The digital platform: A research agenda. *Journal of Information Technology*, *33*(2), 124–135.

Derikx, S., De Reuver, M., & Kroesen, M. (2016). Can privacy concerns for insurance of connected cars be compensated? *Electronic markets*, *26*(1), 73–81.

Dhillon, G. (2015). What to do before and after a cybersecurity breach. *American University, Washington, DC, Kogod Cybersecurity Governance Center*.

Dienlin, T., & Metzger, M. J. (2016). An extended privacy calculus model for snss: Analyzing self-disclosure and self-withdrawal in a representative us sample. *Journal of Computer-Mediated Communication*, *21*(5), 368–383.

Dixon, W. G., Spencer, K., Williams, H., Sanders, C., Lund, D., Whitley, E. A., & Kaye, J. (2014). A dynamic model of patient consent to sharing of medical record data. *bmj*, *348*.

Dwaikat, N. Y., Money, A. H., Behashti, H. M., & Salehi-Sangari, E. (2018). How does information sharing affect first-tier suppliers' flexibility? evidence from the automotive industry in sweden. *Production Planning & Control*, *29*(4), 289–300.

Dwork, C. (2006). Differential privacy. *International Colloquium on Automata, Languages, and Programming*, 1–12.

Dwork, C., Roth, A. et al. (2014). The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, *9*(3-4), 211–407.

Eagly, A. H. (1987). Reporting sex differences.

Faujdar, V. (2019). Customer acceptance of a revenue management platform with multi-party computation: Application of multi-party computation to revenue management in the semiconductor industry.

Fecher, B., Friesike, S., & Hebing, M. (2015). What drives academic data sharing? *PloS one*, *10*(2), e0118053.

Finn, R. L., Wright, D., & Friedewald, M. (2013). Seven types of privacy. *European data protection: Coming of age* (pp. 3–32). Springer.

Fricker, S. A., & Maksimov, Y. V. (2017). Pricing of data products in data marketplaces. *International Conference of Software Business*, 49–66.

Fruhwirth, M., Rachinger, M., & Prlja, E. (2020). Discovering business models of data marketplaces. *Proceedings of the 53rd Hawaii International Conference on System Sciences*.

Gentry, C. (2010). Computing arbitrary functions of encrypted data. *Communications of the ACM*, *53*(3), 97–105.

Gerber, N., Berens, B., & Volkamer, M. (2019). Investigating people's privacy risk perception. *Proceedings on Privacy Enhancing Technologies*, *2019*, 267–288. https://doi.org/10.2478/popets-2019-0047

Ghosh, H. (2018). Data marketplace as a platform for sharing scientific data. *Data science landscape* (pp. 99–105). Springer.

Giaretta, L., Savvidis, I., Marchioro, T., Girdzijauskas, S., Pallis, G., Dikaiakos, M., & Markatos, E. (2021). Pds2: A user-centered decentralized marketplace for privacy preserving data processing. *Third International Workshop on Blockchain and Data Management*.

Goundar, S. (2012). Chapter 3–research methodology and research method.

Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of things (iot): A vision, architectural elements, and future directions. *Future generation computer systems*, *29*(7), 1645–1660.

Haeusermann, T., Greshake, B., Blasimme, A., Irdam, D., Richards, M., & Vayena, E. (2017). Open sharing of genomic data: Who does it and why? *PLoS One*, *12*(5), e0177158.

Haga, S. B., & O'Daniel, J. (2011). Public perspectives regarding data-sharing practices in genomics research. *Public health genomics*, *14*(6), 319–324.

Han, S., Min, J., & Lee, H. (2015). Antecedents of social presence and gratification of social connection needs in sns: A study of twitter users and their mobile and non-mobile usage. *International Journal of Information Management*, *35*(4), 459–471.

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

Hansen, P. (2020). On automotive electronics.

Harbach, M., Fahl, S., & Smith, M. (2014). Who's afraid of which bad wolf? a survey of it security risk awareness. *2014 IEEE 27th Computer Security Foundations Symposium*, 97–110.

Harborth, D., & Pape, S. (2020). How privacy concerns, trust and risk beliefs, and privacy literacy influence users' intentions to use privacy-enhancing technologies: The case of tor. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems*, *51*(1), 51–69.

Hargittai, E. (2007). A framework for studying differences in people's digital media uses. *Grenzenlose cyberwelt?* (pp. 121–136). Springer.

Hauber, A., Arden, N., Mohamed, A., Johnson, F. R., Peloso, P., Watson, D., Mavros, P., Gammaitoni, A., Sen, S., & Taylor, S. (2013). A discrete-choice experiment of united kingdom patients' willingness to risk adverse events for improved function and pain control in osteoarthritis. *Osteoarthritis and cartilage*, *21*(2), 289–297.

Hiratsuka, V., Brown, J., & Dillard, D. (2012). Views of biobanking research among alaska native people: The role of community context. *Progress in community health partnerships: research, education, and action*, *6*(2), 131–139.

Hui, K.-L., Teo, H. H., & Lee, S.-Y. T. (2007). The value of privacy assurance: An exploratory field experiment. *Mis Quarterly*, 19–33.

Javaid, A., Zahid, M., Ali, I., Khan, R. J. U. H., Noshad, Z., & Javaid, N. (2019). Reputation system for iot data monetization using blockchain. *International Conference on Broadband and Wireless Computing, Communication and Applications*, 173–184.

Jen, W., LU, M. L., Wang, W.-T., & Chang, Y.-T. (2013). Effects of perceived benefits and perceived costs on passenger's intention to use self-ticketing kiosk of taiwan high speed rail corporation. *Journal of the Eastern Asia Society for Transportation Studies*, *10*, 215–230.

Jenkins, H. (2006). Game on! the future of literacy education in a participatory media culture," *Threshold*, *4*.

Jernigan, S., Ransbotham, S., & Kiron, D. (2016). Data sharing and analytics drive success with iot-creating business value with the internet of things (global executive study). *MIT Sloan Management Review*.

Jian, G., & Jeffres, L. W. (2006). Understanding employees' willingness to contribute to shared electronic databases: A three-dimensional framework. *Communication Research*, *33*(4), 242–261.

Kaiser, C., Steger, M., Dorri, A., Festl, A., Stocker, A., Fellmann, M., & Kanhere, S. (2018). Towards a privacy-preserving way of vehicle data sharing–a case for blockchain technology? *International Forum on Advanced Microsystems for Automotive Applications*, 111–122.

Kanger, L., & Pruulmann-Vengerfeldt, P. (2015). Social need for secure multiparty computation. *Laud, Peeter*, 43–57.

Karpati, A. (2011). Digital literacy in education. *UNESCO Institute for information technologies in Education*.

Keller, B., Eling, M., Schmeiser, H., Christen, M., & Loi, M. (2018). *Big data and insurance: Implications for innovation, competition and privacy*. Geneva Association-International Association for the Study of Insurance …

Kennedy, T., Wellman, B., & Klement, K. (2003). Gendering the digital divide. *IT & society*, *1*(5), 72–96.

Kerik, L., Laud, P., & Randmets, J. (2016). Optimizing mpc for robust and scalable integer and floating-point arithmetic. *International Conference on Financial Cryptography and Data Security*, 271–287.

Kim, D.-Y., Lehto, X. Y., & Morrison, A. M. (2007). Gender differences in online travel information search: Implications for marketing communications on the internet. *Tourism management*, *28*(2), 423–433.

Kim, D. J., Ferrin, D. L., & Rao, H. R. (2009). Trust and satisfaction, two stepping stones for successful e-commerce relationships: A longitudinal exploration. *Information systems research*, *20*(2), 237–257.

Kim, H.-S. (2016). What drives you to check in on facebook? motivations, privacy concerns, and mobile phone involvement for location-based information sharing. *Computers in Human Behavior*, *54*, 397–406.

Kim, K., Sankar, P., Wilson, M., & Haynes, S. (2017). Factors affecting willingness to share electronic health data among california consumers. *BMC medical ethics*, *18*(1), 1–10.

Kim, S., & Lee, H. (2006). The impact of organizational context and information technology on employee knowledge-sharing capabilities. *Public administration review*, *66*(3), 370–385.

Koch, K., Krenn, S., Pellegrino, D., & Ramacher, S. (2021). Privacy-preserving analytics for data markets using mpc. *arXiv preprint arXiv:2103.03739*.

Koutroumpis, P., Leiponen, A., & Thomas, L. D. (2017). *The (unfulfilled) potential of data marketplaces* (tech. rep.). ETLA Working Papers.

Koutroumpis, P., Leiponen, A., & Thomas, L. D. (2020). Markets for data. *Industrial and Corporate Change*, *29*(3), 645–660.

Kowalewski, S., Ziefle, M., Ziegeldorf, H., & Wehrle, K. (2015). Analyzing user preferences regarding privacy settings in germany. *Procedia Manufacturing*, *3*, 815–822.

KPMG. (2020). Automotive data sharing.

Kramer, R. M. (1999). Trust and distrust in organizations: Emerging perspectives, enduring questions. *Annual review of psychology*, *50*(1), 569–598.

Krasnova, H., Spiekermann, S., Koroleva, K., & Hildebrand, T. (2010). Online social networks: Why we disclose. *Journal of information technology*, *25*(2), 109–125.

Kuhfeld, W. F. (2006). Construction of efficient designs for discrete choice experiments. *The Handbook of Marketing Research: Uses, Misuses, and Future Advances. Sage Publications*.

Kumar, R. S., Pugazhendhi, S., Muralidharan, C., & Murali, S. (2018). An empirical study on effect of information sharing on supply chain performance-the case of indian automotive industry. *International Journal of Logistics Systems and Management*, *31*(3), 299–319.

Kumaraguru, P., & Cranor, L. F. (2005). *Privacy indexes: A survey of westin's studies*. Carnegie Mellon University, School of Computer Science, Institute for …

Lange, J., Stahl, F., & Vossen, G. (2018). Datenmarktplätze in verschiedenen forschungsdisziplinen: Eine Übersicht. *Informatik-Spektrum*, *41*(3), 170–180.

Langer, E. J., & Abelson, R. P. (1983). *The psychology of control*. SAGE Publications, Incorporated.

Lapets, A., Jansen, F., Albab, K. D., Issa, R., Qin, L., Varia, M., & Bestavros, A. (2018). Accessible privacy-preserving web-based data analysis for assessing and addressing economic inequalities. *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, 1–5.

Lapets, A., Volgushev, N., Bestavros, A., Jansen, F., & Varia, M. (2016). Secure mpc for analytics as a web application. *2016 IEEE Cybersecurity Development (SecDev)*, 73–74.

Lee, I., & Lee, K. (2015). The internet of things (iot): Applications, investments, and challenges for enterprises. *Business Horizons*, *58*(4), 431–440.

Lee, S. U., Zhu, L., & Jeffery, R. (2017). Design choices for data governance in platform ecosystems: A contingency model. *arXiv preprint arXiv:1706.07560*.

Leon, P. G., Ur, B., Wang, Y., Sleeper, M., Balebako, R., Shay, R., Bauer, L., Christodorescu, M., & Cranor, L. F. (2013). What matters to users? factors that affect users' willingness to share information with online advertisers. *Proceedings of the ninth symposium on usable privacy and security*, 1–12.

Levy, Y., & Ellis, T. J. (2006). A systems approach to conduct an effective literature review in support of information systems research.

Luo, X. (2002). Trust production and privacy concerns on the internet: A framework based on relationship marketing and social exchange theory. *Industrial Marketing Management*, *31*(2), 111–118.

Lwin, M. O., & Williams, J. D. (2003). A model integrating the multidimensional developmental theory of privacy and theory of planned behavior to examine fabrication of information online. *Marketing Letters*, *14*(4), 257–272.

Malhotra, N. K., Kim, S. S., & Agarwal, J. (2004). Internet users' information privacy concerns (iuipc): The construct, the scale, and a causal model. *Information systems research*, *15*(4), 336–355.

Mamo, L. A., Browe, D. K., Logan, H. C., & Kim, K. K. (2013). Patient informed governance of distributed research networks: Results and discussion from six patient focus groups. *AMIA Annual Symposium Proceedings*, *2013*, 920.

Mandeville, K. L., Lagarde, M., & Hanson, K. (2014). The use of discrete choice experiments to inform health workforce policy: A systematic review. *BMC health services research*, *14*(1), 367.

Mattke, J., Maier, C., Reis, L., & Weitzel, T. (2020). Herd behavior in social media: The role of facebook likes, strength of ties, and expertise. *Information & Management*, *57*(8), 103370.

Maurer, U. (2006). Secure multi-party computation made simple. *Discrete Applied Mathematics*, *154*(2), 370–381.

McFadden, D. et al. (1973). Conditional logit analysis of qualitative choice behavior.

McKinsey. (2016). Monetizing car data. *Information & management*. https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/monetizing-car-data#

Mello, M. M., Lieou, V., & Goodman, S. N. (2018). Clinical trial participants' views of the risks and benefits of data sharing. *New England journal of medicine*, *378*(23), 2202–2211.

Milia, N., Congiu, A., Anagnostou, P., Montinaro, F., Capocasa, M., Sanna, E., & Bisol, G. D. (2012). Mine, yours, ours? sharing data on human genetic variation. *PloS one*, *7*(6), e37552.

Miller, P. (2014). Nurturing the market for data markets (2012). *Last accessed*, 11–24.

Milne, G. R., Rohm, A. J., & Bahl, S. (2004). Consumers' protection of online privacy and identity. *Journal of Consumer Affairs*, *38*(2), 217–232.

Miltersen, P. B., Nielsen, J. B., & Triandopoulos, N. (2009). Privacy-enhancing auctions using rational cryptography. *Annual International Cryptology Conference*, 541–558.

Mišura, K., & Žagar, M. (2016). Data marketplace for internet of things. *2016 International Conference on Smart Systems and Technologies (SST)*, 255–260. https://doi.org/10.1109/SST.2016.7765669

Miyazaki, A. D., & Fernandez, A. (2001). Consumer perceptions of privacy and security risks for online shopping. *Journal of Consumer affairs*, *35*(1), 27–44.

Molin, E. (2017). Lecture note on efficient designs [powerpoint slides].

Moody, D. L., & Walsh, P. (1999). Measuring the value of information-an asset valuation approach. *ECIS*, 496–512.

Mosterd, L., Sobota, V. C., van de Kaa, G., Ding, A. Y., & de Reuver, M. (2021). Context dependent tradeoffs around platform-to-platform openness: The case of the internet of things. *Technovation*, *108*, 102331.

Nanibaa'A, G., Sathe, N. A., Antommaria, A. H. M., Holm, I. A., Sanderson, S. C., Smith, M. E., McPheeters, M. L., & Clayton, E. W. (2016). A systematic literature review of individuals' perspectives on broad consent and data sharing in the united states. *Genetics in Medicine*, *18*(7), 663–671.

Nield, K., & Nordstrom, A. T. (2016). *Response bias in voluntary surveys: An empirical analysis of the canadian census* (tech. rep.).

Noronha, A., Moriarty, R., O'Connell, K., & Villa, N. (2014). Attaining iot value: How to move from connecting things to capturing insights. *White paper, Cisco*.

OECD. (2019). *Enhancing access to and sharing of data reconciling risks and benefits for data re-use across societies*. OECD Publishing.

Otonomo. (2020). What european consumers think about connected car data and privacy. https://info.otonomo.io/sbd-eu-consumer-survey-results-lp

Park, Y. J. (2013). Digital literacy and privacy behavior online. *Communication Research*, *40*(2), 215–236.

Pavlou, P. A. (2002). Institution-based trust in interorganizational exchange relationships: The role of online b2b marketplaces on trust formation. *The Journal of Strategic Information Systems*, *11*(3-4), 215–243.

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153–163.

Petronia, M. (2020). Multiparty computation: The effect of multiparty computation on firms' willingness to contribute protected data.

Pettai, M., & Laud, P. (2015). Combining differential privacy and secure multiparty computation. *Proceedings of the 31st Annual Computer Security Applications Conference*, 421–430.

Pretty, T. (2020). What is sensitive data, sensitive data definition types? https://cipherpoint.com/blog/what-is-sensitive-data/

Prolific. (2020). What are the advantages and limitations of an online sample? https://researcher-help.prolific.co/hc/en-gb/articles/360009501473-What-are-the-advantages-and-limitations-of-an-online-sample-

Rakotonarivo, O. S., Schaafsma, M., & Hockley, N. (2016). A systematic review of the reliability and validity of discrete choice experiments in valuing non-market environmental goods. *Journal of environmental management*, *183*, 98–109.

Rao, R. V., & Selvamani, K. (2015). Data security challenges and its solutions in cloud computing. *Procedia Computer Science*, *48*, 204–209.

Richter, H., & Slowinski, P. R. (2019). The data sharing economy: On the emergence of new intermediaries. *IIC-International Review of Intellectual Property and Competition Law*, *50*(1), 4–29.

Riordan, F., Papoutsi, C., Reed, J. E., Marston, C., Bell, D., & Majeed, A. (2015). Patient and public attitudes towards informed consent models and levels of awareness of electronic health records in the uk. *International journal of medical informatics*, *84*(4), 237–247.

Roman, D., & Vu, K. (2018). Enabling data markets using smart contracts and multi-party computation. *International Conference on Business Information Systems*, 258–263.

Rose, J. M., & Bliemer, M. C. (2013). Sample size requirements for stated choice experiments. *Transportation*, *40*(5), 1021–1041.

Rosenblatt, M., Jain, S. H., & Cahill, M. (2015). Sharing of clinical trial data: Benefits, risks, and uniform principles.

Roth, A. E. (2009). What have we learned from market design? *Innovation policy and the economy*, *9*(1), 79–112.

Salampessy, B. H., Veldwijk, J., Schuit, A. J., Van Den Brekel-dijkstra, K., Neslo, R. E., De Wit, G. A., & Lambooij, M. S. (2015). The predictive value of discrete choice experiments in public health: An exploratory application. *The Patient-Patient-Centered Outcomes Research*, *8*(6), 521–529.

Sánchez, D., & Viejo, A. (2017). Personalized privacy in open data sharing scenarios. *Online Information Review*.

Sanderson, S. C., Brothers, K. B., Mercaldo, N. D., Clayton, E. W., Antommaria, A. H. M., Aufox, S. A., Brilliant, M. H., Campos, D., Carrell, D. S., Connolly, J., et al. (2017). Public attitudes toward consent and data sharing in biobank research: A large multi-site experimental survey in the us. *The American Journal of Human Genetics*, *100*(3), 414–427.

Sarkar, P. (2015). *Data as a service: A framework for providing reusable enterprise data services*. John Wiley & Sons.

Schomakers, E.-M., Lidynia, C., & Ziefle, M. (2020). All of me? users' preferences for privacy-preserving data markets and the importance of anonymity. *Electronic Markets*, *30*(3), 649–665.

Shah, N., Coathup, V., Teare, H., Forgie, I., Giordano, G. N., Hansen, T. H., Groeneveld, L., Hudson, M., Pearson, E., Ruetten, H., et al. (2019). Motivations for data sharing—views of research participants from four european countries: A direct study. *European Journal of Human Genetics*, *27*(5), 721–729.

Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of consumer research*, *16*(2), 158–174.

Sirdeshmukh, D., Singh, J., & Sabol, B. (2002). Consumer trust, value, and loyalty in relational exchanges. *Journal of marketing*, *66*(1), 15–37.

Skatova, A., Johal, J., Houghton, R., Mortier, R., Bhandari, N., Lodge, T., Wagner, C., Goulding, J., Crowcroft, J., & Madhavapeddy, A. (2013). Perceived risks of personal data sharing. *Proc. Digital Economy: Open Digital (Nov. 2013)*.

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*, 333–339.

Solove, D. J., Rotenberg, M., & Schwartz, P. M. (2006). *Privacy, information, and technology*. Aspen Publishers Online.

Sousa, P. R., Antunes, L., & Martins, R. (2018). The present and future of privacy-preserving computation in fog computing. In A. M. Rahmani, P. Liljeberg, J.-S. Preden, & A. Jantsch (Eds.), *Fog computing in the internet of things: Intelligence at the edge* (pp. 51–69). Springer International Publishing. https://doi.org/10.1007/978-3-319-57639-8_4

Spiekermann, M. (2019). Data marketplaces: Trends and monetisation of data goods. *Intereconomics*, *54*(4), 208–216.

Spiekermann, S. (2007). *Perceived control: Scales for privacy in ubiquitous computing*. Auerbach Publications.

Spiekermann, S., & Korunovska, J. (2017). Towards a value theory for personal data. *Journal of Information Technology*, *32*(1), 62–84.

Stahl, F., Löser, A., & Vossen, G. (2015). Preismodelle für datenmarktplätze. *Informatik-Spektrum*, *38*(2), 133–141.

Sun, H. (2013). A longitudinal study of herd behavior in the adoption and continued use of technology. *Mis Quarterly*, 1013–1041.

Sun, Y., Wang, N., Shen, X.-L., & Zhang, J. X. (2015). Location information disclosure in location-based social network services: Privacy calculus, benefit structure, and gender differences. *Computers in Human Behavior*, *52*, 278–292.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., & Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PloS one*, *6*(6), e21101.

Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., Pollock, D., & Dorsett, K. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PloS one*, *10*(8), e0134826.

Thomas, L. D., & Leiponen, A. (2016). Big data commercialization. *IEEE Engineering Management Review*, *44*(2), 74–90.

Tilson, D., Lyytinen, K., & Sørensen, C. (2010). Digital infrastructures: The missing is research agenda. research commentary. *Information Systems Research*, *21*(4), 748–759.

TNO. (2020). Multi-party computation: Zorg optimaliseren door patiëntendata te versleutelen. https://www.tno.nl/nl/aandachtsgebieden/informatie-communicatie-technologie/roadmaps/data-sharing/cybersecurity-mpc/

Toldsepp, K., Pruulmann-Vengerfeldt, P., & Laud, P. (2012). Usable and efficient secure multiparty computation (etla working papers). https://cordis.europa.eu/docs/projects/cnect/1/284731/080/deliverables/001-D12.pdf

Trabelsi, S., Salzgeber, V., Bezzi, M., & Montagnon, G. (2009). Data disclosure risk evaluation. *2009 Fourth International Conference on Risks and Security of Internet and Systems (CRiSIS 2009)*, 35–72.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British journal of management*, *14*(3), 207–222.

Travisi, C. M., & Nijkamp, P. (2008). Valuing environmental and health risk in agriculture: A choice experiment approach to pesticides in italy. *Ecological Economics*, *67*(4), 598–607.

Treiblmaier, H., & Chong, S. (2011). Trust and perceived risk of personal information as antecedents of online information disclosure: Results from three countries. *Journal of Global Information Management (JGIM)*, *19*(4), 76–94.

TU Delft. (n.d.). Educational tooling : Overview of tools used in education. [Accessed: 2021-04-14]. %5Curl%7Bhttps://brightspace-support.tudelft.nl/educational-tooling/%7D

Tucker, C., & Zhang, J. (2011). How does popularity information affect choices? a field experiment. *Management Science*, *57*(5), 828–842.

van de Ven, M., Abbas, A. E., Kwee, Z., & de Reuver, M. (2021). Creating a taxonomy of business models for data marketplaces. *34th Bled eConference: Digital Support from Crisis to Progressive Change*, 313–325.

Venkatesh, V., & Morris, M. G. (2000). Why don't men ever stop to ask for directions? gender, social influence, and their role in technology acceptance and usage behavior. *MIS quarterly*, 115–139.

Verschuren, P., Doorewaard, H., & Mellion, M. (2010). *Designing a research project* (Vol. 2). Eleven International Publishing The Hague.

Veselovská, L., Kožárová, M., & Zavadsky, J. (2018). Relationship between information sharing and flexibility in management of enterprises in automotive industry: An empirical study. *Serbian Journal of Management*, *13*(2), 381–393.

Viereckl, R., Ahlemann, D., Koster, A., & Jursch, S. (2015). Racing ahead with autonomous cars and digital innovation. *Auto Tech Review*, *4*(12), 18–23.

Wang, T., Duong, T. D., & Chen, C. C. (2016). Intention to disclose personal information via mobile applications: A privacy calculus perspective. *International journal of information management*, *36*(4), 531–542.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS quarterly*, xiii–xxiii.

Westin, A., & Interactive, H. (1999). *Ibm-harris multi-national consumer privacy survey* (tech. rep.). Tech. rep., 1999. Approximately 5,000 adults of the US, Britain and Germany.

Westin, A. F. (1968). Privacy and freedom. *Washington and Lee Law Review*, *25*(1), 166.

Whiddett, R., Hunter, I., Engelbrecht, J., & Handy, J. (2006). Patients' attitudes towards sharing their health information. *International journal of medical informatics*, *75*(7), 530–541.

Willem, A., & Buelens, M. (2007). Knowledge sharing in public sector organizations: The effect of organizational characteristics on interdepartmental knowledge sharing. *Journal of public administration research and theory*, *17*(4), 581–606.

Willison, D. J., Steeves, V., Charles, C., Schwartz, L., Ranford, J., Agarwal, G., Cheng, J., & Thabane, L. (2009). Consent for use of personal information for health research: Do people with potentially stigmatizing health conditions and the general public differ in their opinions? *BMC medical ethics*, *10*(1), 1–12.

Xie, E., Teo, H.-H., & Wan, W. (2006). Volunteering personal information on the internet: Effects of reputation, privacy notices, and rewards on online consumer behavior. *Marketing letters*, *17*(1), 61–74.

Xu, H., Dinev, T., Smith, H. J., & Hart, P. (2008). Examining the formation of individual's privacy concerns: Toward an integrative view.

Yao, A. C.-C. (1986). How to generate and exchange secrets. *27th Annual Symposium on Foundations of Computer Science (sfcs 1986)*, 162–167.

Yu, L. (2009). Dap eng l, et al,"survey of research on anonymilization technology in data publication". *Computer Application*, 2361–2364.

Zare-Garizy, T., Fridgen, G., & Wederhake, L. (2018). A privacy preserving approach to collaborative systemic risk identification: The use-case of supply chain networks. *Security and Communication Networks*, *2018*.

Zeelenberg, M., & Pieters, R. (2007). A theory of regret regulation 1.0. *Journal of Consumer psychology*, *17*(1), 3–18.

Zhao, C., Zhao, S., Zhao, M., Chen, Z., Gao, C.-Z., Li, H., & Tan, Y.-a. (2019). Secure multi-party computation: Theory, practice and applications. *Information Sciences*, *476*, 357–372.

Ziefle, M., Halbey, J., & Kowalewski, S. (2016). Users' willingness to share data on the internet: Perceived benefits and caveats. *IoTBD*, 255–265.

Part V

Appendices

A

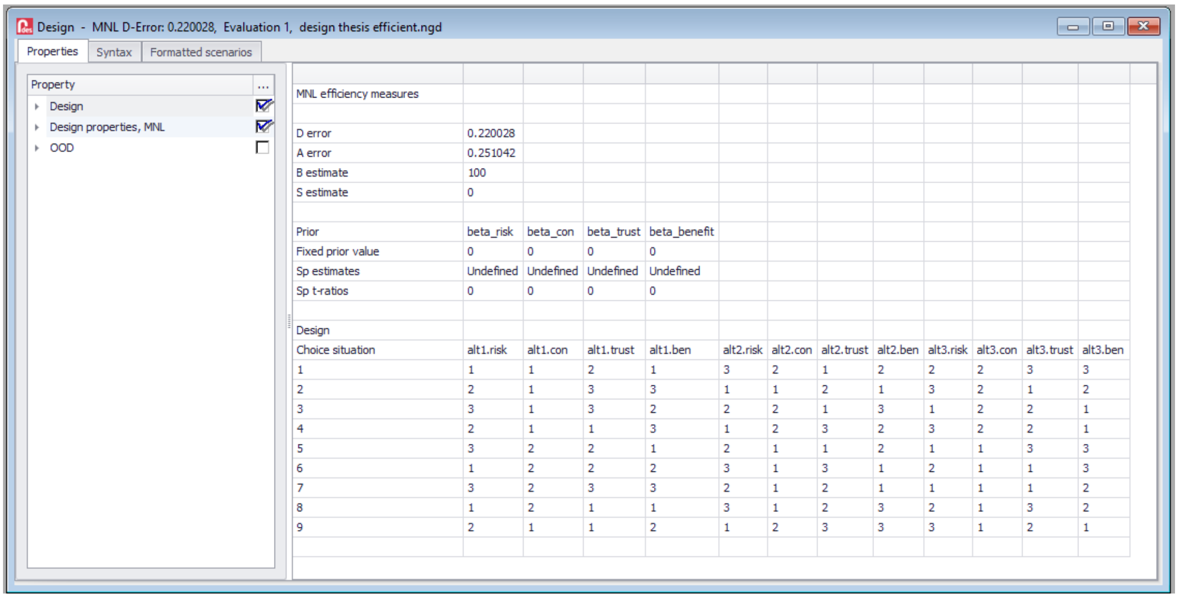
Ngene Syntax

The Ngene syntax is important, as this syntax is the basis of the efficient stated choice design where the choice sets for respondents are based on. In this design, orthogonality and attribute balance is preserved to increase validity and reliability. Furthermore, aimed is for a small design in order to avoid fatigues among respondents. Beneath, the code block is given which is the fundament of the experiment. 3 alternatives are initialized, preferred in 9 choice sets per respondent. Efficient, minimizing the D-error. Risk, trust and benefit have 3 levels. Data control has 2 levels.

```
? Thesis Experiment efficient (for unlabeled experiments)
design
;alts = alt1, alt2, alt3
;rows = 9
;eff = (mnl,d)
;model:
U(alt1) = BETA_RISK*RISK[1,2,3]+BETA_CON*CON[1,2]+BETA_TRUST*TRUST[1,2,3]+BETA_BENEFIT*BEN[1,2,3]/
U(alt2) = BETA_RISK*RISK[1,2,3]+BETA_CON*CON[1,2]+BETA_TRUST*TRUST[1,2,3]+BETA_BENEFIT*BEN[1,2,3]/
U(alt3) = BETA_RISK*RISK[1,2,3]+BETA_CON*CON[1,2]+BETA_TRUST*TRUST[1,2,3]+BETA_BENEFIT*BEN[1,2,3]
$
```

Figure A.1 shows the design which is based on the syntax above.

Figure A.1: Efficient Design



Design - MNL D-Error: 0.220028, Evaluation 1, design thesis efficient.ngd

MNL efficiency measures												
D error	0.220028											
A error	0.251042											
B estimate	100											
S estimate	0											
Prior	beta_risk	beta_con	beta_trust	beta_benefit								
Fixed prior value	0	0	0	0								
Sp estimates	Undefined	Undefined	Undefined	Undefined								
Sp t-ratios	0	0	0	0								
Design												
Choice situation	alt1.risk	alt1.con	alt1.trust	alt1.ben	alt2.risk	alt2.con	alt2.trust	alt2.ben	alt3.risk	alt3.con	alt3.trust	alt3.ben
1	1	1	2	1	3	2	1	2	2	2	3	3
2	2	1	3	3	1	1	2	1	3	2	1	2
3	3	1	3	2	2	2	1	3	1	2	2	1
4	2	1	1	3	1	2	3	2	3	2	2	1
5	3	2	2	1	2	1	1	2	1	1	3	3
6	1	2	2	2	3	1	3	1	2	1	1	3
7	3	2	3	3	2	1	2	1	1	1	1	2
8	1	2	1	1	3	1	2	3	2	1	3	2
9	2	1	1	2	1	2	3	3	3	1	2	1

B

Demographic Analyses

As the following pages are imported pdf. pages via SPSS, it is unable to edit these pages. Therefore, here an overview about which analyses to find on which of the following pages:

- Page 1 shows an One-way Anova between different age groups and the privacy statements
- Page 2 shows an One-way Anova between different education groups and the privacy statements
- Page 3 shows an One-way Anova between different experience in data marketplaces groups and the privacy statements
- Page 4 shows an One-way Anova between different groups based on categorization of Westin (1968) and the 9 conjoint questions

Descriptives

						95% Confidence Interval for Mean			
		N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
Q1 - Consumers have lost all control over how personal information is collected and used by companies.	18 - 24	211	3,6303	,99801	,06871	3,4949	3,7658	1,00	5,00
	25 - 34	147	3,8776	,91336	,07533	3,7287	4,0264	1,00	5,00
	35 - 44	53	3,8868	,86958	,11945	3,6471	4,1265	2,00	5,00
	45 - 54	11	4,1818	1,07872	,32525	3,4571	4,9065	2,00	5,00
	55 - 64	6	3,5000	,83666	,34157	2,6220	4,3780	2,00	4,00
	Total	428	3,7593	,96084	,04644	3,6681	3,8506	1,00	5,00
Q2 - Most businesses handle the personal information they collect about consumers in a proper and confidential way.	18 - 24	211	2,9621	1,00876	,06945	2,8252	3,0990	1,00	5,00
	25 - 34	147	2,7211	1,07768	,08889	2,5454	2,8968	1,00	5,00
	35 - 44	53	2,6604	1,03670	,14240	2,3746	2,9461	1,00	5,00
	45 - 54	11	2,6364	1,02691	,30963	1,9465	3,3263	1,00	4,00
	55 - 64	6	2,5000	,54772	,22361	1,9252	3,0748	2,00	3,00
	Total	428	2,8271	1,03703	,05013	2,7286	2,9256	1,00	5,00
Q3 - Existing laws and organizational practices provide a reasonable level of protection for consumer privacy today.	18 - 24	211	3,3365	,98346	,06770	3,2030	3,4700	1,00	5,00
	25 - 34	147	3,0952	1,04903	,08652	2,9242	3,2662	1,00	5,00
	35 - 44	53	3,0755	1,07147	,14718	2,7801	3,3708	1,00	5,00
	45 - 54	11	2,9091	,94388	,28459	2,2750	3,5432	1,00	4,00
	55 - 64	6	3,3333	1,03280	,42164	2,2495	4,4172	2,00	5,00
	Total	428	3,2103	1,02095	,04935	3,1133	3,3073	1,00	5,00

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Q1 - Consumers have lost all control over how personal information is collected and used by companies.	Between Groups	8,794	4	2,198	2,413	,048
	Within Groups	385,419	423	,911		
	Total	394,213	427			
Q2 - Most businesses handle the personal information they collect about consumers in a proper and confidential way.	Between Groups	8,012	4	2,003	1,878	,113
	Within Groups	451,194	423	1,067		
	Total	459,206	427			
Q3 - Existing laws and organizational practices provide a reasonable level of protection for consumer privacy today.	Between Groups	7,359	4	1,840	1,778	,132
	Within Groups	437,716	423	1,035		
	Total	445,075	427			

Descriptives

						95% Confidence Interval for Mean			
		N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
Q1 - Consumers have lost all control over how personal information is collected and used by companies.	Less than a high school diploma	14	3,5000	1,22474	,32733	2,7929	4,2071	2,00	5,00
	High school degree or equivalent (e.g. GED)	114	3,5877	,92947	,08705	3,4153	3,7602	1,00	5,00
	Some college, no degree	87	3,7471	1,02547	,10994	3,5286	3,9657	1,00	5,00
	Associate degree (e.g. AA, AS)	13	3,7692	,92681	,25705	3,2092	4,3293	2,00	5,00
	Bachelor's degree (e.g. BA, BS)	120	3,8833	,89050	,08129	3,7224	4,0443	1,00	5,00
	Master's degree (e.g. MA, MS, MEd)	69	3,9275	,97496	,11737	3,6933	4,1617	1,00	5,00
	Doctorate or professional degree (e.g. MD, DDS, PhD)	11	3,5455	,93420	,28167	2,9179	4,1731	2,00	5,00
	Total	428	3,7593	,96084	,04644	3,6681	3,8506	1,00	5,00
Q2 - Most businesses handle the personal information they collect about consumers in a proper and confidential way.	Less than a high school diploma	14	3,5000	1,16024	,31009	2,8301	4,1699	1,00	5,00
	High school degree or equivalent (e.g. GED)	114	2,9561	1,06754	,09998	2,7581	3,1542	1,00	5,00
	Some college, no degree	87	2,8276	1,09126	,11700	2,5950	3,0602	1,00	5,00
	Associate degree (e.g. AA, AS)	13	3,0769	1,11516	,30929	2,4030	3,7508	2,00	5,00
	Bachelor's degree (e.g. BA, BS)	120	2,6417	,95966	,08760	2,4682	2,8151	1,00	5,00
	Master's degree (e.g. MA, MS, MEd)	69	2,7536	,94567	,11385	2,5264	2,9808	1,00	5,00
	Doctorate or professional degree (e.g. MD, DDS, PhD)	11	2,8182	1,07872	,32525	2,0935	3,5429	2,00	5,00
	Total	428	2,8271	1,03703	,05013	2,7286	2,9256	1,00	5,00
Q3 - Existing laws and organizational practices provide a reasonable level of protection for consumer privacy today.	Less than a high school diploma	14	3,2857	,99449	,26579	2,7115	3,8599	2,00	5,00
	High school degree or equivalent (e.g. GED)	114	3,3684	1,00673	,09429	3,1816	3,5552	1,00	5,00
	Some college, no degree	87	3,2299	1,08586	,11642	2,9985	3,4613	1,00	5,00
	Associate degree (e.g. AA, AS)	13	3,2308	1,01274	,28088	2,6188	3,8428	2,00	5,00
	Bachelor's degree (e.g. BA, BS)	120	3,1333	1,02024	,09314	2,9489	3,3178	1,00	5,00
	Master's degree (e.g. MA, MS, MEd)	69	3,0725	,97496	,11737	2,8383	3,3067	1,00	5,00
	Doctorate or professional degree (e.g. MD, DDS, PhD)	11	3,0000	1,00000	,30151	2,3282	3,6718	2,00	5,00
	Total	428	3,2103	1,02095	,04935	3,1133	3,3073	1,00	5,00

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Q1 - Consumers have lost all control over how personal information is collected and used by companies.	Between Groups	8,614	6	1,436	1,567	,155
	Within Groups	385,599	421	,916		
	Total	394,213	427			
Q2 - Most businesses handle the personal information they collect about consumers in a proper and confidential way.	Between Groups	13,548	6	2,258	2,133	,049
	Within Groups	445,657	421	1,059		
	Total	459,206	427			
Q3 - Existing laws and organizational practices provide a reasonable level of protection for consumer privacy today.	Between Groups	5,477	6	,913	,874	,514
	Within Groups	439,598	421	1,044		
	Total	445,075	427			

Descriptives

						95% Confidence Interval for Mean			
		N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
Q1 - Consumers have lost all control over how personal information is collected and used by companies.	I have shared data on data markets multiple times	75	3,6400	1,03506	,11952	3,4019	3,8781	1,00	5,00
	I have shared data once on a data marketplace	38	3,5000	1,08429	,17589	3,1436	3,8564	1,00	5,00
	I know what a data marketplace is, but have never shared data on it	162	3,8395	,86988	,06834	3,7045	3,9745	1,00	5,00
	Before this survey, I had never heard of data marketplaces	153	3,7974	,97576	,07889	3,6415	3,9532	1,00	5,00
	Total	428	3,7593	,96084	,04644	3,6681	3,8506	1,00	5,00
Q2 - Most businesses handle the personal information they collect about consumers in a proper and confidential way.	I have shared data on data markets multiple times	75	3,0000	1,15079	,13288	2,7352	3,2648	1,00	5,00
	I have shared data once on a data marketplace	38	3,2368	1,07639	,17461	2,8830	3,5906	1,00	5,00
	I know what a data marketplace is, but have never shared data on it	162	2,7222	,96684	,07596	2,5722	2,8722	1,00	5,00
	Before this survey, I had never heard of data marketplaces	153	2,7516	1,01490	,08205	2,5895	2,9137	1,00	5,00
	Total	428	2,8271	1,03703	,05013	2,7286	2,9256	1,00	5,00
Q3 - Existing laws and organizational practices provide a reasonable level of protection for consumer privacy today.	I have shared data on data markets multiple times	75	3,1867	1,13535	,13110	2,9254	3,4479	1,00	5,00
	I have shared data once on a data marketplace	38	3,6579	,84714	,13742	3,3794	3,9363	2,00	5,00
	I know what a data marketplace is, but have never shared data on it	162	3,2099	1,01805	,07999	3,0519	3,3678	1,00	5,00
	Before this survey, I had never heard of data marketplaces	153	3,1111	,98379	,07953	2,9540	3,2682	1,00	5,00
	Total	428	3,2103	1,02095	,04935	3,1133	3,3073	1,00	5,00

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Q1 - Consumers have lost all control over how personal information is collected and used by companies.	Between Groups	4,887	3	1,629	1,774	,151
	Within Groups	389,326	424	,918		
	Total	394,213	427			
Q2 - Most businesses handle the personal information they collect about consumers in a proper and confidential way.	Between Groups	11,275	3	3,758	3,558	,014
	Within Groups	447,931	424	1,056		
	Total	459,206	427			
Q3 - Existing laws and organizational practices provide a reasonable level of protection for consumer privacy today.	Between Groups	9,160	3	3,053	2,970	,032
	Within Groups	435,915	424	1,028		
	Total	445,075	427			

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Conjoint 1	Between Groups	,828	2	,414	,652	,522
	Within Groups	269,983	425	,635		
	Total	270,811	427			
Conjoint 2	Between Groups	,011	2	,005	,009	,991
	Within Groups	270,176	425	,636		
	Total	270,187	427			
Conjoint 3	Between Groups	,362	2	,181	,349	,705
	Within Groups	219,945	425	,518		
	Total	220,306	427			
Conjoint 4	Between Groups	,244	2	,122	,465	,628
	Within Groups	111,270	425	,262		
	Total	111,514	427			
Conjoint 5	Between Groups	,698	2	,349	,757	,470
	Within Groups	195,891	425	,461		
	Total	196,589	427			
Conjoint 6	Between Groups	1,450	2	,725	,990	,373
	Within Groups	311,305	425	,732		
	Total	312,755	427			
Conjoint 7	Between Groups	,615	2	,307	,396	,673
	Within Groups	329,486	425	,775		
	Total	330,100	427			
Conjoint 8	Between Groups	,301	2	,150	,178	,837
	Within Groups	358,251	425	,843		
	Total	358,551	427			
Conjoint 9	Between Groups	,126	2	,063	,414	,661
	Within Groups	64,909	425	,153		
	Total	65,035	427			

C

Choice behavior modeling example

A extreme simple running example is elaborated and explained to show why choice modeling could present important findings. Suppose that there exists a transport alternative A which is a very slow and cheap route and a new transport alternative B which is very fast but also expensive (e.g. toll-road). Suppose only travel time and cost are observed. Goal:

1. Determine the market share of a new transport alternative.
2. Determine the economic appraisal (welfare analysis). Is the toll road welfare enhancing?
   (a) Depends on aggregate travel time gains on transport network.
   (b) And how much are these worth.

   Both depend on one crucial insight:

   How do travellers weigh (trade-offs) travel times and costs (the factors) ?

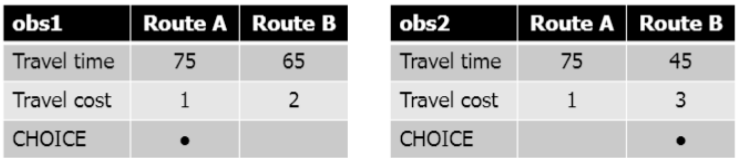
Business perspective:
If one minute of travel time gain is worth very little, then:
Market share of toll road will be low, this raises profitability issues
Prices will have to be kept low, which also raises profitability issues

Government perspective:
If one minute of travel time gain is worth very little, then:
Few travelers benefit from the new service (since small market)
So: small overall benefits (no network effect, low value of time)

The data: choice observations
1. Revealed choice data (observations in real life)
2. Stated choice data (hypothetical observations)

With the choice data, before moving to mathematics, we can use these to:
- Infer trade-off between time and cost ('value of travel time savings'- VoTTS)
- Predict choices for other levels of time and cost

Figure C.1: Simple Example 1



obs1	Route A	Route B
Travel time	75	65
Travel cost	1	2
CHOICE	●	

obs2	Route A	Route B
Travel time	75	45
Travel cost	1	3
CHOICE		●

As you look at the figure above, these are just two observations of a single person in a survey. From the first observation, it is seen that route B is 10 minutes faster and 1 euro more expensive; however, the choice is on A. Hence, VoTTS < 1 euro per 10 minutes (or < 6 euro per hour).

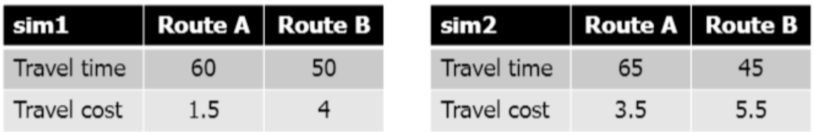
From the second observation, it is seen that route B is 30 minutes faster and 2 euro more expensive; the choice is B. Hence, VoTTS > 2 euro per 30 minutes (or > 4 euro per hour). So now we know for this person that his or her VoTTS is somewhere between 4 and 6 euros per hour. We can 'pinpoint' a value in this way. By increasing observations, we can even better and better pinpoint this value and eventually use these to estimate choices and basing strategies on that.

But why do we not ask respondents to tell their value or in this example their VoTTS directly? That is due to the following:

1. People do not know. Evolution did not program us to explicate trade-offs.
2. In many cases, people hesitate to give true trade-off. E.g. Refugee data: preferred distance to one's home depends on religion
3. Judgment is known to be much more susceptible to bias than choices. E.g. cognitive dissonance, ex-post rationalization, pleasing interviewer
4. Most Revealed Preference data are choices. Not trade-off judgements.
5. Economic theory is based on choices, not judgements. Demand-supply machinery, consumer surplus..

So these choices can be predicted. Suppose that the VoTTS is estimated to be 8 euro per hour. Then, choices for A and B for all combinations of time and cost can be predicted or simulated:

Figure C.2: Simple Example 2



sim1	Route A	Route B
Travel time	60	50
Travel cost	1.5	4

sim2	Route A	Route B
Travel time	65	45
Travel cost	3.5	5.5

Sim1: B gives 10 min gain for 2.50 euro, this is 15 euro per hour > VoTTS; hence choice is for A. Sim2: B gives 20 min gain for 2.00 euro, this is 6 euro per hour < VoTTS; hence choice is for B.

In the next subsections, each model is explained concisely.

D

Multinominal logit model (MNL)

All models are modified course codes of EPA-course SEN1221 by C. Chorus (2017) and Molin (2017).

```
### Load Apollo library
library(apollo)

### Initialise code
apollo_initialise()

### Set core controls
apollo_control = list(
  modelName  ="MNL_1_maintest",
  modelDescr ="First Try 428 respondents on Qualtrics by MNL",
  indivID    ="ID"
)

#### LOAD DATA
database = read.delim("cleaned_data_real.dat",header=TRUE)


### Vector of parameters, including any that are kept fixed in estimation
apollo_beta=c(BETA_RISK   = 0, #risk
              BETA_DATA_CONTROL   = 0,  #data control
              BETA_TRUST  = 0,    #trust
              BETA_BENEFIT   = 0)   #benefit

### Vector with names (in quotes) of parameters to be kept fixed at their starting value in
### apollo_beta, use apollo_beta_fixed = c() if none
apollo_fixed = c()

#### GROUP AND VALIDATE INPUTS
apollo_inputs = apollo_validateInputs()

#### DEFINE MODEL AND LIKELIHOOD FUNCTION

apollo_probabilities=function(apollo_beta, apollo_inputs, functionality="estimate"){

  ### Attach inputs and detach after function exit
  apollo_attach(apollo_beta, apollo_inputs)
  on.exit(apollo_detach(apollo_beta, apollo_inputs))
  ### Create list of probabilities P
  P = list()

  ### List of utilities: these must use the same names as in mnl_settings, order is irrelevant
  V = list()
    V[['A']]  = RISKA * BETA_RISK + CONTA * BETA_DATA_CONTROL + TRUSTA * BETA_TRUST
    + BENA * BETA_BENEFIT
    V[['B']]  = RISKB * BETA_RISK + CONTB * BETA_DATA_CONTROL + TRUSTB * BETA_TRUST
    + BENB * BETA_BENEFIT
    V[['C']]  = RISKC * BETA_RISK + CONTC * BETA_DATA_CONTROL + TRUSTC * BETA_TRUST
    + BENC * BETA_BENEFIT


    ### Define settings for MNL model component
    mnl_settings = list(
      alternatives  = c(A=1, B=2, C=3),
      avail         = list(A=1, B=1, C=1),
      choiceVar     = CHOICE,
      V             = V
    )

  ### Compute probabilities using MNL model
  P[['model']] = apollo_mnl(mnl_settings, functionality)

  ### Take product across observation for same individual
  P = apollo_panelProd(P, apollo_inputs, functionality)

  ### Prepare and return outputs of function
  P = apollo_prepareProb(P, apollo_inputs, functionality)
  return(P)
}

#### MODEL ESTIMATION
model = apollo_estimate(apollo_beta, apollo_fixed, apollo_probabilities, apollo_inputs)

#### MODEL OUTPUTS
apollo_modelOutput(model,modelOutput_settings=list(printPVal=TRUE))

apollo_saveOutput(model)
```

E

Multinominal logit model + Westin interaction (MNL + interaction)

All models are modified course codes of EPA-course SEN1221 by C. Chorus (2017) and Molin (2017).

```
### Load Apollo library
library(apollo)

### Initialise code
apollo_initialise()

### Set core controls
apollo_control = list(
  modelName  ="MNL_1_maintest",
  modelDescr ="First Try 428 respondents on Qualtrics by MNL",
  indivID    ="ID"
)

#### LOAD DATA
database = read.delim("cleaned_and_PS_Male_Westin_I.dat",header=TRUE)


### Vector of parameters, including any that are kept fixed in estimation
apollo_beta=c(BETA_RISK   = 0, #risk
              BETA_DATA_CONTROL   = 0,  #data control
              BETA_TRUST  = 0,    #trust
              BETA_BENEFIT   = 0, #benefit
              int_Risk = 0,
              int_DC = 0,
              int_Trust = 0,
              int_Benefit = 0
              )

### Vector with names (in quotes) of parameters to be kept fixed at their starting value in apollo_b
### use apollo_beta_fixed = c() if none
apollo_fixed = c()


#### GROUP AND VALIDATE INPUTS
apollo_inputs = apollo_validateInputs()
```

```
#### DEFINE MODEL AND LIKELIHOOD FUNCTION
apollo_probabilities=function(apollo_beta, apollo_inputs, functionality="estimate"){

  ### Attach inputs and detach after function exit
  apollo_attach(apollo_beta, apollo_inputs)
  on.exit(apollo_detach(apollo_beta, apollo_inputs))

  ### Create list of probabilities P
  P = list()

  ### List of utilities: these must use the same names as in mnl_settings, order is irrelevant
  V = list()
    V[['A']]  = (int_Risk*W_Index + BETA_RISK)*RISKA + CONTA * (BETA_DATA_CONTROL + int_DC*W_Index)
    + TRUSTA * (BETA_TRUST + int_Trust*W_Index) + BENA * (BETA_BENEFIT + int_Benefit*W_Index)
    V[['B']]  = (int_Risk*W_Index + BETA_RISK)*RISKB + CONTB * (BETA_DATA_CONTROL + int_DC*W_Index)
    + TRUSTB * (BETA_TRUST + int_Trust*W_Index) + BENB * (BETA_BENEFIT + int_Benefit*W_Index)
    V[['C']]  = (int_Risk*W_Index + BETA_RISK)*RISKC + CONTC * (BETA_DATA_CONTROL + int_DC*W_Index)
    + TRUSTC * (BETA_TRUST + int_Trust*W_Index) + BENC * (BETA_BENEFIT + int_Benefit*W_Index)

    ### Define settings for MNL model component
    mnl_settings = list(
      alternatives  = c(A=1, B=2, C=3),
      avail         = list(A=1, B=1, C=1),
      choiceVar     = CHOICE,
      V             = V
    )

  ### Compute probabilities using MNL model
  P[['model']] = apollo_mnl(mnl_settings, functionality)

  ### Take product across observation for same individual
  P = apollo_panelProd(P, apollo_inputs, functionality)

  ### Prepare and return outputs of function
  P = apollo_prepareProb(P, apollo_inputs, functionality)
  return(P)
}

#### MODEL ESTIMATION
model = apollo_estimate(apollo_beta, apollo_fixed, apollo_probabilities, apollo_inputs)

#### MODEL OUTPUTS
apollo_modelOutput(model,modelOutput_settings=list(printPVal=TRUE))

apollo_saveOutput(model)
```

F Mixed-Logit Model (ML)

All models are modified course codes of EPA-course SEN1221 by C. Chorus (2017) and Molin (2017).

```
### Load Apollo library
library(apollo)

### Initialise code
apollo_initialise()

### Set core controls
apollo_control = list(
  modelName  ="ML Test",
  modelDescr ="ML Test data ",
  indivID    ="ID",
  panelData = TRUE,
  mixing    = TRUE,
  nCores=5
)


#### LOAD DATA
database = read.delim("cleaned_and_PS.dat",header=TRUE)

### Vector of parameters, including any that are kept fixed in estimation
apollo_beta=c(BETA_RISK   = 0, #risk
              BETA_DATA_CONTROL   = 0,  #data control
              BETA_TRUST  = 0,    #trust
              BETA_BENEFIT   = 0, #benefit
              Sigma_benefit = 1,
              Sigma_trust = 1,
              Sigma_dc = 1,
              Sigma_risk = 1
             )

### Vector with names (in quotes) of parameters to be kept fixed at their starting value
### in apollo_beta, use apollo_beta_fixed = c() if none
apollo_fixed = c()

### Set parameters for generating draws
apollo_draws = list(
  interDrawsType = "halton",
  interNDraws    = 500,
  interUnifDraws = c(),
  interNormDraws = c("draws_risk", "draws_dc","draws_trust","draws_benefit"),
  intraDrawsType = "halton",
  intraNDraws    = 0,
  intraUnifDraws = c(),
  intraNormDraws = c()
)

### Create random parameters
apollo_randCoeff = function(apollo_beta, apollo_inputs){
  randcoeff = list()

  randcoeff[["heterobenefit"]] =  BETA_BENEFIT + Sigma_benefit * draws_benefit
  randcoeff[["heterotrust"]] =  BETA_TRUST + Sigma_trust * draws_trust
  randcoeff[["heterodc"]] =  BETA_DATA_CONTROL + Sigma_dc * draws_dc
  randcoeff[["heterorisk"]] =  BETA_RISK + Sigma_risk * draws_risk
  return(randcoeff)
}


#### GROUP AND VALIDATE INPUTS
apollo_inputs = apollo_validateInputs()


#### DEFINE MODEL AND LIKELIHOOD FUNCTION
apollo_probabilities=function(apollo_beta, apollo_inputs, functionality="estimate"){

  ### Attach inputs and detach after function exit
  apollo_attach(apollo_beta, apollo_inputs)
  on.exit(apollo_detach(apollo_beta, apollo_inputs))

  ### Create list of probabilities P
  P = list()

  ### List of utilities: these must use the same names as in mnl_settings, order is irrelevant
  V = list()
  V[['A']]  = heterorisk *  RISKA + heterodc * CONTA +  heterotrust * TRUSTA + BENA * heterobenefit
  V[['B']]  = heterorisk *  RISKB + heterodc * CONTB +  heterotrust * TRUSTB + BENB * heterobenefit
  V[['C']]  = heterorisk *  RISKC + heterodc * CONTC +  heterotrust * TRUSTC + BENC * heterobenefit


  ### Define settings for MNL model component
  mnl_settings = list(
    alternatives  = c(A=1, B=2, C=3),
    avail         = list(A=1, B=1, C=1),
    choiceVar     = CHOICE,
    V             = V
  )

  ### Compute probabilities using MNL model
  P[['model']] = apollo_mnl(mnl_settings, functionality)

  ### Take product across observation for same individual
  P = apollo_panelProd(P, apollo_inputs, functionality)
  ### Average across inter-individual draws
  P = apollo_avgInterDraws(P, apollo_inputs, functionality)

  ### Prepare and return outputs of function
  P = apollo_prepareProb(P, apollo_inputs, functionality)
  return(P)
}


#### MODEL ESTIMATION
model = apollo_estimate(apollo_beta, apollo_fixed,
        apollo_probabilities, apollo_inputs, estimate_settings=list(hessianRoutine="maxLik"))


#### MODEL OUTPUTS
apollo_modelOutput(model,modelOutput_settings=list(printPVal=TRUE))

apollo_saveOutput(model)
```

G
Random Regret Minimization Model (RRM)

All models are modified course codes of EPA-course SEN1221 by C. Chorus (2017) and Molin (2017).

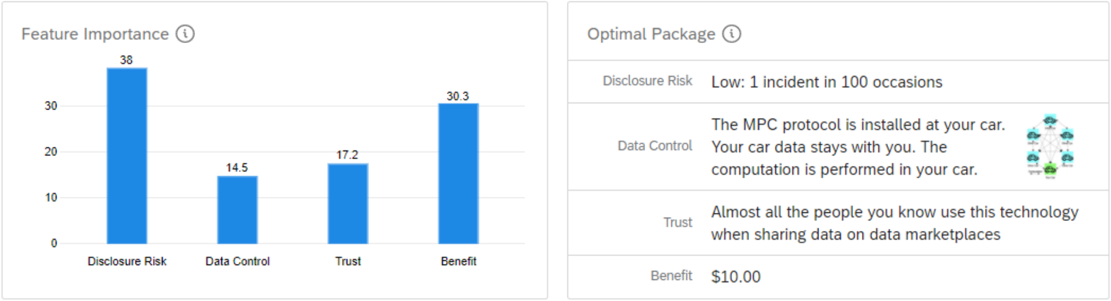
```
### Load Apollo library
library(apollo)

### Initialise code
apollo_initialise()

### Set core controls
apollo_control = list(
  modelName  ="RRM_MNL_test",
  modelDescr ="MNL RRM model of real test data",
  indivID    ="ID"
)

#### LOAD DATA
database = read.delim("cleaned_data_real.dat",header=TRUE)

### Vector of parameters, including any that are kept fixed in estimation
apollo_beta=c(BETA_RISK   = 0, #risk
              BETA_DATA_CONTROL   = 0,  #data control
              BETA_TRUST  = 0,    #trust
              BETA_BENEFIT   = 0)   #benefit

### Vector with names (in quotes) of parameters to be kept fixed at their starting value
### in apollo_beta, use apollo_beta_fixed = c() if none
apollo_fixed = c()

#### VALIDATE INPUTS
apollo_inputs = apollo_validateInputs()

#### DEFINE MODEL AND LIKELIHOOD FUNCTION
apollo_probabilities=function(apollo_beta, apollo_inputs, functionality="estimate"){

  ### Attach inputs and detach after function exit
  apollo_attach(apollo_beta, apollo_inputs)
  on.exit(apollo_detach(apollo_beta, apollo_inputs))
  ### Create list of probabilities P
  P = list()

   ### List of regret functions
  V = list()
  V[['A']]  = -(log(1 + exp(BETA_RISK * (RISKB - RISKA) )) +
               log(1 + exp(BETA_DATA_CONTROL * (CONTB - CONTA) )) +
               log(1 + exp(BETA_TRUST * (TRUSTB - TRUSTA) )) +
               log(1 + exp(BETA_BENEFIT * (BENB - BENA) )) +
               log(1 + exp(BETA_RISK * (RISKC - RISKA) )) +
               log(1 + exp(BETA_DATA_CONTROL * (CONTC - CONTA) )) +
               log(1 + exp(BETA_TRUST * (TRUSTC - TRUSTA) )) +
               log(1 + exp(BETA_BENEFIT * (BENC - BENA) )))

  V[['B']]  = -(log(1 + exp(BETA_RISK * (RISKA - RISKB) )) +
               log(1 + exp(BETA_DATA_CONTROL * (CONTA - CONTB) )) +
               log(1 + exp(BETA_TRUST * (TRUSTA - TRUSTB) )) +
               log(1 + exp(BETA_BENEFIT * (BENA - BENB) )) +
               log(1 + exp(BETA_RISK * (RISKC - RISKB) )) +
               log(1 + exp(BETA_DATA_CONTROL * (CONTC - CONTB) )) +
               log(1 + exp(BETA_TRUST * (TRUSTC - TRUSTB) )) +
               log(1 + exp(BETA_BENEFIT * (BENC - BENB) )))

  V[['C']]  = -(log(1 + exp(BETA_RISK * (RISKA - RISKC) )) +
               log(1 + exp(BETA_DATA_CONTROL * (CONTA - CONTC) )) +
               log(1 + exp(BETA_TRUST * (TRUSTA - TRUSTC) )) +
               log(1 + exp(BETA_BENEFIT * (BENA - BENC) )) +
               log(1 + exp(BETA_RISK * (RISKB - RISKC) )) +
               log(1 + exp(BETA_DATA_CONTROL * (CONTB - CONTC) )) +
               log(1 + exp(BETA_TRUST * (TRUSTB - TRUSTC) )) +
               log(1 + exp(BETA_BENEFIT * (BENB - BENC) )))

  ### Define settings for MNL model component
  mnl_settings = list(
    alternatives  = c(A=1, B=2, C=3),
    avail         = list(A=1, B=1, C=1),
    choiceVar     = CHOICE,
    V             = V
  )

  ### Compute probabilities using MNL model
  P[['model']] = apollo_mnl(mnl_settings, functionality)

  ### Take product across observation for same individual
  P = apollo_panelProd(P, apollo_inputs, functionality)

  ### Prepare and return outputs of function
  P = apollo_prepareProb(P, apollo_inputs, functionality)
  return(P)
}

#### MODEL ESTIMATION
model = apollo_estimate(apollo_beta, apollo_fixed, apollo_probabilities, apollo_inputs)
#### MODEL OUTPUTS
apollo_modelOutput(model,modelOutput_settings=list(printPVal=TRUE))
apollo_saveOutput(model)
```

H

Utility Scores

In this appendix, all utility scores and willingness to pay per attribute level are visualized to check for linearity and what deterioration for each level is worth in terms of money.

Figure H.1: Feature importance 2



Disclosure Risk	Low: 1 incident in 100 occasions
Data Control	The MPC protocol is installed at your car. Your car data stays with you. The computation is performed in your car.
Trust	Almost all the people you know use this technology when sharing data on data marketplaces
Benefit	$10.00

Figure H.2: Risk & Data control utility scores per level

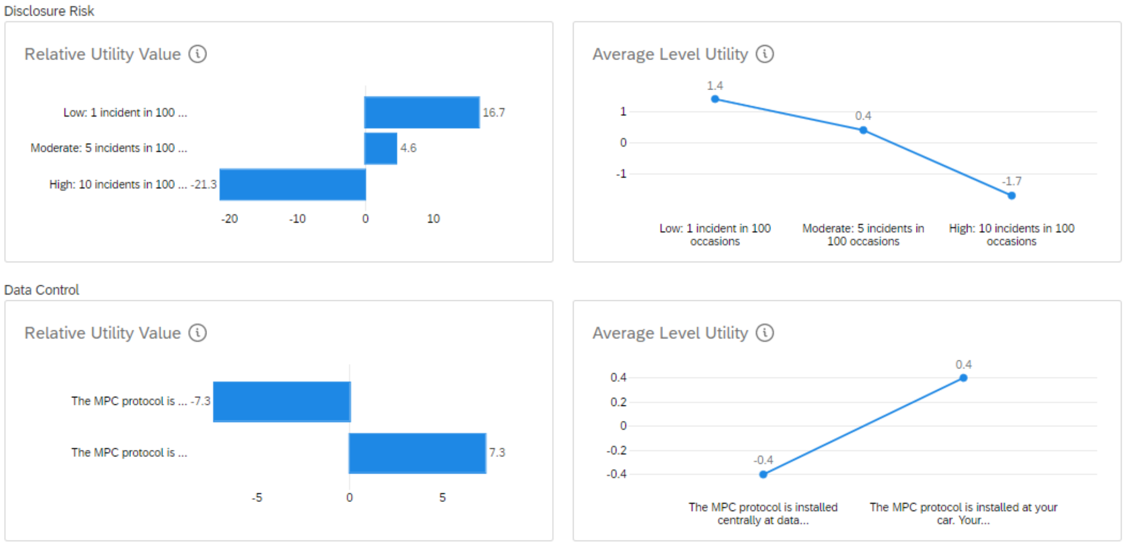


Figure H.3: Trust & Benefit utility scores per level

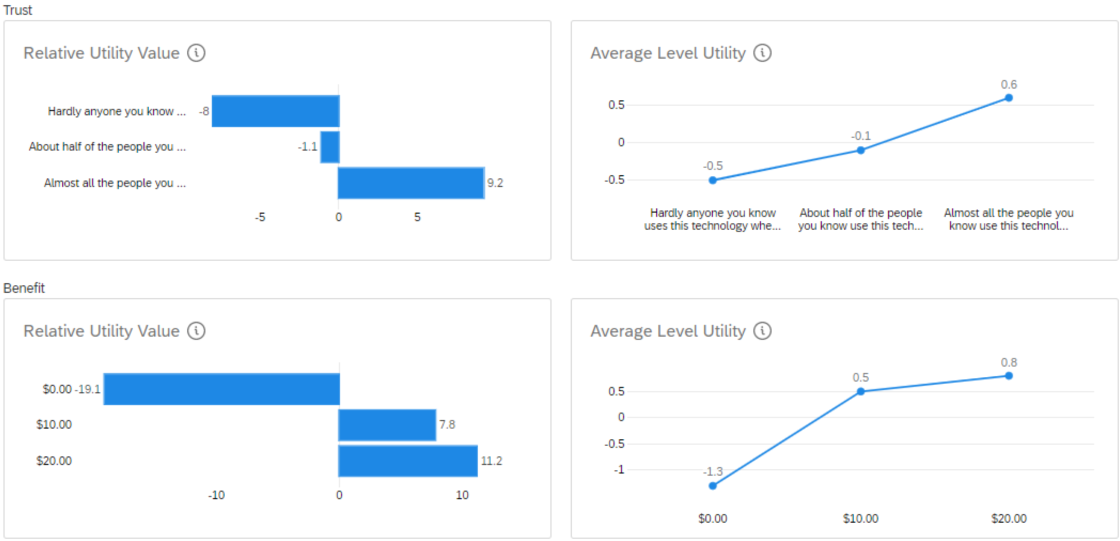
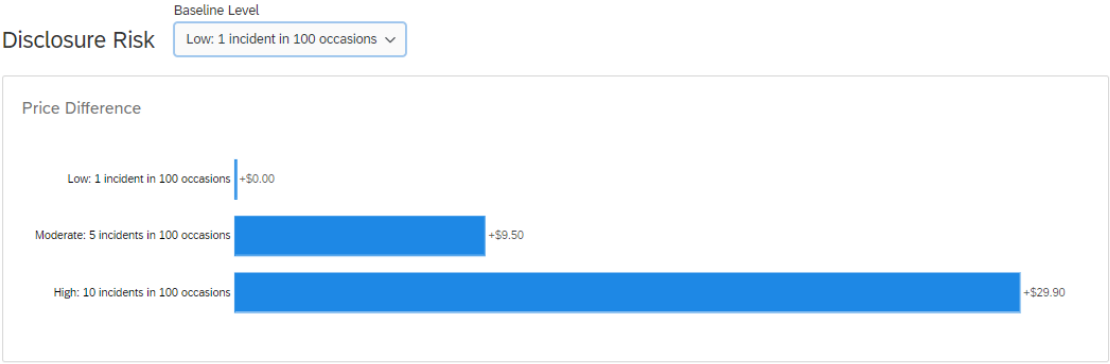
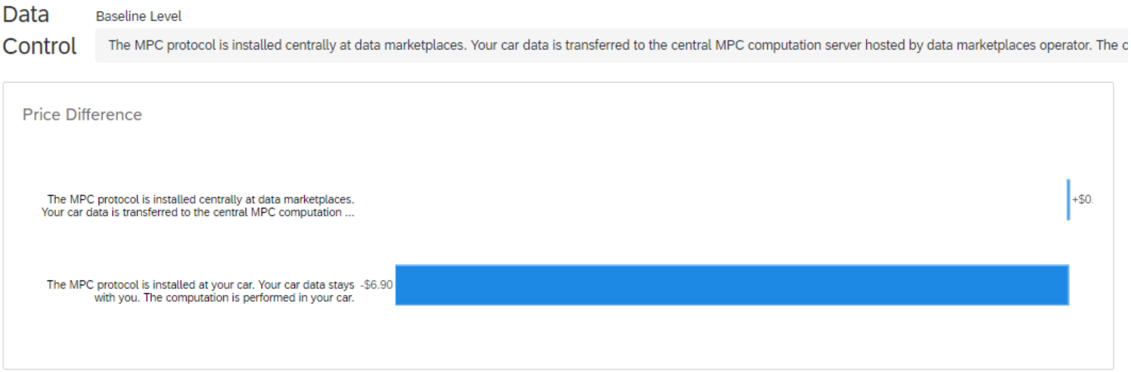


Figure H.4: WtP Risk


Disclosure Risk

Baseline Level: Low: 1 incident in 100 occasions

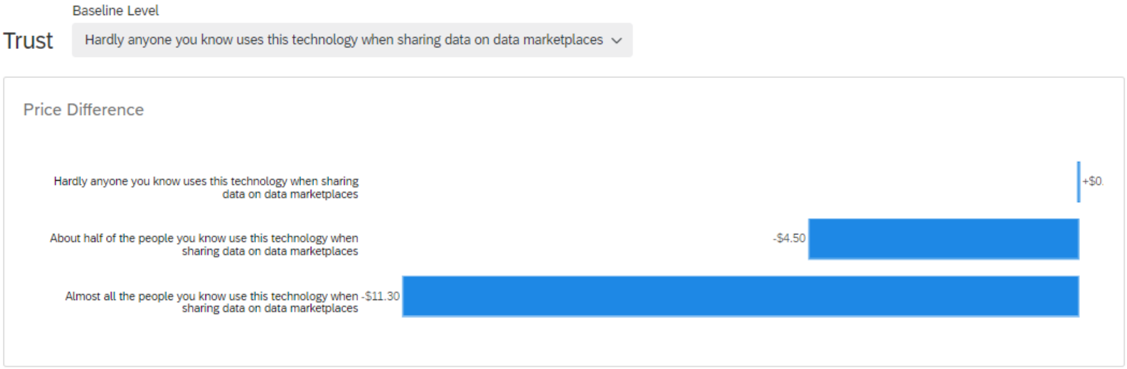
Figure H.5: WtP Data control


Data Control

Baseline Level

The MPC protocol is installed centrally at data marketplaces. Your car data is transferred to the central MPC computation server hosted by data marketplaces operator. The c

Figure H.6: WtP Trust



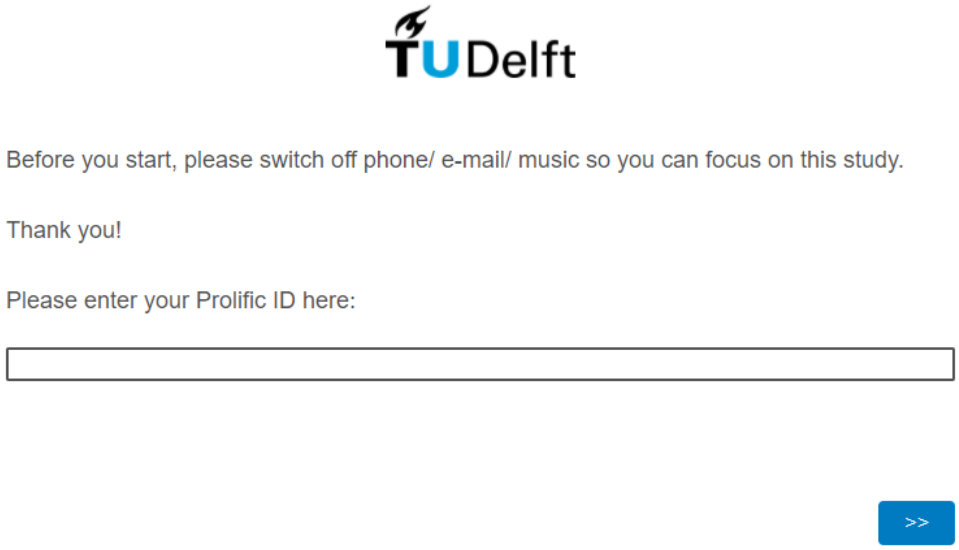
Trust

Baseline Level: Hardly anyone you know uses this technology when sharing data on data marketplaces

Price Difference
Hardly anyone you know uses this technology when sharing data on data marketplaces
+$0
About half of the people you know use this technology when sharing data on data marketplaces
-$4.50
Almost all the people you know use this technology when sharing data on data marketplaces
-$11.30

I Qualtrics Survey

Figure I.1: Survey Page 1

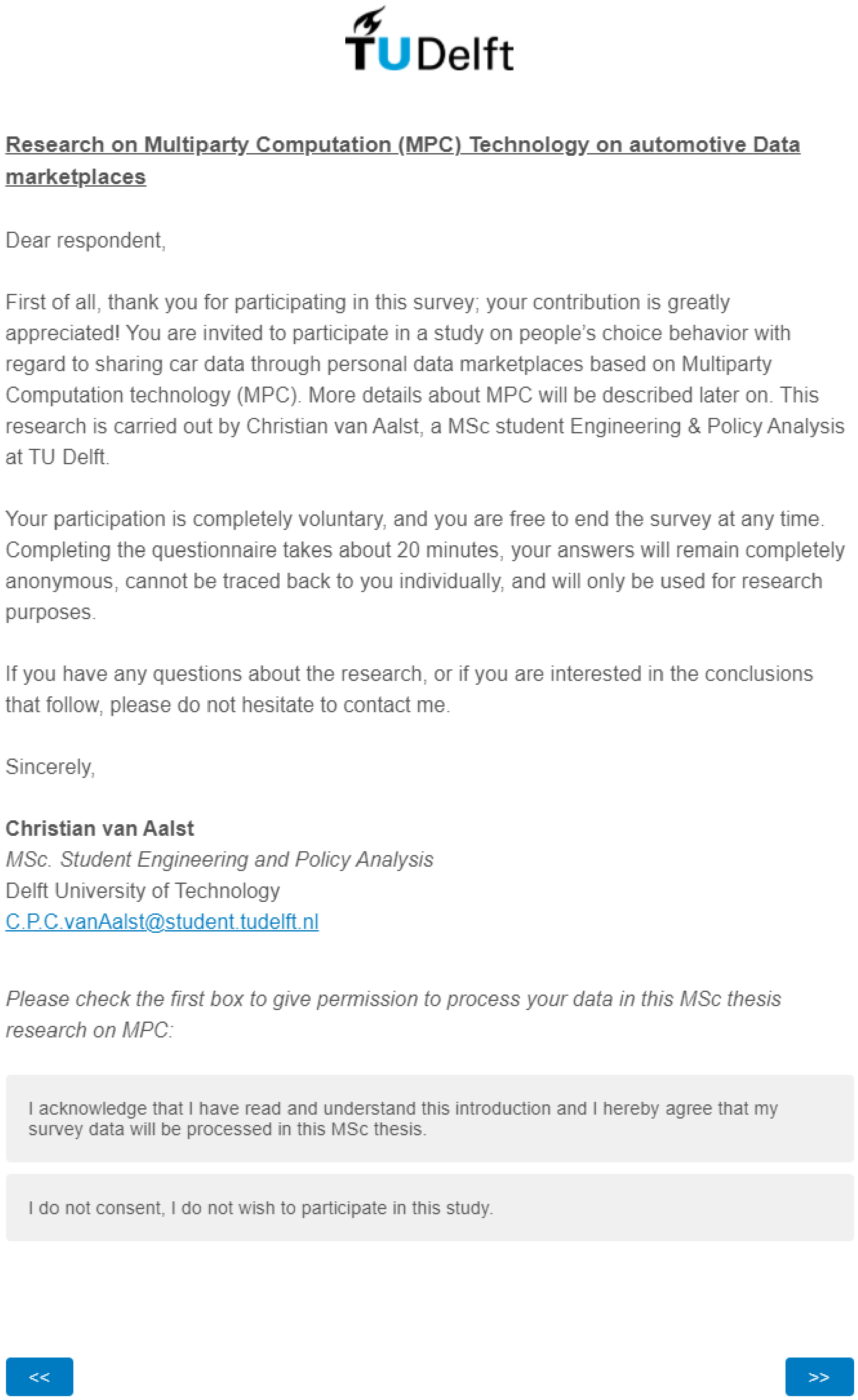


Before you start, please switch off phone/ e-mail/ music so you can focus on this study.

Thank you!

Please enter your Prolific ID here:

Figure I.2: Survey Page 2



Research on Multiparty Computation (MPC) Technology on automotive Data marketplaces

Dear respondent,

First of all, thank you for participating in this survey; your contribution is greatly appreciated! You are invited to participate in a study on people's choice behavior with regard to sharing car data through personal data marketplaces based on Multiparty Computation technology (MPC). More details about MPC will be described later on. This research is carried out by Christian van Aalst, a MSc student Engineering & Policy Analysis at TU Delft.

Your participation is completely voluntary, and you are free to end the survey at any time. Completing the questionnaire takes about 20 minutes, your answers will remain completely anonymous, cannot be traced back to you individually, and will only be used for research purposes.

If you have any questions about the research, or if you are interested in the conclusions that follow, please do not hesitate to contact me.

Sincerely,

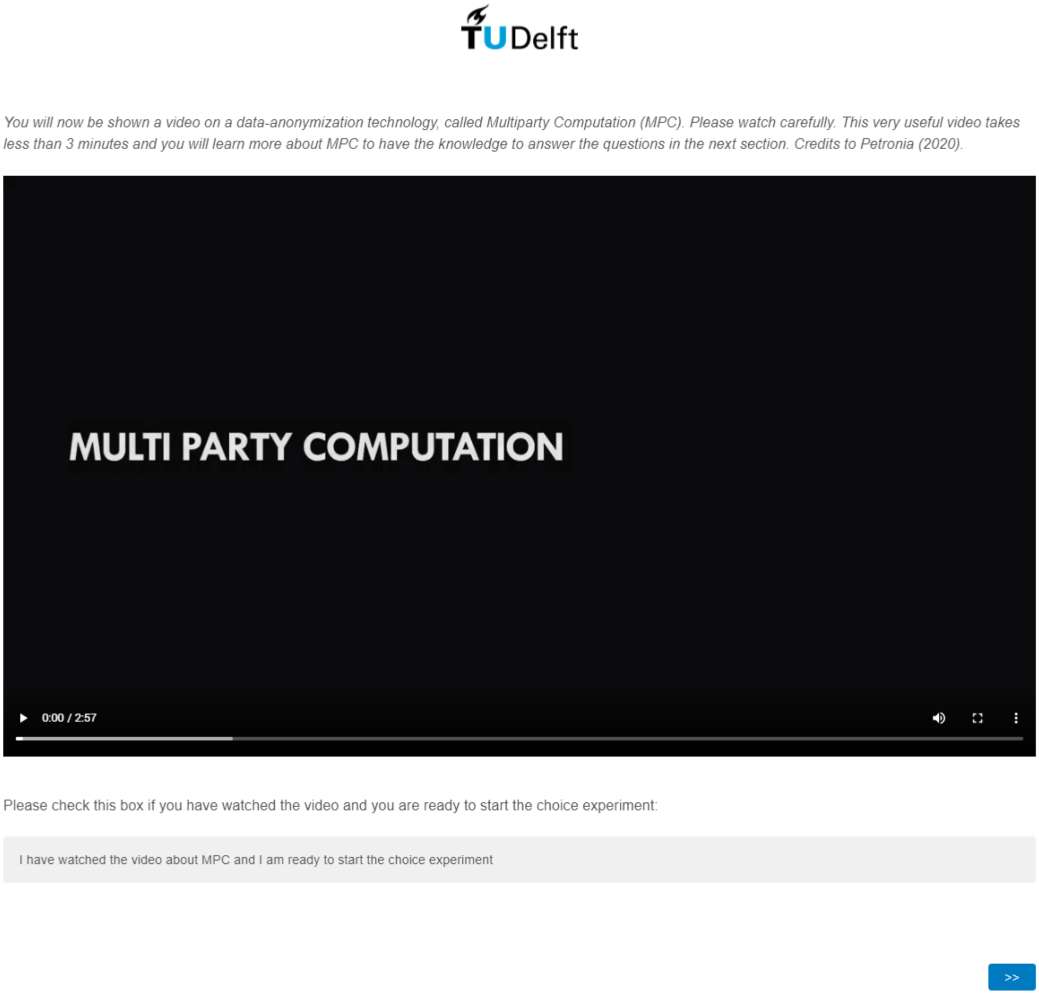
Christian van Aalst
MSc. Student Engineering and Policy Analysis
Delft University of Technology
C.P.C.vanAalst@student.tudelft.nl

Please check the first box to give permission to process your data in this MSc thesis research on MPC:

I acknowledge that I have read and understand this introduction and I hereby agree that my survey data will be processed in this MSc thesis.

I do not consent, I do not wish to participate in this study.

Figure I.3: Survey Page 3

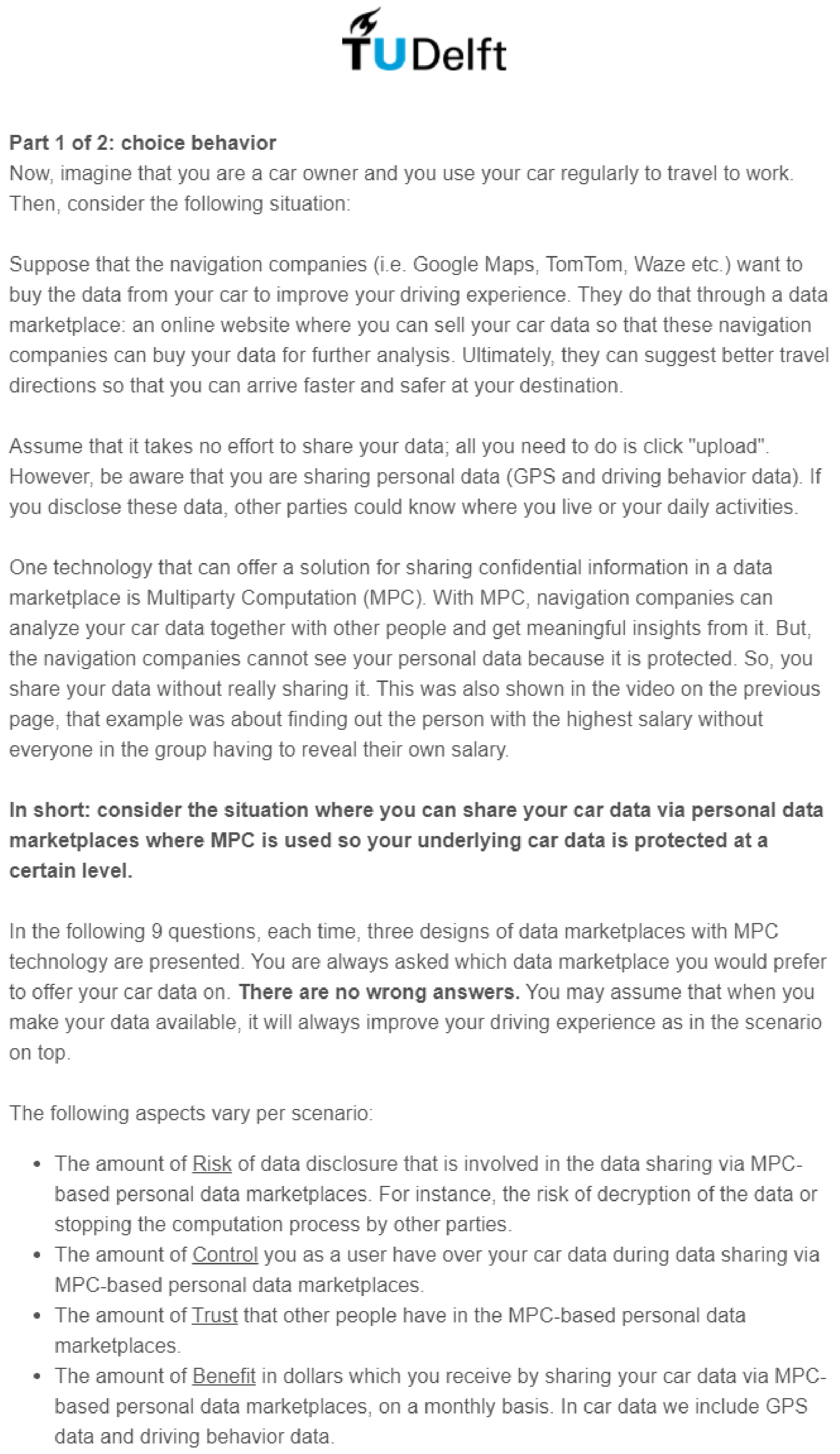


You will now be shown a video on a data-anonymization technology, called Multiparty Computation (MPC). Please watch carefully. This very useful video takes less than 3 minutes and you will learn more about MPC to have the knowledge to answer the questions in the next section. Credits to Petronia (2020).

Please check this box if you have watched the video and you are ready to start the choice experiment:

I have watched the video about MPC and I am ready to start the choice experiment

Figure I.4: Survey Page 4



Part 1 of 2: choice behavior
Now, imagine that you are a car owner and you use your car regularly to travel to work. Then, consider the following situation:

Suppose that the navigation companies (i.e. Google Maps, TomTom, Waze etc.) want to buy the data from your car to improve your driving experience. They do that through a data marketplace: an online website where you can sell your car data so that these navigation companies can buy your data for further analysis. Ultimately, they can suggest better travel directions so that you can arrive faster and safer at your destination.

Assume that it takes no effort to share your data; all you need to do is click "upload". However, be aware that you are sharing personal data (GPS and driving behavior data). If you disclose these data, other parties could know where you live or your daily activities.

One technology that can offer a solution for sharing confidential information in a data marketplace is Multiparty Computation (MPC). With MPC, navigation companies can analyze your car data together with other people and get meaningful insights from it. But, the navigation companies cannot see your personal data because it is protected. So, you share your data without really sharing it. This was also shown in the video on the previous page, that example was about finding out the person with the highest salary without everyone in the group having to reveal their own salary.

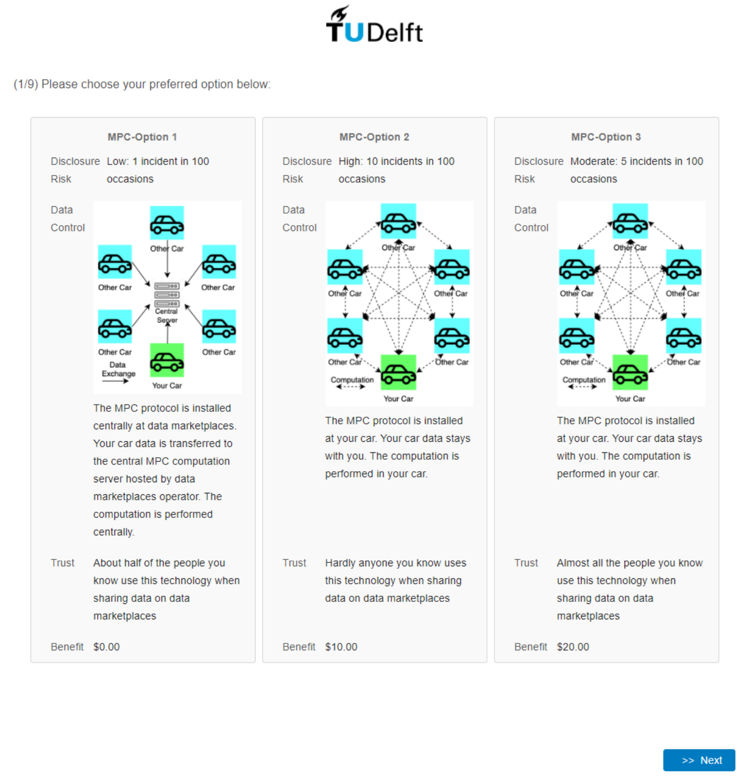
In short: consider the situation where you can share your car data via personal data marketplaces where MPC is used so your underlying car data is protected at a certain level.

In the following 9 questions, each time, three designs of data marketplaces with MPC technology are presented. You are always asked which data marketplace you would prefer to offer your car data on. **There are no wrong answers.** You may assume that when you make your data available, it will always improve your driving experience as in the scenario on top.

The following aspects vary per scenario:

- The amount of Risk of data disclosure that is involved in the data sharing via MPC-based personal data marketplaces. For instance, the risk of decryption of the data or stopping the computation process by other parties.
- The amount of Control you as a user have over your car data during data sharing via MPC-based personal data marketplaces.
- The amount of Trust that other people have in the MPC-based personal data marketplaces.
- The amount of Benefit in dollars which you receive by sharing your car data via MPC-based personal data marketplaces, on a monthly basis. In car data we include GPS data and driving behavior data.

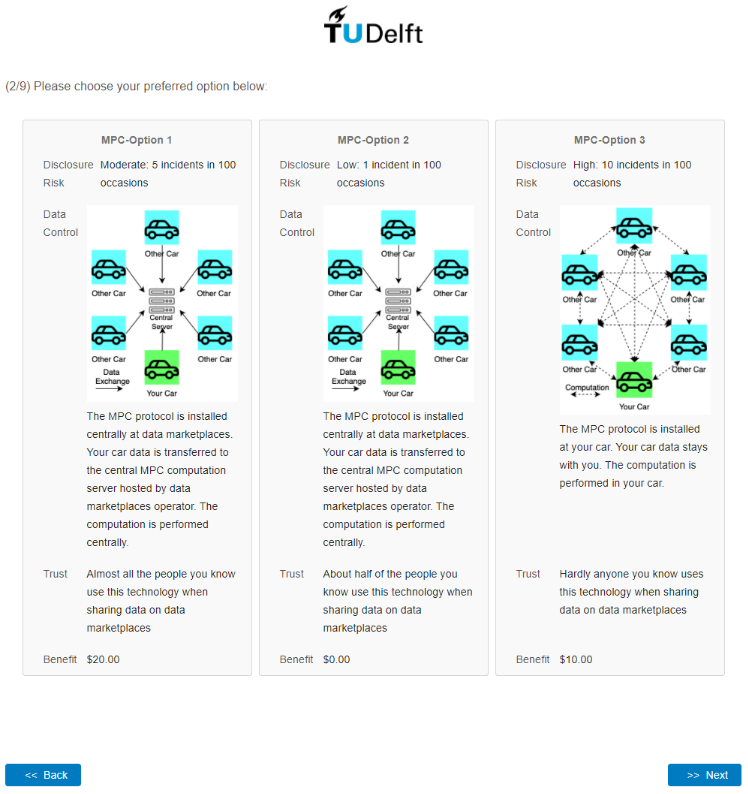
Figure I.5: Survey Page 5



(1/9) Please choose your preferred option below:

	MPC-Option 1	MPC-Option 2	MPC-Option 3
Disclosure Risk	Low: 1 incident in 100 occasions	High: 10 incidents in 100 occasions	Moderate: 5 incidents in 100 occasions
Data Control	The MPC protocol is installed centrally at data marketplaces. Your car data is transferred to the central MPC computation server hosted by data marketplaces operator. The computation is performed centrally.	The MPC protocol is installed at your car. Your car data stays with you. The computation is performed in your car.	The MPC protocol is installed at your car. Your car data stays with you. The computation is performed in your car.
Trust	About half of the people you know use this technology when sharing data on data marketplaces	Hardly anyone you know uses this technology when sharing data on data marketplaces	Almost all the people you know use this technology when sharing data on data marketplaces
Benefit	$0.00	$10.00	$20.00

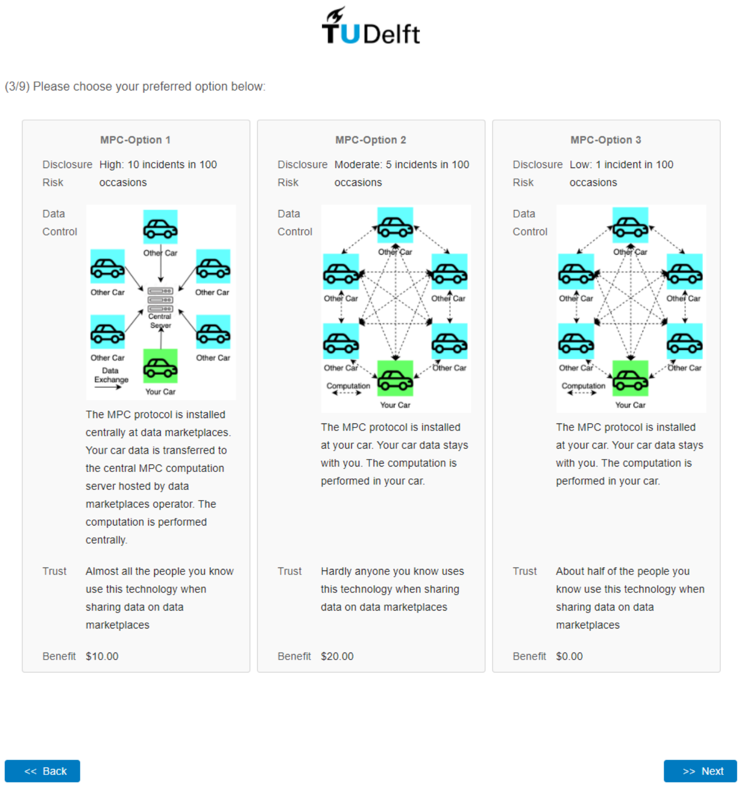
Figure I.6: Survey Page 6



(2/9) Please choose your preferred option below:

	MPC-Option 1	MPC-Option 2	MPC-Option 3
Disclosure Risk	Moderate: 5 incidents in 100 occasions	Low: 1 incident in 100 occasions	High: 10 incidents in 100 occasions
Data Control	The MPC protocol is installed centrally at data marketplaces. Your car data is transferred to the central MPC computation server hosted by data marketplaces operator. The computation is performed centrally.	The MPC protocol is installed centrally at data marketplaces. Your car data is transferred to the central MPC computation server hosted by data marketplaces operator. The computation is performed centrally.	The MPC protocol is installed at your car. Your car data stays with you. The computation is performed in your car.
Trust	Almost all the people you know use this technology when sharing data on data marketplaces	About half of the people you know use this technology when sharing data on data marketplaces	Hardly anyone you know uses this technology when sharing data on data marketplaces
Benefit	$20.00	$0.00	$10.00

Figure I.7: Survey Page 7

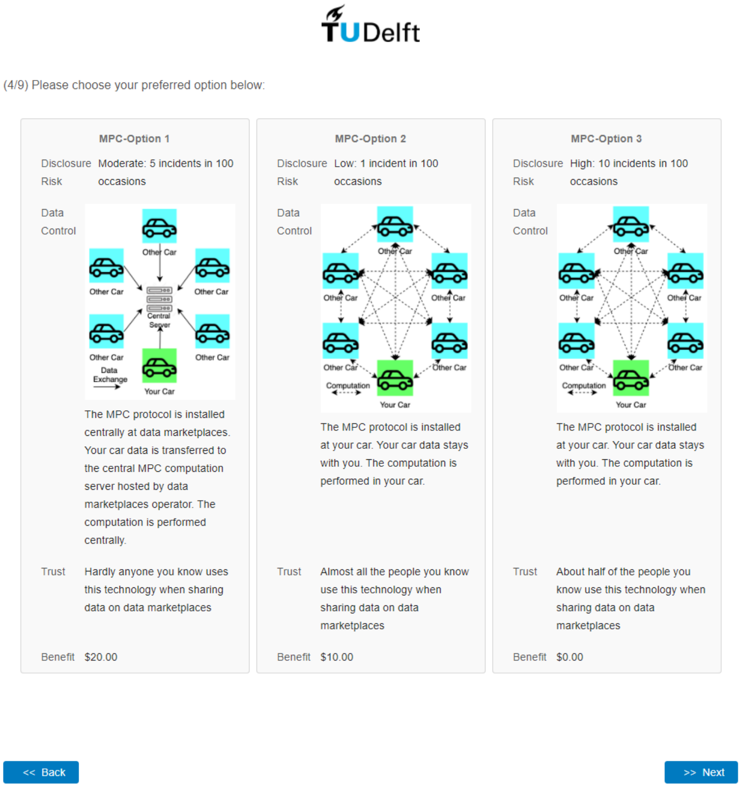


TUDelft

(3/9) Please choose your preferred option below:

	MPC-Option 1	MPC-Option 2	MPC-Option 3
Disclosure Risk	High: 10 incidents in 100 occasions	Moderate: 5 incidents in 100 occasions	Low: 1 incident in 100 occasions
Data Control	The MPC protocol is installed centrally at data marketplaces. Your car data is transferred to the central MPC computation server hosted by data marketplaces operator. The computation is performed centrally.	The MPC protocol is installed at your car. Your car data stays with you. The computation is performed in your car.	The MPC protocol is installed at your car. Your car data stays with you. The computation is performed in your car.
Trust	Almost all the people you know use this technology when sharing data on data marketplaces	Hardly anyone you know uses this technology when sharing data on data marketplaces	About half of the people you know use this technology when sharing data on data marketplaces
Benefit	$10.00	$20.00	$0.00

Figure I.8: Survey Page 8

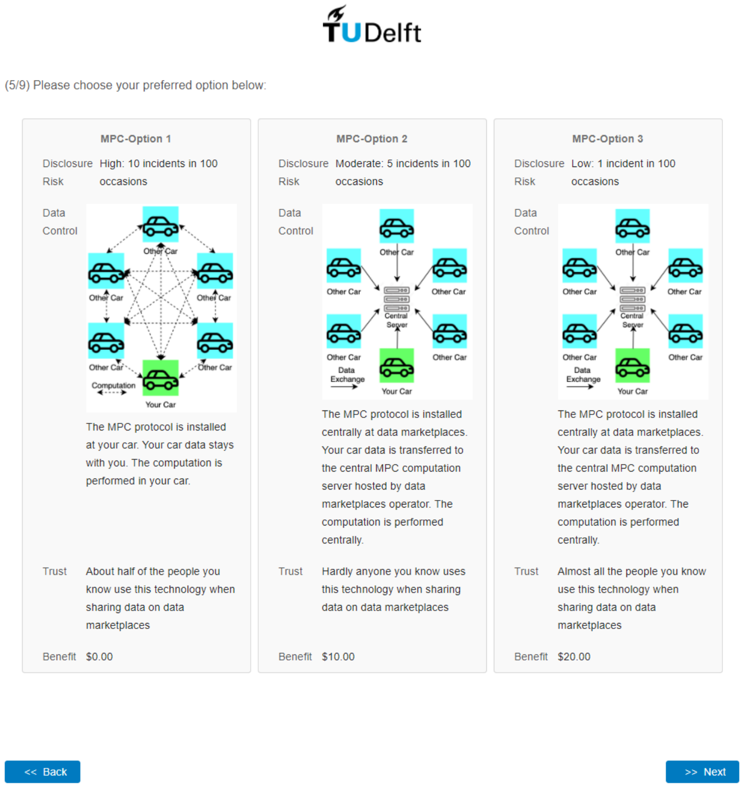


TUDelft

(4/9) Please choose your preferred option below:

	MPC-Option 1	MPC-Option 2	MPC-Option 3
Disclosure Risk	Moderate: 5 incidents in 100 occasions	Low: 1 incident in 100 occasions	High: 10 incidents in 100 occasions
Data Control	The MPC protocol is installed centrally at data marketplaces. Your car data is transferred to the central MPC computation server hosted by data marketplaces operator. The computation is performed centrally.	The MPC protocol is installed at your car. Your car data stays with you. The computation is performed in your car.	The MPC protocol is installed at your car. Your car data stays with you. The computation is performed in your car.
Trust	Hardly anyone you know uses this technology when sharing data on data marketplaces	Almost all the people you know use this technology when sharing data on data marketplaces	About half of the people you know use this technology when sharing data on data marketplaces
Benefit	$20.00	$10.00	$0.00

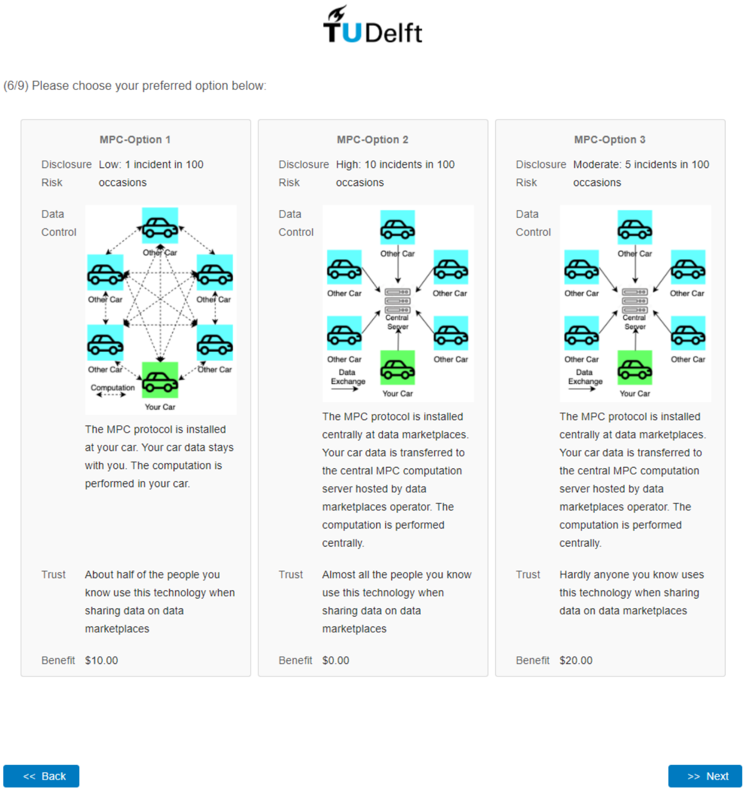
Figure I.9: Survey Page 9



(5/9) Please choose your preferred option below:

	MPC-Option 1	MPC-Option 2	MPC-Option 3
Disclosure Risk	High: 10 incidents in 100 occasions	Moderate: 5 incidents in 100 occasions	Low: 1 incident in 100 occasions
Data Control	The MPC protocol is installed at your car. Your car data stays with you. The computation is performed in your car.	The MPC protocol is installed centrally at data marketplaces. Your car data is transferred to the central MPC computation server hosted by data marketplaces operator. The computation is performed centrally.	The MPC protocol is installed centrally at data marketplaces. Your car data is transferred to the central MPC computation server hosted by data marketplaces operator. The computation is performed centrally.
Trust	About half of the people you know use this technology when sharing data on data marketplaces	Hardly anyone you know uses this technology when sharing data on data marketplaces	Almost all the people you know use this technology when sharing data on data marketplaces
Benefit	$0.00	$10.00	$20.00

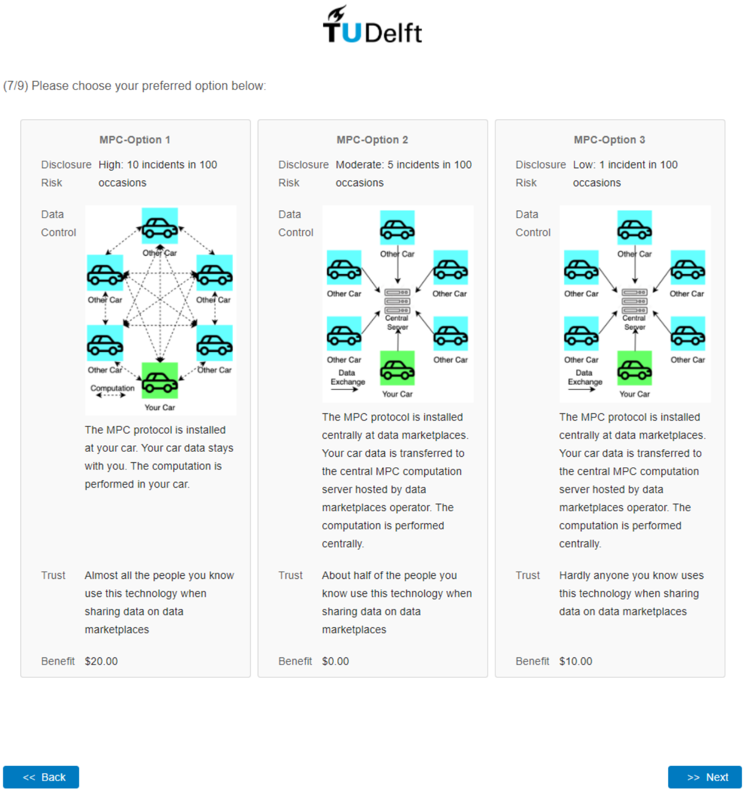
Figure I.10: Survey Page 10



(6/9) Please choose your preferred option below:

	MPC-Option 1	MPC-Option 2	MPC-Option 3
Disclosure Risk	Low: 1 incident in 100 occasions	High: 10 incidents in 100 occasions	Moderate: 5 incidents in 100 occasions
Data Control	The MPC protocol is installed at your car. Your car data stays with you. The computation is performed in your car.	The MPC protocol is installed centrally at data marketplaces. Your car data is transferred to the central MPC computation server hosted by data marketplaces operator. The computation is performed centrally.	The MPC protocol is installed centrally at data marketplaces. Your car data is transferred to the central MPC computation server hosted by data marketplaces operator. The computation is performed centrally.
Trust	About half of the people you know use this technology when sharing data on data marketplaces	Almost all the people you know use this technology when sharing data on data marketplaces	Hardly anyone you know uses this technology when sharing data on data marketplaces
Benefit	$10.00	$0.00	$20.00

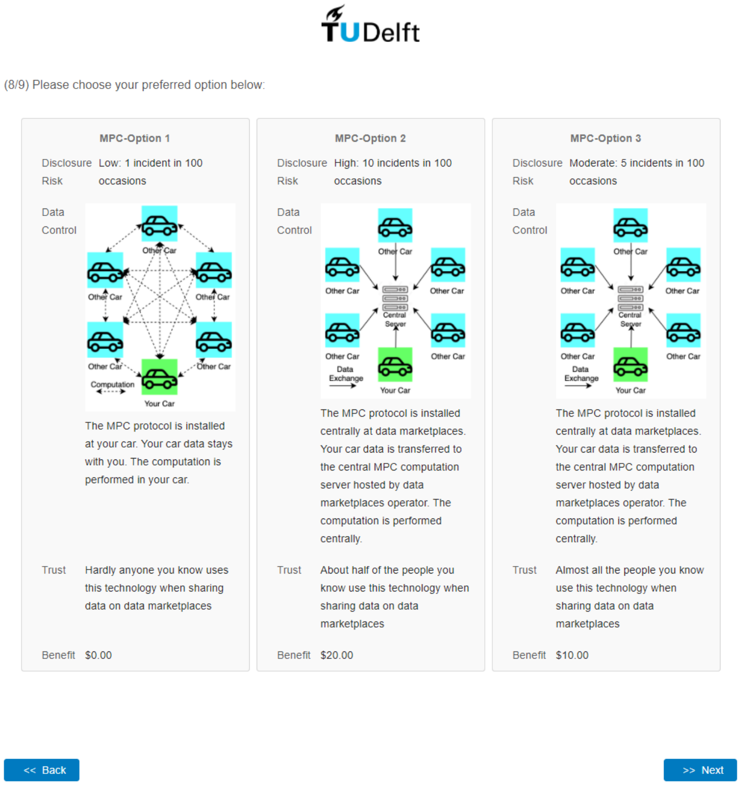
Figure I.11: Survey Page 11



(7/9) Please choose your preferred option below:

	MPC-Option 1	MPC-Option 2	MPC-Option 3
Disclosure Risk	High: 10 incidents in 100 occasions	Moderate: 5 incidents in 100 occasions	Low: 1 incident in 100 occasions
Data Control	The MPC protocol is installed at your car. Your car data stays with you. The computation is performed in your car.	The MPC protocol is installed centrally at data marketplaces. Your car data is transferred to the central MPC computation server hosted by data marketplaces operator. The computation is performed centrally.	The MPC protocol is installed centrally at data marketplaces. Your car data is transferred to the central MPC computation server hosted by data marketplaces operator. The computation is performed centrally.
Trust	Almost all the people you know use this technology when sharing data on data marketplaces	About half of the people you know use this technology when sharing data on data marketplaces	Hardly anyone you know uses this technology when sharing data on data marketplaces
Benefit	$20.00	$0.00	$10.00

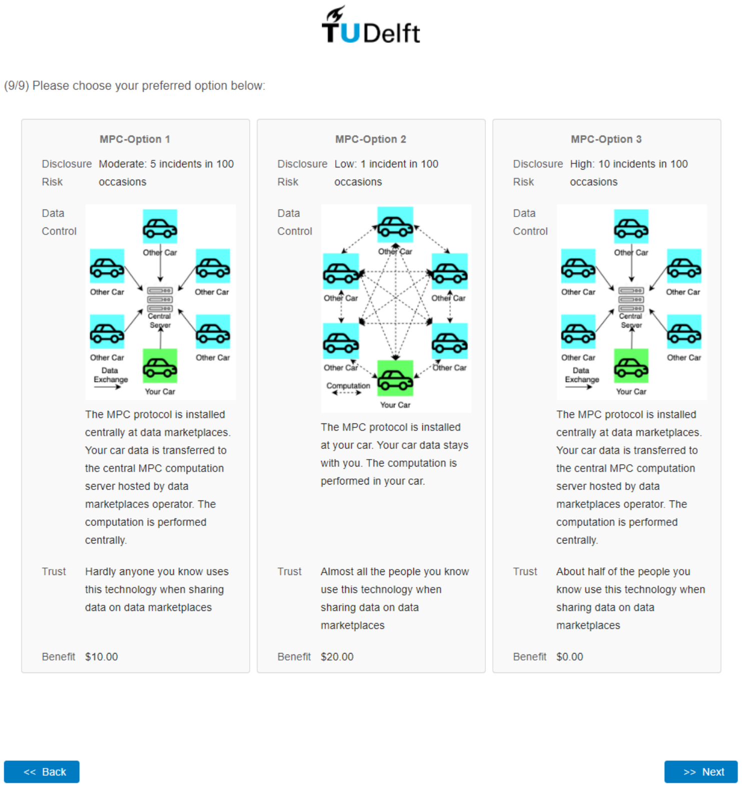
Figure I.12: Survey Page 12



(8/9) Please choose your preferred option below:

	MPC-Option 1	MPC-Option 2	MPC-Option 3
Disclosure Risk	Low: 1 incident in 100 occasions	High: 10 incidents in 100 occasions	Moderate: 5 incidents in 100 occasions
Data Control	The MPC protocol is installed at your car. Your car data stays with you. The computation is performed in your car.	The MPC protocol is installed centrally at data marketplaces. Your car data is transferred to the central MPC computation server hosted by data marketplaces operator. The computation is performed centrally.	The MPC protocol is installed centrally at data marketplaces. Your car data is transferred to the central MPC computation server hosted by data marketplaces operator. The computation is performed centrally.
Trust	Hardly anyone you know uses this technology when sharing data on data marketplaces	About half of the people you know use this technology when sharing data on data marketplaces	Almost all the people you know use this technology when sharing data on data marketplaces
Benefit	$0.00	$20.00	$10.00

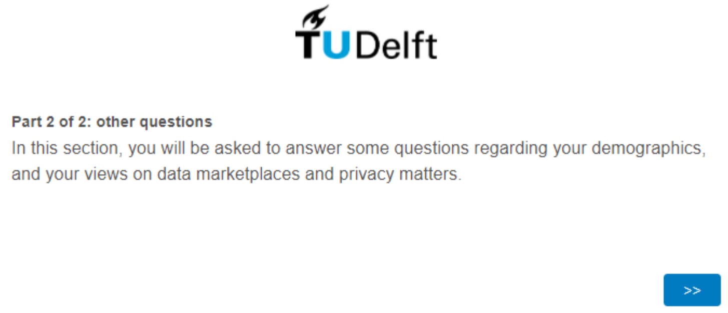
Figure I.13: Survey Page 13



(9/9) Please choose your preferred option below:

	MPC-Option 1	MPC-Option 2	MPC-Option 3
Disclosure Risk	Moderate: 5 incidents in 100 occasions	Low: 1 incident in 100 occasions	High: 10 incidents in 100 occasions
Data Control	The MPC protocol is installed centrally at data marketplaces. Your car data is transferred to the central MPC computation server hosted by data marketplaces operator. The computation is performed centrally.	The MPC protocol is installed at your car. Your car data stays with you. The computation is performed in your car.	The MPC protocol is installed centrally at data marketplaces. Your car data is transferred to the central MPC computation server hosted by data marketplaces operator. The computation is performed centrally.
Trust	Hardly anyone you know uses this technology when sharing data on data marketplaces	Almost all the people you know use this technology when sharing data on data marketplaces	About half of the people you know use this technology when sharing data on data marketplaces
Benefit	$10.00	$20.00	$0.00

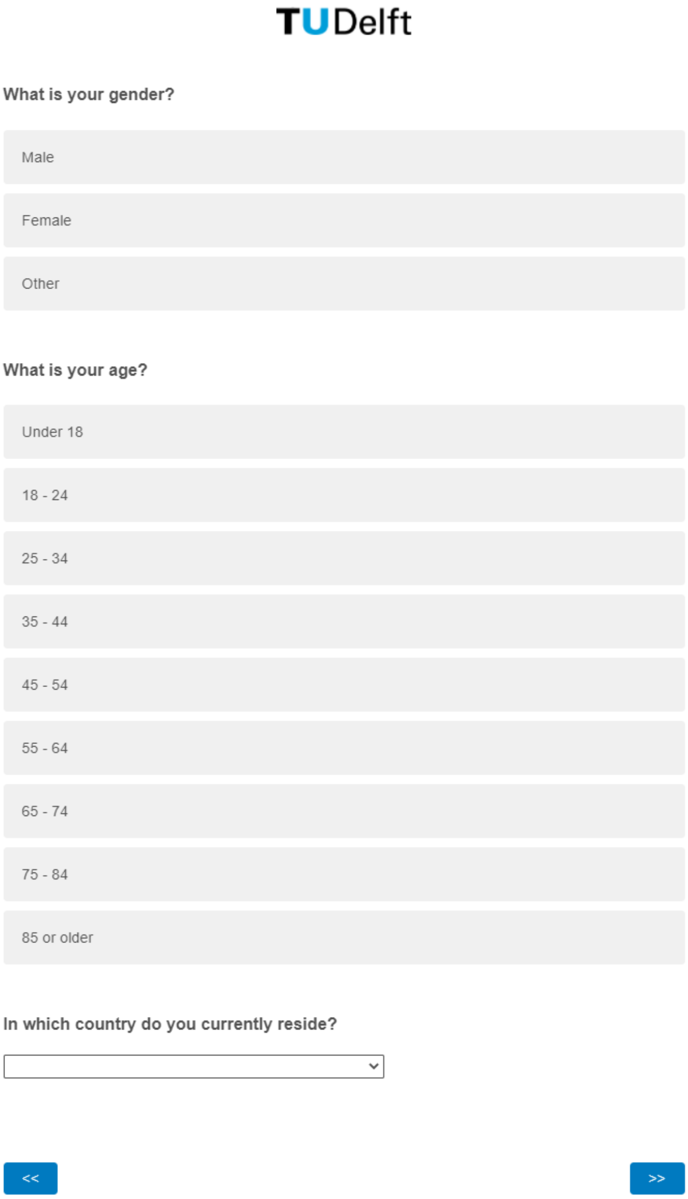
Figure I.14: Survey Page 14



Part 2 of 2: other questions

In this section, you will be asked to answer some questions regarding your demographics, and your views on data marketplaces and privacy matters.

Figure I.15: Survey Page 15



What is your gender?

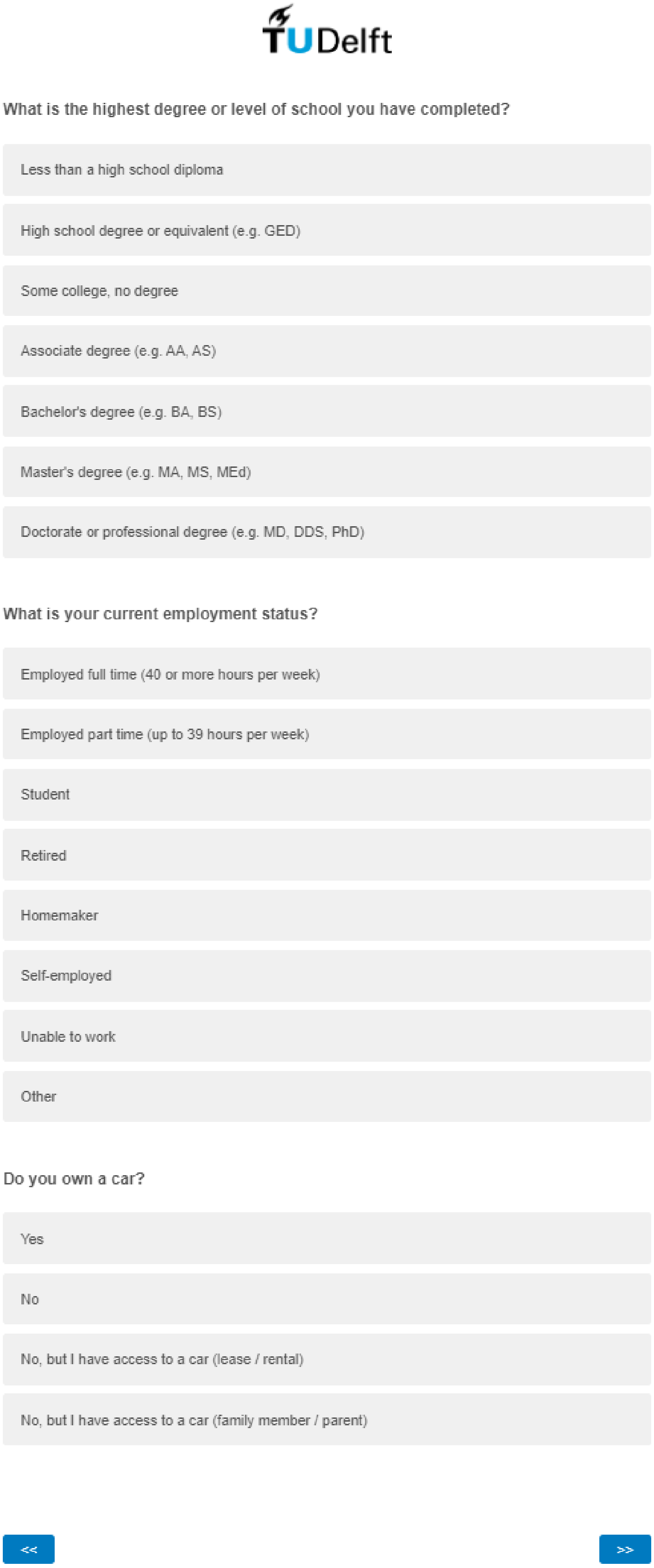
- Male
- Female
- Other

What is your age?

- Under 18
- 18 - 24
- 25 - 34
- 35 - 44
- 45 - 54
- 55 - 64
- 65 - 74
- 75 - 84
- 85 or older

In which country do you currently reside?

Figure I.16: Survey Page 16



What is the highest degree or level of school you have completed?

- Less than a high school diploma
- High school degree or equivalent (e.g. GED)
- Some college, no degree
- Associate degree (e.g. AA, AS)
- Bachelor's degree (e.g. BA, BS)
- Master's degree (e.g. MA, MS, MEd)
- Doctorate or professional degree (e.g. MD, DDS, PhD)

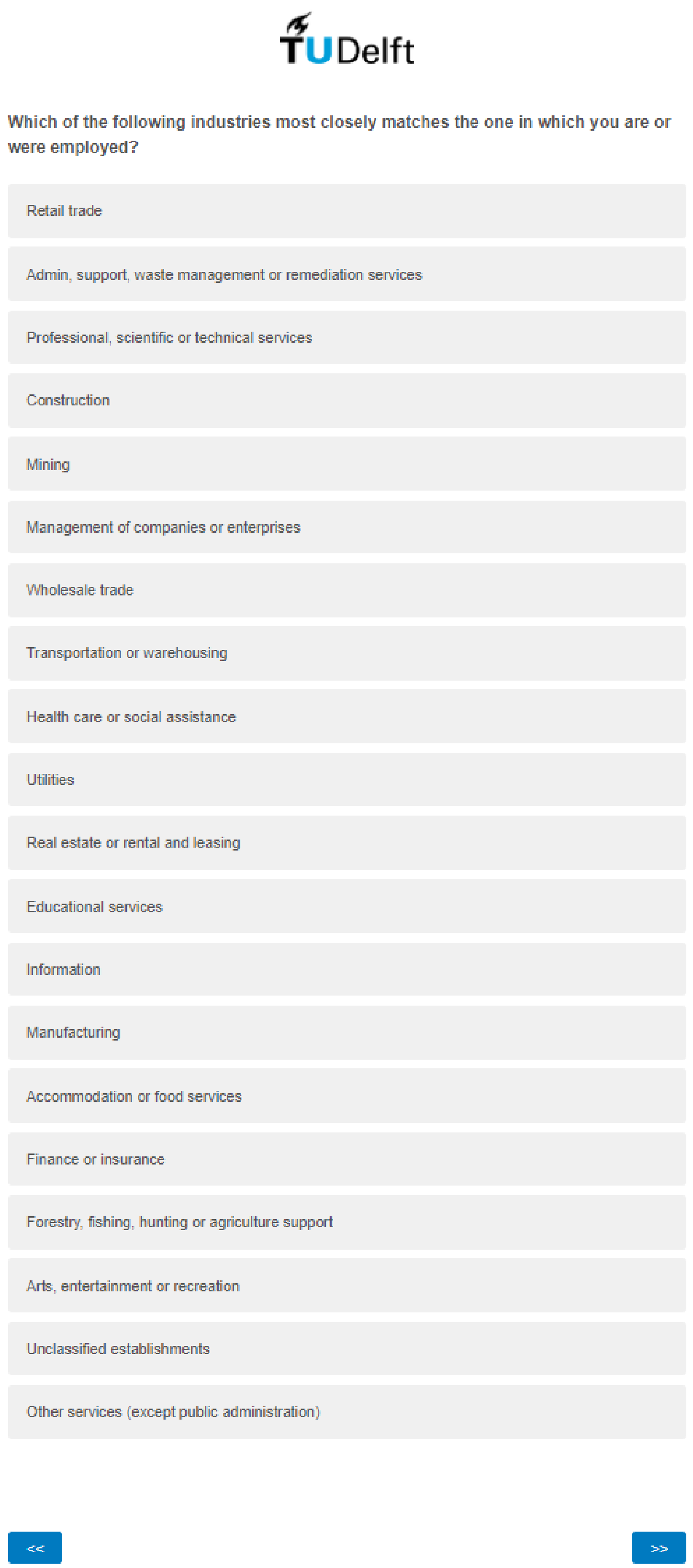
What is your current employment status?

- Employed full time (40 or more hours per week)
- Employed part time (up to 39 hours per week)
- Student
- Retired
- Homemaker
- Self-employed
- Unable to work
- Other

Do you own a car?

- Yes
- No
- No, but I have access to a car (lease / rental)
- No, but I have access to a car (family member / parent)

Figure I.17: Survey Page 17



Which of the following industries most closely matches the one in which you are or were employed?

Retail trade

Admin, support, waste management or remediation services

Professional, scientific or technical services

Construction

Mining

Management of companies or enterprises

Wholesale trade

Transportation or warehousing

Health care or social assistance

Utilities

Real estate or rental and leasing

Educational services

Information

Manufacturing

Accommodation or food services

Finance or insurance

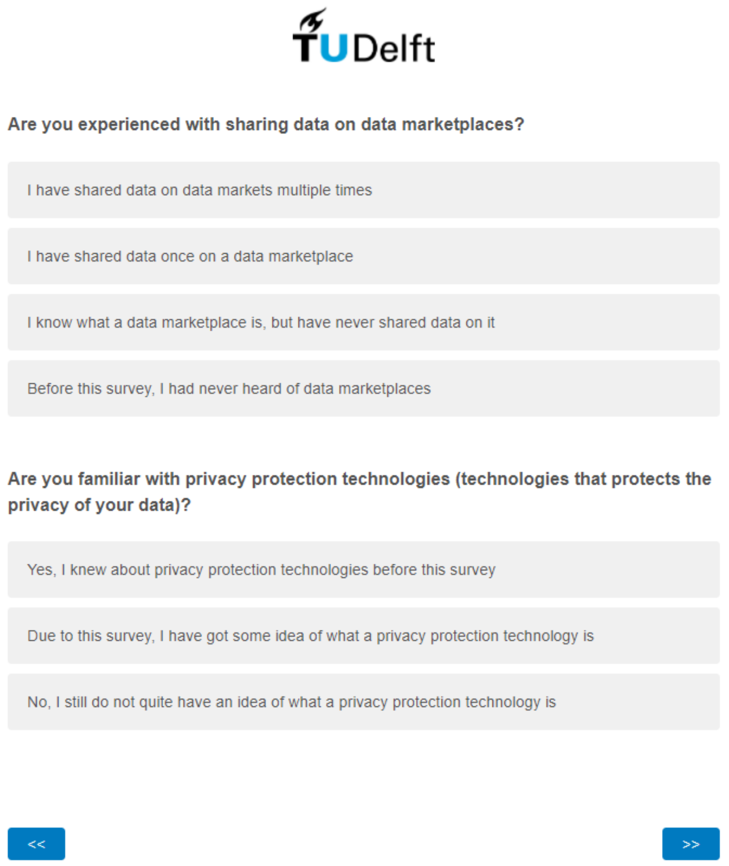
Forestry, fishing, hunting or agriculture support

Arts, entertainment or recreation

Unclassified establishments

Other services (except public administration)

Figure I.18: Survey Page 18



Are you experienced with sharing data on data marketplaces?

I have shared data on data markets multiple times

I have shared data once on a data marketplace

I know what a data marketplace is, but have never shared data on it

Before this survey, I had never heard of data marketplaces

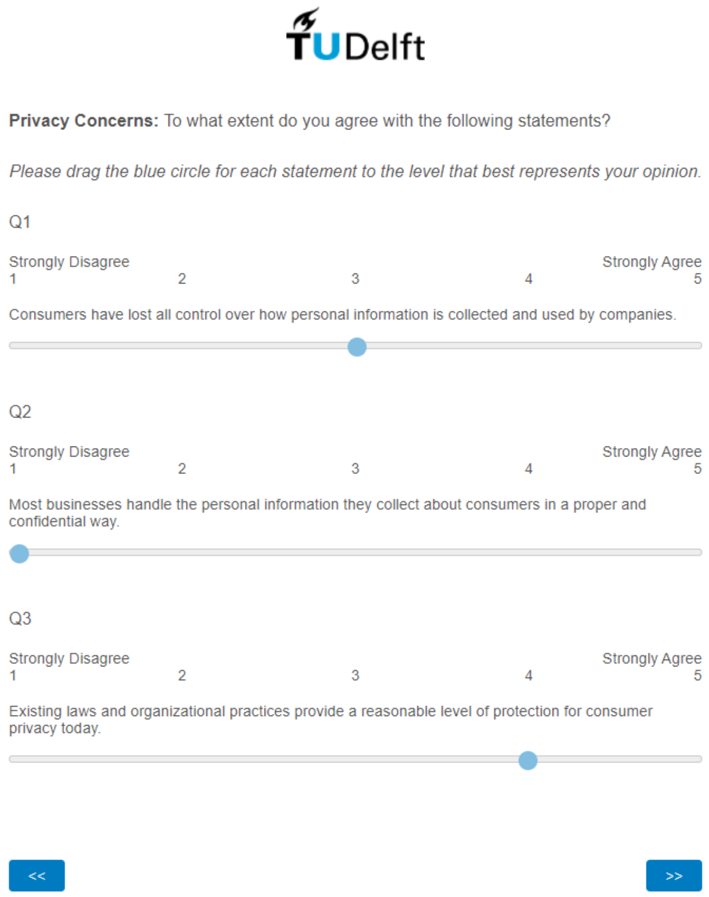
Are you familiar with privacy protection technologies (technologies that protects the privacy of your data)?

Yes, I knew about privacy protection technologies before this survey

Due to this survey, I have got some idea of what a privacy protection technology is

No, I still do not quite have an idea of what a privacy protection technology is

Figure I.19: Survey Page 19



Privacy Concerns: To what extent do you agree with the following statements?

Please drag the blue circle for each statement to the level that best represents your opinion.

Q1

Strongly Disagree 1 — 2 — 3 — 4 — Strongly Agree 5

Consumers have lost all control over how personal information is collected and used by companies.

Q2

Strongly Disagree 1 — 2 — 3 — 4 — Strongly Agree 5

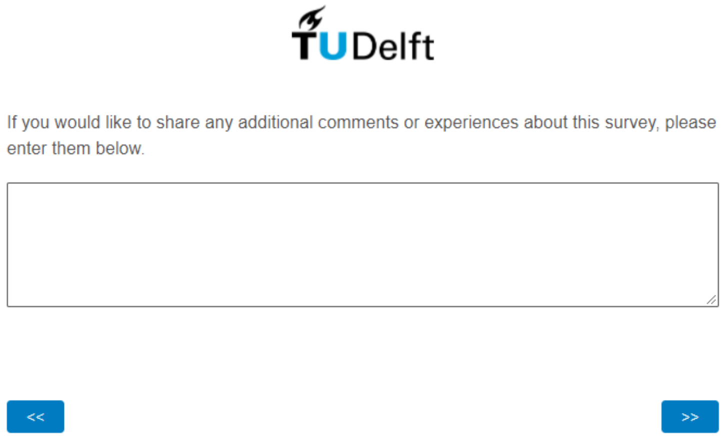
Most businesses handle the personal information they collect about consumers in a proper and confidential way.

Q3

Strongly Disagree 1 — 2 — 3 — 4 — Strongly Agree 5

Existing laws and organizational practices provide a reasonable level of protection for consumer privacy today.

Figure I.20: Survey Page 20



If you would like to share any additional comments or experiences about this survey, please enter them below.