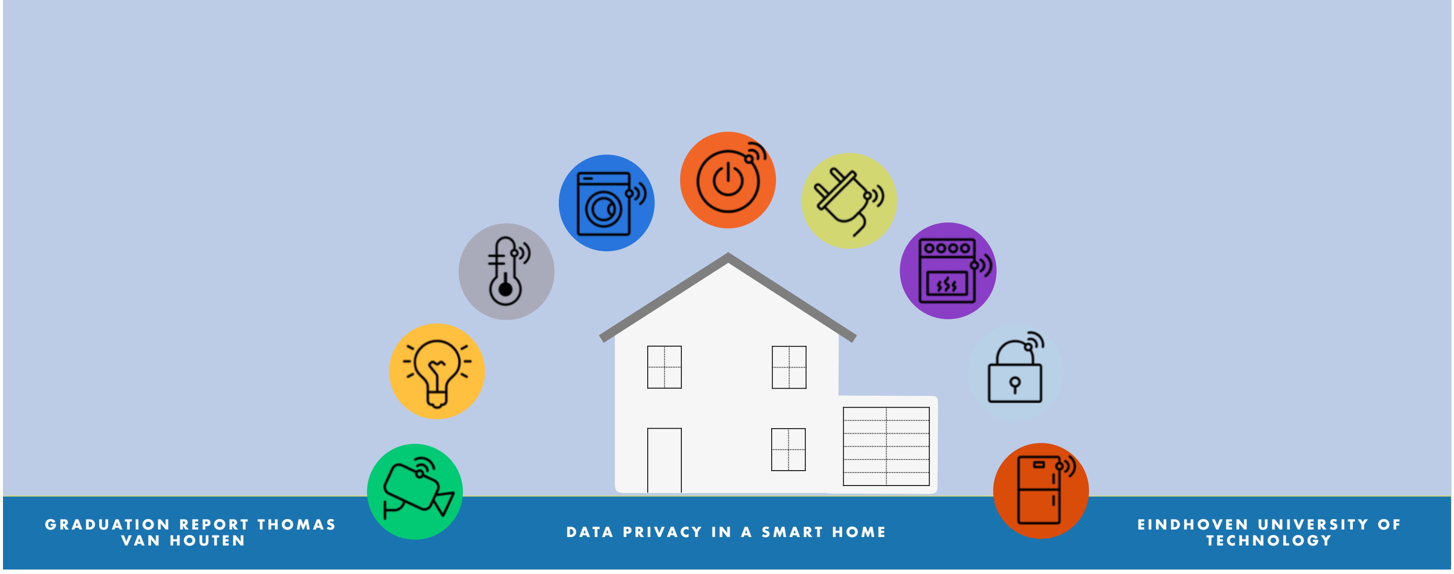
MASTER

Data privacy in a smart home

a stated choice experiment on the trade-off between data privacy and the benefits of smart home appliances on energy consumption

van Houten, T.

Award date:
2020



DATA PRIVACY IN A SMART HOME

A stated choice experiment on the trade-off between data privacy and the benefits of smart home appliances on energy consumption.

Graduation Report

Author: T (Thomas van Houten)

Master Construction Management & Engineering

Eindhoven University of Technology

Study-year 2019 -2020

Final Colloquium date: 21-01-2020

DATA PRIVACY IN A SMART HOME

A stated choice experiment on the trade-off between data privacy and the benefits of smart home appliances on energy consumption.

Author

T. (Thomas) van Houten (BSc)

Student Construction Management & Engineering

Department of the Built Environment

Eindhoven University of Technology

Student ID - 1032517

E-mail (TU/e) thomas.van.houten@student.tue.nl

E-mail (private) thomashout@gmail.com

Graduation Committee

Chairman - Prof. Dr. Ir. B (Bauke) de Vries

Chair of Information Systems in the Built Environment (ISBE) group

Full professor in the department of Built Environment

First Supervisor - Qi Han

Professor in the Information Systems in the Built Environment (ISBE) group
Full Professor in the department of Built Environment

Second Supervisor - Dujuan Yang

Professor in the Information Systems in the Built Environment (ISBE) group
Full Professor in the department of Built Environment

Technische Universiteit **Eindhoven** University of Technology

“Arguing that you don't care about the right to privacy because you have nothing to hide is no different than saying you don't care about free speech because you have nothing to say.”

Edward Snowden

Preface

This report is written as the final report for my master Construction Management & Engineering (CME) and the Eindhoven University of Technology (TU/e). This research will discuss the trade-off between data privacy and the benefits of smart home appliances.

Over the past months I have discovered the world of data privacy. With social, economic and governmental activities increasingly carried out online, the flow of personal data is expanding rapidly. The topic became a public conversation in 2013, when Edward Snowden leaked highly classified information from the NSA about numerous global surveillance programs. Since then, a lot has changed in the public (and my personal) perception about data privacy. Privacy is more often discussed in the news, literature and new regulations have been actuated. With this research, I hope to increase the attention to this topic in the build environment. Especially since we are inviting new smart devices often referred to as smart home appliances into our most intimate and private place, our home.

It is expected that in 2022 on a population of 7.7 billion people, approximately 29 billion interconnected devices are actively used. Consequently, we are allowing companies to track our personal behavior on a day-to-day base. Even if you physically own a device, others own the data it collects. Business are challenged by protecting identifiable information and managing the risks without restricting the potential benefits of the data. This raises an ethical discussion about the difference between data privacy and data protection. While data privacy is a distinct human right, the right for data protection is restricted in several situations. For example, in case of national security or public safety. Data privacy is one of the biggest topics of this era. Some say that this is an inevitable development that otherwise restrains innovation. I believe we need to find the balance between data privacy and data protection. I would like to boost this conversation by researching the privacy behavior of consumers towards smart home appliances.

I would like to thank my supervisors Qi Han and Dujuan Yang for the professional support and trust during my research. Also, I would also like to thank my family and friends for their mental support over the past months. It has been a difficult time with ups and downs, but it was worth the struggle. Lastly, I would like to mention that we should take more care about your data privacy. As stated in the quote on the previous page, data privacy should be a cherished value. Saying you don't care about data privacy because you have nothing to hide seems a misplaced argument to me. There is nothing wrong with sharing data as long it is beneficial to you. Be aware of what you share, why you share and with whom you're sharing with.

Happy reading,

Thomas van Houten

This page is Intentionally left blank

Summary

In the last century, the volume of data has grown exponentially to uncountable proportions. Due to advanced algorithms and computable power, data processing has changing the activities and lifestyle of every individual. Personal data is becoming a tradeable asset and new markets are emerging rapidly not only through the web, but increasingly through wearables and smart devices. A related example are smart home appliances. The market for smart home appliances has grown significantly in the past year. Worldwide, over 1.2 billion smart home devices are connected in 2018, which was an increase of approximately 45 percent since the year before. Smart home appliances such as a smart thermostat or a smart speaker collect 'personal' data about the user to improve the capabilities of the product. The increasing amount of data entering our lives brings new innovative markets but also raise questions about data protection. Data of smart home appliances need protection since it captures accurate information about the individual's whereabouts, financial information, medical information, personal preferences, energy use, habits, etcetera.

The decision of an individual to share private data is determined by a trade-off between the risks of sharing privacy sensitive data and the benefits the users receives from the products and/or services. If the benefits outweigh the risks, individuals are willing to 'sacrifice' their data privacy. Researches in the fields of online shopping and social media suggest that individuals are willing to share their personal data for a relatively low benefit. This inconsistency between privacy attitudes and privacy behavior is frequently referred to as the "privacy paradox" (Kokolakis, 2017) (Williams, Nurse, & Creese, 2018). This paradox is however not researched extensively in relation to smart home appliances. At this moment, there is no evidence that is focused on the trade-off's individuals are willing to make between data privacy and the benefits of smart home appliances. Hence, a better insight is crucial since more smart appliances are entering the most private place of an individual, our home. Most studies about privacy are focused on how individuals think about data privacy instead of how they act. The goal of this research is to obtain insight in the trade-offs that individuals are willing to make between sharing privacy sensitive data and the benefits of smart home appliances. Since there are many appliances available on the market, it is focused on appliances that are beneficial to reduce the user's energy consumption.

To determine the trade-off between the consumers data privacy and the benefits of smart home appliances, an online survey has been used. The respondents were questioned about several socio-demographic characteristics and about their concerns regarding data privacy and energy consumption. Lastly, the survey contained a stated choice experiment. A stated choice experiment is a statistical technique that looks at the choices that individuals make between alternatives. By decomposing the alternatives into different attributes, the value of how respondents perceive the value can be measured (Louviere, Flynn, & Carson, 2010). A stated choice experiment does not measure actual behavior but the behavioral intention since respondents are asked about theoretical choice situations. Nevertheless, an experiment of this type is considered as a hybrid approach that enables testing behavior while using a survey. The goal of this survey is to test to what extent individuals are willing to trade-off privacy sensitive data to achieve benefits of smart home appliances including energy efficiency?

A total of 354 respondents have started the online survey. After noise reduction, a total of 256 surveys were used for data analysis. In the stated choice experiment part of the survey, the respondents were asked in eight choice situations to choose between two theoretical smart home appliances. All smart home appliances contained of 6 attributes explaining what kind of data is processed by the appliance,

for what purposes, with whom this data is shared, how frequently the data is shared and when the data is removed. Lastly, a financial benefit has been proposed as a compensation for the shared data.

The results of the survey have been analyzed using a multinomial and mixed logit models performed with the mlogit package of R-statistics. The trade-off between sharing privacy sensitive data and the benefits of smart home appliances is mainly determined by three attributes, the type of data that is processed, the reason why this data is processed and the financial benefit that can be obtained by the smart home appliance. This means that individuals are showing less interest in the actor that is processing the data, the frequency of data processing and the retention time (when is the data removed). The choice experiment showed that the type of data was considered more important than the financial benefit, indicating that individuals are showing a real interest in their privacy. The capabilities that individuals prefer in a smart home appliance are remotely controlling the appliance or appliances that work automatically. Individuals prefer a smart home appliance that work remotely or automatically indicating that most individuals choose a smart home appliance because it needs less attention while having some sort of benefit in return. It also suggests that individuals rather choose a smart home appliance that is beneficial for themselves instead of an appliance that is beneficial for the environment. Thus, energy reduction should be recognized as a side benefit rather than the main goal for smart home appliances.

In this research, the privacy paradox is investigated by testing the difference between how concerned individuals are and the difference in choice behavior. It is found out that individuals with high concern about data privacy are more likely to choose a smart home appliance that processes the least amount of data. Thus, respondents with high concerns are narrowing their perception of these concerns by sharing fewer data. Also, if respondents have a higher concern about data privacy, they are less likely to choose a smart home appliance without a financial benefit. Thus, respondents are demanding financial compensation for the privacy harm they experience. The privacy paradox is apparent when looking at the interaction with socio-demographic variables and choice behavior. First, the type of education does not significantly influence the behavioral intentions of individuals at all. Also, the influence of age and income on the choice behavior is contradictory to the concerns of the respondents. Both older individuals and individuals with a high income are significantly less likely choosing a smart home appliance that processes the least amount of data. Thus, they are not as worried about sharing more data than younger and lower-income individuals. On the other hand, Women are more likely to choose a smart home appliance that processes the least amount of data while they are also more concerned about their data privacy. Thus, women are showing more privacy-protective behavior.

It is found that there is a clear difference between what people claim about data privacy and how they behave. The confirmation of the privacy paradox should be recognized as the main scientific relevance of this research. This research should be seen as a startup for more (privacy-specific) research. It is recommended to collect more academic evidence to make the conclusions of this research more useful in practice. Also, this research was mainly focused on the trade-off between data and a financial benefit. This is a theoretical value and therefore unrepresentative when testing actual existing products. Future research should consider additional benefits such as comfort, safety, health and reliability. Also, more insight is demanded in the privacy-sensitive locations within a smart home. Applying smart home appliances in the bedroom might be considered as more privacy-invasive than implementing a smart washing machine in the garage.

Samenvatting

In de laatste jaren, de hoeveelheid data dat wordt verzameld is exponentieel gegroeid tot ontelbare proporties. Door rekenkracht en geavanceerde algoritmes heeft big data een grote invloed gekregen op het uitvoeren van onze dagelijkse activiteiten. Onze persoonlijke data is handelswaar geworden voor nieuwe markten. Deze data is niet alleen gerelateerd aan het internet, maar is steeds vaker verwerkt in draagbare producten en slimme apparaten in het Smart home concept. De markt voor slimme apparaten is significant gegroeid in het laatste jaar. Wereldwijd zijn er meer dan 1,2 miljard slimme apparaten verbonden wat een toename was van circa 45 procent was ten opzichte van het jaar ervoor. Slimme apparaten zoals een slimme thermostaat of een slimme speaker verzamelen persoonlijke data van de gebruiker om de toepassingen te vergroten. De groeiende hoeveelheid data moet daarom goed beschermd worden aangezien het accurate informatie kan bevatten zoals zijn of haar verblijfplaats, financiële information, persoonlijke voorkeuren, energiegebruik en/of gewoontes etc.

De keuze dat een individu maakt tussen het delen van privacygevoelige informatie wordt bepaald door de afweging tussen de risico's van het delen van de privacygevoelige informatie en de voordelen die aan het product of dienst zijn verbonden. Als de voordelen zwaarder wegen dan het risico zijn individuen bereid om hun privacy 'op te geven'. Onderzoeken in de studierichtingen van online shopping en sociale media hebben geconcludeerd dat individuen tamelijk snel bereid zijn om hun persoonlijke information in te ruilen voor een voordeeltje. Deze tegenstrijdigheid wordt in literatuur regelmatig benoemd als een 'privacy paradox' (Kokolakis, 2017) (Williams et al., 2018). Of deze paradox ook geldt in relatie met slimme apparaten uit het smart home concept is reeds onbekend. Om het moment van schrijven is er geen wetenschappelijk bewijs gevonden dat de afweging tussen dataprivacy en de voordelen van slimme apparaten heeft onderzocht. Een betere inkijk is daarvoor van groot belang. Vooral omdat slimme apparaten zich gaan nestelen in ons meest persoonlijke plek, namelijk ons huis. Veel studies zijn gefocust op hoe individuen denken over dataprivacy in plaats van hoe mensen zich gedragen. Het doel van dit onderzoek is het verkrijgen van een beter inzicht in de afweging data wordt gemaakt tussen het delen van privacygevoelige informatie en de voordelen van slimme apparaten. Aangezien veel verschillende slimme apparaten op de markt zijn, focust dit onderzoek zich op slimme apparaten dat ook een energievoordeel kunnen realiseren.

Om de afweging tussen dataprivacy en de voordelen van slimme apparaten te testen wordt een online enquête gebruikt. De respondenten worden gevraagd over verschillende sociaal-demografische karakteristieken en hun bezorgdheid over dataprivacy en energiegebruik. Tenslotte bevat de enquête een stated choice experiment. Een stated choice experiment is een statistische techniek dat de keuzes tussen alternatieven observeert. Door het onderscheiden van de alternatieven in verschillende attributen kan worden getest hoe de respondenten de attributen waarderen (Louviere et al., 2010). A stated choice experiment meet geen 'echt' gedrag van de respondent maar de keuze intentie. Desondanks wordt dit type experiment gezien als een hybride methodiek dat zowel gedrag kan meten als gebruikt kan worden in een online enquête. Het uiteindelijke doel van de enquête is het testen in hoeverre individuen bereid zijn om privacygevoelige informatie om te ruilen voor de voordelen van een slim apparaat dat tevens een energievoordeel kan bereiken.

Totaal hebben 354 respondenten de online enquête gestart. Na het verwijderen van de incomplete en verkeerde ingevulde enquêtes zijn er 256 enquêtes gebruikt voor verdere data-analyse. In het onderdeel van de stated choice experiment werden de respondenten acht keuzesituaties voorgelecht waarin ze moesten kiezen tussen twee theoretische slimme apparaten. Deze slimme apparaten waren onderverdeelt in 6 attributen dat karakteristieken van het apparaat beschreef. De 6 attributen beschrijven: het doel van

de verwerking, met wie de data wordt gedeeld, hoe frequent de data wordt gedeeld en verwijderd. Tevens werd er een financiële tegencompensatie voorgesteld voor het delen van de data.

De resultaten van de enquête zijn verwerkt met een multinomiaal en mixed logit modellen, uitgevoerd door de mlogit uitbreiding van het programma R-statistics. De modelberekeningen toonde aan dat er drie attributen de grootste impact hebben op de totale score. Het type data dat wordt verwerkt door een slim apparaat, de reden waarom de data is verwerkt en het financiële voordeel dat kan worden verkregen bij het gebruik van het slimme apparaat. Dit betekent dat individuen minder geïnteresseerd zijn in de data verwerker, de frequentie van data verwerken en het moment van data verwijderen. Het keuze experiment toonde tevens aan dat het de data type belangrijker werd gevonden dan een mogelijk financieel voordeel. Dit geeft aan dat de enquêteurs een serieuze interesse toonde in hun dataprivacy. Gemiddeld hadden de respondenten een voorkeur voor slimme apparaten dat op afstand bestuurd kunnen worden of apparaten dat geheel automatisch functioneren. Ofwel, er wordt gekozen voor slimme apparaten die weinig aandacht nodig hebben. Dit toont aan dat de meeste individuen liever kiezen voor een slim apparaat dat de gebruiker een voordeel biedt dan een apparaat dat voordelig is voor het milieu. Dus energie reductie moet worden geïnterpreteerd als een bijkomend voordeel in plaats van het hoofddoel van het slimme apparaat.

In dit onderzoek, de privacy paradox is onderzocht door de afweging tussen privacy en de voordelen van slimme apparaten te meten. Uit het onderzoek kwam naar voren dat individuen met een hoge bezorgdheid over privacy eerder zijn geneigd om slimme apparaten te kiezen dat zo min mogelijk data verwerkt. Dus, ze verkleinen de ervaring van privacy risico's door te kiezen voor het delen van minder data. Individuen met hoge bezorgdheid zijn ook minder geïnteresseerd in slimme apparaten dat geen financieel voordeel bevat. Dit toont aan dat men een financiële compensatie verwacht voor de schade die zij ervaren. Tevens is de privacy paradox aanwezig als de interactie tussen het de keuzegedrag en sociaal demografische variabelen wordt onderzocht. Ten eerste heeft de opleiding van een individu geen invloed op het keuzegedrag, maar wel op zijn/haar bezorgdheid over dataprivacy. Ook is er een duidelijk verschil tussen het keuzegedrag van de respondenten op basis van leeftijd en inkomen. Oudere en individuen met een hoger inkomen zijn significant minder bereid om slimme apparaten te kiezen dat weinig data verwerkt. Dus, deze groepen zijn minder bezorgd over data delen dan jongere en respondenten met een lager inkomen. Tevens is gevonden dat vrouwen meer bereid zijn om slimme apparaten te kiezen dat weinig data verwerk terwijl ze meer bezorgd zijn over dataprivacy. Dus, vrouwen tonen privacybeschermend gedrag.

Dit onderzoek heeft aangetoond dat er een duidelijk verschil zit tussen wat mensen beweren over dataprivacy en hoe zij zich uiteindelijk gedragen. De bevestiging van de privacy paradox moet worden gezien als de wetenschappelijke relevantie van dit onderzoek. Dit onderzoek is een eerste aanzet naar meer (privacy gerelateerd) onderzoek dat focust op het gedrag van de respondenten. Het is aanbevolen om meer praktische en wetenschappelijk bewijs te verzamelen over dataprivacy. Toekomstig onderzoek moet meer nadruk leggen op extra toegevoegde waarde van slimme apparaten zoals comfort, veiligheid, gezondheid en betrouwbaarheid. Tevens moet worden onderzocht hoe mensen reageren op slimme apparaten in verschillende privacygevoelige locaties in een slimme woning. Het toepassen van deze apparaten in een slaapkamer kan worden gezien als een grotere inbraak op de privacy dan een slimme wasmachine in de garage van de woning.

Abstract

Data privacy in smart homes is becoming a more sensitive topic due to the increasing number of smart appliances that are present in our homes. Research has shown that there is a large discrepancy in how individuals think about data privacy instead of how they act. This is called the privacy paradox. Consequently, it is unknown how data privacy will influence to implementation of these smart home appliances. This study aims to determine the trade-offs that individuals are willing to make between sharing privacy-sensitive data and the benefits of smart home appliances. With the use of a survey including a stated choice experiment it is tested which attributes are having the biggest influence on the choice of an individual. It is found out that the trade-off is mainly determined by three attributes, the type of data that is processed, the reason why this data is processed and the financial benefit that can be obtained by the smart home appliance. Individuals are showing less interest in the actor that is processing the data, the frequency of data processing and the retention time (when is the data removed). Furthermore, it was found that individuals care more about the content that is shared rather than with whom the data is shared and for how long indicating that individuals with high concerns are narrowing their perception of these concerns by sharing fewer data. Individuals are also demanding a (theoretical) financial compensation for the privacy harm they experience. This research also confirmed the existence of the privacy paradox. Individuals with high concern about data privacy are more likely to choose a smart home appliance that processes the least amount of data while demanding a financial compensation for the privacy harm they experience. Also women are showing more privacy-protective behavior. The results of this research can be used to improve the knowledge of privacy behavior in the built environment. This research should be seen as a startup for more (privacy-specific) research. It is recommended to collect more academic evidence to make the conclusions of this research more useful in practice.

Keywords:
Data Privacy, Smart Home Appliances, Stated Choice Modeling, Privacy Paradox

List of abbreviations / Glossary

A summary of the important definitions, notions, classifications etc.

Topic	Description
Alternative	One of the variations that is tested in a stated choice experiment. One alternative consists of one level variation of every attribute that is tested.
Attribute	The independent factors in a stated choice experiment. Also, a combination of factor levels.
Big Data	A massive volume of both structured and unstructured data that is so large it is difficult to process using traditional database and software techniques.
Choice Setting	A situation in stated choice experiment where the respondent is asked to choose between two or more alternatives.
Data Controller	The natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data.
Data Privacy	Having control over the personal information, including the transfer and exchange of that data (Ginosar & Ariel, 2017).
Data processing	Any operation or set of operations which is performed on personal data. Examples of data processing are: collection, structuring, storing, adapting, combining or erasing of personal data.
Data Processor	The party that performs data processing.
GDPR	European data regulation abbreviation for General Data Protection Regulation.
Choice level	Values that express a range of actual or potential variations in the attribute.
Personal Data	Any information relating to an identified or identifiable natural person.
Smart Home Appliances	A smart home is equipped with a high-tech network, linking sensors and domestic devices, appliances, and features that can be remotely monitored, accessed or controlled, and provide services that respond to the needs of its inhabitants" (Balta-Ozkan et al., 2013).
Stated Choice Experiment	A statistical technique that looks at the choices that individuals make between alternatives. By decomposing the alternatives into different attributes, the value of how respondents perceive the value can be measured (Louviere, Flynn, & Carson, 2010).

List of Figures

List of Tables

This page is Intentionally left blank

Table of Contents

This page is Intentionally left blank

1 INTRODUCTION

Chapter one explains the background information that is necessary to understand this research, it also describes the goals and the importance of this research. Lastly, a conceptual framework is established to describe how the different topics in this research are connected.

1.1 Background

1.1.1 Smart homes and smart home appliances

In the last century, the volume of data has grown exponentially to uncountable proportions. Due to advanced algorithms and computable power, data processing is changing the activities and lifestyle of every individual. Personal data is becoming a tradeable asset and new markets are emerging rapidly not only through the web, but increasingly through wearables and smart devices in people's homes. The rise of data processing entering our lives does bring new innovative markets, but also raise questions regarding the protection of personal data and the individuals' concerns over its privacy.

A 'traditional home' has manually operated appliances, activated by flipping a switch or pushing a button. "A smart home is equipped with a high-tech network, linking sensors and domestic devices, appliances, and features that can be remotely monitored, accessed or controlled, and provide services that respond to the needs of its inhabitants" (Balta-Ozkan, Davidson, Bicket, & Whitmarsh, 2013). With the increasing attention in smart appliances, everyday household products are becoming smarter by implementing computer chips. Examples of smart home appliances are a smart washing machine, dishwasher, stove, refrigerator and many others (Paetz, Becker, Fichtner, Schmeck, & Methods, 2011). The market for smart home appliances has grown significantly in the past year (figure 1). Worldwide, over 1.2 billion smart home devices are connected in 2018, which was an increase of approximately 45 percent since the year before (Statista, 2019). Most smart home appliances ensures the achievement of energy reduction, it is therefore that smart homes are one of the EU's 10 priority action areas in its Strategic Energy Technology Plan: This plan focuses on: "Creating technologies and services for smart homes that provide smart solutions to energy consumers" (Directorate-General for Research and Innovation (European Commission), 2018)

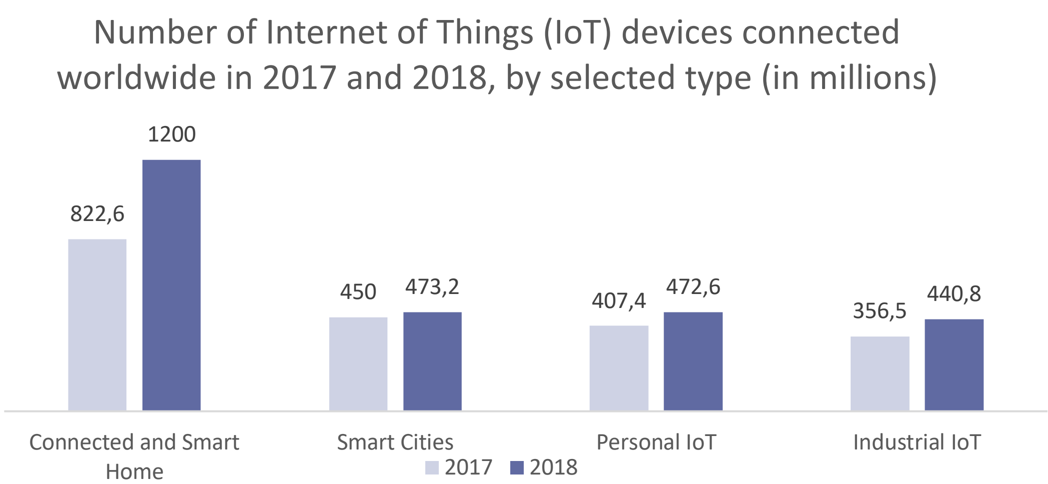


Figure 1 - Number of IoT devices worldwide in 2017 and 2018 (Statista, 2019)

1.1.2 Data privacy

Smart home appliances collect 'personal' data of the user to improve the capabilities of these products. Organizations are using data to indicate target groups, personalize their products, services and create future business strategies. Thus, this data is valuable and companies are willing to invest in data collection, processing and storing. The data needs protection since it captures accurate information of an individual's whereabouts, financial information, medical information, energy use, habits, etcetera. Since this data is so personal, there are potential risks when this data falls into the wrong hands. Examples of potential risks are data theft, data loss, data corruption, identity theft and economic crimes such as credit card stealing (Potoglou, Dunkerley, Patil, & Robinson, 2017).

From the perspective of the consumer, data privacy is a rising topic. It is more frequently discussed in scientific papers and in the news. Individuals develop deeper concerns about the privacy and security of their personal information. Based on a survey conducted by the Economist Intelligence Unit, 90 percent of the respondents are concerned about identity theft and fraud. Also, 89 percent is concerned that third parties are being allowed to access personal data without consent (ForgeRock, 2018). The potential issues about data privacy have been acknowledged before. Consequently, since May 2018, new European privacy legislation has been actuated. This legislation named "General Data Protection Regulation" (GDPR) puts control of personal data back in the customers' hands. Under GDRP, data must be collected transparently, used only for its stated purpose, accurate and up to date, protected and deleted when the relationship ends. Due to the GDRP, organizations have increased responsibilities of securing individual's data while also asking permission for using and/or sharing this information with third parties. The six main principles of the renewed GDPR are displayed in figure 2.

Figure 2 - The six main principles of the GDPR

Contradictory to the high percentages of concerns about privacy, individuals tend to act differently. "It is a documented fact that users have a tendency towards privacy-compromising behavior which eventually results in a dichotomy between privacy attitudes and actual behavior" (Barth & de Jong, 2017). This inconsistency between privacy attitudes and privacy behavior is frequently referred to as the "privacy paradox" (Kokolakis, 2017) (Williams et al., 2018). The phenomenon of the privacy paradox is not (yet) tested extensively in combination with smart homes. Since this is a growing market, this topic is becoming more relevant today.

1.2 Research Question

The new regulations in the GDRP; growing attention towards data privacy and the increasing number of smart home appliances makes data privacy one of the most important topics for the upcoming years. In the build environment, this topic is hard to ignore, however easy to burn your fingers on. Real estate developers, smart home appliance manufacturers and energy companies all need to make significant investments in smart home technologies to attract new customers. A deeper understanding of data privacy might be the difference between the success or failure of new technologies as smart home appliances.

The decision of an individual to share private data is determined by a trade-off between the risks of sharing privacy-sensitive data and the benefits the users receive from the products and/or services. Potential risks are the are data theft, data loss, data corruption, identity theft and economic crimes such as credit card stealing (Potoglou et al., 2017). Potential benefits of smart home appliances are increase in comfort, safety, reliability and reduction in energy consumption. If the benefits outweigh the risks, individuals are willing to 'sacrifice' their privacy. Researches in the fields of online shopping and social media suggests that individuals are willing to share their personal data for a relatively low benefit. This is however not researched extensively in relation to smart home appliances. At this moment, there is no research focused on the trade-off's individuals are willing to make between data privacy and the benefits of smart home appliances including energy efficiency. Most studies about privacy are focused on how individuals think about data privacy instead of how they act. The goal of this research is to obtain insight into the trade-offs that individuals are willing to make between sharing privacy-sensitive data and the benefits of smart home appliances. The main research question that will be answered within the graduation thesis is:

1.2.1 Research Questions:

1	To what extent are individuals willing to trade-off privacy-sensitive data to obtain the benefits of smart home appliances including energy efficiency?

1.2.2 Sub Questions:

1	What type of data are smart home appliances processing to achieve energy efficiency and what additional benefits can be achieved?
2	Which types of data do individuals experience as privacy sensitive and how does this affect their choice behavior?
3	How do individuals experience data privacy in relation to smart home appliances and which personal characteristics play a role in their choice behavior?
4	For what benefits of smart home appliances including energy efficiency are individuals willing to trade-off for privacy sensitive data?

1.3 Research Approach

Privacy is a regularly discussed topic in media. Thus, individuals might react differently nowadays when they are surveyed about data privacy. Consequently, qualitative research might result in a misleading view on the topic. A quantitative approach is more suitable for this research. To do this, a discrete choice experiment (DCE) will be executed. DCE is a quantitative technique that enables to track down individual preferences. It allows researchers to uncover how individuals value attributes by asking them to state their choice over different hypothetical alternatives. A DCE is suitable for this research since it allows testing different characteristics of theoretical smart home appliances while it comes as close as possible to measuring actual behavior. Also, DCE's are implementable in a survey which means that it is relatively easy to obtain a large amount of observations. The most discussed theoretical base of DCE is the random utility theory which suggests that individuals strive for the maximization of total utility or total satisfaction received from consuming goods or service. Based on this theory, model estimations will be made using the multinomial logit (MNL) model and the mixed logit (ML) model.

This research has two important topics of literature research that need to be addressed, namely: Data Privacy and Smart Home Appliances. As stated in the research problem, there is no comprehensive research found that combines these two topics. This research aims to find common grounds between these two topics. Firstly, the definition, theory and regulations regarding data privacy are mentioned. Secondly, data privacy will be compared to smart home appliances. At the end of the literature research, the input for the DCE experiment is clarified and the first Sub-Question (SQ1) is answered. When the literature part is finalized, the discrete choice experiment (DCE) will be prepared. The goal of the experiment is to determine the willingness of individuals to trade-off privacy for the benefits of smart home appliances including energy efficiency. In general, the quality of the results is depended on the quality of the experiment. Thus, all decisions regarding the experiment are considered carefully. Important considerations are dealing with the number of attributes, levels, choice situations and the execution of the survey. Also, the methods of data analysis will be explained. After the choice experiment is executed and has reached the required number of respondents, the results will be analyzed. This will be accomplished using the Multinomial Logit (MNL) modeling and Mixed Logit (ML). The modeling will be used to test which types of data will be experienced as privacy sensitive (SQ2) which personal characteristics affect the choice behavior (SQ3) and for what benefits individuals are willing to trade-off privacy sensitive information (SQ4). After the sub-questions are answered, the main research question will be answered in the conclusion of this research.

1.3.1 Scientific Relevance and limitations

This research will provide an insight into which data individuals are willing to share, with whom and for what benefits. Because discrete choice modeling is used, the respondent's answers are considered as meaningful since it comes closer to their actual behavior. Because there is a lack of scientific evidence that discusses choice behavior in relation to smart home appliances, this research hopes to increase the attention on the subject. Also, the results might assist both academics as organizations with the implementation of the GDPR while having an insight in the actual behavior of individuals. The research covers fast-evolving topics such as smart homes, smart technology and data privacy. It is therefore that the research boundaries are of importance. This research is focused on the smart home appliances that will be beneficial to reduce energy consumption. Furthermore, the data that will be discussed is the data where the user needs to give consent according to the GDPR. This research will not discuss the implementation of the GDPR for organizations. Also, this research does not include the anonymization of data and the security of data within the technology of smart home appliances.

1.4 Conceptual Framework

To demonstrate the link between the different topics in this research, a conceptual framework is constructed (Figure 3). The first step involved is the literature research about the potential risks of data privacy and the potential benefits of smart home appliances (SQ1). Using the DCE, it is tested what types of data is experiences as privacy sensitive (SQ2), what personal characteristics affect the individuals' choices (SQ3) and it examines the difference between the privacy concerns that individual experience and the actual choices that individuals make (SQ4).

Individuals strive for a maximization of the total utility that can be achieved in a choice situation. The utility is determined by an individual by comparing the potential risks and benefits and deciding which choice situation is more valuable (higher utility) (Li, 2012). The utility score is tested with a Discrete Choice Experiment (DCE). This is a theoretical experiment where individuals will be surveyed about their behavioral intentions. A DCE do not test real-world observations. However, this type of experiment comes closest to test actual behavior while having the benefit of being able to reach a large number of observations. Still, it is not the actual choice that will be tested, but the individual's behavioral intention.

Every individual has a different interpretation of the utility. The choices that individuals make are different, but certain groups of individuals might have similar behavioral intentions. For example, students might not care as much for the potential risks than the elderly. To test the behavioral intention of the respondents, different socio-demographic characteristics will be questioned and compared to the choices that have been made. Also, the privacy paradox suggests that there is a contradictory between the perception of privacy concerns and the behavioral intentions individuals have. Thus, it is tested if similar groups of individuals have similar concerns and similar behavioral intentions. Thus, both socio-demographic characteristics and perception of privacy concerns are affecting how respondents experience the risks and benefits of a smart home appliance. As shown in the framework, the behavioral intention can be tested with several variables like socio-demographic characteristics and perception of privacy concerns of individuals.

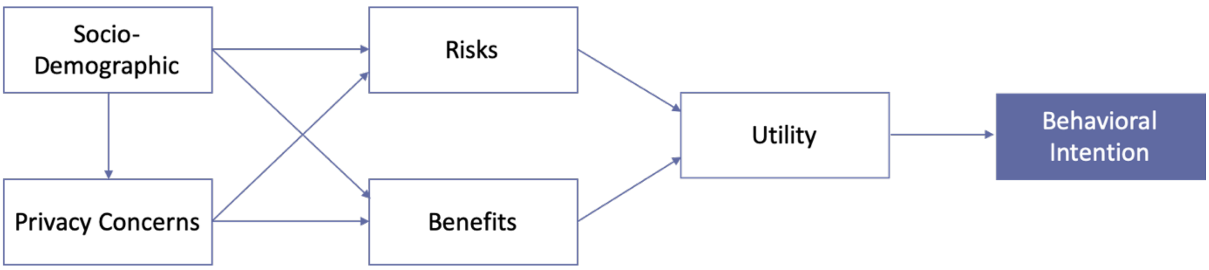


Figure 3 - Conceptual Framework

1.5 Reading Guide

In the next chapter, the literature that is related to data privacy and smart home appliances will be examined. The goal of this chapter is to theorize the topics and establish the input for the discrete choice experiment. Chapter three explains the methodology of this research. It discusses several considerations such as the type of survey instrument, Choice experiment considerations and data analysis methods. In the fourth chapter, the data collected from the survey will be analyzed. Firstly, the descriptive statistics of the survey are discussed. Thereafter, the behavioral intentions are analyzed using the multinomial and mixed logit models. Lastly, behavioral intentions are tested by adding interaction terms in the mixed logit model. The fifth chapter of this research provides the overall conclusions of this research. The main research questions will be answered based on the conclusions of the sub-questions. Additionally, the scientific relevance, the project evaluation and the recommendations for further research are discussed in this chapter. After the conclusion, the bibliography is provided with the reference to all documentation that have been used for this research. Lastly, additional information is provided in appendixes I to VI.

2 LITERATURE RESEARCH

Chapter two reviews the literature that is related to data privacy and smart home appliances. The goal of this chapter is to theorize the topics and establish three literature matrices that are used in further analysis.

2.1 Data Privacy – General Data Protection Regulation

Since May 2018, new regulations are actuated in the European Union including the Netherlands. These regulations have been upgraded to fit into the 21st century while also creating uniformity in the regulation throughout Europe. The previous regulations were suggested as unclear and lacking guidelines for future regulation and standardization (Ginosar & Ariel, 2017). The renewed European regulation regarding data privacy is named the General Data Protection Regulation (GDPR). The GDPR aims to protect all EU citizens from privacy and data breaches in today's data-driven world (EUGDPR.org, 2018).

2.1.1 Relevant definitions in the GDPR

The GDPR works via the direct effect of European law which means that it enables individuals to immediately invoke a European provision before a national or European court. In Dutch, the GDPR is translated to '*Algemene verordening gegevensbescherming'* (AVG). Since the GDPR leaves space for local regulations, the *'Uitvoeringswet Algemene verordening gegevensbescherming'* (UAVG) describes the day to day execution of the GDPR in the Netherlands. Even though the regulations are European, the GDPR does not only apply to Europe-based organizations. It also applies to an organization that is based outside the EU that offer goods and services to EU citizens. Every piece of 'personal data' that is collected, stored or shared needs to be processed according to the GDPR. The concept of 'personal data' has a broad definition, namely.

> *'personal data' means any information relating to an identified or identifiable natural person; an identifiable natural person is one who can be (in)directly be identified, by reference to an identifier such as a name, identification number, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person*

The concept of data processing is critical to understand the regulations correctly. According to the renewed GDPR, processing of data means any operation or set of operations which is performed on personal data. Examples of data processing are collection, structuring, storing, adapting, combining or erasing of personal data. The party that performs data processing is named a 'data processor'. Another legal person or body is named the 'data controller'. The controller determines the purposes and means of the processing personal data. The GDPR states different regulations for the controller and processor. The 'data controller' is responsible for appointing a suitable 'data processor' who can provide guarantees that they have implemented safeguards (both technical as organizational) that meet the requirements of the GDPR. Regularly, the controller and processor are the same actor. Consequently, the controller has to meet both requirements that are stated in the GDPR. Since organizations cannot control themselves, the controller and processor should be different departments within the organization. Thus, the implementation of the GDPR and the distinction between the two actors has especially a severe impact on smaller businesses (Spiekermann, Acquisti, Böhme, & Hui, 2015).

2.1.2 Impact of the renewed GDPR

Although the key principles of the GDPR haven't changed since the previous regulations, multiple changes have been proposed. The change with the largest impact in the GDPR 2018 will be the right of giving consent and the right to revoke consent. *Article 4, (11)* describes consent as follows:

> *'Consent' of the data subject means any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her' (European Parliament, 2016).*

'freely given' implies that the user has a free choice of accepting, refusing or withdrawing their consent at any time. This must be done by an *'affirmative action'* which implies that the user should tick a box or sign a contract to accept or decline the consent. A silence pre-ticked box is insufficient. This principle is noticeable on websites where they ask visitors for an affirmative action to accept or refuse cookies. A related topic is the 'right to access' and the 'right to be forgotten' which indicates that the user should have a clear insight in data they are sharing and have the capability to erase their data when demanded. Also, the data to the subject must be '*specific, informed and unambiguous* indication'. *Recital 39* describes that the controller must demonstrate that the data to subject has consented to the processing of his or her personal data. This suggests (in plain text) that the processor of the data is no longer allowed to hide data processing behind generic statements like *"we may process your personal data to improve our services"*. Instead, the data controller must spell out:

1. What type of personal data will be processed? (name, email, and/or browsing behavior?)
2. Why are such data processed? (remember preferred language choice, to share with third parties)
3. Who will be processing the data? (the identity of the controller, processor, and any third-party partners)
4. When will processing take place? (including data expiration date)

This more detailed explanation that is required, have a severe impact on the day to day operation of every organization. It is experienced as an administrative burden, especially those operating in different countries. Worldwide, there are different definitions of personal data in place. Also, there are different regulations regarding the collecting and use of the data (Spiekermann et al., 2015). Consequently, there are multiple definitions of data privacy in place. Where previously data privacy was defined as having control over the personal information, including the transfer and exchange of that information. This definition is used for a long time but lost power since the rise of online information. There are multiple additions to the definition of data privacy. First, data must be collected for relevant purposes only. Also, data may only be used when the user grants access. Most importantly, the data must be collected consistently and stored securely (Ginosar & Ariel, 2017). These additions on the definition are similar to the definitions that are used in the GDPR. The definitions of personal data and privacy consent as explained in the previous paragraph is used as the leading definition for this research.

2.2 Data Privacy – Theory and the missing link with smart homes

2.2.1 Privacy theory

There are two types of trade-offs that influence an individual's privacy behavior. The privacy calculus trade-off (i.e., the trade-off between expected benefits and privacy risks) and the risk calculus trade-off (i.e., the trade-off between privacy risks and efficacy of coping mechanisms Within the privacy calculus, there are multiple methods of predicting an individual's choice. Of the 15 privacy theories that were identified by Li (2012), one of the most discussed theory is the utility maximization. The utility maximization theory describes that the choice that individuals tent to make is based on a behavior that is guided towards the maximization of total utility or total satisfaction received from consuming a good or service. The total utility is a function encompassing aspects of the goods or service and an optimal level is pursued by an individual (Li, 2012). In other words, an individual makes a risk-benefit analysis where the negative consequences are rationally weighed against outcomes, aiming to minimize the risks of information disclosure and maximize the potential benefits (Barth & de Jong, 2017).

Every individual has a different interpretation of the privacy risks and benefits. In the research of (Dinev & Hart, 2006) the willingness to provide personal information is determined by the perception of risks, concerns, trust and interest. In the interdisciplinary review of Smith et al (2011) is found out that over time, there has been a movement toward the measurement of privacy concerns as the central construct to measure the privacy risks (Jeff Smith, Dinev, & Xu, 2011). Privacy concerns can be seen as both a dependent and/or and independent variable. When 'concerns' are used as the dependent variable in the research, it is related to consumer awareness, socio-demographic characteristics and cultural differences. In the research of Potoglou et al., 2015, several antecedents (dependent variables) are listed such as socio-demographic characteristics (e.g. age, gender); personality differences such as social awareness; past distressing experiences related to disclosing personal information or privacy awareness (Potoglou, Palacios, & Feijóo, 2015).

When 'concerns' is the independent variable, most dependent variables are related to behavioral reactions and the willingness to disclose information (Jeff Smith et al., 2011). In the research of Dinev & Hart., (2006), it is found that there is a relationship between personal interest, internet trust and the willingness to provide personal data to transact on the internet. While the perceived internet risks and internet privacy concerns have a negative relationship towards the willingness to share personal information. In the research of Smith et al., (2011), privacy concerns and the privacy calculus theory are combined to construct the foundation of a framework called the APCO (Antecedents – Privacy Concerns – Outcomes). This framework demonstrates that there two streams of research have the largest impact on the behavioral reactions (Figure 4).

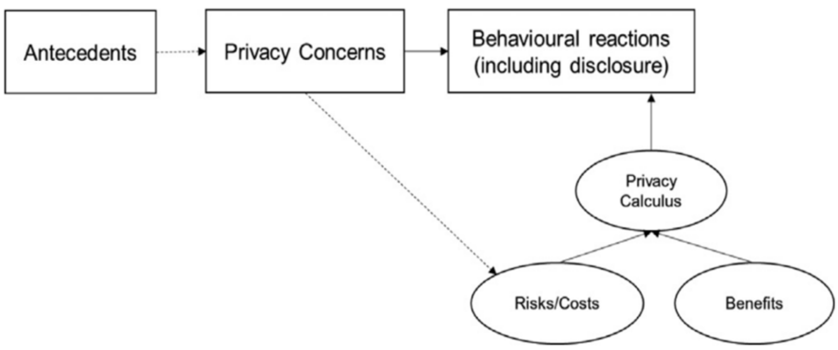


Figure 4 - ACPO Framework as used in Potoglou et al (2015), Adapted from Smith et al (2011)

The ACPO is a frequently mentioned framework that explains that antecedents such as demographic characteristics are influencing the privacy concerns of an individual. The concerns will have a direct influence on how induvial perceive the risks involved in the privacy transaction. There is also a direct influence from privacy concerns on the behavioral intention of an individual. The privacy calculus theory is used to observe the trade-offs. As mentioned before, there is a significant contradictory between how individuals think about privacy (i.e. concerns, trust, interest) and how individuals act. This phenomenon is called the privacy paradox. Even though individuals state they have severe concerns about data privacy, this rarely translates into actual protective behavior (Barth & de Jong, 2017). There are multiple explanations for the privacy paradox. Most of them are listed in the research of Kokolakis, (2017) and further categorized in the systematical literature review of Barth & de Jong (2017). The most prominent explanations are: biased individuals, habitual use, social status and the privacy calculus theory. The privacy paradox has been studied in various research settings, falling into one of two categories: social situations and transactional situations. Social situations mostly explain the paradox in relation to the use of social network sites (Kokolakis, 2017). The transactional situation where the user should choose between the risks/costs of data privacy fits the ACPO framework. The transactional situation is used in multiple contexts like e-commerce, smartphone usage, online shopping. It has a close relationship with the privacy calculus theory which is frequently used as a theoretical foundation of the research method (Kokolakis, 2017).

2.2.2 Research methodologies of data privacy

Data privacy research is often focused on the privacy attitude or the behavioral intention of respondents. This paragraph explains the different methodologies that are used in related literature and connects the topic of data privacy with smart home appliances and energy efficiency.

Privacy Concerns

When discussing the user's concerns regarding data privacy, the most frequently used methodology is an (online) survey. In surveys, multiple-choice questions are easily implemented to measure the respondent's demographics such as age, occupation, gender, income, etcetera. Additionally, concerns can be tested with ratings such as a Likert scale to measure the strength of the concern. The results of the surveys are used to assist organizations and governmental bodies with market development and policy making (Wilson, Hargreaves, & Hauxwell-Baldwin, 2017).

Different demographics characteristics affect the individual's attitude towards data privacy. These characteristics are shown in figure 1. In the literature review of Lee et al. (2019), a survey is executed to test the relationship between individuals' demographic characteristics and internet privacy concerns. They found out that gender, age, education and income have the most significant influence on the privacy concerns (Lee, Wong, Oh, & Chang, 2019). According to Potoglou et al. (2015), mostly women and older consumers are more privacy concerned when it comes to their personal information (Potoglou et al., 2015). Similar demographic characteristics are used to express the energy-saving behavior of households. In the research of Yue et al., (2013), it was concluded that the energy-saving activities are more likely to be performed by women and older individuals as well.

Similar to the demographic characteristics, there are different statements used to test the individual's attitude towards data privacy (Table1 1). For example, Naus et al., (2015) used statements to measure the acceptations of smart home appliances, smart technology and data sharing. They found that in general, respondents were supportive of energy management practices, but specific privacy concerns shaped how supportiveness of the respondents. (Naus, Van Vliet, & Hendriksen, 2015). The research of Ohler (2014) is used to describe the factors affecting energy-saving behaviors. They found that self-interest has a significant impact on energy-saving behavior. In other words, respondents that have a greater concern about energy costs utilize less electricity (Ohler & Billger, 2014). Statements that are used to explain data privacy are mostly referencing to internet privacy and not data privacy in relation to smart homes. For example, Lee et al. (2019) used 6 statements regarding Internet Privacy Concerns that was firstly proposed by Buchanan et al. (2017). Based on their statements, they concluded that women and high-income respondents show higher privacy concerns, while respondents age older than 50 had a relative low IPC (Lee et al., 2019). The research of Patil et al. (2016) focused their privacy statements on the misuse of data is widely used in relation to other privacy issues (Patil, Patruni, Potoglou, & Robinson, 2016).

Table 1 - Literature matrix attitude towards data privacy and smart homes

1st Author	Year	Field of Study	Demographic characteristics	Privacy Attitude	Smart Home / Energy Attitude
Dinev	**2006**	**Information systems**	Race, Gender, Age, Occupation, Education, Income	Misuse of information, Sensitivity of information, Third-party usage, Unknown usage	
Lee	**2019**		Gender, Age, Income, Education, Marriage	Concerns about submission, misuse, illegal use and data quantity	
Naus	**2015**	**Energy transition**		Possible privacy violations Openness towards privacy issues	Statements regarding the Acceptation of smart meter services, smart technology, data sharing, collective arrangements.
Ohler	**2014**	**Energy economics**	Socio-demographic & Household & Dwelling variables		Statements regarding Energy conservation, Environmental concern and global warming
Patil	**2016**	**Public Transportation**	Demographic & Geographic variables	Privacy threats, security concerns, privacy concerns	
Potoglou	**2015**	**E-commerce**	Age, Income, Region, Gender, Occupation	Likert Scale for General caution, Technical Protection, Privacy Concern	
Potoglou	**2017**	**Human Behavior**	Age, Gender, Income, internet use, Country	Likert Scale for data protection, internet surveillance, security concern	
Yue	**2013**	**Energy Policy**	Age. Gender, education, household composition		Energy Saving awareness, Reduction behavior, Promotion behavior, behavioral ability,

Behavioral Intentions

While a survey is suitable to explore the user's attitude, it is less suitable when you want to measure the user's behavior. The only relevant example is Ohler et al. (2014) where the respondents were asked to report the behavioral actions that are beneficial to lower energy consumption (Ohler & Billger, 2014). According to Kokolakis (2017), such self-reports are however an unreliable approach to measure the behavior accurately. Thus, an experiment is the most suitable approach to measure privacy behavior (Kokolakis, 2017). An example of a privacy experiment is the research of Preibusch et al., (2013). They tested privacy behavior by offering the respondents the option to purchase a DVD from one of two online stores where one of the stores asked for more invasive personal data (Preibusch, Kübler, & Beresford, 2013). Another example of an experiment in relation to a smart home is the research of Paetz et al., (2013). They tested actual behavior by letting test residents move into a dwelling and track their energy consumption. Although both experiments provide qualitative results, these types of experiments are not frequently used since it is costly and time-consuming. Also, the results are only interpretable for specific situations and it is often chosen to test the behavior for certain groups such as students or in certain countries.

To solve the issues, several studies do not capture privacy behavior but instead, capture behavioral intentions. This means that the respondents are surveyed about what they would do in certain (privacy-invasive) situations. Measuring behavioral intentions is suggested as a hybrid approach that enables testing of privacy behavior while using some sort of survey. A commonly used 'hybrid' methodology is named choice modeling. Choice modeling is a statistical technique that looks at the choices that individuals make between alternatives of products and/or services. An example of privacy-related choice modeling is the research of (Patil et al., 2016). They implemented a stated preference experiment to test the preferences for various privacy settings in the context of security and surveillance of train/metro facilities in Europe. Also, Potoglou et al., (2015 & 2017) used a stated preference experiment to test the behavioral intention in relation to internet surveillance and e-commerce (Potoglou et al., 2017, 2015). Measuring behavioral intentions instead of actual behavior has downsides too. First of all, there is a clear difference between theoretical and actual choice situations. When a real-life situation occurs, individuals might react differently. Also, the term behavioral intention is more open for interpretation. Researchers have used various disciplines to theorize behavioral intentions such as social theory, behavioral economics and psychology (Kokolakis, 2017).

The examples of capturing behavioral intentions in relation to smart homes are scarce. The research of Broberg & Persson, (2016) a used web-based choice experiment where respondents were faced with three hypothetical electricity contracts (Broberg & Persson, 2016). Ohler used both behavioral intentions and actual behavior to determine the factors that affect energy-saving behaviors and electricity usage (Ohler & Billger, 2014). While both researches provide valuable insights, data privacy was not mentioned. When privacy is present in smart home research, the attitude towards privacy and energy consumption is discussed, not behavioral intentions. For example, Naus et al., (2015) examined the participation of Dutch households in a smart and sustainable energy transition while including horizontal and vertical privacy concerns. They found out that the participation was shaped impeded or even obstructed by privacy considerations (Naus et al., 2015).

Research that combines the behavioral intentions within the context of smart home appliances and data privacy is not found at all. Thus, this can be seen as a research gap that needs more attention. This attention is already established in the fields of e-commerce and social networking and public transportation. Table 2 shows a literature matrix that provides evidence of the missing relationship. The most relatable examples are the researches of Potoglou et al., (2015 &2017). In both papers, they offered insights about multiple relevant privacy dimensions including data storage, retention of data and the trade-off between data privacy and enhancing technologies (Potoglou et al., 2017).

Table 2 - Literature matrix behavioral intention towards privacy and smart homes

1st Author	Year	Field of Study	Privacy Behavior	Smart Home Behavior
Broberg	**2016**	**Energy Economics**		Behavioral intention related to demand management of Swedish households' energy use
Dinev	**2006**	**Information systems**	The attitude that influences the behavioral intention to provide personal information to conduct transactions on the Internet.	
Ohler	**2014**	**Energy economics**		Behavioral intention & actual behavior related to the factors the affected energy-saving behaviors and electricity usage.
Paetz	**2011**	**Smart Home appliances**		Actual behavior by test-residents to move into the smart home and experience the technologies on a daily basis.
Patil	**2016**	**Public Transportation**	Behavioral intention regarding privacy and surveillance in PT	
Preibusch	**2013**	**E-commerce**	Actual behavior between privacy and online shopping	
Potoglou	**2015**	**E-commerce**	Behavioral intention and the role of privacy concerns in e-commerce	
Potoglou	**2017**	**Human Behavior**	Behavioral intention regarding privacy implications of internet surveillance	
Tsai	**2011**	**E-commerce**	Actual behavior between privacy and online purchasing decisions.	

2.3 Smart Home Appliances – Relevant Actors

2.3.1 Data processing by energy companies

There are three major actors that distribute energy to the consumers (Energiekamer, 2019). The two actors that are closest to the consumer are the energy suppliers and retailers. They have direct contact with consumer which makes them most familiar. The largest energy suppliers in the Netherlands are Innogy (Essent & Energiedirect), Eneco and Vattenfall (Nuon). Together these companies have approximately 7.5 million customers in the Netherlands. In recent years, the energy market became saturated. “Processes of liberalization, privatization and environmental activism have given rise to more fragmented, competitive energy networks with a diversity of energy providers” (Naus et al., 2015). This diversity resulted in an increase in smaller energy suppliers entering the market that focused on the smart home concept. Especially suppliers with special attention on green energy are upcoming (Energievergelijk, 2019).

The energy supplier is not the actor that is responsible for energy distribution. The first in line is the Transmission System Operator (TSO). The two TSO’s are ‘Tennet’ for the high voltage network and ‘Gasunie’ for the gas network. Between the TSO and the suppliers are the Distribution system operators (DSO) which provide the connection to houses and maintenance of the electricity and gas networks. Figure 5 shows a simplified value chain of electricity production, distribution and consumption (Rodríguez-Molina, Martínez-Núñez, Martínez, & Pérez-Aguiar, 2014).

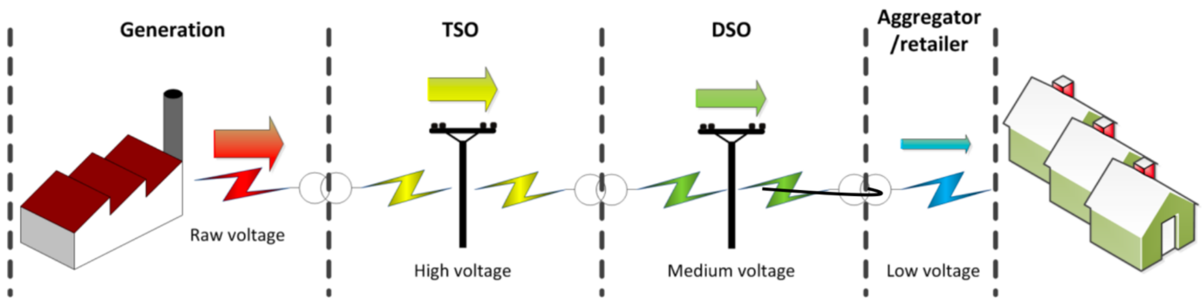


Figure 5 - Value chain of the Dutch energy industry Rodríguez-Molina (2014)

Consumers are having the strongest relation with the energy suppliers and the DSO’s. Since the energy market has evolved, DSO’s took an active role in facilitating and managing consumer data. DSO’s are aiming for a resilient network with flattened consumption peaks. Energy suppliers are focusing on a larger customer base by luring consumers with a price reduction into a new contract. The implementation of smart home appliances has a significant role since it may decide the consumer to change to another supplier.

Data processing by the energy industry

For all energy-related data, the Dutch energy market has established a central organization to smoothen the administrative processes (Van Aubel & Poll, 2019). This organization named ‘Energie Data Services Nederland’ (EDSN) takes responsibilities in providing metering data to energy suppliers irrespective of the responsible party. Thus, the EDSN is the processor for all energy-related data in the Netherlands. In the current set-up, metering data is inactively collected by EDSN into a central database. The metering data is only stored in the meter located in a consumer’s dwelling. When an energy supplier requires metering data of one of its customers, it first has to request the data from EDSN; EDSN forwards this request to the responsible DSO, which in turn retrieves the data from the customer’s meter and sends it to EDSN. EDSN caches the data, and the energy supplier has to contact EDSN again, the next day, to retrieve the data. For all energy-related data, ESDN works as the data processor which creates continuity in the process and makes easier collaboration with other parties (Van Aubel & Poll, 2019).

Code of Conduct

All actors in the value chain of electricity distribution have signed a code of conduct where they legally obliged to conform with. In the code of conduct named '*Gedragscode slimme meters voor netbeheerders'* is specified for how frequent data from smart meters can be collected and for how long this data can be stored. Table 3 shows the periodicity and retention time as described in the code of conduct. It explains that if there is an increase in frequency, there will be a shorter retention time. Thus, DSO's are restricted in the quantity of data they can store about their consumers. The data that is used by DSO's are used to estimate consumption peaks and to predict consumption. The Energy suppliers are using this information for billing the consumers and price estimation.

Table 3 - Periodicity and retention time according to code of conduct of Dutch DSO's.

Periodicity	**Retention time**
Monthly	13 Months
Daily	40 days
Hourly	10 days
15 min	10 days

Energy Cost

For 17 years in a row the average energy costs is increased (Rijkdienst voor ondernemend Nederland, 2018). In 2001 a household paid an average of €820 for gas and €415 for electricity. In 2018 the costs were increased to €1048 and €544 for gas and electricity (Nibud, 2019). This implies that on average, a Dutch household pays €133 per month for energy. There are two variables that affect the average energy consumption most. These variables are the household composition and the type of dwelling. Due to these variables, comparing smart technology is complicated since the parameters are different. Also, it is hard to measure what percentage of energy reduction is achieved by the technology and what percentage is achieved by the changing behavior of the user. According to Paetz et al. (2011), there is proof of an energy reduction of up to 27 percent. Darby (2010) mentioned that an energy reduction between 5 and 15 percent is possible with direct feedback options. And the more direct the feedback is, the higher the potential savings are (Grønhøj & Thøgersen, 2011). In the pilot study executed by Van Dam et al., (2010) there were households that saved up to 42.6 percent. However, there were also households that showed an increased energy consumption of up to 40.2 percent indicating that there are many more parameters with a significant influence on the consumption (Van Dam, Bakker, & Van Hal, 2010).

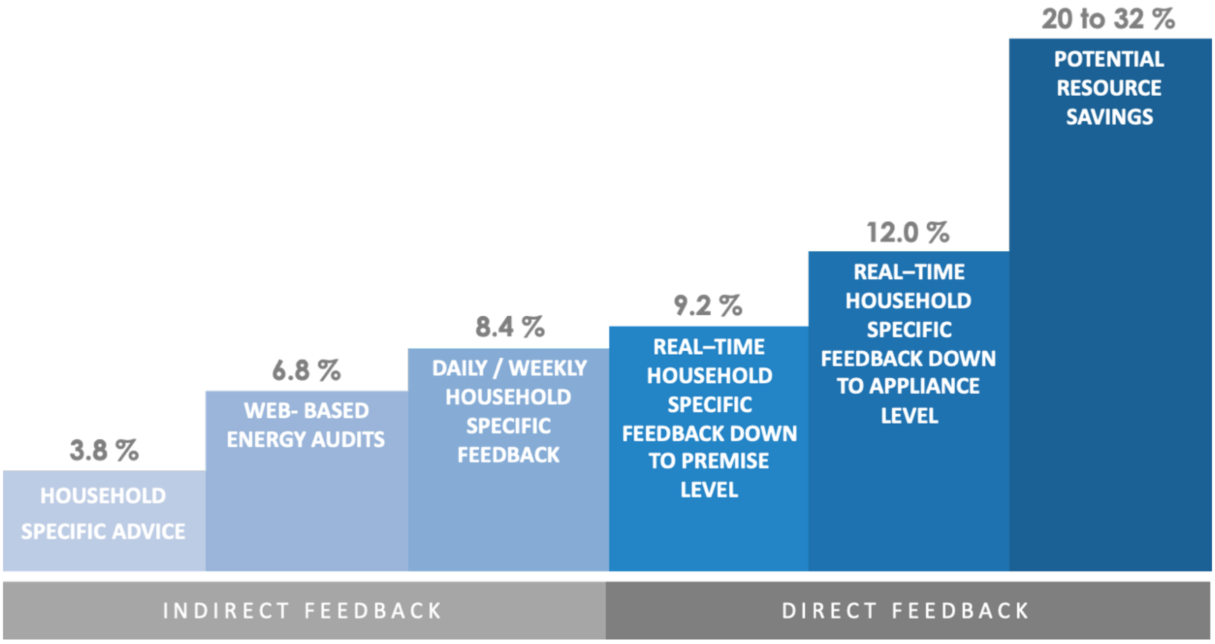


Figure 6 - Potential annual savings of feedback approaches Murray & Hawley, (2016)

Based on €133 monthly costs in the Netherlands and the potential energy reductions mentioned in the literature, the financial benefits may vary between monthly savings roughly between €5 and €36. Based on the average reduction of 8,1 percent as shown in Grønhøj & Thøgersen (2011), the monthly saving would be just under €10,-. However, this is an inaccurate estimation that cannot be executed accurately without context-specific research. Besides the financial benefit, smart home appliances have a high(er) installation cost. Thus, investments for smart home appliances are attractive for homeowners that have necessary funds, automatically excluding those on low incomes and tenants living in privately rented properties (Balta-Ozkan et al., 2013). Also, there is no demonstrable rate of return since the investments are significant while the savings are mentioned as meaningless (Balta-Ozkan et al., 2013). Energy suppliers that offer a smart thermostat with an energy contract is therefore an interesting option for many consumers.

2.3.2 *Data processing by technology companies*

Smart home appliances are integrated in the daily life of individuals. As a result, companies show a growing interest in the smart home concept. The two largest actors of smart home appliances are Google and Amazon. Both offer a variety of devices with their brands named Google Nest and Amazon Ring. Examples of the products they offer are doorbells, thermostats, cameras, WIFI hubs and smoke detectors. Another large manufacturer of smart home appliance is Samsung. They produce different types of appliances such as washing machines and refrigerators. The goal of tech companies is mainly focused on creating an eco-system of smart devices (phone, watch, glasses, cars and smart home appliances). The union of previously separated devices removes the physical boundaries between homes and redefines the concept of smart technology (Marikyan, Papagiannidis, & Alamanos, 2019). Similar to Google and Amazon, most tech companies that focus on smart home appliances are multinationals. Consequently, these companies have huge databases of processed data that is valuable for whoever has access. There are multiple examples of governmental bodies that have reached out to tech companies for data sharing of its users (Bélanger & Crossler, 2011). Sometimes multinationals have shared their data, but there are also examples where they openly refused this request. There is little empirical evidence when it comes to the acceptance and adoption rates of smart home technology that is provided by tech companies. The number of smart devices is rapidly growing and so does the market share of these multinationals. Even though it is known by the users that 'big brother' is watching them, the revenues of Google nest grew with 69 percent between the first and last quartile of 2017.

Technology companies are unlike the energy industry, solely focused on their revenues by increasing product sales and services. They try to find the balance between enhancing the consumer's trust while turning data into value. Since the rise of big data, organizations are in the need of systems that not only perform data analysis but then also extract the results in such a way they find clear benefits in the data. Since technology companies see data as a revenue model, the data itself has a different price tag. Data is worth money since organizations can use data for marketing or to improve their products and services. Valuating this data is challenging since in the current situation this data is used freely.

The research of Acquisti et al., (2009) questioned what privacy is worth. With a series of experiments, the research examined the valuations of individuals privacy. They found out that the value individuals assign to data privacy depends on the question itself. "When people assign to protect a piece of information is very different from the price they assign to sell the same piece of information" (Acquisti, John, & Loewenstein, 2009). In other words, they proved that individuals are not willing to spend even a few cents to protect their data privacy but the same individuals reject offers of several dollars to sell the same data. The research of Tsai et al., (2011) summarized several papers that discuss the valuation of privacy. They concluded that the value of selling their data is very dependent on the context of the research. They showed evidence of data valuations between 28 and 55 US dollars (Tsai, Egelman, Cranor, & Acquisti, 2011).

2.4 Smart Home Appliances – Potential Benefits

2.4.1 Processing to achieve energy reduction.

A smart home is equipped with multiple devices that collaborate as a homogeneous system to monitor electronic appliances and promote efficient energy management and sustainability. Energy-focused smart home appliances aim to achieve an energy reduction while also flattening consumption peaks. Additionally, it is aimed to achieve a resilient energy network while also promoting environmental sustainability. Energy efficiency with smart home appliances is enabled through the implementation of four services named by Marikyan et al., (2019), namely:

1) Monitoring the information on energy consumption
2) Controlling the consumption patterns through remote devices and direct control
3) Management of the service, aimed at achieving efficiency and optimization
4) Consultancy

There is an abundant of appliances available on the market that assists with monitoring, controlling, managing and consulting energy consumption. The most common appliances are the smart meter and smart thermostat. A smart thermostats permit the user to turn on the heat or air conditioning when going home which means it is a consumer focused appliance (Luor, Lu, Yu, & Lu, 2015). A smart meter allows for a radical change in customer–utility relations. It communicates back to the utility for monitoring and billing purposes with the possibility of remote change of tariff, and the allowance of adaptive peak demands (Darby, 2010). Also, smart meters make it easier to switch energy supplier while there is no cost of having the meters read (Van Aubel & Poll, 2019).

The main reason for consumers to invest in a smart thermostat is to have an insight in their energy consumption. Alternative reasons are consumption awareness and contribution to the environment (Rijkdienst voor ondernemend Nederland, 2018). The smart meter is beneficial for the energy supplier since it automatically shares the energy consumption. This improves the accuracy of energy billing while also improving reliability. Literature regularly combines the topic of energy consumption with a smart grid. Smart grids are expected to promote the production of renewable energy sources while also improving energy management through detailed monitoring and intensive two-way communication between sites of production and consumption (Naus et al., 2015). The investment in smart home appliances reduces energy due the improved coordination between electricity, thermal and gas grids (Lund, Østergaard, Connolly, & Mathiesen, 2017). As a result, it has a significant change to replace the relative old infrastructure of current grids in the Netherlands.

Smart home appliances contain modern technology that is focused on reducing energy consumption. This does not mean that the technology itself is more energy-efficient than the non-smart equivalent. Smart devices might contain extra features as large touchscreens and additional sensors that increase the energy consumption of the device itself. Also, smart appliances are sometimes new in the household (e.g. smart doorbell) which also increases the energy costs. The energy reduction is only achieved by a collaboration between the user and the technology. Thus, data processing is required to achieve an energy reduction. In the research of Wilson et al., (2017) the benefits and risks of using smart home technology have been researched. They state that smart home appliances are designed to achieve an energy reduction with three approaches, namely (1) The product provides information to the household (2) The product enables the household to control their smart home (3) The product controls the house on behalf of the household. If a smart home appliance provides information as a service, it is uncertain it will reduce energy. But it empowers users to remotely control household appliances and decrease the burden of everyday household activities. Thus, it may stimulate the user in changing their lifestyle. When a product takes over responsibilities, an individual has to put trust into the smart home appliance. The product demands data sharing otherwise, the service will be at the expense of comfort. For example: without specific data about the home-owner, a smart thermostat will remain a low temperature. When the product controls the dwelling on behalf of the household, the data quantity increases. The product should anticipate based on the users' activities, preferences and habits while also focusing on saving energy (Marikyan et al., 2019).

Change in energy behavior

Energy reduction is not only achieved by the appliances but also by a changing attitude of the user. Research on household energy use and energy-saving behavior has found that several types of factors can influence energy-usage behavior. For example, socio-demographics, moral norms, various incentives and barriers, energy-saving awareness and attitudes, regulations and policies, informational and promotional activities (Yue, Long, & Chen, 2013). Ohler & Bilinger (2014) found out that most consumers do not have an energy-saving attitude by themselves. Simple behavioral adjustments as 'lowering water heater temperature' and unplugging battery chargers are barely done (Ohler & Billger, 2014). The research also states that a pro-active attitude does not lead to pro-environmental behavior. This implies that technology should enhance individuals' behavior and take control where possible. However, when technology takes control, an increasing amount of private data will be stored and shared.

Since energy savings can be achieved by partly the user's attitude and partly the product itself, the actual reduction is hardly measurable. The total energy savings are depended on the type of dwelling, type of product, household size, etcetera. This suggests that the impact on the energy demand of smart home appliances once adopted is unclear. Research states that attitudinal energy reduction is hard to remain. By testing the possible effects of a smart home, Peatz et al., (2011) found out that an energy reduction is achievable. However, consumers are unwilling to sacrifice their comfort (Paetz et al., 2011). According to the research of van Dam et al., (2010) consumers in the Netherlands were unable to maintain their attitudinal energy reduction for a longer time (test period of 15 months). The longer the smart thermostat is installed, the lower the attention to the smart meter, the lower the attention to their consumption behavior. Participants who kept the smart thermostat were unable to sustain their electricity savings any better than those without a monitor (Dam et al., 2010). For a larger reduction, smart products should have a focus on persuasive technology which assists the user towards an energy-saving behavior for a longer time.

2.4.2 Processing for additional benefits

There are several other potential benefits of smart home appliances. In this paragraph comfort, safety, health, and reliability are addressed.

Comfort

In the paper of Park et al (2018), the acceptance of technology is described as a combination of the comfort value, hedonic value, security value and economic value (Park, Kim, Kim, & Kwon, 2018). This distinction displays that besides the economic value, all benefits are mostly intangible and therefore service-orientated. The most frequently discussed service in literature is comfort. This is a different experience for every individual which makes is complicated to express. Marikyan et al. (2019) categorized 3 types of comfort of smart home appliances. First the automation of daily routines. This category is explained with the example of a smart dishwasher that turns itself on when the energy prices are reduced. For this category, not much user-specific data is required. Secondly, 'remote home management' which assist the user by locking doors automatically or sending an alert when windows are left open. For this category, often a phone connection is made to assist the user. The last category is smart appliances which detect the environment and will act accordingly. For example, when the lights are turned on when an individual enters the room. The latter category tracks the user activities, which makes it more data privacy sensitive.

Since smart home appliances are usually the modern equivalent of an already existing product, most appliances are capable to execute tasks that were previously carried out by a human. This is regularly named 'home automation' or 'controllability' (Yang, Lee, & Zo, 2017). Another benefit is referred to as interoperability which indicates that smart home appliances use data from different smart appliances and work together for optimized service. As a result, the devices can be managed via a smartphone tablet or central touchscreen in the house. Since most products are modernizations of existing products, there is a limited learning curve.

Safety

Another reason for data processing is security and physical safety. Home security and physical safety fall partly outside the scope of this research. However, this topic is relevant to user perception and acceptance rates. The literature focuses on two aspects of security, namely, the security of a smart home itself and the security of the data within a smart home (Luor et al., 2015). Literature discussing data security mentions the avoidance of data hacking, data encryption and data anonymization. Research discusses that data security and safety can be supported by smart technology. Nevertheless, smart home appliances are increasingly vulnerable to data hacking in comparison to their old-fashioned equivalent. Since smart appliances are connected to the internet, hackers find new approaches to retrieve personal data from the user which can be used for espionage, theft or even terrorist attacks. Safety also implies to physically being secure from intruders. The focus of these products lies in detecting unusual behavior. This is done with pressure and motion sensors to detect open/closed windows and smart cameras (Balta-Ozkan et al., 2013). In the findings of Luor et al. (2015), it is concluded that respondents usually trust the security function of smart homes. They are willing to share privacy-sensitive data in exchange for physical privacy. The biggest barrier with the implementation of these products is the high costs of the technology. Although the price of the technology is becoming cheaper in the future (Luor et al., 2015).

Health

The homecare concept uses technology to provide cost-effective solutions for aging and vulnerable users (Marikyan et al., 2019). Smart home appliances are capable to remotely monitor health, detect emergencies and even provide medical care from distance. Thus, smart home appliances can achieve a reduction in hospital admission and practitioner visits. As a result, elderly individuals are allowed to maintain healthy and independent living for as long as possible. The cost of healthcare is rapidly increasing, these appliances can reduce costs while increasing the quality of care. Contradictory to the last statement is the stated that technology has been the principal driver of the increase in health care costs in the last 50 years (Chan, Campo, Estève, & Fourniols, 2009). Since numerous appliances are focusing on the elder population, new challenges occur. In general, the acceptance rate of smart technology is lower for the elderly since they are uncomfortable using these products. Thus, these smart appliances need to work seamlessly without additional tasks for the user.

Reliability

Smart home appliances are also focusing on an increase in reliability. When data is used to control energy consumption and appropriately manage maintenance, products will have a longer lifespan. At this moment, implementing innovative smart home appliances will however lower the reliability of the product itself since technology is relatively new. Also, the collaboration between devices is not optimal. Balta-Ozkan et al., (2013) exampled that boiler designers and home computer developers work under different assumptions about the appropriate tolerance level for crashes. Combining the two different products introduces room for complications and potentially cause dangerous malfunctions. Even when there are no technical malfunctions, there may still be an unreliable service because the system is lacking intelligence to correctly understand or anticipate the needs of its occupiers (Balta-Ozkan et al., 2013). These technological misunderstandings are expected to reduce which increases reliability.

2.5 Smart Home Appliances – Data Processing

There are many smart home appliances available on the market which result in countless data controllers and data processors. As a result, products are unable to communicate because there are different agreements in place. For example, your mobile phone manufacturer cannot see your energy consumption without clear consent. In the research of Wilson et al. (2017), three methods of controlling a smart home appliance are mentioned. First, with pre-set schedules or profiles. This means that there is a fixed moment where the device will be activated. Secondly, technologies will be controlled on automatic responses which means that the device reacts when certain behavior is detected. The last category is called 'in the spur of the moment input or adjustments' which means that appliances are controlled manually by the user. With all three categories of controllability comes new approaches to process data. Also, processing takes place at different moments (Wilson et al., 2017).

Two types of data storage are used by a smart home appliance. First the internal storage of the smart appliances and secondly and most importantly the connectivity with the 'cloud'. Cloud computing provides scalable computing power, storage space and accessibility of the smart home appliances. Due to the increasing number of smart appliances in the IoT, the amount of stored data in the cloud is rapidly increasing too. The longer this data is stored, the more detailed organizations are capable to build a consumer profile about the user. Building a consumer profile is accepted by most users (Naus et al., 2015). The GDPR 2018 provides the outlines of data processing with smart appliances. If the user requests it, data controllers are legally obligated to stop data processing and delete all collected data. Organizations are also obligated to inform the user about the duration of storage. However, there is no limitation in the duration of storage meaning that if the user gives consent, organizations are allowed to store data until the product is out of use. Technology companies are using the legal boundaries and store and process as much data as legally is allowed.

2.5.1 Type of data that is detected

Data processing has three general stages, (1) Collection, (2) Analysis and (3) Usage. Most of these processes are atomized which means that it converts (or transmits) input data into output data without human interaction. This transmission is done by computer algorithms. Recently, there is an increasing focus on Artificial Intelligence (AI) which supports the optimization of algorithms. Due to AI, algorithms perform have increased performance the longer they are in operation. Since there is an abundance of data that is used by smart home appliances, individuals are unaware of how this data might affect their lives and what the potential consequences there are. Individuals might easily share personal data due to ignorance on this topic.

Data detection is generally a seamless process. In other words, there are no physical interactions with the devices that (in)activates data collection. Smart home appliances contain detection mechanisms. such as sensors, camera's or actuators that are embedded in the structural fabric of the smart home (Chan et al., 2009). Most data detection is done by sensors. There are three general types of sensors namely, Proximity sensors, Motion sensors and Image sensors. In smart homes, sensors are capable of the detection of data like temperature, moisture, sunlight, time, energy usage, power quality, voltage quality, etcetera. Especially the last-mentioned examples are related to energy consumption. Of these, the actual energy usage and power quality and considered as privacy-sensitive since it is related to energy usage behavior. Voltage quality and information about the meter itself is are not considered privacy-sensitive (Van Aubel & Poll, 2019).

Data types such as temperature or power quality are examples of raw data since the data needs considerable processing before it is useful and/or privacy sensitive (Darby, 2010). As a result, mistakes and irregularities may occur. Even though a smart home appliance might be full of advanced technology, the detected data like energy consumption might be inaccurate. Explanations for incorrect data can be obvious such as a change in season, weather, household size, number of devices, etcetera. But sometimes reasons can be unexpected as explained by the University of Twente. They found out that in the Netherlands approximately 750.000 smart meters are detecting a higher consumption than the actual consumption. This is mainly because the detection of the smart meter was influenced by modern (energy efficient) smart home appliances and the use of light dimmers. As a result, in 5 of the 9 smart meters, the measured values were significantly higher than the actual values, up to six times the expected energy usage (Leferink, Keyer, & Melentjev, 2016).

In the research of Balta-Ozkan et al., (2013), the social barriers to the adoption of smart homes were tested. They found out that if respondents are asked about sharing their daily routines or household activities, they reacted differently than sharing 'innocent data'. However, innocent data does not exist according to their research. When data falls into the 'wrong hands', that one piece of 'innocent data' combined with a second piece of 'innocent data' becomes a piece of 'non-innocent' data (Balta-Ozkan et al., 2013). Individuals lack an overview of what can be accomplished with their data and are unaware of potential privacy threats of sharing sensitive data. What appears to be 'innocent data' according to the user, is not innocent at all. Balta-Ozkan (2013) proposed 4 categories of smart home services and the type of data that is required to achieve an energy reduction. These four categories provide a distinction in what data will be processed while also identifying the reasons for collection. In general, with every level, there will be more information collected about the user. Thus, every level will be more privacy-sensitive than the previous level. These four categories are:

(1) Daily routines to discover an individual's behavior patterns.
(2) Identity of a person to activate personal preferences
(3) The current location of an individual to determine someone's location
(4) Households' activities to predict and detect (unusual) behavior and anticipate when necessary.

To improve the services of smart home appliances, the user also needs to provide additional data. This is performed by the user itself. Examples are sharing your temperature preferences, registering your age, household composition or e-mail address. In most cases, the user's smartphone can operate as a hub to manage smart home appliances. Smartphone applications are capable to control lights, turn on dishwashers and change the room's temperature. The smartphone has also an important role in sharing a GPS signal with smart appliances. According to research, of Balta-Ozkan et al., (2014) the use of tracking devices such as GPS was frequently considered a no-go for their survey participants (Balta-Ozkan, Amerighi, & Boteler, 2014).

2.6 Literature Matrix

The literature in this chapter has been summarized in the literature matrix of table 4. The matrix shows what type of data there are processed by smart home appliances, what kind of benefits there can be achieved, the relevant actor that are mentioned and the frequency of data processing and data removal of smart home appliances. This literature matrix provides a systematic overview that assist the setup of the experimental design as discussed in the next chapter. In the first three columns of table 4, the author, year of publication and field of study are mentioned of the relevant publication. The remaining columns provide the necessary content that has been found in that particular source.

Table 4 - Literature matrix smart home appliances

1st Author	Year	Field of Study	What is processed	Why are such data processed?		Who's processing?	When does processing take place?		Trade-Off
			Type of Data	**Energy related benefit**	**Additional benefits**	**Actors**	**Frequency of sharing**	**Retention Time**	**Financial Benefits**
Aubel	**2019**	**Smart Energy**	Metering data		Automation, remote control	DSO's	Monthly, daily, hourly	13 months, 40 days, 10 days	
Balta-Ozkan	**2013**	**Smart Home**	Movements, energy usage	Controlling energy Monitoring energy	Entertainment, convenience, comfort	Operators, distribution, retailers	Day-by-day, minute-by-minute		
Balta-Ozkan	**2014**	**Smart Home**	Daily activities, house occupancy		Safety, lifestyle support, energy management				
Chan	**2009**	**Smart Home**	Raw data, sensor data	Control, detect, remind	Comfort				
Darby	**2010**	**Smart Energy**	Billing data	Supplier switching, real-time feedback, demand reduction,	Fraud reduction, accurate billing		Real time frequency		5-12% of energy costs
Gronhoj	**2011**	**Smart Home**	Historic data, everyday behavior, household size	Consumer motivation and awareness		Energy supplier	Daily, weekly, monthly, directly	User removal	Average of 8.1%
Lund	**2017**	**Smart Energy**		Flexible storage, Resilience, energy exchange		Governments			
Luor	**2015**	**Smart Home**	Camera's, data storage, fingerprints, energy bills	Flexibility, transparency	Entertainment, security, automation				
Marikyan	**2019**	**Smart Home**	Usage patterns	Control, manage, support / assist, anticipate / respond	Comfort, emotional, security, healthcare, sustainability, QoL	Apple, Google			
Molina	**2014**	**Smart Energy**	Sensors, devices, systems	Lowering peak hours, flattening peaks		TSO, DSO, Retailer, Prosumers,			
Paetz	**2011**	**Smart Home**	Voltage, Active Power, Behavioral Data, Real-time Data	Conservation, direct feedback,	Integration, flexibility, transparency		Direct feedback, fixed time schemes		No benefit up to 27% of energy cost.
van Dam	**2010**	**Smart Energy**	Raw data e.g. Watt, Volt, M^3, Celsius	Habit development		Commercial parties, gas and electricity supplier			Between 40.6% more and 42.6% less energy consumption
Wilson	**2017**	**Smart Home**	Temperature, light, motion, humidity, etc.	Demand reduction, alleviating peak loads	Convenient, easy, comfortable,		Pre-set schedules, fixed moments, during active usage		

2.7 Conclusion

This paragraph summarized the conclusions that can be drawn from analyzing the literature regarding data privacy, smart home appliances and the data regulations. It also answers the first sub-question.

The literature shows that energy reduction is barely achieved by the smart appliances itself. The largest reduction is achieved with (in)direct feedback options to the user. Data is used to predict behavior and consult the user in making energy-efficient decisions. Since consumers don't have an energy-saving attitude by themselves, smart technology should take one additional step, namely: Smart home appliances should enhance consumer behavior and automatically controls features of the smart home where possible. Still, smart home appliances can achieve an (in)direct energy reduction. The potential reduction varies between 4 and 12 percent with an average of 8.1 percent (Grønhøj & Thøgersen, 2011). Based on the Dutch average energy costs of €133,- per month, the potential benefits vary roughly between 5 and 15 euros.

The data that is collected often referred to as raw data, needs considerable processing before it is beneficial. Examples of raw data types are temperature, energy usage, power quality, voltage quality, etc. The literature suggests that individuals are unknown about the sensitivity of data. What appears to be 'innocent data' according to the user, is not innocent at all Balta-Ozkan (2013). In general, there are four types of data that need to be processed to achieve an energy reduction, namely: daily routines, the user's identity, their current GPS location and activities. The bigger the data quantity, the stronger the data sensitivity. Likewise, the more frequent the data is collected and the longer the data is stored, the more accurate a consumer profile is built about the user. Although privacy sensitive has potential privacy consequences, the data is better in assisting the user in achieving an energy reduction. It is up to the consumer to consider the trade-off between privacy-sensitive data and energy efficiency.

This page is Intentionally left blank

3 METHODOLOGY

Chapter three explains the methodology of this research. It discusses several considerations such as the type of survey instrument, choice experiment considerations and data analysis methods.

3.1 Introduction

3.1.1 Choice Modeling

To investigate the tradeoffs between data privacy and the benefits of smart home appliances, a Discrete Choice Experiment (DCE) is used. Choice modeling methods such as DCE have been used extensively in the field of marketing, transport demand and environmental science. Choice modeling is a statistical technique that looks at the choices that individuals make between alternatives of products and/or services. By decomposing the alternatives into different attributes, the value of how respondents perceive the value can be measured (Louviere et al., 2010). This makes choice modeling different from most survey data which include depended and explanatory variables. SC modeling technique allows examining the impact of product configurations including pricing and promotions on different attributes (Kroes & Sheldon, 1988). By executing a choice experiment, stated intentions involving choices across multiple dimensions can be captured. This will provide a detailed insight into how choices are associated with varying aspects of privacy and how these choices are related to privacy concerns of individuals.

There is an important distinction between revealed choice preference and stated choice preference methods (Figure 7). Revealed preference methods are based on data retrieved from real market conditions while stated preference models are based on respondent's observations from the experimental environment. Kemperman (2000) concluded that revealed preference methods are the most appropriate tool for deriving utilities and estimating demands. However, it has several limitations. First, it can be difficult to obtain sufficient variation in the revealed preference data. Second, when strong correlations are expected, it does not provide proper trade-off ratios. Thirdly it cannot evaluate concepts unexisting yet (Kemperman, 2000). Stated preference methods on the other hand, offer different characteristics. In this approach, choice situations are used to construct hypothetical products or services. This increases the control over the existing alternatives and the attributes that are tested. Because the designs are hypothetical, the trade-offs between attributes can be measured without bias. With stated preference methods, there is also a low correlation between the attributes. Negatively, hypothetical concepts may not apply to the 'real world' conditions.

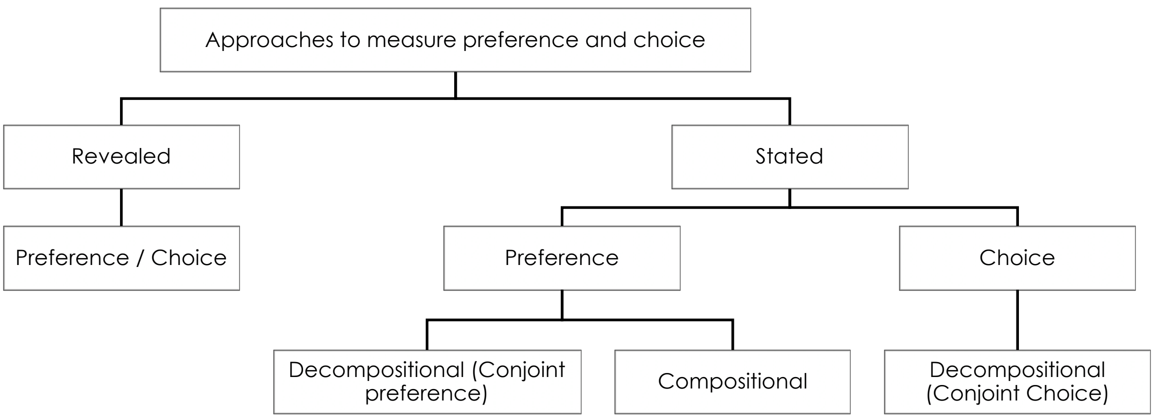


Figure 7 - An overview of preference and choice measurement approaches Kemperman (2000)

3.1.2 Experimental design process

The foundation for any discrete choice experiment is an experimental design. The experimental design assures that all stages from design until the execution of the experiment are well thought out. For discrete choice experiments, the results are very dependent on the choices that have been made before execution. To guarantee that all stages of the experimental design process are taken, the experimental design framework of Hensher (2005) is used (Figure 8). Additionally, this research will also adopt their terminology. Thus, this choice experiment consists of multiple 'choice settings' per observation. A choice setting consists of three alternatives the respondent needs to choose between. One of these alternatives will be the 'no preference' option. The two other alternatives contain a finite number of attributes. An attribute has several levels that may vary per alternative.

The first stage of the experimental design process is already executed in previous chapters. The research problem is defined in the first chapter and potential attributes and levels are investigated in the literature research. The stages 2-8 will be discussed in this chapter. Starting with the attribute identification in the next paragraph. At the end of this process, the experimental design is which means that the experiment can be executed (Hensher, Rose, Greene, et al., 2015).

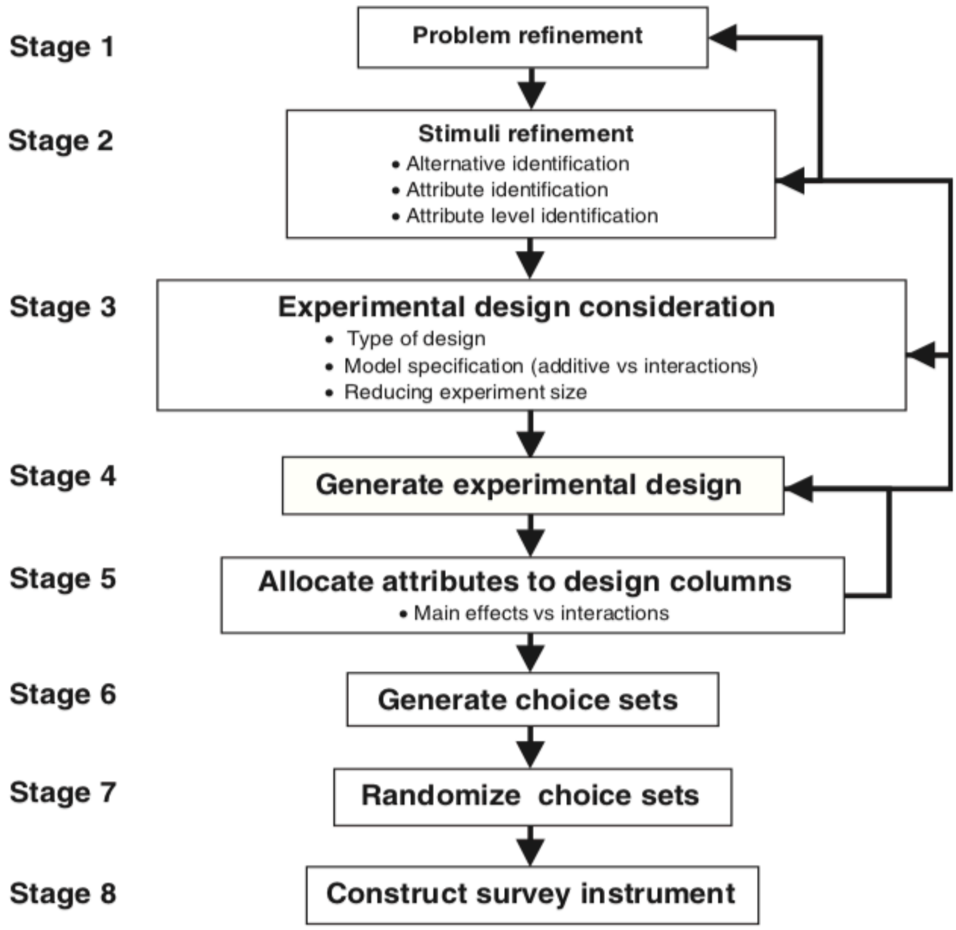


Figure 8 - Experimental design process Hensher et al., (2015)

3.2 Attribute Identification

The alternatives of a stated choice experiment are decomposed into attributes and levels that are discussed in this paragraph. Additionally, this research will contain several socio-demographic questions and statements that are discussed in this paragraph.

3.2.1 Input for the Choice Experiment (Stage 2)

The most frequently discussed method to measure the behavioral intentions of an individual is the risk, benefit analysis. In the stated choice experiment, the benefits and risks will not be specifically mentioned. It is up to the respondents to decide the risks and benefits based on the alternatives in the choice setting. For the stated choice experiment it is important that the set of attributes and levels are mutually exclusive and exist of a finite number of alternatives while also being a realistic representation of the situation. A higher number of levels and attributes indicate a higher accuracy of the results. However, it will also result in a larger experiment and consequently a larger sample size. All attributes consist of 2 or 4 attributes since there is no middle level to choose from. For every attribute, there is one level assigned as the 'reference level'. This level is best relatable to the current situation and will be used in the introduction of the survey to inform the respondent about the current situation. In this paragraph, all attributes and levels are discussed individually and summarized in table 5.

Attribute 'Type' – The type of data that is processed

The types of data that are processed are related to the types of smart home appliances. Thus, an abundant number of data types can be named. Sometimes, the data should be filled in by the consumer. In other situations, the data is detected by the product itself. For this attribute, it is chosen not to enlist raw data levels (Time, temperature, energy consumption, etcetera) but instead what can be done with the type of data. There are two reasons that legitimize this decision. First, the number of raw data levels is limitless which makes it impossible to choose 2 or 4 representative levels. Secondly, the respondent might not understand the consequences of what happens when these types of data are processed. As a result, it is chosen for a 4-level attribute in which every new level contains additional data. In the first level, 'just' data regarding the user's energy consumption, in the higher levels it also incorporates detailed data which is also of a larger quantity. In the third level, also data that is processed by smart home appliances is added. Smart home appliances mostly collect personal data such as name, address, age, household composition, etcetera. The last level adds GPS data which means that location of the user is shared with the smart home appliance. Thus, the first two levels are focused on energy data while the last two are focused on the data that is processed by the smart home appliance.

Attribute 'Why' – The reason why the data is being processed

There is a distinction between data that is beneficial for the users and data beneficial for the data controller and processor. These latter actors are processing data for billing the consumer or optimizing their products and services. For this research, the levels will be focused on the users of smart home appliances. Since this research is focused on energy consumption, the levels are focused on informing, managing and controlling and automating energy consumption. With every next level, the respondent benefits by letting the technology take additional responsibilities. Since these levels might be hard to understand for the respondents, an example will be provided.

Attribute 'Act' – The responsible actor for data processing

Different actors are capable of processing the data from smart home appliances. Actors close to the energy network use data to optimize their energy distribution and for billing purposes. For technology companies, the data can be used for optimizing their products and services. Their ultimate goal of data processing is to increase their profit on these services. Based on the literature research it is found out that users are less trustworthy to (inter)national technology companies than to energy providers. But if this will change their decisions regarding data privacy is unknown. his attribute has two levels which means that the attribute will be tested twice as much.

Attribute 'Share' – The frequency of data sharing

Most smart home appliances need a local WIFI connection for optimal use of the product. This connection might be used to transfer data to a local server or to share the data with numerous actors and smart home appliances. The higher the frequency of data processing, the higher the change of sharing private data. Actors in the energy industry are allowed to share data in a frequency of one's per month for billing purposes. For advanced appliances, processing every hour, minute or second is regularly used. The higher the data volume, the better smart home appliances are capable to adapt to the consumers' demands. Eventually, this will result in higher product approval and energy efficiency.

Attribute 'Remove' – The frequency of data removal

When processing has taken place, the data controller typically stores consumer data in their own databases (Van Aubel & Poll, 2019). With the new regulations of the GDPR, the user has the right to revoke data storage at all times. When data is unavailable, smart home appliances are unable to perform optimally. Contrarily, smart home appliances have an increasing vulnerability to data breaches since they contain large quantities of consumer information. Since smart home appliances are mostly connected to the internet, hackers will find new approaches to steal personal data. If individuals have concerns regarding data privacy, data removal might have an influence on their choice behavior. The retention times are set after 10 days, 1 month, 1 year and after the product is out of use.

Attribute 'Trade' –The trade-off of data processing

The previous attributes are discussing the capabilities of the smart home appliances. With this attribute, it is tested if the respondent is willing to trade-off the data collected from the smart home appliances for an environmental benefit and financial compensation. The literature shows that there is no clear evidence of what percentage can be achieved by implementing smart home appliances. The potential compensation is dependent on the size of the dwelling, household composition, income, energy prices, etcetera. Thus, the financial compensation that is expressed in this attribute are based on potential incentives, not on market prizes and energy reductions. Nevertheless, it is crucial that the hypothetical compensations are realistic and reasonable from a cognitive perspective (Broberg & Persson, 2016). Thus, the reference level of this attribute is solely focused on an environmental benefit without financial compensation. The remaining three levels are both representing an environmental benefit and financial compensation. It is chosen for three values on a ratio scale, namely €5,- €10,- and €15,-. These values are well rounded and related to the values that are mentioned in the literature research. Additionally, the prices are comparable to the prices for digital services as Spotify, Netflix, PlayStation Network or a sim-only phone contract.

3.2.2 Input for survey questions

The survey does not only consist of a stated choice experiment. Prior to the choice experiment, the respondents will be asked about their socio-demographic characteristics. Also, their perception on privacy concerns and energy conservation are surveyed with 10 statements.

Socio-demographic variables

As found in the literature, there are several socio-demographic characteristics that have significantly affect the individuals' attitude towards data privacy and smart home appliances. The most relevant characteristics are age, gender, income, education, household composition and occupation (i.e. student or working). With the length of the survey in mind, there are no additional variables tested. To create the correct level of measurements, the Dutch census called "Centraal Bureau voor de Statistieken" (CBS) is consulted. As the national statistical office, CBS provides reliable statistical information and data freely available. With their data, the survey results can be compared to the Dutch average and concluded if there is are any under- or overrepresented categories in the survey results.

Perception of privacy concerns

According to the privacy theory, individuals are willing to sacrifice their privacy if there is a positive trade-off available. Thus, it is important to have an insight into the development of these privacy concerns. To have an insight into the respondents' perceived risk, the often-cited privacy calculus model of Dinev & Hart (2006) is consulted. In their calculus model, the perceived risk of an individual is based on trust and concerns (Dinev & Hart, 2006). Since this model is focused on internet use, it cannot be copied for this research. However, the questions that indicate the perception of privacy concerns can be modified to fit the purpose of this research. This approach is similar to the modification proposed by Lee et al. (2019). Thus, the perception of privacy concerns will be measured with 5 survey questions with a 5-point Likert scale ongoing from 'Strongly Disagree' to 'Strongly Agree'. A 5-point scale is chosen since it is the most widely used size. The survey questions are presented in table 6.

Opinion of energy conservation

Similar to the privacy concern, it is important to develop an understanding of how the respondents think about sustainability and the attitude they have regarding their energy conservation. Based on these questions it can be concluded if there is a significant difference between the respondent's attitude and behavior after the research is conducted. The research of Ohler (2014) is used to describe the factors affecting energy-saving behaviors. There are three questions that focus on the respondent's opinion about energy consumption. Also, two questions discuss the respondents' general opinion on sustainability. The levels of measurement of the privacy concerns are also based on a 5-point Likert scale ongoing from 'Strongly Disagree' to 'Strongly Agree'. The survey questions are presented in table 7.

Table 5 - Experimental design attribute and level identification

Attribute	Attribute Levels
(A) What type of data is processed?	1. Sensor data of your total energy consumption *[Reference Level]* 2. Sensor data of your total energy consumption + Specified for every individual electronical product 3. Sensor data of your total energy consumption + Specified for every individual electronical product + Personal data collected from smart home appliances 4. Sensor data of your total energy consumption + Specified for all electronical products in your house + Personal data collected from smart home appliances + Real-time GPS location data of household members
(B) Why are such data processed?	1. To inform you about the energy usage of products in your house *[Reference Level]* (Example: Inform about the energy usage in your living room) 2. To remotely manage the products in your house. (Example: Manage the room temperature from distance) 3. To control daily routines in your house (Example: Control the dishwasher so that it turns itself on when energy usage is low) 4.To automate smart home appliances that detect and act (Example: Automate so that lights are turned off when you leave the room)
(C) Who has access to your data?	1. Energy Provider *[Reference Level]* 2. Technology company
(D) When will processing take place?	1. Every Month *[Reference Level]* 2. Every Day 3. Every Hour 4. Every Minute
(E) When will data be removed?	1. After 10 days *[Reference Level]* 2. After 1 Month 3. After 1 year 4. After the product is out of use
(F) Trade-off	1. Environmental benefit and no financial benefit *[Reference Level]* 2. Environmental benefit and financial benefit of € 5 per month 3. Environmental benefit and financial benefit of € 10 per month 4. Environmental benefit and financial benefit of € 15 per month

Table 6 – Socio-demographic variables

	Socio-Demographic Characteristics	Levels of measurement	CBS Percentages
A	Age Category	1. Age < 19 2. Age – 19 – 29 3. Age – 30 – 45 4. Age – 46 - 65 5. Age > 65	21,9 12,7 18,1 28,0 19,2
B	Gender	1. Men 2. Women	49,6 50,4
C	Current Occupation	1. Student 2. Employed (Fulltime) 3. Employed (Part-time) 4. Unemployed 5. Retired	09,0 44,5 26,0 02,1 18,4
D	Household Composition	1. Single Person 2. Two Person 3. Family with children 4. Single Parent	38,3 28,8 25,5 07,4
E	Gross Income	1. < €20.000 2. €20.000 - €30.000 3. €30.000 - €40.000 4. €40.000 - €50.000 5. > €50.000	30.8 32.4 21.2 08.9 06.7
F	Highest Finished Education	1. Secondary Education (VMBO) 2. Secondary Education (HAVO, VWO) 3. MBO 4. HBO 5. University (Bachelor) 6. University (Masters)	21,6 17,6 27.7 18.0 10.8 04.3

Table 7 – Statements about perception on privacy concerns

	Statements regarding data privacy	Measurement
G	I'm concerned about third parties being able to access my personal data	Likert Scale
H	I'm concerned that parties are not keeping my personal information secure	Likert Scale
I	I'm concerned that the information I submit on the internet could be misused	Likert Scale
J	I'm concerned about parties building a profile of me to predict my consumer behavior	Likert Scale
K	I'm concerned that I have insufficient control over the data that is collected about me	Likert Scale

Table 8 – Statements about perception on energy conservation

	Statements regarding energy-saving and sustainability	Measurement
L	It is important to me to reduce my energy consumption.	Likert Scale
M	I'm interested in having a better insight in my energy consumption.	Likert Scale
N	I'm interested in smart technology that helps me reducing energy consumption.	Likert Scale
O	I'm interested in the latest technology and gadgets.	Likert Scale
P	I'm concerned about the environmental effects of Global Warming.	Likert Scale

3.3 Experimental Design

In this paragraph, the attributes and levels of the choice experiment will be transformed into choice situations. The choices regarding the experimental design have a significant influence on the outcome of the experiment. It is therefore important that these choices do not constrain the results.

3.3.1 *Generate experimental choice design (Stage 3)*

To create an experimental design, Statistical Analysis Software (SAS) is used. SAS is capable of making two types of choice experiments namely a generic and branded choice experiment. In this choice experiment, the alternatives contain bundles of attributes where all options are possible. Thus, all attributes are independent of all other attributes which is why it is called a generic choice experiment (Kuhfeld, 2010). Since the attributes are independent, there is zero correlation in the experiment. This is frequently named or orthogonality in the correlation structure. In SAS, four main macros are used to code the experiment. The %mktruns, %MktEx, %MktLab and the %ChoicEff macro. In Appendix I, the coding including the results are shown. The first macro to use is the %mktruns macro to evaluate the generated design. From the Stimuli Refinement, it is known that there are 6 alternatives, 1 attribute is composed of 2 levels, while 5 attributes are composed into 4 levels. The first part of the output shows this (Table 9)

Table 9 - Output %Mktruns Macro 1

```
                    Design Summary
                     Number of
          Levels                   Frequency
            2                          1
            4                          5
```

3.3.2 *Reducing experiment size*

In the second part of the output of the %mktruns macro several statistics are shown (table 10). If all combinations of attributes and levels will be tested, a total of 2,048 choice situations will be tested (Full factorial). This is considered too large for a research of this size. Thus, a fractional factorial design will be used. The macro already suggests several options for 100 percent efficient fractional factorial designs. These designs can be created by the use of the MktEx macro. It is suggested to use a design size of 32, 48 or 64. All three options have no limitations since the values can be divided by 4, 8 and 16. In this research, a design size of 32 is chosen.

Table 10 - Output %Mktruns Macro 2

```
                     Saturated      = 17
                   Full Factorial = 2,048
     Some Reasonable                        Cannot Be
      Design Sizes        Violations     Divided By

              32 *             0
              48 *             0
              64 *             0
              24              10              16
              40              10              16
              56              10              16
              20              15               8 16
              28              15               8 16
              36              15               8 16
              44              15               8 16
              17 S            21               2   4   8 16
   * - 100% Efficient design can be made with the MktEx macro.
  S - Saturated Design - The smallest design that can be made.
```

Since this experiment is generic, all levels appear randomly over the choice settings. To make sure that the levels are not similar in a choice setting, the experiment needs to be blocked. In SAS, this approach is called flagging. Two flags are added since every choice situation contains of two alternatives. Since the experiment size is set on 32, and 2 alternatives per choice setting, a total of 16 choice setting will be created. Table 11 shows the first 4 choice settings that have been created by the %MktLab macro. The next step is to randomize these choice settings while remaining an efficient design.

Table 11 - Output of %MktLab macro

Choice Set	Flag 1	Flag 2	Att. A	Att. B	Att. C	Att. D	Att. E	Att. F	Att. G
1	1	1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	1	2	2
3	1	1	1	1	2	1	2	3	4
4	1	1	1	2	1	2	2	4	3

3.3.3 Generate Experimental Design (Stage 4)

The %ChoicEff macro is used to create efficient choice designs and also to evaluate the goodness of the choice model design. The macro constructs a covariance matrix and displays the parameters, variances, standard errors and the relative D-efficiency. As can be seen in table 12, the results of the final design have a relative D-efficiency is 55.62 on a 0 to 100 scale. This indicates a goodness of this design compared to a hypothetical optimal design. A score of 0 suggests that one or more levels cannot be estimated. When D-efficiency is 100, the design is balanced and orthogonal. A 100 percent efficiency cannot be achieved for this research since it is no full fractional experiment and the number of levels is unequal to the number of alternatives. All scores in between 1 and 100 imply that all of the levels can be estimated but without optimal precision. The D-efficiency can be improved by multiple design adjustments such as a reduction in the number of levels an increase in the number of choices settings. If it is chosen to reduce the number of levels, the research loses detail since there is a lower number of relevant attributes that can be tested. If it is chosen to increase the number of choice setting, the total size of the research increases. Ultimately, it is chosen to remain the current size of the experiment since all attributes can be estimated. Also, a D-efficiency of roughly 56 percent can be considered as an average result (Kuhfeld, 2010). Also, the covariance matrix and did not reveals significant errors.

Table 12 - Output %ChoicEff macro

```
                 Final Results
Design                    59
Choice Sets               16
Alternatives               2
Parameters                16
Maximum Parameters        16
D-Efficiency              8.8987
Relative D-Eff            55.6172
D-Error                   0.1124
1 / Choice Sets           0.0625
```

Now that the choice settings are created, the next step is to allocate the attributes to the experimental design (Stage 5). In other words, the numeric values as shown in table 13 have to be changed into the attribute and level description as shown in table 14. The %MktLab is used for that. Since most choice experiments have their information presented vertically, (attributes vertically and alternatives horizontally), the information in these tables is transposed. To do that, the information of table 14 is used to create figure 9. This will be the presentation that will be used in the survey.

After all choice settings are generated, Stage 6 of the experimental design process has been completed. It is however important to mention that during the experimental design process, there is no special attention towards the 'no preference' option. This option provides the respondent a third choice-option in every choice setting if they cannot choose between the presented options. Since this choice-option doesn't contain attribute and levels, it is excluded in the design generation but included in the survey.

Table 13 - Choice Setting 1, efficiency, probability, flags and coding.

Set	Design	Efficiency	Index	Prob	n	f1	f2	x1	x2	x3	x4	x5	x6
1	59	8.89875	17	0.5	1825	1	1	3	3	1	3	3	3
	59	8.89875	13	0.5	1826	1	1	2	4	1	2	4	1

Table 14 - Choice Setting 1, attribute and level description

Set	(x1) What type of data is processed?	(x2) Why are such data processed?	(x3) Who will be processing the data?	(x4) When will processing take place?	(x5) When will data be removed?	Financial Benefits per month
1	Data of your total energy consumption collected by sensors ++	To control daily routines in your house	Energy companies	Every Hour	After 1 year	Environmental benefit and $10
	Data of your total energy consumption collected by sensors +	To automate smart home appliances that detect and act	Energy companies	Every Day	After the product is out of use	Environmental benefit and no financial benefit

*

Which appliance do you prefer?

Zoom in for more details

Characteristic	Smart appliance A	Smart appliance B
Type of data that is processed	Sensor data about your energy consumption + Data is specified for all electronical products + Personal data collected from smart home appliances	Sensor data about your energy consumption + Data is specified for all electronical products + Personal data collected from smart home appliances + Real time location data (GPS) of household members
Reason of data processing	To remotely manage the products in your house. (Example: Manage your room temperature from distance)	To remotely manage the products in your house. (Example: Manage your room temperature from distance)
Responsible data actors	Technology companies	Energy companies
Frequency of data sharing	Every day	Every Minute
Frequency of data removal	After 1 year	After 1 month
Trade-off	Environmental benefit and no financial benefit	Environmental benefit and financial benefit of €10 per month

	Example A	Example B	No Preference
Which Example do you prefer?	○	○	○

Figure 9 - Choice Setting 1 as used in the Survey

3.4 Survey Instrument

The last step of the experimental design is the construction of the survey instrument. The tool that is used for this is called LimeSurvey. There are 16 choice setting created. All settings contain three choice alternatives with a theoretical probability of 30 percent that an alternative will be chosen.

3.4.1 Randomize choice settings (Stage 7)

Since the experiment contains multiple parts a phenomenon called survey fatigue must be considered. Survey fatigue suggests if the survey takes too much time or effort according to the respondent, they will no longer be accurate with their answers. This will eventually reduce the quality of the experiment without noticing. It is therefore suggested let the respondents perform a maximum of 9-16 choice situations per respondent (Sanko, 2001). The experiment contains lots of reading and there are multiple questions prior to the experiment. Eventually, it is chosen to have all respondents answer 8 choice settings in total. LimeSurvey is used to randomly select 8 out of the 16 choice settings (Stage 7). This will result that there is no bias in choice order. Consequently, there might be a discrepancy in the survey results since some choice settings can appear in more surveys than others. This issue will be solved prior to closure of the survey by manually composing the last couple of surveys.

3.4.2 Privacy of the survey instrument

For the survey, it is chosen to anonymize the results. Thus, no traceable data will be stored and saved such as name, ip-address or e-mail. Since this research is about data privacy, it is considered as the only correct decision. In practice, all respondents are allocated a random id-number that is untraceable. The consequence of choosing these privacy settings is limited. The only noticeable difference for the respondents is that they cannot save the survey and continue at a later time since there are no cookies saved. This might result that when respondents accidentally close the survey, they have to start over. For data processing, there are no significant limitations at all. Since the results will be time-stamped, it is still possible to analyze noise in the data results. The respondents are informed about data privacy too. It will be explained that their data will only be used for data analysis of this research and removed once the analysis has been completed. This means that the results (statistics, analysis, conclusions) will be present in this thesis. However, the specific individual responses will be deleted from LimeSurvey as well as the Dropbox where all data is stored during the writing of this thesis.

3.4.3 Survey Introduction

On the first page of the survey, the respondents were invited to choose between a Dutch or English version of the survey. Having two languages will increase the target audience while also improving the quality since an increasing number of respondents had the opportunity to fill in the survey in their preferred language. After selecting a language, the survey introduction will be shown. The introduction is importance since all respondents need to have a certain level of knowledge about data privacy and smart home appliances prior to the survey. To inform the respondents, three descriptive illustrations are presented in the introduction It highlights the most important aspects of the research and introduces the terms of 'Smart Home Appliances' ,'Data privacy' and the 6 principles of the GDPR.

Since most respondents have never executed a stated choice experiment before, there might be a certain learning curve to understand the operation of this type of experiment. To solve this, a second introduction is present before the respondent starts with the choice experiment. In this introduction, there will be one example choice situation. This will help the respondent to understand the attributes and corresponding levels. It also explains the current situation since the example questions will present the reference levels as stated in table 5. In appendix II, the complete survey is presented.

3.4.4 Pre-testing the survey instrument

When the survey was considered ready for activation, it is tested with a small test panel. The test was subdivided into two rounds containing close family in the first round and study mates in the second round. It is chosen to ask close family first since they are generally unaware of smart home appliances and data privacy. Thus, they may face more severe issues that wouldn't be seen by fellow classmates. It was therefore expected to solve the most noticeable issues first. In the first round, there were two severe issues found. First, the survey was unsuitable for the execution on a mobile phone. An unnecessary amount of scrolling was required which resulted in uncareful reading. Consequently, the choice experiment was considered as complicated by several members of the test panel. To maintain accurate and consistent results, this issue needed to be solved. It is also expected that most respondents will perform the survey on a mobile device such as a phone or tablet. The issue is solved by making the survey primarily suitable for a mobile phone. Consequently, this resulted that the survey became less suitable (but still operable) on an average size computer screen. Secondly, the respondents experienced an issue with the terminology that is used. Several words were experienced as complicated or unknown by the members of the test panel. It was expected that the introduction would solve this issue in the first place. Still, several words were considered as complicated to understand. The issue is solved by expanding the explanation in the introduction. Also, the term 'attribute' has been changed to characteristic since it was unclear for the Dutch survey panel. In the second test run, several study friends were used to fill in the survey. They did not face additional issues, except of a spelling error. Thus, the survey was ready for execution

3.4.5 Noise reduction in survey results

Although it is expected that an average respondent is able to correctly fill in the survey, there might be inaccurate or inconsistent results that need to be solved prior to data analysis. Several noise-reduction protocols are set in place (table 15). When the survey is closed, only fully answered surveys will be downloaded from LimeSurvey and checked individually. This already establishes that there are no missing observations within the survey results. The second protocol is analyzing duplicate records which means that it will be tested if two consecutive surveys will not have the same results. If it turns out both choice experiments are identical while filled in on the same day, one survey will be removed. Thirdly, the 'research outliers will be evaluated. This protocol suggests that if there is an observable pattern in the survey results, the survey will be nullified. For example, all questions are answered A, no preference, or all questions are answered on the middle score of the 5-point Likert scale. Lastly, the survey will be nullified when the survey is finished under three minutes. This is considered as the bare minimum to read through the information and answer the survey.

Table 15 - Noise reduction protocol including identification method.

No.	Noise	Description	Identification Method
1	Missing Records	Missing observations on one or more questions in the survey	Inspect rows, columns and next 5 survey's
2	Duplicate Records	Identical observations on all questions in the survey	Inspect rows, columns, date and time
3	Research Outliers	An observable pattern on all questions in the survey	Inspect rows, columns
4	Short survey time	The survey is finished under three minutes.	Check timestamp LimeSurvey

3.4.6 Sample Size calculation

The survey needs to achieve a minimum number of respondents before it is suitable for accurate data analysis. There are multiple approaches to calculate the sample size for a choice experiment. Most calculations are using a rule of thumb formula. In the research of Orme (1998) the minimum sample size is calculated with the following formula where (n) is the number of respondents, (t) is the number of tasks, (a) is number of alternatives per task (excluding the none alternative), and (c) is the number of analysis cells (Orme, 1998).

Formula (I)	$$\frac{nta}{c} > 500$$

the number of tasks (t) for one respondent has been set on 8 and the number of alternatives (a) in a choice setting is 2. The number of analysis cells (c) is the maximum number of levels in an attribute, which is 4. This calculation suggests that there are at least 125 full results are needed. Since one respondent does 8 out of 16 tasks, the calculation is doubled resulting in a minimum of 250 respondents. Another calculation is explained in the paper of Bekker-Grob et al., (2015). They show that the minimum sample size requirements can be calculated by the statistical program R. To perform this calculation, a small sample is pre-tested for an initial belief of the parameter values. Also, the statistical power level, significance level, type of model and DCE design are demanded (de Bekker-Grob, Donkers, Jonker, & Stolk, 2015). The calculations predicted that several parameters are insignificant even with a very large sample size (>1000). The remaining parameters were significantly below the indicated 125 surveys indicating that the sample size of 250 respondents will be sufficient. The results of these calculations are provided in Appendix III.

3.4.7 Effect Coding

Before the results of the choice experiment can be analyzed, the results are coded. This means that the results are changed from a textual result into a numerical result. Coding allows for non-linear effects in the different levels of the attributes which means it is necessary for accurate data analysis. There are several methods of coding, each method is called a coding scheme. The two most knowns coding schemes are called the dummy variable coding scheme and the effects coding scheme. The difference between dummy and effect coding is that with effect coding, the last value in the newly coded parameter will be valued (-1) instead of a (0). Thus, dummy and effects coding differ only in how the last level. With effect coding, every attribute will be subdivided into multiple new parameters. The number of new parameters is equivalent to the number of levels of the attribute being coded, minus one. Thus, for a 4-leveled attribute, there will be three parameters (see table 16).

Table 16 - Effect Coding Scheme including Derived part-worth utility

No. Levels		Parameter 1	Parameter 2	Parameter 3	Derived part-worth utility
2	**Level 1**	1			$\beta_1 * 1$
	Level 2	-1			$\beta_1 * -1$
4	**Level 1**	1	0	0	$\beta_1 * 1 + \beta_2 * 0 + \beta_3 * 0$
	Level 2	0	1	0	$\beta_1 * 0 + \beta_2 * 1 + \beta_3 * 0$
	Level 3	0	0	1	$\beta_1 * 0 + \beta_2 * 0 + \beta_3 * 1$
	Level 4	-1	-1	-1	$\beta_1 * -1 + \beta_2 * -1 + \beta_3 * -1$

3.5 Data Analysis Methods

The last paragraph of this chapter explains the different theories, methodologies and formulas that are used to analyze the survey data correctly.

3.5.1 Choice theory

Choice theory is used to observe the decisions of larger groups. The data will be used to predict the behavioral intentions of the respondents for specific circumstances. Every individual makes their choices based on different standpoints, which is often an internal process sometimes called the decision rule. The decision rule is the internal (cognitive) process used by individuals to process the available information in such a way that a unique choice is made. There are multiple interpretations of the decision rule. Four frequently used decision rule theories are mentioned below:

1. Dominance – One alternative is better than the other alternative(s) when at least one of the attributes is better.
2. Satisfaction – Every attribute of an alternative provides a level of satisfaction. The levels are set internal by the choice maker.
3. Lexicographical rule – The attributes are ordered by importance by the decisionmaker. The attribute they value the most determines the decision.
4. Utility – A vector defines the attractiveness of an alternative. This attractiveness is referred to as the utility. The utility is a measure that the decisionmaker tries to maximize.

Of these four, the utility theory is most known and used. Thus, the modeling approaches that will be discussed are based on the utility theory. To be specific the Random Utility Theory (RUT) or Random Utility Maximization (RUM) which describes that individuals aim to maximize the utility first introduced by McFadden in 1974. When RUM is discussed in combination with privacy, the literature speaks of the privacy calculus theory. The utility is usually subdivided into two separate components, an observed component named V_{nsj} and an unobserved component named ε_{nsj} as presented in the following formula (McFadden, 1974):

Formula (II)

$$\mathbf{U_{nij} = V_{nij} + \varepsilon_{nij}}$$

U = Utility value
V = Observed component
ε = Stochastic unobserved error component
n = The decision maker
i = Alternative
j = Consuming or possessing the alternative

Since the unobserved component ε_{nij} is a stochastic error component, the utility is mainly calculated by the observed component. The observed utility can be defined as the sum of the parameter representing the weight of attribute (ßj) multiplied by the score of alternative (i) on attribute (j) of individual (n) as stated in the following formula: (Hensher, Rose, & Greene, 2015).

Formula (III)

$$\mathbf{V_{nij} = \sum B_j * x_{nij}}$$

B = Utility weight attribute (j)
x = Score of the alterative
n = The decision maker
i = Alternative
j = Consuming or possessing the alternative

3.5.2 Multinomial Logit (MNL) Model

The most regularly applied method to estimate the utility value of a choice situation is the multinomial logit (MNL) model. With MNL modeling, the probability is determined by the utility which is expressed by the utility of decisionmaker (n) in choice situation (s) will derive from consuming or possessing alternative (j) (Hensher, Rose, & Greene, 2015). The result of the MNL calculation is a positive or negative (partial) utility value for each of the parameters. Since the calculations are on a logit scale, the utility estimates typically range between −2 and 2. A positive value indicates a positive assessment while a negative value indicates a negative assessment. The stronger the value (positive or negative), the heavier the rating counts in the choice. Thus, the stronger the influences on the overall utility.

Formula (IV)

$$P_{ni} = \frac{exp\ V_{nsj}}{\sum_{j=1}^{J_{ns}} Exp(V_{nsj})}$$

P = Probability
Exp = 2.718x
V = Observed component
n = The decisionmaker
i = Alternative
j = Consuming or possessing alternative

3.5.3 Mixed Logit (ML) Model

The random parameters or mixed logit (ML) model differs from the MNL model in the assumption that there is a taste variation among the respondents. Thus, when respondents have comparable characteristics, they share the same preferences but attach different utility values to the attribute levels. This is called taste heterogeneity. ML models will consider taste heterogeneity by estimating the standard deviation of the attribute parameters. Also, ML takes panel effects into account which implies that the choices that individuals make can be correlated since all individuals have multiple observations in the survey. ML models will account for the correlations across the choice of an individual by estimating all sequences of choices made by one respondent (Train, 2003). In general, a higher number of repetitions in ML will result in a higher accuracy of the results and a higher explanation power (stronger utility scores). Hensher et al. (2015) even named this model the most promising state of the art discrete choice model currently available (Hensher, Rose, & Greene, 2005).

Formula (V)

$$P(Choice_{ns} = \mathrm{j} | x_{nsj}, Z_n, V_n) = \frac{exp\ V_{nsj}}{\sum_{j=1}^{J_{ns}} Exp(V_{nsj})}$$

P = Probability
Xnsj = the K attribute of alternative (j) in choice situation (c) faced by individual (n)
Zn = a set of M characteristics of individual (n) that influence the mean of the taste parameters
Vn = a vector of K random variables with zero means and known variances and zero covariances

---

3.5.4 Log-Likelihood

Before the results can be evaluated, the model performance must be checked. Two tests will be performed to prove the level of performance: the McFadden Rho Squared Test and the Adjusted Rho Squared Test. Both tests are using the log-likelihood value to determine the model performance. The log-likelihood is estimated (by software packages) using the formula as shown on the next page.

Formula (VI)

$$LL(\beta) = \sum_{n=1}^{N} \sum_{i} y_{ni} \ln (P_{ni})$$

LL(β) = The log-likelihood of the proposed model with the estimated parameters of β
N = Total Sample Size used in the model
Y_{ni} = The choice that individual n made of alternative (i) (value 1 if chosen, 0 if not chosen)
P_{ni} = The probability that individual (n) choose alternative (i)
Ln = Natural logarithm

3.5.5 McFadden Rho Square Test

Total fit of the model can be determined by McFadden's rho^2 formula. To calculate the goodness of fit, the log-likelihood of the estimated model function needs to be calculated and divided by the log-likelihood of the null model. Finally, the result subtracted from 1 for the R^2. According to Hensher et al (2005) values for R^2 values between the 0.2 and 0.4 represent sufficient goodness of fit. Values higher than 0.5 are considered as unrealistic for behavioral experiments.

Formula (VII)

$$\mathbf{R^2 = 1 - \frac{LL_{Estimated\ Model}}{LL_{Null\ Model}}}$$

The log-likelihood calculations of the estimated model(s) contain many calculations since the estimated parameters (β) are be estimated individually for every choice that is made. This is why usually software power is used. The estimation of the log-likelihood of the null model is relatively easy since all parameters of β are equal to 0, which indicates that Y_{ni} is irrelevant to the equation. Also, since this is an unlabeled choice experiment (Choice A,B, No Preference), There exists no behavioral reason why one alternative would differ from the second and third unlabeled alternative (Hensher, Rose, Greene, et al., 2015). This results that that the probability (P_{ni}) is one-third since there are three possible options a respondent can choose from. Thus, the following formula applies: LL(0) = Total Sample Size * ln (1/3) which is written out below:

Formula (VIII)

$$\mathbf{LL(0) = \sum_{n=1}^{N} \sum_{i} \ln \frac{1}{3}}$$

LL(0) = The log-likelihood of the null model with the estimated parameters of (β = 0)
N = Total Sample Size used in the model
Y_{ni} = The choice that individual (n) made of alternative (i) ($Y_{ni} = 0$)
P_{ni} = The probability that individual (n) choose alternative (i) ($P_{ni} = 1/3$)
Ln = Natural logarithm

3.5.6 Adjusted Rho Square Test

The result from the Rho^2 (Formula VII) is generally considered as ambiguous since there are multiple predictors in the models. Hence, the adjusted Rho^2 (Formula IX) is suggested. It provides an unbiased estimation of the explained variance in the model. The adjusted Rho Square takes the degrees of freedom (n-1) and the number of respondents (k) into account. Thus, it can be discovered if a model scores a higher Rho^2 because it is the superior model or because it has a higher number of predictors (independent variables). The adjusted Rho^2 is therefore important when sub-models are being used. The adjusted Rho^2 value is lower than the 'regular' Rho^2 value, unless only one parameter is used.

Formula (IX)

$$\mathbf{R^2_{Adj} = 1 - \frac{(1 - R^2) * (n - 1)}{n - k - 1}}$$

4 DATA ANALYSIS

In the fourth chapter, the data collected from the survey will be analyzed. Firstly, the descriptive statistics of the survey are discussed. Thereafter, the behavioral intentions are analyzed using the multinomial and mixed logit models. Lastly, behavioral intentions are tested by adding interaction terms in the mixed logit model.

4.1 Exploratory Analysis

4.1.1 Survey administration

This research has no specific target group, which means that it is accessible by everyone that has the correct web address. Because there is no target group, it is tried to achieve an equal distribution over all socio-demographic variables. A total of three approaches to reach respondents have been used. First, personal contacts were reached to fill in the survey. This is executed via social media such as Facebook and WhatsApp. Secondly, work and school-related contacts have been reached via LinkedIn and personal messages. After both methods had a sufficient number of respondents, it was found out that there was an overrepresentation of males working fulltime. To solve this issue, there is more actively searched for individuals that were underrepresented in the results and approached individually. Because the choice sets of the experiment were randomly selected, some choice sets appeared more often than others. To equal this, the last 15 surveys were printed out. After surveying the last 15 respondents, the last 15 surveys were added manually.

A total of 354 respondents have started the survey, of these respondents a total of 257 have answered all questions. It took an average of 13 minutes (excluding the 15 printed surveys) to fill in the survey. This means that the survey had the correct length (within the 10-15 minutes timeframe). The pages of the survey that took the longest where the introduction page and the choice experiment which was as expected since it contains the most careful reading. On the last page of the survey, it is asked to send a personal email if the respondents requested more information after the research is finished. A total of 8 respondents have sent an email and requested more information after the research. Additionally, some respondents approached to tell that the survey was interesting and 'very challenging' to make the trade-offs in the choice experiment. A total of 214 individuals finished in the survey in Dutch, which is 83,6 percent. Since the survey contains respondents with different backgrounds and therefore a different understanding of the English language, choosing to make this survey available in the Dutch languages has turned out to be the right decision.

Noise Reduction

With noise reduction, the total quality of the survey is increased. Since LimeSurvey was able to distinguish all partially filled surveys, the first type of noise reduction was relatively easy to execute. Next, all cases were manually checked for outliers. Manually, one full survey is removed since it was finished in just 48 seconds, resulting in a total of 256 full surveys. Some of these surveys had some inconsistencies in the results that needed to be adjusted. Approximately 5 respondents filled in to be a housewife which they did not consider as unemployed. Also, 8 respondents filled in to live in a student apartment, which falls under the one-person household category. After the noise reduction, a total of 256 full surveys were considered as representative. All choice sets were filled in 128 times which was above the threshold that was set. The survey success rate was 72 percent.

4.1.2 Frequency Distributions

The frequency distributions are analyzed using descriptive analysis approach in SPSS. All frequency distributions are shown on the next two pages in table 17 and figures 10-12. Table 17 contains three types of percentages. The predicted percentage is based on the Dutch census data. The observed percentages represent the survey results, the adjusted percentages show percentages without the 'other' option included.

A total of 130 are male and 126 are female respondents who have completed the survey (Figure 10). This means that the gender ratio is approximately 50:50. In no survey, the gender 'other' has not been chosen and is therefore removed for further analysis. The age categories in the survey are not as equally distributed as the Dutch averages. In general, children (<19) and especially elderly (>65) are underrepresented in the survey. Thus, further analysis on these two categories is impossible unless merged with other categories. The survey has relatively many respondents in the age categories 19-29 and 46-65.

As expected, most respondents are having a full-time job or are still studying (Figure 11). In the observed distributions, there are more full-time workers and less part-time workers than the Dutch average. However, around 72 percent has a job (full-time or part-time) as their main occupation. This is comparable to the percentage of the total working population in the Netherlands according to the statistics of the CBS. A total of 22 percent of the respondents is currently studying which is above the Dutch average. This household composition category represents a lot of families with children (53%). This is understandable since this may contain students as well as fulltime working parents that have children. Of all respondents that have filled in 'Family with children', almost 10 percent are <19 years while 70 percent falls in the age categories (30-45 and 46-64). Also, 20 out of the 58 students are living with their parents (34%). Of the students not living with their parents, 40 percent live in student housing (single-person household). One respondent has filled in 'other' without a specific explanation (Centraal Bureau voor de Statistieken, 2019).

Looking at the descriptive of 'income', a total of 29 respondents were unwilling to share their income which means that almost 89 percent of the respondents were willing to share this type of personal information. The most often selected category was >50.000. This is surprising at first, but understandable since the average education level of the respondents is also above average. Besides, it might be possible that some respondents have misunderstood the question and have filled in their household income instead of their personal income. This cannot be tested. Many respondents have an income < € 20.000. In this category, most respondents were students or unemployed. Since there are relatively many students in the survey, it is surprising that the observed percentage is still below the Dutch average. This suggest that there are not many non-students that fall into this category. The average Dutch income is around €35.000, which is represented in the category (€30 – €40.000). This category is relatively infrequently chosen (19%). The number of individuals that have an above-average income (income categories >3) are almost equal to the number of respondents that have an under average income (categories <3). This indicates that the income distribution is fairly equal considering that category 1 and 5 are infrequently chosen.

Looking at the distribution of the highest finished education in figure 12, Most respondents in the survey have finished an MBO or HBO education. In general, the category HBO is overrepresented in this survey. One-third of all respondents have finished this type of education. The respondents are generally higher educated than the Dutch average since the lowest two categories are underrepresented while there is also an overrepresentation of respondents that have finished their education at the university (Bachelor and Masters). The latter is however not very significantly. In general, the representative women in this survey have a higher degree than men. A total of 42 women have finished a university degree at the university while 27 men have finished a Bachelor or Master degree. On the other hand, more men have finished an HBO study (degree at applied university). Almost 70 percent of all respondents that have selected the option HBO are male.

Table 17 - Descriptive statistics socio-demographic characteristics

	Socio-Demographic Characteristics	Levels of measurement	CBS Percentage	Sample Count	Sample Percentage
A	Age	1. Age < 19	21,9	22	08,5
		2. Age – 19 – 29	12,7	78	30,0
		3. Age – 30 – 45	18,1	56	22,9
		4. Age – 46 - 65	28,0	96	37,5
		5. Age > 65	19,2	4	01,6
B	Gender	1. Men	49,6	130	50,8
		2. Women	50,4	126	49,2
		3. Other, namely	-	0	0
C	Occupancy	1. Student	09,0	58	22,6
		2. Employed (Fulltime)	44,5	142	55,5
		3. Employed (Part-time)	26,0	44	17,2
		4. Unemployed	02,1	8	03,1
		5. Retired	18,4	4	01,6
D	Household Composition	1. Single Person	38,3	39	15,3
		2. Two Person	28,8	77	30,2
		3. Family with children	25,5	134	52,5
		4. Single Parent	07,4	6	02,0
E	Income	1. < €20.000	30.8	63	24,6
		2. €20.000 - €30.000	32.4	26	10,2
		3. €30.000 - €40.000	21.2	44	17,2
		4. €40.000 - €50.000	08.9	27	10,5
		5. > €50.000	06.7	67	26,2
		6. Rather not say	-	29	11,3
F	Highest Finished Education	1. Secondary Education (VMBO)	21,6	9	3,5
		2. Secondary Education (HAVO, VWO)	17,6	30	11,7
		3. MBO	27.7	62	24,2
		4. HBO	18.0	86	33,6
		5. University (Bachelor)	10.8	37	14,5
		6. University (Masters)	04.3	32	12,5

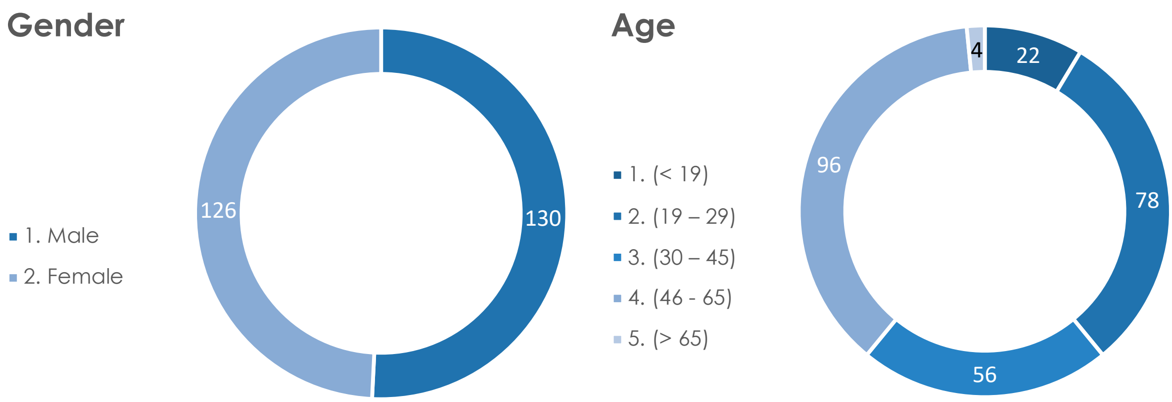


Figure 10 - Frequency distributions of Gender (Left) and Age (Right)

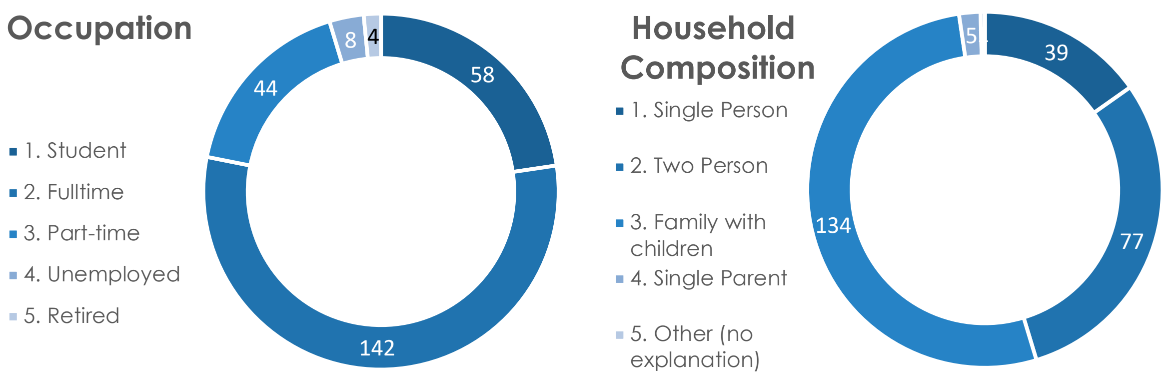


Figure 11 - Frequency distributions of Occupation (Left) and Household Composition (Right)

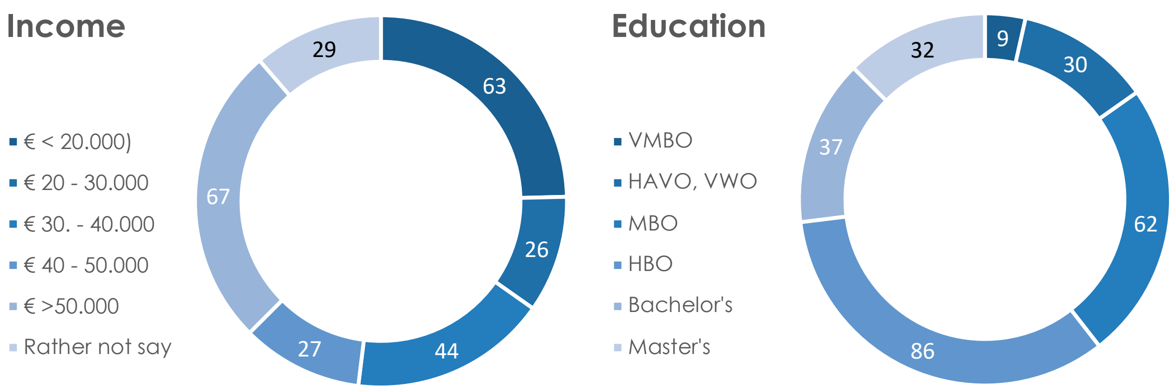


Figure 12 - Frequency distributions of Income (Left) and Education (Right)

4.2 Analyzing Concerns

In this paragraph, the relationship between the demographic variables and the statements is evaluated using the reliability analysis and ANOVA test in SPSS. It is tested which variables have a significant effect on the statements regarding privacy and energy. The statements are recoded into variables from 1 to 5 where 1 represents 'Totally Disagree' and 5 represents 'Totally Agree'. This means that a score of three is average.

4.2.1 Statements regarding data privacy

Firstly, it is tested if the results from the 5 privacy statements are reliable. The Cronbach's Alpha (α) test is used for that. The Cronbach alpha tests the consistency between self-made scales such as the Likert scale, where a higher reliability is better. Doing the reliability analysis, on the statements, the Cronbach's Alpha (α) is 0.86 which is fairly over the recommended level of reliability of 0.70. The results conclude that there is a relatively high inter-correlation among the statements. Thus, all 5 statements will be used for further analysis.

Looking at the statements discussing data privacy it stands out that all means are above average (3.00) indicating that in general, the respondents are worried about their data privacy. The total overall mean of the 5 statements is 3,69. Especially statements 5 has a high mean of 4.00 and a standard deviation of 0.89 (Figure 13). A high score on this statement suggests that the respondents are worried they have insufficient control over their data. The results also reveal that the impact of third-party' access is influential in a lower degree. The respondents are less worried about third party access of their data (mean 3.57) while respondents are not agreeing upon the impact of third parties building a consumer profile. The latter has the lowest mean (3.54) while having the highest standard deviation of 1.04.

Previous studies have argued that women tend to feel more concerned about their privacy than men (Lee et al., 2019). Using ANOVA analysis is tested if there are statistically significant differences between the means in this research. The hypothesis is partly confirmed in this research since all mean values are higher and standard deviations lower for women. The mean score of all 5 statements is 3.59 for men and 3.78 for women (Appendix IV). However, only statements 3 and 5 are significant at the 0.05 level. Women are most worries about the lack of control over their data (mean 4.10). Men are not very worrisome about the fact that companies are making a consumer profile about them (mean 3.43). There is a strong relationship between privacy and level of education since 4 out of 5 statements are significant. Statement 1 and 2 on 0.05 level, statement 3 on 0.01 level and Statement 4 on .01 level (See appendix IV for the ANOVA tables). However, the mean values are not as expected from the research of Lee et al., (2019). The mean value of the lowest level of education (VMBO) is 4,07 which is fairly above the total mean of 3,69 while the highest level of education has a mean value of 3,87. This means that this hypothesis is not confirmed.

There is found a correlation between age and perception of privacy concerns for statements 1 and 2 (significant at 0.01). The difference in means between the age categories is noticeable. Where the category <19 years has an average mean value of 3.47, the age category 50-65 has a mean value of 3.84. The age category >65 even has a has mean value of 3.95 but an unrepresentative number of respondents. Still the results indicate that older respondents have more concerns about their privacy than younger respondents. This is also according to the literature review of (Lee et al., 2019). Thus, the hypothesis can be partially confirmed. The category income is very similar to age. Both statements 1 and 2 age significant at the 0.01 level and the means for the higher incomes are notably above average. So, again the assumption that higher income groups have higher privacy concerns can be partially confirmed since not all statements are significant.

4.2.2 Statements regarding energy consumption

Performing a reliability analysis on the statements, the Cronbach's Alpha (α) of the energy statements is 0.64 which is under the recommended level of reliability of 0.7. To reach the recommended level of reliability (higher is better), SPSS makes suggestions to remove statements that constrains the overall reliability. As can be seen in table 18, the Cronbach's Alpha (α) is increased when Energy Statements 4 and 5 are removed. Removing just one statement is not sufficient to reach the threshold. After removal, the Cronbach's Alpha (α) is 0.78 which is indicated as an acceptable to good model fit. Thus, it can be concluded that there is a relatively high intercorrelation among the statements 1, 2 and 3, but a relatively low intercorrelation among statements 4 and 5. For this analysis, only statements 1,2 and 3 are used.

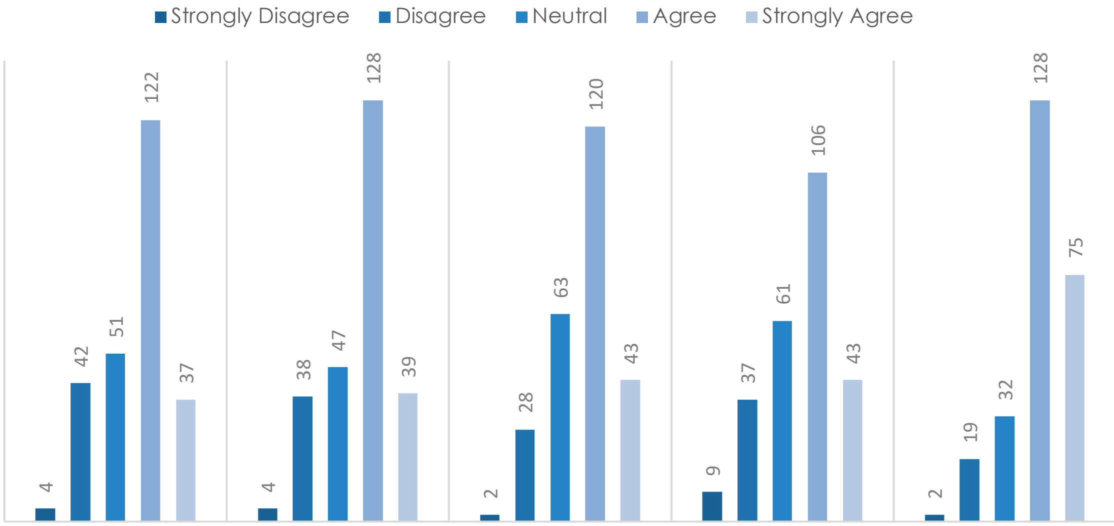
Table 18 - Reliability Analysis in SPSS of energy statements

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
EnergyStatement1	14,82	5,272	,501	,454	,529
EnergyStatement2	14,96	4,959	,551	,514	,498
EnergyStatement3	14,85	5,157	,572	,400	,497
EnergyStatement4	14,82	5,834	,237	,161	,664
EnergyStatement5	14,51	6,408	,161	,030	,687

Alike the statements referencing data privacy, all the statements concerning energy consumption have a mean above 3.00. This suggests that the respondents are generally are willing to take care of the environment by showing more interest in their energy consumption. Statement 2 has the lowest mean score of 3.53. This suggests that not all respondents are interested in having a better insight into their energy consumption. The respondents were generally more interested in technology that might assist them in reducing energy (mean 3.68). This indicates that the respondents prefer to reduce energy, but are less willing to change their behavior accordingly. The respondents rather have the technology take care of an energy reduction. Looking at the gender differences between the statements. As expected, men are more interested in the technology that assists with reducing energy consumption. (mean 3.77 vs 3.51). Women om the other hand, have slightly more interest in having an overview of their energy consumption (3.55 vs. 3.52).

For the demographic variable 'age' all three remaining statements are significant. (Appendix IV). Where the mean of individuals < 19 is 3,44, the mean of the respondents in age-category 46-65 is 3,86. This indicate that the older the respondent, the more they see the importance of energy reduction. Also, older respondents are more interested in having a consumption overview and having more interest in energy technology. For 'income' there is a significant correlation for statements 1 and 3. This indicates that the higher the income, the more interest in energy reduction and interest in energy technology. This is slightly counterintuitive since the financial impact of high-income respondents is relatively lower.

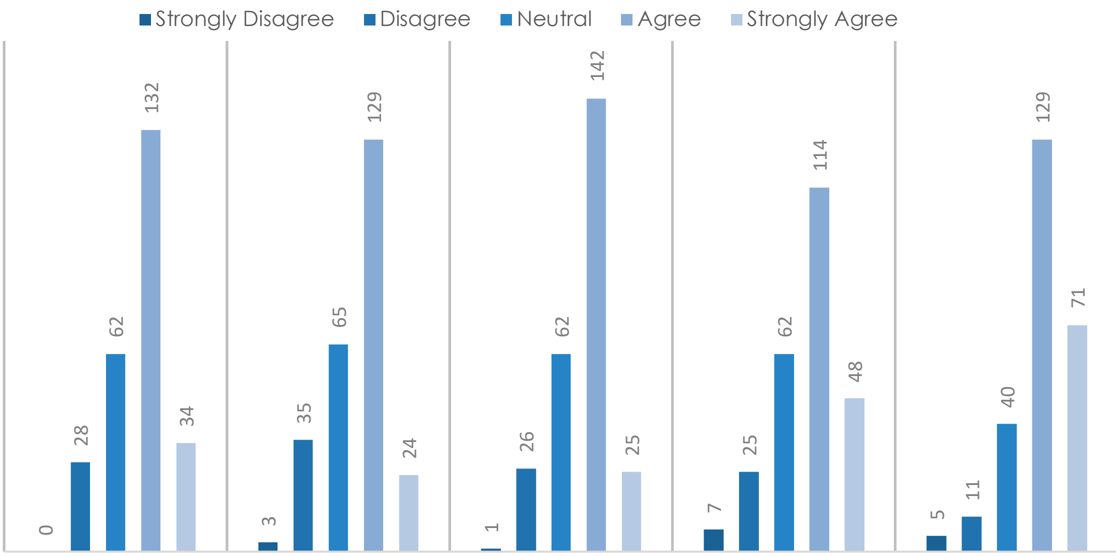
STATEMENT REGARDING DATA PRIVACY



	Statement 1	Statement 2	Statement 3	Statement 4	Statement 5
Topic	Third-party access	Data Security	Misuse of data	Consumer profile	Control over data
Mean	3.57	3.63	3.68	3.54	4.00
S.D.	0.98	0.97	0.91	1.04	0.89

Figure 13 – Descriptive statistics of Statements regarding data privacy

STATEMENT REGARDING ENERGY CONSUMPTION



	Statement 1	Statement 2	Statement 3	Statement 4	Statement 5
Topic	Important to reduce energy	Consumption overview	Interest in energy technology	Interest in the newest gadgets	Concerns about Global Warming
Mean	3.63	3.53	3.64	3.67	3.98
S.D.	0.84	0.86	0.81	0.98	0.89

Figure 13 - Descriptive statistics of Statements regarding energy consumption

4.3 Analyzing behavioral intentions

The behavioral intentions are analyzed using the Multinomial and Mixed logit analysis. R-statistics including the mlogit package is used for analysis. The package requires the data to be in a special 'wide' data format, which is executed before the analysis. Since the utility estimates are on the logit scale, the results typically range between −2 and 2.

4.3.1 Multinomial logit analysis

In table 19, the results from the multinomial logit model are presented including the utility score, standard error and significance. Table 20 presents the statistics and model performance of the multinomial logit model including the log-likelihood and McFadden Rho^2.

Table 19 - Results MNL Analysis

Parameters	Utility	Std. Error	t-value	Pr(>|t)	Significance
Constant	-1,501	0,082	-18,337	2.20-e -16	****
Type 1: Share sensor data	0,400	0,056	7,089	1.35e -12	****
Type 2 + Data specified per product	0,308	0,054	6,662	1,50e -08	****
Type 3 + Data collected by Home Appliance	-0,020	0,055	-0,373	0.708	-
Type 4 + GPS data household members	-0,688	-	-	-	-
Why 1: Inform user of energy usage	-0,055	0,059	-0,932	0.351	-
Why 2: Remotely manage energy consumption	-0,201	0,060	-3,318	0.000	****
Why 3: Remotely control daily routines	0,200	0,056	3,515	0.000	****
Why 4: Automate Smart Home Appliances	0,056	-	-	-	-
Act 1: Energy Companies	0,055	0,060	-0,091	0.927	-
Act 2: Technology Companies	-0,055	-	-	-	-
Share 1: Every Month	-0,034	0,053	-0,641	0,522	
Share 2: Every Day	0,184	0,056	3,304	0,000	****
Share 3: Every Hour	-0,055	0,058	-0,949	0,343	-
Share 4: Every Minute	-0,095	-	-	-	-
Remove 1: After 10 days	0,086	0,068	1,277	0,202	-
Remove 2: After 1 month	0,054	0,061	0,878	0,380	-
Remove 3: After 1 year	-0,084	0,059	-1,424	0,154	-
Remove 4: After de product is out of use	-0,056	-	-	-	-
Trade 1: Environmental benefit	-0,440	0,057	-7,654	1,95e -14	****
Trade 2: Environmental benefit + €5	-0,053	0,057	-0,938	0.348	-
Trade 3: Environmental benefit + €10	0,121	0,058	2,094	0.036	**
Trade 4: Environmental benefit + €15	0,372	-	-	-	-

Significance codes: (0 = '****') (0.001 = '***') (0.01 = '**') (0.05 = '*')

Table 20 – Model performance Multinomial Logit Model

Statistics Multinomial Logit Model	
Number of observations	2048
Number of parameters	17
Log-likelihood of the zero model LL(0)	-2249.95
Log-likelihood of estimated parameters LL(β)	-1776.4
McFadden's Rho^2	0,210
Adjusted Rho^2	0,204

The utility of the parameters should be interpreted relative to the base levels of each attribute which makes it tricky to analyze. For example, Type1 measures the attractiveness of sharing sensor data of your energy consumption opposed to the reference level (i.e. 'no preference' option). It is important to notice that an increase in the level 'Type' means more private data will be shared. The positive utility tells us that, the odds of a person that chose Type 1, relative to a person that chose the no preference option, increases by a factor of EXP(0.400), while holding all other parameters in the estimation (Why, Act, Share, Remove and Trade) constant. This means that Type1 has a positive influence on the total utility, which means it is more preferred than the 'no preference' option. When the utility estimates are of a larger magnitude indicate a stronger preference. Another way of interpreting the utility is: If 'Type1' is present in an alternative, the probability this alternative is being chosen, increases. Other mentioned statistics in table 19 are the std. error, t-value and significance. The std. error column indicates how precise the estimate is. The smaller the error, the more certain the estimation of the utility is. The Pr(>|t|) column shows the two-tailed t-values testing the null hypothesis that the coefficient is equal to zero (i.e. no significant effect). The usual threshold for a significant value is 0.05. A non-significant test indicates that there is no significant difference in preference for that level in comparison with the 'no preference' option. The more stars, the more significant the results.

Looking at the goodness of fit of the model (Table 20), the model scores a McFadden Rho^2 of 0.21. The Rho^2 has a value between 0 and 1. The higher the value the better the fit of the model. Statistical models with a value between 0,2 and 0,4 are considered as an excellent fit for choice experiments (McFadden, 1974). Although a higher score is always better, it should be noted that a value above 0,4 can be considered as unrealistically for behavioral researches. Since this goodness of fit is met, it can be stated that the model that is estimated provides better results than the model without parameters, also called the null-model. Figure 14 shows the utility scores of the multinomial logit model schematically. Green values indicate a significant utility, red an insignificant utility value.

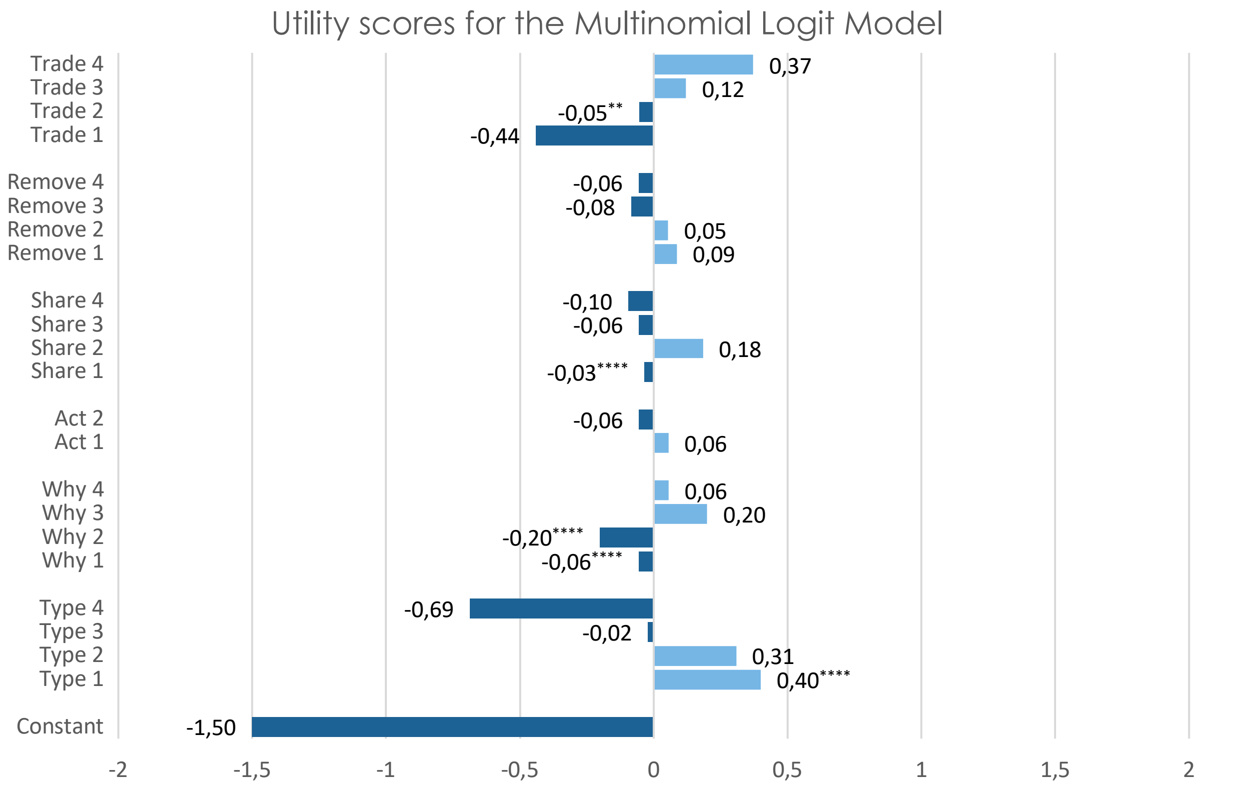


Figure 14 - Utility Scores of the Multinomial Logit Model

Utility Scores

In figure 14, the utility scores from the multinomial logit model are visualized. The utility weight (β) of the constant is negative (-1,50). The more negative the constant, the less often the 'no preference' option is chosen by the respondents. In this case, the 'no preference' option is chosen in less than 1 percent of all choice situations. The sum of the least attractive utility values is (-1,57). Thus, the respondents preferred the 'no preference' option over the least attractive choice situation. However, for most choice situations, the sum of the utility values is higher than the negative constant, which explains the high utility value.

When looking at the utility scores of the attribute Type, it can be concluded that the first two levels are having a positive utility, while the last two levels are having a negative utility. This means that there is a negative tendency in the attribute. Between level 3 and 4 there is a rather significant difference of 0.67 suggesting that the respondents are not very willing to share their GPS data with their smart home appliances. Also, respondents are willing to let the smart home appliances process data regarding energy consumption, which may also be specified per product. The moment that more personal data will be collected by smart home appliances itself, the respondents shift their behavior from a positive to a negative behavioral intention.

Analyzing at the reasons why data is processed, it strikes out that there is no specific positive or negative tendency. This is surprising since with every next level shows more advanced capabilities that may have benefits for the respondent. Nonetheless, there are some discoveries in the results. First, it can be stated that processing for energy-related benefits results in a negative influence on the total utility. Processing data to inform the user about their energy consumption has a negative utility of -0,055, processing for energy management has a negative utility of -0,201. On the other hand, the respondents are in favor of data processing for the usage of Smart Home Appliances. The MNL model also shows that there is a positive significant utility for remotely controlling the smart home appliance, indicating that this feature is more appealing by the respondents. Automatically controlling by smart home appliances has a (non-significant) positive utility too. This demonstrates that respondents prefer smart home appliances that are user-orientated. However, they also prefer to remain control over the devices.

The attribute 'actor' has no significant results. This proves that the type of actor has no forceful effect on the total utility. This is surprising since literature discusses a distrust in energy suppliers. However, the respondents do net change their behavioral intention based on the type of actor. Furthermore, the attributes 'Share' and 'Remove' also have low utility values. The only significant attribute is Share2, which indicates that the respondents have a positive attitude towards sharing data once a day.

The last utility score to discuss is the trade-off that the respondents make. Since theoretical values are used for this attribute, the values cannot be used for accurate willingness to pay estimations. Based on the results, it can be concluded that there is a positive tendency in the attribute. This means that the higher the financial benefit is, the higher the probability that this level will be chosen. In choice situations where no financial benefit is offered, the utility is negatively -0,440. With the second level, the utility is almost equal to zero (-0,05). This suggests that the turning point from a negative to a positive utility lies just above €5,- monthly benefit (between level 2 and 3). Since the respondents are aware of the price range in this attribute, the turning point lies as expected roughly in the middle.

Part worth utilities

In figure 15, the Part Worth Utilities (PWU) are shown. A higher part-worth utility means that the attribute has a stronger effect on the total utility. The PWU is calculated by using the range of all attributes dividing by the total sum of all ranges. It is chosen to take both significant and insignificant attributes into account. What can be seen is that the type of data that is processed has the biggest influence on the total utility (38%), followed by the trade-off (28%). Thus, the behavioral intention of the respondents is mostly based on these two attributes. Still, since type shows a larger impact than trade, it shows that the respondents were aware of the potential consequences of data sharing.

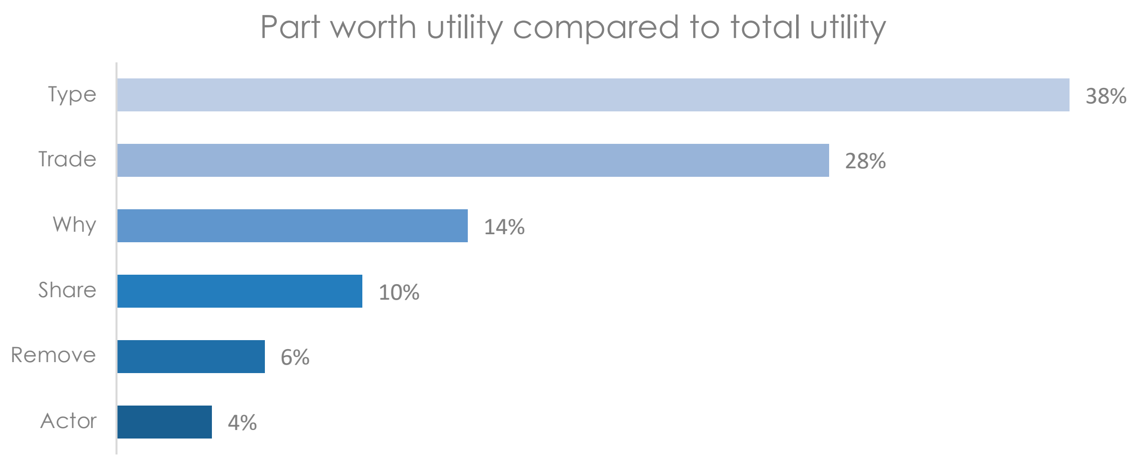


Figure 15 – Part Worth Utilities Multinomial Logit Model

The least attractive attributes according to the PWU are the type of actor, the frequency of data sharing and the retention time before data removal. A low PWU suggests that the respondents were not basing their decision on these three attributes. This is further confirmed with the lack of significance utility values. The least attractive attribute is actor, which affects the total utility for just 4 percent.

4.3.2 Mixed Logit Analysis

In table 22, the results from the mixed logit model are presented including the utility score, standard and significance and the standard deviation and the significance of the standard deviation. Table 23 presents the statistics and model performance of the multinomial logit model including the log-likelihood and McFadden Rho2.

To run the mixed logit model, additional input is required for accurate results. First, the mixed model takes panel data into account. This means that since one respondent has multiple observations, there might be an unobserved factor that persisted the respondent to choose for a particular choice. Also, it is assumed that all parameters are normally distributed across the population. Thus, it is expected that there is a correlation between the random parameters. Lastly, a set of Halton draws is used. A Halton sequence is most easily understood with the example as described in Train (2003). Consider the prime of 3. The Halton sequence for 3 is created by dividing the unit interval into three parts with breaks at 1/3 and 2/3. Then each of the three segments is divided again into thirds. Even though it is clear that Halton draws provide better results than random draws for mixed logit estimations, it is unclear what number of draws is best (Train, 2003). In the research of Bhat (2001) it is found out that 100 draws provided lower estimation errors than 1000 random draws. The research concluded that 125 draws provided the best balance between results and processing time (Bhat, 2001). For this research, different number of Halton draws are tested. As shown in Table 21, 125 draws provided the highest (least negative) results on the log-likelihood while having the highest number of significant parameters at 95% or higher. On the other hand, 50 draws resulted in the most significant parameters at the 0.000 level while also having two parameters significant at the 0.001 level. It is chosen to use 125 draws for further analysis since it provides the most significant results while having the lowest log-likelihood.

Table 21 - Number of Halton Draws Mixed Logit Analysis

	10	50	100	125	250	500	1000
Log-Likelihood (LL)	-1768.5	-1767.8	-1766.9	-1766.7	-1767.5	-1769.9	-1770.2
0.000 (****)	5	5	5	4	0	2	1
0.001 (***)	2	2	1	2	1	4	0
0.010 (**)	0	1	4	4	2	3	5
0.050 (*)	2	2	0	1	4	1	3
-	8	7	7	6	10	7	8

For the mixed logit analysis, double the number of parameters is estimated. The first set of parameters describes the average part-worth coefficient across the population. The second set of parameters shows how the part-worth coefficient vary across the population. This is reported as sd. (standard deviations). The standard deviation could be considered as a measurement to what extent respondents have similar results. A small standard deviation suggests that the respondents have similar results, while a large sd. suggests a lot of taste differences between the respondents. With the inclusion of standard deviation parameters for all attributes, only Share and Remove showed significant deviations. Thus, the choices of the respondents are mainly homogeneous. After the deletion of the insignificant standard deviation parameters, the model performance is evaluated.

Looking at table 22, it is discovered that there is a small difference in the results when compared to the multinomial logit model. This suggests that the multinomial logit model already calculated a suitable model. The McFadden's Rho^2 and adjusted Rho^2 confirmed this assumption since they are both above the threshold of 0,2. The mixed logit model scores a McFadden Rho^2 of 0.213 and an adjusted Rho^2 of 0,204. since this goodness of fit is met, it can be stated that the model that is estimated provides better results than the model without parameters (null model) but not better than the multinomial logit model. This indicates that the respondents have similar utility values. Thus, there is a small taste difference within the population.

Table 22 - Results ML Analysis

Coefficients	**Estimate**	**Sign.**	**Std. Dev**	**Sign.**
Constant	-1,561	****		
Type 1: Share sensor data	0,602	****		
Type 2 + Data specified per product	0,458	****		
Type 3 + Data collected by Home Appliance	0,092	-		
Type 4 + GPS data household members	-1,152	-	-	
Why 1: Inform user of energy usage	-0,123	-		
Why 2: Remotely manage energy consumption	-0,285	**		
Why 3: Remotely Control daily routines	0,455	***		
Why 4: Automate Smart Home Appliances	-0,047	-	-	
Act 1: Energy Companies	-0,072	-		
Act 2: Technology Companies	0,072	-	-	

Share 1: Every Month	-0,065	-	0,872	
Share 2: Every Day	0,336	***	0,450	*
Share 3: Every Hour	-0,202	*	0,825	
Share 4: Every Minute	-0,069	-	-	
Remove 1: After 10 days	0,259	**	-1,242	***
Remove 2: After 1 month	0,061	-	0,384	
Remove 3: After 1 year	-0,236	**	1,130	**
Remove 4: After de product is out of use	-0,084	-	-	
Trade 1: Environmental benefit	-0,706	****		
Trade 2: Environmental benefit + €5	-0,065	-		
Trade 3: Environmental benefit + €10	0,257	**		
Trade 4: Environmental benefit + €15	0,514	-	-	

Significance codes: (0 = '****') (0.001 = '***') (0.01 = '**') (0.05 = '*')

Table 23 - Model performance of Mixed Logit Model

Statistics Mixed Logit Model	
Number of observations	2048
Number of parameters	23
Number of Halton Draws	125
Log-likelihood of the zero model LL(0)	-2249,95
Log-likelihood of estimated parameters LL(β)	-1770,3
McFadden's Rho-square	0,213
Adjusted Rho-square	0,204

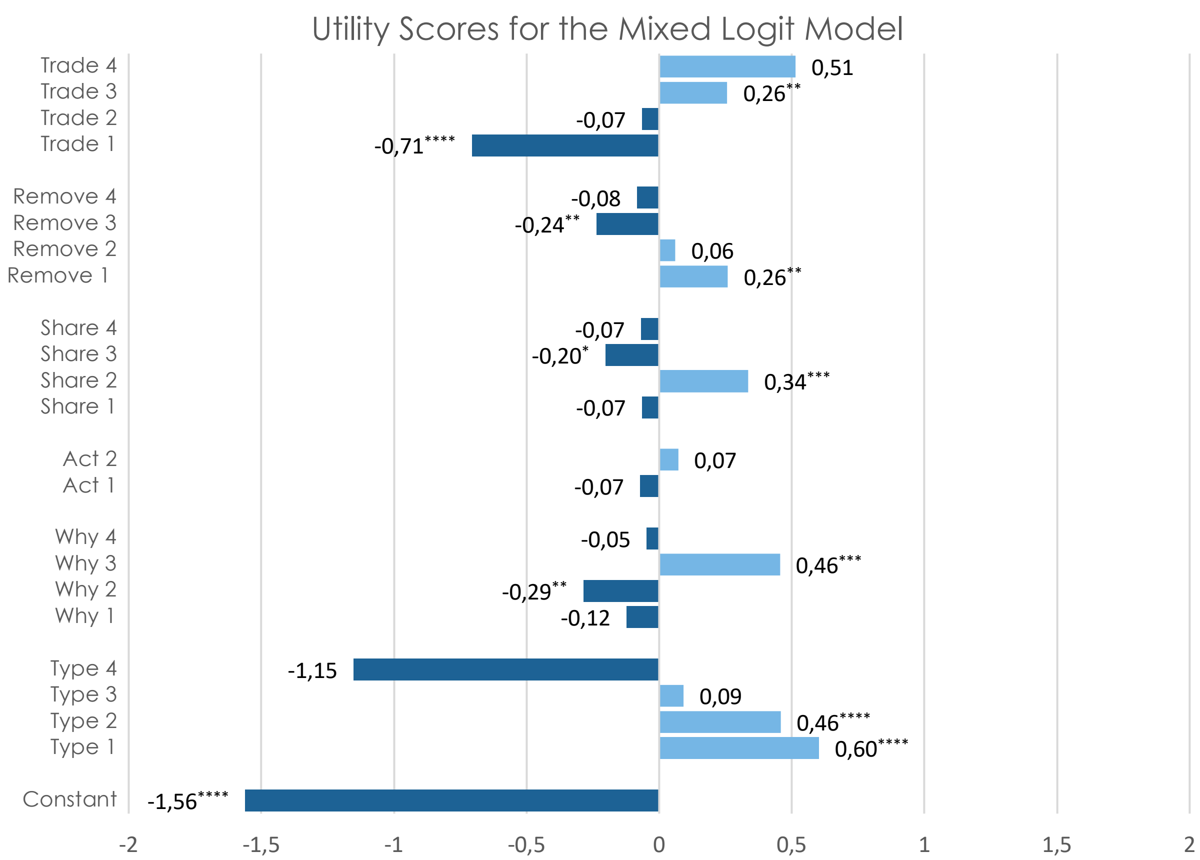


Figure 16 – Utility Values of the Mixed Logit Analysis

Utility Values

When comparing the constant utility value with the multinomial logit model, there is a slight increase from (-1,50) to (-1,56). Although this is a limited change, the interpretation has a significant different since all other utility values have been changed too. The sum of the least attractive utility values in the mixed logit model is (-2,50). When compared to the least attractive option in the multinomial logit model, there is a noticeable difference. Thus, the mixed logit model shows that when the homogeneity of the respondents is considered, more choice situations are experienced as less attractive than the 'no preference' option. Nevertheless, the constant value of -1,56 indicates that generally speaking, the respondents prefer to choose one of the two alternatives rather than the 'no preference' option. Thus, the respondents are willing to choose a smart home appliance indicating that most respondents are not against the implementation of smart appliances in their homes.

For the attribute 'Type', the first three attributes are positive, while the fourth level is severely negative (-1,08). This indicates that respondents are generally not against sharing data. However, sharing GPS data is considered a bridge too far. The difference between the two outer levels (1 and 4) in this attribute is the largest. This suggests that the impact of the type of data on the overall utility is very impactful. There are no significant standard deviations for this attribute indicating that the coefficients do not vary in the population. Thus, homogeneous.

For the attribute 'Why', the results are very similar to the multinomial logit model. The second and third level are most chosen. This indicates that respondents prefer to remotely control smart home appliances. For the level regarding remotely managing your energy usage, there is a significant standard deviation of (0,92). Compared to the parameter estimate of (-0,30) it can be concluded that there is heterogeneity in the way the respondents react to this attribute.

In line with the multinomial logit model, the attribute 'Actor' is not significant and close to 0 indicating a limited impact. The impact of data sharing and data removal is fairly similar too. However, both attributes have one significant standard deviation. For sharing data once a day, there is a standard deviation of (0,73) compared to the utility value of (0,36). Also, data removal after 10 days has a significant standard deviation of (1,08) compared to the utility value of (0,22). Thus, the behavioral intention between the respondents varies significantly for these levels. Looking at the trade-off attribute, the results do not vary much to the multinomial logit model. Neither of the attributes has significant standard deviations too.

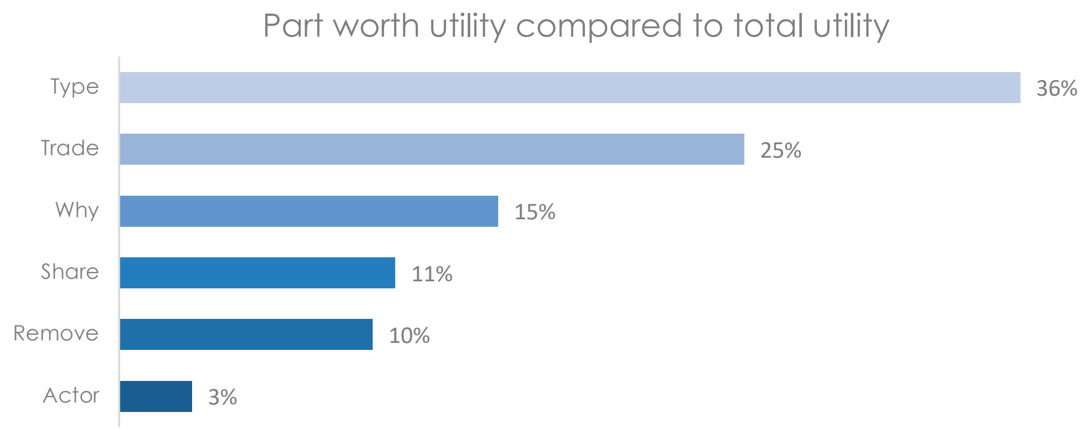


Figure 17 – Part Worth Utilities Mixed Logit Model

4.4 Analyzing behavioral intentions with interactions

With the use of ML analysis, the model performance with different interaction variables is tested. This analysis tests the choice difference between various types of respondents that have completed in the survey. According to Hensher et al., (2015) an interaction term represents the multiplication of two or more variables. Adding the interaction terms is considered the most accepted approach to include demographics into the utility functions of unlabeled models (Hensher, Rose, Greene, et al., 2015).

4.4.1 Interaction with socio-demographic variables.

With the implementation of interaction terms, it will be tested which personal characteristics have a significant influence on the respondent's behavioral intention. Based on previous analysis, it is decided to focus on the interaction between the personal characteristics and the attributes Type and Trade. This is chosen based on three reasons (1) the attributes are best representing the trade-off between data and the benefits of smart home appliances; (2) they are both homogeneous according to the mixed logit model; (3 The attributes combined are representing 61 percent of the part-worth utility. The socio-demographic variables that will be tested are gender, age, income and education since these variables proved to have a significant correlation with the perception of privacy concerns. Since not all categories are having a representative number of respondents, the categories have been merged where possible. For the variable 'age', the number of categories is reduced from 5 to 3 by merging the first two and last two categories. The income categories have been reduced from five to four by merging the third and fourth categories into a middle- income category. Lastly, for education, the categories have been reduced from 6 to 4 by merging the category 1 and 2 and category 5 and 6. To evaluate the significant interactions between the attribute levels and the demographic variables, all models are executed twice. One where the highest category is represented by the highest value, and one where the category is represented by the lowest value. In table 24, the interactions are shown including the model performances. All models are presented in Appendix V. Since there are no significant interaction with the variable education, the results are not displayed in table 24. Also, this suggests that the educational level of the respondent does not significantly affect the behavioral intentions of the respondents.

Gender

The mixed logit model shows a significant difference between gender and the type of data that is processed by the smart home appliances. Women are significantly more interested in sharing sensor data while men are significantly less interested in sharing this data. For men, the more data is shared, the more positive the utility value. For the female respondents, it is exactly the opposite. Since Type1 indicates that the least amount of data is shared, it can be concluded that women are more concerned about sharing data than men. Looking at Type4, the utility values are also showing strong values indicating that men are more interested and women are less interested in smart home appliances that process sensor data, specified per product, including personal data such as identity and GPS location. Since this fourth level is manually calculated (sum of the first three levels), it is not tested whether these interactions are significant. However, the utility values are indicating that is level has a strong (non) preference. Looking at the adjusted Rho^2, it strikes out that the additional interaction parameters are having a relatively large impact on the model performance. The model scores 0,211 which is higher than the previously used mixed logit model (0,204). This also confirms that gender has a prominent role in the behavioral intention of the respondents. Looking at the attribute Trade, it can be concluded that there is no significant taste difference between males and females when a financial benefit is considered. The adjusted Rho^2 value is equal to the regular mixed logit model (0,204) indicating that these interactions are not significantly improving the model.

Table 24 – Interaction terms between Type, Trade and the socio-demographic variables

Interaction between	Male	Female	Age < 29	Age 30 - 45	Age > 46	Income < € 20.000	Income € 30 – 50.000	Income > € 50.000
Type 1: Share data	-0,323****	0,323****	0,059	0,326***	-0,385****	0,14	0,264**	-0,269**
Type 2: Share data +	-0,06	0,06	-0,011	0,056	-0,045	0,242*	0,042	0,006
Type 3: Share data ++	0,026	-0,026	0,098	-0,093	-0,005	0,017	-0,092	-0,034
Type 4: Share data +++	0,357	-0,357	-0,146	-0,289	0,435	-0,399	-0,214	0,297
No. Halton draws	125		125			125		
No. Parameters	26		29			32		
Log-likelihood LL(0)	-2249,96		-2249,96			-2249,96		
Log-likelihood LL(β)	-1753,1		-1756,9			-1758,9		
Rho-square	0,221		0,220			0,219		
Adjusted Rho-square	0,211		0,207			0,203		
Interaction between	**Male**	**Female**	**Age < 29**	**Age 30 - 45**	**Age > 46**	**Income < € 20.000**	**Income € 30 – 50.000**	**Income > € 50.000**
Trade 1: € 0,-	-0,068	0,068	-0,123	0,018	0,105	-0,065	0,079	0,001
Trade 2: € 5,-	-0,038	0,038	0,117	-0,214*	0,097	0,084	0,009	0,123
Trade 3: € 10,-	-0,026	0,026	0,024	0,147	-0,171*	0,104	-0,017	-0,061
Trade 4: € 15,-	0,132	-0,132	-0,018	0,049	-0,031	-0,123	-0,071	-0,063
No. Halton draws	125		125			125		
No. Parameters	26		29			32		
Log-likelihood LL(0)	-2249,96		-2249,96			-2249,96		
Log-likelihood LL(β)	-1768,3		-1768,1			-1768,8		
Rho-square	0,214		0,214			0,218		
Adjusted Rho-square	0,204		0,204			0,206		

Significance codes: (0 = '****') (0.001 = '***') (0.01 = '**') (0.05 = '*')

Age

Looking at the interactions between age and the attribute 'Type', there are two significant interactions. The middle-age category (30 – 45 y/o) are showing a positive interaction with the Type1. This indicates that this age group is more likely to choose a smart home appliance that processes as little data as possible. The model also shows a relatively large affection against smart home appliances that processed the most data (Type 4). Thus, this middle-age category is the most careful about the amount of data they allow to be processed by a smart home appliance. The youngest and oldest age categories are showing different behavioral intentions. The youngest age category has no significant utility values. Thus, respondents in this age category are not showing significant similar choice behavior. The oldest age category (> 46) showed opposite behavior to the middle-age category. The first three levels are showing a negative utility value, while the 4^{th} level showed a positive utility value. This indicates that compared to the other two age categories, the oldest respondents are more likely choosing a smart home appliance that processes the most data. Thus, this age category is more willing to share (potentially) privacy-sensitive data. The adjusted Rho^2 value of the model is slightly above the regular mixed logit model indicating that the addition of the interactions improves the goodness of fit of the model.

There are two significant interactions between age and the attribute 'Trade' apparent. First, there is a significant interaction between the middle-age category (30 – 45 y/o) and Trade2 at the 0.05 level. Additionally, there is a significant interaction between age older than 45 and Trade3 at the 0.05 level. Looking at the enclosing values, it can be concluded that the age categories are not showing a specific preference for a high or low financial compensation. Thus, there is no apparent choice difference between the categories when the respondents are categorized by age. When looking at the model performances, it strikes out that the adjusted Rho^2 value is equal to the regular mixed logit model. Thus, the addition of the interaction variables in the model does not affect the goodness of fit of the model.

Income

Evaluating the interactions between income and the type of data, there are three significant interactions noticeable. The respondents with a middle-income prefer the level that processes the least data. The oldest group of respondents is showing a negative utility value significant at the 0.01 level for the same level. Taking also the lowest income category into account, it can be concluded that the higher the income, the more likely the respondents are choosing the alternatives where a lot of data is processed. This is opposite to what was expected since the high-income respondents are more concerned about their data privacy according to the survey results. However, since this mixed model has a lower adjusted Rho^2 value than the regular mixed logit model, it can be concluded that this observation does not have a strong foundation.

Between income and the attributes of 'Trade', there are no significant interactions. This is remarkable since the categories are strongly related. Looking at the non-significant values, it can be observed that the highest two income categories are showing a negative utility function against the highest financial compensations. But generally speaking, there is no significant difference in the behavioral intentions between the three groups. Examining the adjusted Rho^2 value, it is observed that the model performs better than the regular mixed logit model. Thus, there is a joint effect between income and the attribute Trade.

4.4.2 Interaction with the perception of privacy concerns and energy conservation

With the implementation of the interaction terms, it will be tested if the respondent's perception of privacy conservation and energy conservation will have a significant influence on the behavioral intention of the respondents. Since the perception of the respondents is measured using 5 statements on a Likert scale, it is tested which statements are contributing to the overall perception. Therefore, principle component analysis is used. The purpose of the principle component analysis is to reduce the number of statements into a fewer number of variables.

For this analysis, the results on the 5-point Likert scale are factorized using the principal components factor analysis in SPSS. Based on the Scree plot and corresponding Eigenvalues, it can be confirmed that all 5 privacy statements are contributing to one underlying factor (Appendix VI). The KMO test provided an adequate performance value of 0.813, which is fairly above the threshold of 0.6. For the energy conservation statements, the three relevant statements are contributing to one underlying factor. The KMO value for this factor is 0.638 and therefore above the threshold as well. Energy Statements 4 and 5 are not contributing to one underlying factor. Therefore, these two statements are removed with further analysis. Since both variables are factorized using SPSS, they can be used as interaction terms with the attributes 'Type' and 'Trade' just as done with the socio-demographic variables.

Table 25 - Interaction variables with Privacy Concerns and Energy Conservation

Interaction between	**Privacy Concerns**	**Energy Conservation**
Type 1: Share data	0,177**	0,063
Type 2: Share data +	0,029	-0,064
Type 3: Share data ++	0,041	0,04
Type 4: Share data +++	-0,247	-0,039
No. Halton draws	125	125
No. Parameters	26	26
Log-likelihood LL(0)	-2249,96	-2249,96
Log-likelihood LL(β)	-1762,8	-1769,4
Rho-square	0,217	0,214
Adjusted Rho-square	0,206	0,203
Trade 1: € 0,-	-0,152**	-0,128*
Trade 2: € 5,-	-0,115	-0,027
Trade 3: € 10,-	0,203***	0,120*
Trade 4: € 15,-	0,064	0,035
No. Halton draws	125	125
No. Parameters	26	26
Log-likelihood LL(0)	-2249,96	-2249,96
Log-likelihood LL(β)	-1765,7	-1768,2
Rho-square	0,215	0,214
Adjusted Rho-square	0,205	0,204

Significance codes: (0 = '****') (0.001 = '***') (0.01 = '**') (0.05 = '*')

Evaluating the interactions with the factorized privacy value, it can be noted that there is a significant interaction at the 0.01 level with Type1. This indicates that respondents with relatively high concerns about their privacy are more likely to choose a smart home appliance that processes the least amount of data. Thus, respondents with high concerns are narrowing their perception of these concerns by sharing fewer data. The highest level (Type 4) shows a negative utility value which confirms this conclusion.

However, this interaction term is not tested on significance since this level is the sum of the previous three levels. Looking at the interaction with the attribute 'Trade' it can be concluded that if respondents have a higher concern about data privacy, they are less likely to choose a smart home appliance without a financial benefit. Thus, respondents are demanding financial compensation for the privacy harm they experience. The respondents are not necessarily looking for the highest potential financial trade-off. The level Trade3, indicating a financial benefit of € 10,- has the strongest positive utility function. This level is also significant at the 0.001 level. Looking at the adjusted Rho2 values it can be concluded that both models are performing slightly better than the regular mixed logit model. This indicates that there is a noticeable relation between privacy concerns and the behavioral intention of the respondents.

Looking at the factorized energy conservation value, there is no significant interaction found with the attribute 'Type'. On the other hand, for the attribute "Trade', there are two significant interactions. These values a similar to the interactions as found between privacy concerns and the attribute 'Trade'. Thus, the stronger the factor value on energy conservation, the less likely they are choosing a smart home appliance without a financial benefit. Also, the stronger they value energy conservation, the more likely they are willing to trade-off their data for a financial benefit of € 10. Since the adjusted Rho2 values are similar to the regular mixed logit model, it should be concluded that the influence of energy conservation on the behavioral intention of the respondents is however marginal.

4.5 Conclusion

This paragraph summarized the conclusions that can be drawn from analyzing the data from the survey including the stated choice experiment. The paragraph also answers sub-questions 2,3 and 4

It can be concluded that three aspects affect the respondent's choice behavior the most. Firstly, the type of data that is processed by a smart home appliance. Generally, the respondents are in favor of sharing sensor data about their energy consumption, which may also be specified per product. The respondents are less in favor of sharing data that can track the user's behavior. Especially sharing GPS data to track the user's behavior has a negative influence on the total utility. Secondly, the respondents based their behavioral intention on the reason why the data is processed. What strikes out is that most respondents choose a smart home appliance that is beneficial for themselves. They prefer appliances that perform without human interaction such as smart lighting that reacts when entering the room. Less preferable are appliances that only process data to achieve an energy reduction. This suggests that individuals choose a smart home appliance because it provides the user several benefits such as comfort, rather than environmental benefits. Thirdly, the respondents changed their behavioral intentions for a financial benefit that might be attached to a smart home appliance. Although accurate willingness to pay calculations are not possible, it is suggested that on average an individual is willing to trade-off their data for just over 5 euro. The attributes act, share and remove are showing a relatively small influence on the total utility. This suggests that the individuals care more about the content that is shared rather than with whom the data is shared and for how long.

It is found that several types of socio-demographic variables have a significant effect on the perceptions of privacy concerns and energy conservation. Firstly, women are more concerned about their privacy and more interested in achieving an energy reduction. Also, older respondents, higher educated and higher-income respondents generally have a stronger perception of privacy concerns. Looking at the interactions between these variables and the behavioral indentions, not all interactions show significant results. First, there are no significant interactions found between education and the attributes. Thus, education does not affect the behavioral intention of the respondents. Gender on the other hand does show a significant difference. Woman are more likely to choose a smart home appliance that processes the least amount of data. The influence of age is visible for both the type of data and the financial compensation. The middle-age category is the most careful about the amount of data they disclose to be processed by a smart home appliance. The oldest group of respondents are more likely to share a larger quantity of data. This is opposite to what was expected based on the perceptions of this category on privacy concerns. The variable income also shows opposite results. The higher the income, the more likely the respondents are choosing the alternatives where a lot of data is processed. However, since the model performance is slightly below the regular mixed logit model, it is doubtful how accurate this finding is.

Lastly, respondents with relatively high concerns about their privacy are more likely to choose a smart home appliance that processes the least possible amount of data. Thus, the high-concerned respondents are actively trying to minimize the amount of information they disclose to the smart home appliances. Additionally, respondents with a relatively high concern about their privacy are demanding a financial compensation for the risks they experience. However, this compensation is a theoretical value which means they are not receiving this compensation when this smart home appliance will be used. Respondents with a strong perception of energy conservation are not showing a significant interaction with the amount of data that is processed by the smart home appliance. On the other hand, respondents with a high factor value for energy conservation are significantly more interested in a financial compensation.

5 CONCLUSION

The fifth chapter provides the overall conclusions of this research. The main research questions will be answered based on the conclusions of the sub-questions. Additionally, the scientific relevance, the project evaluation and the recommendations for further research are discussed in this chapter.

5.1 Conclusion

Smart home appliances are everyday household devices that are implemented with smart technology. Examples of smart home appliances are a smart washing machine, dishwasher, thermostats, refrigerator and many others (Paetz et al., 2011). These appliances can be remotely monitored, accessed or controlled, and provide services that respond to the needs of the user. An essential benefit of smart home appliances is the potential for energy reduction. The literature shows that energy reduction is barely achieved by the smart appliances itself. The largest reduction is achieved with (in)direct feedback options to the user. Data is used to predict behavior and consult the user in making energy-efficient decisions. Smart home appliances are processing a large quantity of data. Literature listed the processing of raw data types such as temperature, time, energy usage etcetera. But also the processing of more personal types of data such as name, age, e-mail address, GPS data, etcetera. The more data that is processed, the larger the potential privacy consequences are. When this data falls into the 'wrong hands', that one piece of 'innocent data' combined with a second piece of 'innocent data' becomes a piece of 'non-innocent' data (Balta-Ozkan et al., 2013). However, the usability of data is essential for smart home appliances. When individuals decide to invest in smart home appliances they should consider the benefits, but also the potential risks of having personal data stored and processed by a product that is connected to the internet. It is therefore that the usage of these products is considered a trade-off between sharing privacy-sensitive data and the benefits of smart home appliances. In his research, the following main research question has been researched.

> To what extent are individuals willing to trade-off privacy-sensitive data to obtain the benefits of smart home appliances including energy efficiency?

Based on the choice experiment it is found out that the trade-off is mainly determined by three attributes, the type of data that is processed, the reason why this data is processed and the financial benefit that can be obtained by the smart home appliance. This means that individuals are showing less interest in the actor that is processing the data, the frequency of data processing and the retention time (when is the data removed). The choice experiment showed that the type of data was considered more important than the financial benefit, indicating that individuals are showing a real interest in their privacy. The capabilities that individuals prefer in a smart home appliance are remotely controlling the appliance or appliances that work automatically. Additionally, Individuals rather choose a smart home appliance that operates remotely or automatically. This indicates that most individuals choose a smart home appliance because of its usability. It also suggests that individuals rather choose a smart home appliance because it is beneficial for themselves rather than a potential environmental benefit. Thus, energy reduction should be recognized as a side benefit rather than the main goal for smart home appliances.

Not every individual has the same opinion about data privacy. In this research, it is found that several personal characteristics influence the concerns individuals have about their data privacy. Firstly, women are more concerned about their privacy and more interested in an energy reduction. Also, older individuals, higher educated individuals and high-income individuals generally have more concerns about their data privacy. These results are consistent with the findings from previous researches. However, when individuals are concerned about their data privacy does not mean that they are showing more privacy-protective behavior. This contradiction is called the privacy paradox.

In this research, the privacy paradox is investigated by testing the difference between how concerned individuals are and the difference in choice behavior. It is found out that individuals with high concern about data privacy are more likely to choose a smart home appliance that processes the least amount of data. Thus, respondents with high concerns are narrowing their perception of these concerns by sharing fewer data. Also, if respondents have a higher concern about data privacy, they are less likely to choose a smart home appliance without a financial benefit. Thus, respondents are demanding financial compensation for the privacy harm they experience. The privacy paradox is apparent when looking at the interaction with socio-demographic variables and choice behavior. First, the type of education shows significant interactions with privacy concerns. However, it does not significantly influence the behavioral intentions of individuals. Also, the influence of age and income on the choice behavior is contradictory to the concerns of the respondents. Both older individuals and individuals with a high income are significantly less likely choosing a smart home appliance that processes the least amount of data. Thus, they are not as worried about sharing more data than younger and lower-income individuals. On the other hand, Women are more likely to choose a smart home appliance that processes the least amount of data while they are also more concerned about their data privacy. Thus, women are showing more privacy-protective behavior.

5.2 Scientific relevance and Recommendation

This research aimed to provide more insight into the choice behavior of individuals in the context of a smart home. This topic arose from several trends such as the increasing attention towards data privacy and the rising number of smart home appliances that are available on the market. It is found that there is a clear difference between what individuals claim about data privacy and how they behave. The confirmation of the privacy paradox should be recognized as the main scientific relevance of this research. When future research is examining the barriers and benefits of a smart home, it should be recognized that privacy cannot be tested on a Likert scale. It might represent how individuals think about privacy, it does not represent how individuals behave. Thus, the conclusions of such researches might be misinterpreted. Future research should not only discuss the risks of data privacy, but it should also implement privacy issues as a general construct in their research. When new products, services or concepts are developed, the potential privacy risks should be considered as important as the benefits.

This research should be seen as a startup for more (privacy-specific) research. It is recommended to collect more academic evidence to make the conclusions of this research more useful in practice. In this research, theoretical smart home appliances were surveyed which was sometimes experienced as confusing by the respondents. At this moment, it is unknown if individuals are showing different behavior when they are more familiar with these products. Future research should consider the privacy trade-off for several realistic products. Also, this research was mainly focused on the trade-off between data and a financial benefit. This is a theoretical value and therefore unrepresentative when testing actual existing products. Future research should consider additional benefits such as comfort, safety, health, and reliability. Also, more insight is demanded in the privacy-sensitive locations within a smart home. Applying smart home appliances in the bedroom might be considered as more privacy-invasive than implementing a smart washing machine in the garage.

6 BIBLIOGRAPHY

Acquisti, A., John, L. K., & Loewenstein, G. (2009). What is privacy worth ? *What Is Privacy Worth?* 1–39. Phoenix, AZ.

Balta-Ozkan, N., Amerighi, O., & Boteler, B. (2014). A comparison of consumer perceptions towards smart homes in the UK, Germany and Italy: reflections for policy and future research. *Technology Analysis and Strategic Management*, *26*(10), 1176–1195. https://doi.org/10.1080/09537325.2014.975788

Balta-Ozkan, N., Davidson, R., Bicket, M., & Whitmarsh, L. (2013). Social barriers to the adoption of smart homes. *Energy Policy*, *63*, 363–374. https://doi.org/10.1016/j.enpol.2013.08.043

Barth, S., & de Jong, M. D. T. (2017). The privacy paradox – Investigating discrepancies between expressed privacy concerns and actual online behavior – A systematic literature review. *Telematics and Informatics*, *34*(7), 1038–1058. https://doi.org/10.1016/j.tele.2017.04.013

Bélanger, F., & Crossler, R. E. (2011). Privacy in the Digital Age: A Review of Information Privacy Research in Information Syst. *MIS Quarterly Vol. 35*, *35*(4), 1017–1041.

Bhat, C. R. (2001). Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B: Methodological*, *35*(7), 677–693. https://doi.org/10.1016/S0191-2615(00)00014-X

Broberg, T., & Persson, L. (2016). Is our everyday comfort for sale? Preferences for demand management on the electricity market. *Energy Economics*, *54*, 24–32. https://doi.org/10.1016/j.eneco.2015.11.005

Centraal Bureau voor de Statistieken. (2019). Statline Open Data. Retrieved from https://opendata.cbs.nl/statline/#/CBS/nl/?fromstatweb

Chan, M., Campo, E., Estève, D., & Fourniols, J. Y. (2009). Smart homes - Current features and future perspectives. *Maturitas*, *64*(2), 90–97. https://doi.org/10.1016/j.maturitas.2009.07.014

Dam, S. S. Van, Bakker, C. A., Hal, J. D. M. Van, Bakker, C. A., Home, J. D. M. V. H., Dam, S. S. Van, … Hal, J. D. M. Van. (2010). *Home energy monitors : impact over the medium- term Home energy monitors : impact over the medium-term*. *3218*. https://doi.org/10.1080/09613218.2010.494832

Darby, S. (2010). Smart metering: What potential for householder engagement? *Building Research and Information*, *38*(5), 442–457. https://doi.org/10.1080/09613218.2010.492660

de Bekker-Grob, E. W., Donkers, B., Jonker, M. F., & Stolk, E. A. (2015). Sample Size Requirements for Discrete-Choice Experiments in Healthcare: A Practical Guide. *Patient*, *8*(5), 373–384. https://doi.org/10.1007/s40271-015-0118-z

Dinev, T., & Hart, P. (2006). An extended privacy calculus model for e-commerce transactions. *Information Systems Research*, *17*(1), 61–80. https://doi.org/10.1287/isre.1060.0080

Directorate-General for Research and Innovation (European Commission). (2018). The Strategic Energy Technology (SET) Plan. In *Joint Research Centre*. https://doi.org/10.2777/48982

Energiekamer. (2019). Wat doen netbeheerders? Retrieved from https://www.energiekamer.nl/informatie/wat-doen-netbeheerders/

Energievergelijk. (2019). Grote spelers op de energiemarkt. Retrieved from https://www.energievergelijk.nl/wp-content/uploads/2018/06/Energieleveranciers_April2019.jpg

EUGDPR.org. (2018). GDPR Key Changes. Retrieved May 4, 2019, from https://eugdpr.org/the-regulation/

European Parliament. (2016). *REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL*. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679

Ginosar, A., & Ariel, Y. (2017). An analytical framework for online privacy research: What is missing? *Information and Management*, *54*(7), 948–957. https://doi.org/10.1016/j.im.2017.02.004

Grønhøj, A., & Thøgersen, J. (2011). Feedback on household electricity consumption: Learning and social influence processes. *International Journal of Consumer Studies*, *35*(2), 138–145. https://doi.org/10.1111/j.1470-6431.2010.00967.x

Hensher, D. A., Rose, J. M., & Greene, W. H. (2015). Applied Choice Analysis 2nd. *Cambridge University Press*, (May 2019), 1243. https://doi.org/10.1017/CBO9781107415324.004

Hensher, D. A., Rose, J. M., Greene, W. H., Hensher, D. A., Rose, J. M., & Greene, W. H. (2015). Experimental design and choice experiments. In *Applied Choice Analysis*. https://doi.org/10.1017/cbo9781316136232.008

Jeff Smith, H., Dinev, T., & Xu, H. (2011). Information privacy research: An interdisciplinary review. *MIS Quarterly: Management Information Systems*.

Kemperman, A. D. A. M. (2000). *Temporal aspects of theme park choice behavior : modeling variety seeking, seasonality and diversification to support theme park planning Eindhoven*. https://doi.org/10.6100/IR542240

Kokolakis, S. (2017). Privacy attitudes and privacy behavior: A review of current research on the privacy paradox phenomenon. *Computers and Security*, *64*, 122–134. https://doi.org/10.1016/j.cose.2015.07.002

Kroes, E. P., & Sheldon, R. J. (1988). Stated preference methods, an introduction. *Journal of Transport Economics and Policy*, Vol. 22, pp. 11–25. Retrieved from http://www.jstor.org/stable/20052832

Kuhfeld, W. F. (2010). Marketing research methods in SAS experimental design, choice, conjoint, and graphical techniques. *... Graphical Techniques. Cary, NC, SAS-Institute TS-722*.

Lee, H., Wong, S. F., Oh, J., & Chang, Y. (2019). Information privacy concerns and demographic characteristics: Data from a Korean media panel survey. *Government Information Quarterly*, *36*(2), 294–303. https://doi.org/10.1016/j.giq.2019.01.002

Leferink, F., Keyer, C., & Melentjev, A. (2016). Static energy meter errors caused by conducted electromagnetic interference. *IEEE Electromagnetic Compatibility Magazine*, *5*, 49–55. https://doi.org/10.1109/MEMC.2016.7866234

Li, Y. (2012). Theories in online information privacy research: A critical review and an integrated framework. *Decision Support Systems*. https://doi.org/10.1016/j.dss.2012.06.010

Louviere, J. J., Flynn, T. N., & Carson, R. T. (2010). Discrete choice experiments are not conjoint analysis. *Journal of Choice Modelling*, *3*(3), 57–72. https://doi.org/10.1016/S1755-5345(13)70014-9

Lund, H., Østergaard, P. A., Connolly, D., & Mathiesen, B. V. (2017). Smart energy and smart energy systems. *Energy*, *137*, 556–565. https://doi.org/10.1016/j.energy.2017.05.123

Luor, T., Lu, H. P., Yu, H., & Lu, Y. (2015). Exploring the critical quality attributes and models of smart homes. *Maturitas*, *82*(4), 377–386. https://doi.org/10.1016/j.maturitas.2015.07.025

Marikyan, D., Papagiannidis, S., & Alamanos, E. (2019). A systematic review of the smart home literature: A user perspective. *Technological Forecasting and Social Change*, *138*(August 2018), 139–154. https://doi.org/10.1016/j.techfore.2018.08.015

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics*.

Naus, J., Van Vliet, B. J. M., & Hendriksen, A. (2015). Households as change agents in a Dutch smart energy transition: On power, privacy and participation. *Energy Research and Social Science*, *9*, 125–136. https://doi.org/10.1016/j.erss.2015.08.025

Nibud. (2019). Energie en Water. Retrieved from https://www.nibud.nl/consumenten/energie-en-water/

Ohler, A. M., & Billger, S. M. (2014). Does environmental concern change the tragedy of the commons? Factors affecting energy saving behaviors and electricity usage. *Ecological Economics*, *107*, 1–12. https://doi.org/10.1016/j.ecolecon.2014.07.031

Orme, B. (1998). Sawtooth Software Sample Size Issues for Conjoint Analysis Studies. *Measurement*.

Paetz, A., Becker, B., Fichtner, W., Schmeck, H., & Methods, F. D. (2011). With Smart Home Technologies –. *30th USAEE/IAEE North American Conference Online Proceedings*, *1*, 1–19.

Park, E., Kim, S., Kim, Y. S., & Kwon, S. J. (2018). Smart home services as the next mainstream of the ICT industry: determinants of the adoption of smart home services. *Universal Access in the Information Society*, *17*(1), 175–190. https://doi.org/10.1007/s10209-017-0533-0

Patil, S., Patruni, B., Potoglou, D., & Robinson, N. (2016). Public preference for data privacy – A pan-European study on metro/train surveillance. *Transportation Research Part A: Policy and Practice*, *92*, 145–161. https://doi.org/10.1016/j.tra.2016.08.004

Potoglou, D., Dunkerley, F., Patil, S., & Robinson, N. (2017). Public preferences for internet surveillance, data retention and privacy enhancing services: Evidence from a pan-European study. *Computers in Human Behavior*, *75*, 811–825. https://doi.org/10.1016/j.chb.2017.06.007

Potoglou, D., Palacios, J. F., & Feijóo, C. (2015). An integrated latent variable and choice model to explore the role of privacy concern on stated behavioral intentions in e-commerce. *Journal of Choice Modelling*, *17*, 10–27. https://doi.org/10.1016/j.jocm.2015.12.002

Preibusch, S., Kübler, D., & Beresford, A. R. (2013). Price versus privacy: An experiment into the competitive advantage of collecting less personal information. In *Electronic Commerce Research* (Vol. 13). https://doi.org/10.1007/s10660-013-9130-3

Rijkdienst voor ondernemend Nederland. (2018). *Monitor energiebesparing gebouwde omgeving 2017*. (0). Retrieved from www.RVO.nl./onderwerpen/duurzaam-ondernemen/gebouwen/ energiecijfers

Rodríguez-Molina, J., Martínez-Núñez, M., Martínez, J. F., & Pérez-Aguiar, W. (2014). Business models in the smart grid: Challenges, opportunities and proposals for prosumer profitability. *Energies*, *7*(9), 6142–6171. https://doi.org/10.3390/en7096142

Sanko, N. (2001). *Guidelines for Stated Preference Experiment Design (Professional Company Project in Association with RAND Europe) School of International Management Ecole Nationale des Ponts et Chaussées (Class 2000/01)*. Retrieved from http://www.b.kobe-u.ac.jp/~sanko/pub/Sanko2001_1.pdf

Spiekermann, S., Acquisti, A., Böhme, R., & Hui, K. L. (2015). The challenges of personal data markets and privacy. *Electronic Markets*, *25*(2), 161–167. https://doi.org/10.1007/s12525-015-0191-0

Statista. (2019). Number of Internet of Things (IoT) devices connected worldwide in 2017 and 2018, by selected type (in millions).

Train, K. E. (2003). Discrete choice methods with simulation. In *Discrete Choice Methods with Simulation*. https://doi.org/10.1017/CBO9780511753930

Tsai, J. Y., Egelman, S., Cranor, L., & Acquisti, A. (2011). The effect of online privacy information on purchasing behavior: An experimental study. *Information Systems Research*, *22*(2), 254–268. https://doi.org/10.1287/isre.1090.0260

Van Aubel, P., & Poll, E. (2019). Smart metering in the Netherlands: What, how, and why. *International Journal of Electrical Power and Energy Systems*, *109*(January), 719–725. https://doi.org/10.1016/j.ijepes.2019.01.001

Van Dam, S. S., Bakker, C. A., & Van Hal, J. D. M. (2010). Home energy monitors: Impact over the medium-term. *Building Research and Information*, *38*(5), 458–469. https://doi.org/10.1080/09613218.2010.494832

Williams, M., Nurse, J. R. C., & Creese, S. (2018). *The Perfect Storm : The Privacy Paradox and the Internet-of-Things*. 1–9.

Wilson, C., Hargreaves, T., & Hauxwell-Baldwin, R. (2017). Benefits and risks of smart home technologies. *Energy Policy*, *103*(January), 72–83. https://doi.org/10.1016/j.enpol.2016.12.047

Yang, H., Lee, H., & Zo, H. (2017). User acceptance of smart home services: An extension of the theory of planned behavior. *Industrial Management and Data Systems*, *117*(1), 68–89. https://doi.org/10.1108/IMDS-01-2016-0017

Yue, T., Long, R., & Chen, H. (2013). Factors influencing energy-saving behavior of urban households in Jiangsu province. *Energy Policy*, *62*, 665–675. https://doi.org/10.1016/j.enpol.2013.07.051

7 APPENDIXES

7.1 Appendix I – Input & Results SAS

The content of this appendix is mentioned in § 3.3.1 on page 45 of this thesis.


```
title Evaluate Generic Candidate Designs;
%mktruns(4 4 2 4 4 4);

title Create Candidate Design;
%mktex(4 4 2 4 4 4, n=32, seed=123);

title Add Alternatives;
%mktlab(data=design, int=f1-f2);
proc print; run;

title Find Efficient Choice Designs;
%choiceff(data=final,    model=class(x1-x6 / sta),    nsets=16,
                         maxiter=60,                  flags=f1-f2,
                         Seed=123,
                         options=relative,            beta=zero);

title Variance-Covariance Matrix;
proc print data=bestcov;
    id __label;
    label __label = '0'x;
    format _numeric_ zer5.2;
    var x:; run;

title Choice Sets;
proc print; run;

title Choice Sets by code;
proc print;
    by set;
    id set;
    var x:; run;

title Choice Sets including Statistics;
proc print;
    by set;
    id set; run;

title Choice Sets including Attributes and Levels description;
proc print label;
   label x1 = 'What type of data is processed?'
         x2 = 'Why are such data processed?'
         x3 = 'Who will be processing the data?'
         x4 = 'When will processing take place? '
         x5 = 'When will data be removed?'
         x6 = 'Financial Benefits per month'
         x7 = 'Other achieved Benefits';

   format x1 x1f. x2 x2f. x3 x3f. x4 x4f. x5 x5f. x6 x6f. x7 x7f.;
   by set;
   id set;
   var x:; run;

proc format;
   value x1f     1='Data of your total energy consumption collected by sensors '
                 2='Data of your total energy consumption collected by sensors +'
                 3='Data of your total energy consumption collected by sensors ++'
                 4='Data of your total energy consumption collected by sensors +++';

   value x2f     1='To inform you about the energy usage of products in your house'
                 2='To remotely manage the products in your house. '
```

```
                3='To control daily routines in your house'
                4='To automate smart home appliances that detect and act';

    value x3f   1='Energy companies'
                2='Technology companies';

    value x4f   1='Every Month'
                2='Every Day'
                3='Every Hour'
                4='Every Minute';

    value x5f   1='After 10 days'
                2='After 1 month'
                3='After 1 year'
                4='After the product is out of use';

    value x6f   1='Environmental benefit and no financial benefit'
                2='Environmental benefit and $5'
                3='Environmental benefit and $10'
                4='Environmental benefit and $15';
```

Evaluate Generic Candidate Designs

```
        Design Summary
    Number of
    Levels       Frequency
         2           1
         4           5
```

Evaluate Generic Candidate Designs

```
       Saturated      = 17
       Full Factorial = 2,048
       Some Reasonable                      Cannot Be
          Design Sizes       Violations     Divided By
                    32 *              0
                    48 *              0
                    64 *              0
                    24               10     16
                    40               10     16
                    56               10     16
                    20               15      8 16
                    28               15      8 16
                    36               15      8 16
                    44               15      8 16
                    17 S             21      2  4  8 16
   * - 100% Efficient design can be made with the MktEx macro.
   S - Saturated Design - The smallest design that can be made.
       Note that the saturated design is not one of the
       recommended designs for this problem.  It is shown
       to provide some context for the recommended sizes.
```

Evaluate Generic Candidate Designs

n	Design	Reference
32	2 ** 16 4 ** 5	Fractional-Factorial
32	2 ** 13 4 ** 6	Fractional-Factorial
32	2 ** 10 4 ** 7	Fractional-Factorial
32	2 ** 9 4 ** 5 8 ** 1	Fractional-Factorial
32	2 ** 7 4 ** 8	Fractional-Factorial
32	2 ** 6 4 ** 6 8 ** 1	Fractional-Factorial
32	2 ** 4 4 ** 9	Fractional-Factorial
32	2 ** 3 4 ** 7 8 ** 1	Fractional-Factorial
48	2 ** 32 4 ** 5	Orthogonal Array
48	2 ** 29 4 ** 6	Orthogonal Array
48	2 ** 26 4 ** 7	Orthogonal Array
48	2 ** 25 3 ** 1 4 ** 5	Orthogonal Array
48	2 ** 23 4 ** 8	Orthogonal Array
48	2 ** 23 4 ** 5 6 ** 1	Orthogonal Array
48	2 ** 22 3 ** 1 4 ** 6	Orthogonal Array
48	2 ** 21 4 ** 5 12 ** 1	Orthogonal Array
48	2 ** 20 4 ** 9	Orthogonal Array
48	2 ** 20 4 ** 6 6 ** 1	Orthogonal Array
48	2 ** 19 3 ** 1 4 ** 7	Orthogonal Array
48	2 ** 18 4 ** 6 12 ** 1	Orthogonal Array
48	2 ** 17 4 ** 10	Orthogonal Array
48	2 ** 17 4 ** 7 6 ** 1	Orthogonal Array
48	2 ** 16 3 ** 1 4 ** 8	Orthogonal Array
48	2 ** 15 4 ** 7 12 ** 1	Orthogonal Array
48	2 ** 14 4 ** 11	Orthogonal Array
48	2 ** 14 4 ** 8 6 ** 1	Orthogonal Array
48	2 ** 13 3 ** 1 4 ** 9	Orthogonal Array

64	2 ** 13 4 ** 5 8 ** 5	Fractional-Factorial
64	2 ** 12 4 ** 17	Fractional-Factorial
64	2 ** 12 4 ** 12 16 ** 1	Fractional-Factorial
64	2 ** 12 4 ** 10 8 ** 3	Fractional-Factorial
64	2 ** 11 4 ** 15 8 ** 1	Fractional-Factorial
64	2 ** 11 4 ** 8 8 ** 4	Fractional-Factorial
64	2 ** 10 4 ** 13 8 ** 2	Fractional-Factorial
64	2 ** 10 4 ** 6 8 ** 5	Fractional-Factorial
64	2 ** 9 4 ** 18	Fractional-Factorial
64	2 ** 9 4 ** 13 16 ** 1	Fractional-Factorial
64	2 ** 9 4 ** 11 8 ** 3	Fractional-Factorial
64	2 ** 8 4 ** 16 8 ** 1	Fractional-Factorial
64	2 ** 8 4 ** 9 8 ** 4	Fractional-Factorial
64	2 ** 7 4 ** 14 8 ** 2	Fractional-Factorial
64	2 ** 7 4 ** 7 8 ** 5	Fractional-Factorial
64	2 ** 6 4 ** 19	Fractional-Factorial
64	2 ** 6 4 ** 14 16 ** 1	Fractional-Factorial
64	2 ** 6 4 ** 12 8 ** 3	Fractional-Factorial
64	2 ** 6 4 ** 5 8 ** 6	Fractional-Factorial
64	2 ** 5 4 ** 17 8 ** 1	Fractional-Factorial
64	2 ** 5 4 ** 10 8 ** 4	Fractional-Factorial
64	2 ** 4 4 ** 15 8 ** 2	Fractional-Factorial
64	2 ** 4 4 ** 8 8 ** 5	Fractional-Factorial
64	2 ** 3 4 ** 20	Fractional-Factorial
64	2 ** 3 4 ** 15 16 ** 1	Fractional-Factorial
64	2 ** 3 4 ** 13 8 ** 3	Fractional-Factorial
64	2 ** 3 4 ** 6 8 ** 6	Fractional-Factorial

Create Candidate Design

Algorithm Search History

```
                           Current          Best
Design    Row,Col   D-Efficiency  D-Efficiency  Notes
----------------------------------------------------------
     1      Start       100.0000      100.0000  Tab
     1        End       100.0000
```

Add Alternatives

Obs	f1	f2	x1	x2	x3	x4	x5	x6
1	1	1	1	1	1	1	1	1
2	1	1	1	1	2	2	2	4
3	1	1	1	2	1	3	4	1
4	1	1	1	2	2	4	3	4
5	1	1	1	3	1	1	3	2
6	1	1	1	3	2	2	4	3
7	1	1	1	4	1	3	2	2
8	1	1	1	4	2	4	1	3
9	1	1	2	2	1	2	2	2
10	1	1	2	2	2	1	1	3
11	1	1	2	1	1	4	3	2
12	1	1	2	1	2	3	4	3
13	1	1	2	4	1	2	4	1

14	1	1	2	4	2	1	3	4
15	1	1	2	3	1	4	1	1
16	1	1	2	3	2	3	2	4
17	1	1	3	3	1	3	3	3
18	1	1	3	3	2	4	4	2
19	1	1	3	4	1	1	2	3
20	1	1	3	4	2	2	1	2
21	1	1	3	1	1	3	1	4
22	1	1	3	1	2	4	2	1
23	1	1	3	2	1	1	4	4
24	1	1	3	2	2	2	3	1
25	1	1	4	4	1	4	4	4
26	1	1	4	4	2	3	3	1
27	1	1	4	3	1	2	1	4
28	1	1	4	3	2	1	2	1
29	1	1	4	2	1	4	2	3
30	1	1	4	2	2	3	1	2
31	1	1	4	1	1	2	3	3
32	1	1	4	1	2	1	4	2

Find Efficient Choice Designs

n	Name	Beta	Label
1	x11	0	x1 1
2	x12	0	x1 2
3	x13	0	x1 3
4	x21	0	x2 1
5	x22	0	x2 2
6	x23	0	x2 3
7	x31	0	x3 1
8	x41	0	x4 1
9	x42	0	x4 2
10	x43	0	x4 3
11	x51	0	x5 1
12	x52	0	x5 2
13	x53	0	x5 3
14	x61	0	x6 1
15	x62	0	x6 2
16	x63	0	x6 3

Find Efficient Choice Designs

```
Design   Iteration  D-Efficiency        D-Error
----------------------------------------------
     1       0          3.08442 *      0.32421
             1          7.91328 *      0.12637
             2          8.36127 *      0.11960
             3          8.40318 *      0.11900
             4          8.40318        0.11900


Design   Iteration  D-Efficiency        D-Error
----------------------------------------------
     2       0                0              .
             1          8.32136        0.12017
             2          8.44183 *      0.11846
             3          8.46029 *      0.11820
```

```
                2          8.51072       0.11750
                3          8.57155       0.11667
                4          8.76685       0.11407
                5          8.76685       0.11407


Design   Iteration  D-Efficiency       D-Error
----------------------------------------------
    57       0           2.82843       0.35355
             1           7.68168       0.13018
             2           8.25623       0.12112
             3           8.37126       0.11946
             4           8.37126       0.11946


Design   Iteration  D-Efficiency       D-Error
----------------------------------------------
    58       0                 0             .
             1           7.15183       0.13982
             2           7.96630       0.12553
             3           8.40905       0.11892
             4           8.54936       0.11697
             5           8.54936       0.11697


Design   Iteration  D-Efficiency       D-Error
----------------------------------------------
    59       0                 0             .
             1           8.03226       0.12450
             2           8.52081       0.11736
             3           8.89875 *     0.11238
             4           8.89875       0.11238


Design   Iteration  D-Efficiency       D-Error
----------------------------------------------
    60       0           3.36359       0.29730
             1           7.54501       0.13254
             2           8.49478       0.11772
             3           8.71979       0.11468
             4           8.71979       0.11468
```

Find Efficient Choice Designs

```
      Final Results
Design                 59
Choice Sets            16
Alternatives            2
Parameters             16
Maximum Parameters     16
D-Efficiency       8.8987
Relative D-Eff    55.6172
D-Error            0.1124
1 / Choice Sets    0.0625
```

Find Efficient Choice Designs

n	Variable Name	Label	Variance	DF	Standard Error
1	x11	x1 1	0.12000	1	0.34641
2	x12	x1 2	0.12889	1	0.35901
3	x13	x1 3	0.12778	1	0.35746
4	x21	x2 1	0.17500	1	0.41833
5	x22	x2 2	0.12778	1	0.35746
6	x23	x2 3	0.13056	1	0.36132
7	x31	x3 1	0.09667	1	0.31091
8	x41	x4 1	0.10250	1	0.32016
9	x42	x4 2	0.11806	1	0.34359
10	x43	x4 3	0.11111	1	0.33333
11	x51	x5 1	0.19833	1	0.44535

12	x52	x5 2	0.15944	1	0.39930
13	x53	x5 3	0.13889	1	0.37268
14	x61	x6 1	0.15417	1	0.39264
15	x62	x6 2	0.09972	1	0.31579
16	x63	x6 3	0.13444	1	0.36667
				16	

Variance-Covariance Matrix

__label	x11	x12	x13	x21	x22	x23	x31	x41	x42	x43	x51	x52	x53	x61	x62	x63
x1 1	0.12	-0.00	-0.00	-0.01	-0.00	-0.02	0.01	0.01	0.00	-0.01	0.01	0.03	0.01	0.00	0.02	-0.02
x1 2	-0.00	0.13	0.01	-0.00	-0.02	0.01	0.03	0.01	-0.01	0.01	-0.07	-0.03	0.02	0.01	-0.00	0.03
x1 3	-0.00	0.01	0.13	-0.02	0.01	0.00	0.00	-0.01	-0.01	0.02	-0.06	0.00	0.04	0.04	-0.01	0.01
x2 1	-0.01	-0.00	-0.02	0.18	0.00	0.02	0.00	0.01	0.01	-0.01	0.02	0.01	-0.03	0.02	0.01	-0.03
x2 2	-0.00	-0.02	0.01	0.00	0.13	-0.05	0.00	-0.00	0.01	-0.02	0.01	0.01	-0.01	-0.01	-0.02	0.02
x2 3	-0.02	0.01	0.00	0.02	-0.05	0.13	-0.00	-0.02	-0.00	0.00	-0.02	-0.02	0.02	0.03	0.01	-0.01
x3 1	0.01	0.03	0.00	0.00	0.00	-0.00	0.10	0.01	0.01	-0.01	-0.02	-0.00	0.01	0.01	-0.00	0.01
x4 1	0.01	0.01	-0.01	0.01	-0.00	-0.02	0.01	0.10	-0.01	0.01	0.01	0.01	-0.02	-0.01	-0.00	-0.02
x4 2	0.00	-0.01	-0.01	0.01	0.01	-0.00	0.01	-0.01	0.12	0.01	0.02	0.00	-0.01	0.00	0.00	0.00
x4 3	-0.01	0.01	0.02	-0.01	-0.02	0.00	-0.01	0.01	0.01	0.11	-0.03	-0.01	0.01	-0.00	-0.00	-0.01
x5 1	0.01	-0.07	-0.06	0.02	0.01	-0.02	-0.02	0.01	0.02	-0.03	0.20	0.04	-0.06	-0.03	0.00	-0.04
x5 2	0.03	-0.03	0.00	0.01	0.01	-0.02	-0.00	0.01	0.00	-0.01	0.04	0.16	0.03	-0.01	0.00	-0.02
x5 3	0.01	0.02	0.04	-0.03	-0.01	0.02	0.01	-0.02	-0.01	0.01	-0.06	0.03	0.14	0.02	0.00	0.02
x6 1	0.00	0.01	0.04	0.02	-0.01	0.03	0.01	-0.01	0.00	-0.00	-0.03	-0.01	0.02	0.15	0.03	-0.03
x6 2	0.02	-0.00	-0.01	0.01	-0.02	0.01	-0.00	-0.00	0.00	-0.00	0.00	0.00	0.00	0.03	0.10	-0.03
x6 3	-0.02	0.03	0.01	-0.03	0.02	-0.01	0.01	-0.02	0.00	-0.01	-0.04	-0.02	0.02	-0.03	-0.03	0.13

Choice Sets

Obs	Design	Efficiency	Index	Set	Prob	n	f1	f2	x1	x2	x3	x4	x5	x6
1	59	8.89875	17	1	0.5	1857	1	1	3	3	1	3	3	3
2	59	8.89875	13	1	0.5	1858	1	1	2	4	1	2	4	1
3	59	8.89875	19	2	0.5	1859	1	1	3	4	1	1	2	3
4	59	8.89875	4	2	0.5	1860	1	1	1	2	2	4	3	4
5	59	8.89875	24	3	0.5	1861	1	1	3	2	2	2	3	1
6	59	8.89875	29	3	0.5	1862	1	1	4	2	1	4	2	3
7	59	8.89875	7	4	0.5	1863	1	1	1	4	1	3	2	2
8	59	8.89875	14	4	0.5	1864	1	1	2	4	2	1	3	4
9	59	8.89875	30	5	0.5	1865	1	1	4	2	2	3	1	2
10	59	8.89875	2	5	0.5	1866	1	1	1	1	2	2	2	4
11	59	8.89875	5	6	0.5	1867	1	1	1	3	1	1	3	2
12	59	8.89875	22	6	0.5	1868	1	1	3	1	2	4	2	1
13	59	8.89875	26	7	0.5	1869	1	1	4	4	2	3	3	1
14	59	8.89875	18	7	0.5	1870	1	1	3	3	2	4	4	2
15	59	8.89875	23	8	0.5	1871	1	1	3	2	1	1	4	4
16	59	8.89875	12	8	0.5	1872	1	1	2	1	2	3	4	3
17	59	8.89875	9	9	0.5	1873	1	1	2	2	1	2	2	2
18	59	8.89875	1	9	0.5	1874	1	1	1	1	1	1	1	1
19	59	8.89875	31	10	0.5	1875	1	1	4	1	1	2	3	3
20	59	8.89875	16	10	0.5	1876	1	1	2	3	2	3	2	4
21	59	8.89875	6	11	0.5	1877	1	1	1	3	2	2	4	3
22	59	8.89875	11	11	0.5	1878	1	1	2	1	1	4	3	2
23	59	8.89875	25	12	0.5	1879	1	1	4	4	1	4	4	4
24	59	8.89875	10	12	0.5	1880	1	1	2	2	2	1	1	3
25	59	8.89875	27	13	0.5	1881	1	1	4	3	1	2	1	4

26	59	8.89875	8	13	0.5	1882	1	1	1	4	2	4	1	3
27	59	8.89875	20	14	0.5	1883	1	1	3	4	2	2	1	2
28	59	8.89875	3	14	0.5	1884	1	1	1	2	1	3	4	1
29	59	8.89875	21	15	0.5	1885	1	1	3	1	1	3	1	4
30	59	8.89875	28	15	0.5	1886	1	1	4	3	2	1	2	1
31	59	8.89875	15	16	0.5	1887	1	1	2	3	1	4	1	1
32	59	8.89875	32	16	0.5	1888	1	1	4	1	2	1	4	2

Choice Sets by code

Set	x1	x2	x3	x4	x5	x6
1	3	3	1	3	3	3
	2	4	1	2	4	1

Set	x1	x2	x3	x4	x5	x6
2	3	4	1	1	2	3
	1	2	2	4	3	4

Set	x1	x2	x3	x4	x5	x6
3	3	2	2	2	3	1
	4	2	1	4	2	3

Set	x1	x2	x3	x4	x5	x6
4	1	4	1	3	2	2
	2	4	2	1	3	4

Set	x1	x2	x3	x4	x5	x6
5	4	2	2	3	1	2
	1	1	2	2	2	4

Set	x1	x2	x3	x4	x5	x6
6	1	3	1	1	3	2
	3	1	2	4	2	1

Set	x1	x2	x3	x4	x5	x6
7	4	4	2	3	3	1
	3	3	2	4	4	2

Set	x1	x2	x3	x4	x5	x6
8	3	2	1	1	4	4
	2	1	2	3	4	3

Set	x1	x2	x3	x4	x5	x6
9	2	2	1	2	2	2
	1	1	1	1	1	1

Set	x1	x2	x3	x4	x5	x6
10	4	1	1	2	3	3
	2	3	2	3	2	4

Set	x1	x2	x3	x4	x5	x6

11	1	3	2	2	4	3
	2	1	1	4	3	2

Set	x1	x2	x3	x4	x5	x6
12	4	4	1	4	4	4
	2	2	2	1	1	3

Set	x1	x2	x3	x4	x5	x6
13	4	3	1	2	1	4
	1	4	2	4	1	3

Set	x1	x2	x3	x4	x5	x6
14	3	4	2	2	1	2
	1	2	1	3	4	1

Set	x1	x2	x3	x4	x5	x6
15	3	1	1	3	1	4
	4	3	2	1	2	1

Set	x1	x2	x3	x4	x5	x6
16	2	3	1	4	1	1
	4	1	2	1	4	2

Choice Sets including Statistics

Set	Design	Efficiency	Index	Prob	n	f1	f2	x1	x2	x3	x4	x5	x6
1	59	8.89875	17	0.5	1857	1	1	3	3	1	3	3	3
	59	8.89875	13	0.5	1858	1	1	2	4	1	2	4	1

Set	Design	Efficiency	Index	Prob	n	f1	f2	x1	x2	x3	x4	x5	x6
2	59	8.89875	19	0.5	1859	1	1	3	4	1	1	2	3
	59	8.89875	4	0.5	1860	1	1	1	2	2	4	3	4

Set	Design	Efficiency	Index	Prob	n	f1	f2	x1	x2	x3	x4	x5	x6
3	59	8.89875	24	0.5	1861	1	1	3	2	2	2	3	1
	59	8.89875	29	0.5	1862	1	1	4	2	1	4	2	3

Set	Design	Efficiency	Index	Prob	n	f1	f2	x1	x2	x3	x4	x5	x6
4	59	8.89875	7	0.5	1863	1	1	1	4	1	3	2	2
	59	8.89875	14	0.5	1864	1	1	2	4	2	1	3	4

Set	Design	Efficiency	Index	Prob	n	f1	f2	x1	x2	x3	x4	x5	x6
5	59	8.89875	30	0.5	1865	1	1	4	2	2	3	1	2
	59	8.89875	2	0.5	1866	1	1	1	1	2	2	2	4

Set	Design	Efficiency	Index	Prob	n	f1	f2	x1	x2	x3	x4	x5	x6
6	59	8.89875	5	0.5	1867	1	1	1	3	1	1	3	2
	59	8.89875	22	0.5	1868	1	1	3	1	2	4	2	1

Set	Design	Efficiency	Index	Prob	n	f1	f2	x1	x2	x3	x4	x5	x6

7	59	8.89875	26	0.5	1869	1	1	4	4	2	3	3	1
	59	8.89875	18	0.5	1870	1	1	3	3	2	4	4	2

Set	Design	Efficiency	Index	Prob	n	f1	f2	x1	x2	x3	x4	x5	x6
8	59	8.89875	23	0.5	1871	1	1	3	2	1	1	4	4
	59	8.89875	12	0.5	1872	1	1	2	1	2	3	4	3

Set	Design	Efficiency	Index	Prob	n	f1	f2	x1	x2	x3	x4	x5	x6
9	59	8.89875	9	0.5	1873	1	1	2	2	1	2	2	2
	59	8.89875	1	0.5	1874	1	1	1	1	1	1	1	1

Set	Design	Efficiency	Index	Prob	n	f1	f2	x1	x2	x3	x4	x5	x6
10	59	8.89875	31	0.5	1875	1	1	4	1	1	2	3	3
	59	8.89875	16	0.5	1876	1	1	2	3	2	3	2	4

Set	Design	Efficiency	Index	Prob	n	f1	f2	x1	x2	x3	x4	x5	x6
11	59	8.89875	6	0.5	1877	1	1	1	3	2	2	4	3
	59	8.89875	11	0.5	1878	1	1	2	1	1	4	3	2

Set	Design	Efficiency	Index	Prob	n	f1	f2	x1	x2	x3	x4	x5	x6
12	59	8.89875	25	0.5	1879	1	1	4	4	1	4	4	4
	59	8.89875	10	0.5	1880	1	1	2	2	2	1	1	3

Set	Design	Efficiency	Index	Prob	n	f1	f2	x1	x2	x3	x4	x5	x6
13	59	8.89875	27	0.5	1881	1	1	4	3	1	2	1	4
	59	8.89875	8	0.5	1882	1	1	1	4	2	4	1	3

Set	Design	Efficiency	Index	Prob	n	f1	f2	x1	x2	x3	x4	x5	x6
14	59	8.89875	20	0.5	1883	1	1	3	4	2	2	1	2
	59	8.89875	3	0.5	1884	1	1	1	2	1	3	4	1

Set	Design	Efficiency	Index	Prob	n	f1	f2	x1	x2	x3	x4	x5	x6
15	59	8.89875	21	0.5	1885	1	1	3	1	1	3	1	4
	59	8.89875	28	0.5	1886	1	1	4	3	2	1	2	1

Set	Design	Efficiency	Index	Prob	n	f1	f2	x1	x2	x3	x4	x5	x6
16	59	8.89875	15	0.5	1887	1	1	2	3	1	4	1	1
	59	8.89875	32	0.5	1888	1	1	4	1	2	1	4	2

Choice Sets including Attributes and Levels description

Set	What type of data is processed?	Why are such data processed?	Who will be processing the data?	When will processing take place?	When will data be removed?	Financial Benefits per month
1	Data of your total energy consumption collected by sensors ++	To control daily routines in your house	Energy companies	Every Hour	After 1 year	Environmental benefit and $10
	Data of your total energy consumption collected by sensors +	To automate smart home appliances that detect and act	Energy companies	Every Day	After the product is out of use	Environmental benefit and no financial benefit

Set	What type of data is processed?	Why are such data processed?	Who will be processing the	When will processing take	When will data be removed?	Financial Benefits per

			data?	place?		month
2	Data of your total energy consumption collected by sensors ++	To automate smart home appliances that detect and act	Energy companies	Every Month	After 1 month	Environmental benefit and $10
	Data of your total energy consumption collected by sensors	To remotely manage the products in your house.	Technology companies	Every Minute	After 1 year	Environmental benefit and $15

Set	What type of data is processed?	Why are such data processed?	Who will be processing the data?	When will processing take place?	When will data be removed?	Financial Benefits per month
3	Data of your total energy consumption collected by sensors ++	To remotely manage the products in your house.	Technology companies	Every Day	After 1 year	Environmental benefit and no financial benefit
	Data of your total energy consumption collected by sensors +++	To remotely manage the products in your house.	Energy companies	Every Minute	After 1 month	Environmental benefit and $10

Set	What type of data is processed?	Why are such data processed?	Who will be processing the data?	When will processing take place?	When will data be removed?	Financial Benefits per month
4	Data of your total energy consumption collected by sensors	To automate smart home appliances that detect and act	Energy companies	Every Hour	After 1 month	Environmental benefit and $5
	Data of your total energy consumption collected by sensors +	To automate smart home appliances that detect and act	Technology companies	Every Month	After 1 year	Environmental benefit and $15

Set	What type of data is processed?	Why are such data processed?	Who will be processing the data?	When will processing take place?	When will data be removed?	Financial Benefits per month
5	Data of your total energy consumption collected by sensors +++	To remotely manage the products in your house.	Technology companies	Every Hour	After 10 days	Environmental benefit and $5
	Data of your total energy consumption collected by sensors	To inform you about the energy usage of products in your house	Technology companies	Every Day	After 1 month	Environmental benefit and $15

Set	What type of data is processed?	Why are such data processed?	Who will be processing the data?	When will processing take place?	When will data be removed?	Financial Benefits per month
6	Data of your total energy consumption collected by sensors	To control daily routines in your house	Energy companies	Every Month	After 1 year	Environmental benefit and $5
	Data of your total energy consumption collected by sensors ++	To inform you about the energy usage of products in your house	Technology companies	Every Minute	After 1 month	Environmental benefit and no financial benefit

Set	What type of data is processed?	Why are such data processed?	Who will be processing the data?	When will processing take place?	When will data be removed?	Financial Benefits per month
7	Data of your total energy consumption collected by sensors +++	To automate smart home appliances that detect and act	Technology companies	Every Hour	After 1 year	Environmental benefit and no financial benefit
	Data of your total energy consumption collected by sensors ++	To control daily routines in your house	Technology companies	Every Minute	After the product is out of use	Environmental benefit and $5

Set	What type of data is processed?	Why are such data processed?	Who will be processing the data?	When will processing take place?	When will data be removed?	Financial Benefits per month
8	Data of your total energy consumption collected by sensors ++	To remotely manage the products in your house.	Energy companies	Every Month	After the product is out of use	Environmental benefit and $15
	Data of your total energy consumption collected by sensors +	To inform you about the energy usage of products in your house	Technology companies	Every Hour	After the product is out of use	Environmental benefit and $10

Set	What type of data is processed?	Why are such data processed?	Who will be processing the data?	When will processing take place?	When will data be removed?	Financial Benefits per month
9	Data of your total energy consumption collected by sensors +	To remotely manage the products in your house.	Energy companies	Every Day	After 1 month	Environmental benefit and $5
	Data of your total energy consumption collected by sensors	To inform you about the energy usage of products in your house	Energy companies	Every Month	After 10 days	Environmental benefit and no financial benefit

Set	What type of data is processed?	Why are such data processed?	Who will be processing the data?	When will processing take place?	When will data be removed?	Financial Benefits per month
10	Data of your total energy consumption collected by sensors +++	To inform you about the energy usage of products in your house	Energy companies	Every Day	After 1 year	Environmental benefit and $10
	Data of your total energy consumption collected by sensors +	To control daily routines in your house	Technology companies	Every Hour	After 1 month	Environmental benefit and $15

Set	What type of data is processed?	Why are such data processed?	Who will be processing the data?	When will processing take place?	When will data be removed?	Financial Benefits per month
11	Data of your total energy consumption collected by sensors	To control daily routines in your house	Technology companies	Every Day	After the product is out of use	Environmental benefit and $10
	Data of your total energy consumption collected by sensors +	To inform you about the energy usage of products in your house	Energy companies	Every Minute	After 1 year	Environmental benefit and $5

Set	What type of data is processed?	Why are such data processed?	Who will be processing the data?	When will processing take place?	When will data be removed?	Financial Benefits per month
12	Data of your total energy consumption collected by sensors +++	To automate smart home appliances that detect and act	Energy companies	Every Minute	After the product is out of use	Environmental benefit and $15
	Data of your total energy consumption collected by sensors +	To remotely manage the products in your house.	Technology companies	Every Month	After 10 days	Environmental benefit and $10

Set	What type of data is processed?	Why are such data processed?	Who will be processing the data?	When will processing take place?	When will data be removed?	Financial Benefits per month
13	Data of your total energy consumption collected by sensors +++	To control daily routines in your house	Energy companies	Every Day	After 10 days	Environmental benefit and $15
	Data of your total energy consumption collected by sensors	To automate smart home appliances that detect and act	Technology companies	Every Minute	After 10 days	Environmental benefit and $10

Set	What type of data is processed?	Why are such data processed?	Who will be processing the data?	When will processing take place?	When will data be removed?	Financial Benefits per month
14	Data of your total energy consumption collected by sensors ++	To automate smart home appliances that detect and act	Technology companies	Every Day	After 10 days	Environmental benefit and $5
	Data of your total energy consumption collected by sensors	To remotely manage the products in your house.	Energy companies	Every Hour	After the product is out of use	Environmental benefit and no financial benefit

Set	What type of data is processed?	Why are such data processed?	Who will be processing the data?	When will processing take place?	When will data be removed?	Financial Benefits per month
15	Data of your total energy consumption collected by sensors ++	To inform you about the energy usage of products in your house	Energy companies	Every Hour	After 10 days	Environmental benefit and $15
	Data of your total energy consumption collected by sensors +++	To control daily routines in your house	Technology companies	Every Month	After 1 month	Environmental benefit and no financial benefit

Set	What type of data is processed?	Why are such data processed?	Who will be processing the data?	When will processing take place?	When will data be removed?	Financial Benefits per month
16	Data of your total energy consumption collected by sensors +	To control daily routines in your house	Energy companies	Every Minute	After 10 days	Environmental benefit and no financial benefit
	Data of your total energy consumption collected by sensors +++	To inform you about the energy usage of products in your house	Technology companies	Every Month	After the product is out of use	Environmental benefit and $5

7.2 Appendix II – Example Stated Choice Experiment

The content of this appendix is mentioned in § 3.4.3 on page 48 of this thesis

Appendix II – Example Choice Experiment

Note: The size of the appendix is reduced by shrinking the size of the images.

Introduction to the survey

Technische Universiteit
Eindhoven
University of Technology

Language: English

Introduction to the survey

Welcome

My name is Thomas van Houten and I would like to thank you in advance.
This survey is part of my master thesis for the study Construction Management & Engineering (CME) at the Eindhoven University of Technology (TU/e).
The topic is about the willingness to trade-off privacy sensitive data for the benefits of smart home appliances.

This questionnaire consist of four parts:

1. Introduction to the topic
2. Respondent Information
3. Statements about data privacy
4. Choice Experiment

The completion of this questionnaire will take approximately 10 minutes.
This survey works best on a mobile phone since you might need to zoom in/out.
Your personal data will be processed anonymously and only used for my masterthesis.
All questions are mandatory.

Since this research is anonymous. I will not ask you to leave personal information such as your e-mail.
If you are interested in the results of my research or have any questions, please contact me at t.v.houten@student.tue.nl

Again, thank you in advance,

Thomas van Houten

Part 1 - Introduction to the topic

Technische Universiteit Eindhoven University of Technology

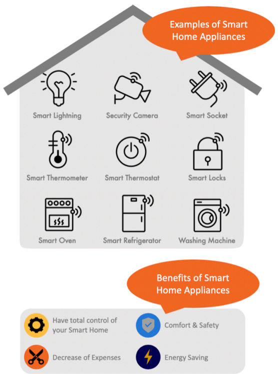
Language: English

Part 1 - Introduction to the topic

Have you ever heard of Smart Homes?

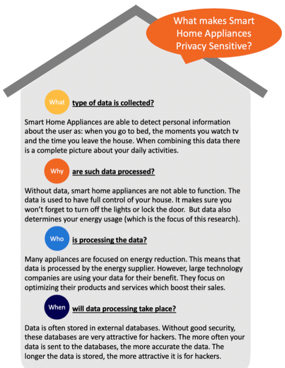
Smart homes are filled with smart appliances (devices) that can monitor, control and automate everything that happens in and around your house.

This research is focused on the appliances that are able to achieve an energy reduction. Here are some examples of smart home appliances:



What about data privacy?

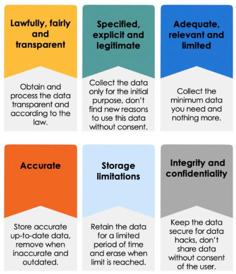
Have you ever considered the privacy consequences of using smart home appliances? Below, four important questions are answered for you:



General Data Protection Regulation (GDPR)

Since may 2018, there are new European privacy regulations actuated with big consequences.

The so called General Data Protection Regulation (GDPR) has 6 main principles:



Visit the website below for additional information

Website about the GDPR

Part 2 – Respondent Information

TU/e Technische Universiteit Eindhoven University of Technology

Language: English

Part 2 - Respondent Information

*
What is your gender?

- Male
- Female
- Other

*
What is your age category?

- < 19
- 19 - 29
- 30 - 45
- 46 - 65
- > 65

*
What is your occupation?

- Student
- Employed (Fulltime)
- Employed (Part-time)
- Unemployed
- Retired
- Other:

*
How does your household composition look like?

- Single Person household
- Two person household
- Family with Child(ren)
- Single parent
- Other:

Student housing falls under the single person household category.

*

What is your estimated gross income?

- < € 20.000
- € 20.000 - € 30.000
- € 30.000 - € 40.000
- € 40.000 - € 50.000
- > €50.000
- Rather not say
- Other:

*

What is your highest finished education?

- High School (VMBO)
- High School (HAVO, VWO)
- Vocational Education (MBO)
- Applied University (HBO)
- University (Bachelor's / Undergrad)
- University (Master's / Postgrad)
- Other:

If you are unfamiliar with the Dutch Educational System, please fill in 'other' including a description

Part 3 – Statements regarding data privacy and Energy Consumption

TU/e Technische Universiteit Eindhoven University of Technology

Language: English

Part 3 - Statements about your privacy concerns

*

Statements regarding data privacy

The following 5 questions are about potential concerns regarding data privacy. Please fill in to what extent you agree with this statement.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
I'm concerned about third parties being able to access my personal data	○	○	○	○	○
I'm concerned that parties are not keeping my personal information secure	○	○	○	○	○
I'm concerned that the information I submit on the internet could be misused	○	○	○	○	○
I'm concerned about parties building a profile of me to predict my consumer behavior	○	○	○	○	○
I'm concerned that I have insufficient control over the data that is collected about me	○	○	○	○	○

*

Statements regarding energy consumption and the environment

The following 5 questions are about your attitude towards energy consumption and the environment. Please fill in to what extent you agree with these statements.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
It is important to me to reduce my energy consumption.	○	○	○	○	○
I'm interested in having a better insight in my energy consumption.	○	○	○	○	○
I'm interested in smart technology that helps me reducing energy consumption.	○	○	○	○	○
I'm interested in the latest technology and gadgets.	○	○	○	○	○
I'm concerned about the environmental effects of Global Warming.	○	○	○	○	○

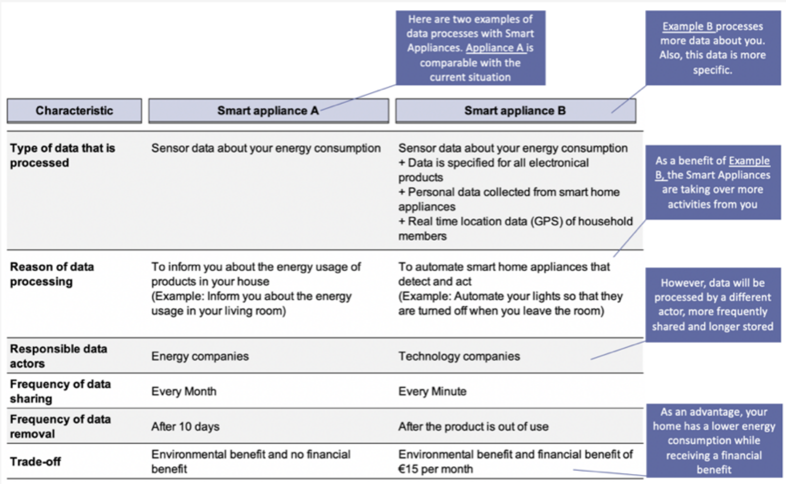
Part 4 - Choice Experiment

Example question: Which Smart Home Appliance do you prefer?

In the last part of this survey I would like you to compare two examples.
If you would live in a hypothetical smart home, are you willing to share more privacy sensitive data for an environmental or financial benefit? Or would you rather protect your data?

Check the example question below

Zoom in for more details



Characteristic	Smart appliance A	Smart appliance B
Type of data that is processed	Sensor data about your energy consumption	Sensor data about your energy consumption + Data is specified for all electronical products + Personal data collected from smart home appliances + Real time location data (GPS) of household members
Reason of data processing	To inform you about the energy usage of products in your house (Example: Inform you about the energy usage in your living room)	To automate smart home appliances that detect and act (Example: Automate your lights so that they are turned off when you leave the room)
Responsible data actors	Energy companies	Technology companies
Frequency of data sharing	Every Month	Every Minute
Frequency of data removal	After 10 days	After the product is out of use
Trade-off	Environmental benefit and no financial benefit	Environmental benefit and financial benefit of €15 per month

Now, 8 of these sets will be shown to you. Take a look at the characteristics and decide which example suits you best.

If the examples are equal to you, please select the 'no preference' option.

*

Which appliance do you prefer?

Zoom in for more details

Characteristic	Smart appliance A	Smart appliance B
Type of data that is processed	Sensor data about your energy consumption + Data is specified for all electronical products + Personal data collected from smart home appliances	Sensor data about your energy consumption + Data is specified for all electronical products + Personal data collected from smart home appliances + Real time location data (GPS) of household members
Reason of data processing	To remotely manage the products in your house. (Example: Manage your room temperature from distance)	To remotely manage the products in your house. (Example: Manage your room temperature from distance)
Responsible data actors	Technology companies	Energy companies
Frequency of data sharing	Every day	Every Minute
Frequency of data removal	After 1 year	After 1 month
Trade-off	Environmental benefit and no financial benefit	Environmental benefit and financial benefit of €10 per month

	Example A	Example B	No Preference
Which Example do you prefer?	○	○	○

*

Which appliance do you prefer?

Zoom in for more details

Characteristic	Smart appliance A	Smart appliance B
Type of data that is processed	Sensor data about your energy consumption + Data is specified for all electronical products + Personal data collected from smart home appliances	Sensor data about your energy consumption + Data is specified for all electronical products
Reason of data processing	To control daily routines in your house (Example: Control your dishwasher so that it turns itself on when energy usage is low)	To automate smart home appliances that detect and act (Example: Automate your lights so that they are turned off when you leave the room)
Responsible data actors	Energy companies	Energy companies
Frequency of data sharing	Every Hour	Every day
Frequency of data removal	After 1 year	After the product is out of use
Trade-off	Environmental benefit and financial benefit of €10 per month	Environmental benefit and no financial benefit

	Example A	Example B	No Preference
Which Example do you prefer?	○	○	○

*

Which appliance do you prefer?

Zoom in for more details

Characteristic	Smart appliance A	Smart appliance B
Type of data that is processed	Sensor data about your energy consumption + Data is specified for all electronical products + Personal data collected from smart home appliances + Real time location data (GPS) of household members	Sensor data about your energy consumption
Reason of data processing	To remotely manage the products in your house. (Example: Manage your room temperature from distance)	To inform you about the energy usage of products in your house (Example: Inform you about the energy usage in your living room)
Responsible data actors	Technology companies	Technology companies
Frequency of data sharing	Every Hour	Every day
Frequency of data removal	After 10 days	After 1 month
Trade-off	Environmental benefit and financial benefit of €5 per month	Environmental benefit and financial benefit of €15 per month

	Example A	Example B	No Preference
Which Example do you prefer?	○	○	○

*

Which appliance do you prefer?

Zoom in for more details

Characteristic	Smart appliance A	Smart appliance B
Type of data that is processed	Sensor data about your energy consumption + Data is specified for all electronical products + Personal data collected from smart home appliances + Real time location data (GPS) of household members	Sensor data about your energy consumption + Data is specified for all electronical products + Personal data collected from smart home appliances
Reason of data processing	To automate smart home appliances that detect and act (Example: Automate your lights so that they are turned off when you leave the room)	To control daily routines in your house (Example: Control your dishwasher so that it turns itself on when energy usage is low)
Responsible data actors	Technology companies	Technology companies
Frequency of data sharing	Every Hour	Every Minute
Frequency of data removal	After 1 year	After the product is out of use
Trade-off	Environmental benefit and no financial benefit	Environmental benefit and financial benefit of €5 per month

	Example A	Example B	No Preference
Which Example do you prefer?	○	○	○

*

Which appliance do you prefer?

Zoom in for more details

Characteristic	Smart appliance A	Smart appliance B
Type of data that is processed	Sensor data about your energy consumption + Data is specified for all electronical products + Personal data collected from smart home appliances + Real time location data (GPS) of household members	Sensor data about your energy consumption + Data is specified for all electronical products
Reason of data processing	To inform you about the energy usage of products in your house (Example: Inform you about the energy usage in your living room)	To control daily routines in your house (Example: Control your dishwasher so that it turns itself on when energy usage is low)
Responsible data actors	Energy companies	Technology companies
Frequency of data sharing	Every day	Every Hour
Frequency of data removal	After 1 year	After 1 month
Trade-off	Environmental benefit and financial benefit of €10 per month	Environmental benefit and financial benefit of €15 per month

	Example A	Example B	No Preference
Which Example do you prefer?	○	○	○

*

Which appliance do you prefer?

Zoom in for more details

Characteristic	Smart appliance A	Smart appliance B
Type of data that is processed	Sensor data about your energy consumption + Data is specified for all electronical products + Personal data collected from smart home appliances + Real time location data (GPS) of household members	Sensor data about your energy consumption + Data is specified for all electronical products
Reason of data processing	To automate smart home appliances that detect and act (Example: Automate your lights so that they are turned off when you leave the room)	To remotely manage the products in your house. (Example: Manage your room temperature from distance)
Responsible data actors	Energy companies	Technology companies
Frequency of data sharing	Every Minute	Every Month
Frequency of data removal	After the product is out of use	After 10 days
Trade-off	Environmental benefit and financial benefit of €15 per month	Environmental benefit and financial benefit of €10 per month

	Example A	Example B	No Preference
Which Example do you prefer?	○	○	○

*

Which appliance do you prefer?

Zoom in for more details

Characteristic	Smart appliance A	Smart appliance B
Type of data that is processed	Sensor data about your energy consumption + Data is specified for all electronical products + Personal data collected from smart home appliances + Real time location data (GPS) of household members	Sensor data about your energy consumption
Reason of data processing	To control daily routines in your house (Example: Control your dishwasher so that it turns itself on when energy usage is low)	To automate smart home appliances that detect and act (Example: Automate your lights so that they are turned off when you leave the room)
Responsible data actors	Energy companies	Technology companies
Frequency of data sharing	Every day	Every Minute
Frequency of data removal	After 10 days	After 10 days
Trade-off	Environmental benefit and financial benefit of €15 per month	Environmental benefit and financial benefit of €10 per month

	Example A	Example B	No Preference

*

Which appliance do you prefer?

Zoom in for more details

Characteristic	Smart appliance A	Smart appliance B
Type of data that is processed	Sensor data about your energy consumption + Data is specified for all electronical products + Personal data collected from smart home appliances	Sensor data about your energy consumption + Data is specified for all electronical products + Personal data collected from smart home appliances + Real time location data (GPS) of household members
Reason of data processing	To inform you about the energy usage of products in your house (Example: Inform you about the energy usage in your living room)	To control daily routines in your house (Example: Control your dishwasher so that it turns itself on when energy usage is low)
Responsible data actors	Energy companies	Technology companies
Frequency of data sharing	Every Hour	Every Month
Frequency of data removal	After 10 days	After 1 month
Trade-off	Environmental benefit and financial benefit of €15 per month	Environmental benefit and no financial benefit

	Example A	Example B	No Preference
Which Example do you prefer?	○	○	○

7.3 Appendix III – Calculations Sample Size (R-statistics)

The content of this appendix is mentioned in § 3.4.6 on page 50 of this thesis

```
test_alpha=0.05
z_one_minus_alpha<-qnorm(1-test_alpha)
test_beta=0.2
z_one_minus_beta<-qnorm(1-test_beta)
parameters<-c(-1.501, 0.500, 0.308, -0.020, -0.668, -0.055, -0.201, 0.200, 0.056, 0.055,-0.055
           -0.034, 0.184, -0.055, -0.095, -0.440, -0.053, 0.121, 0.372)
ncoefficients=17
nalts=3
nchoices=16

# load the design information
design<-as.matrix(read.table(file.choose(),header=FALSE));

#compute the information matrix
# initialize a matrix of size ncoefficients by ncoefficients filled with zeros.
info_mat=matrix(rep(0,ncoefficients*ncoefficients), ncoefficients, ncoefficients)

# compute exp(design matrix times initial parameter values)
exputilities=exp(design%*%parameters)

# loop over all choice sets
for (k_set in 1:nchoices) {

 # select alternatives in the choice set
 alternatives=((k_set-1)*nalts+1) : (k_set*nalts)

 # obtain vector of choice shares within the choice set
 p_set=exputilities[alternatives]/sum(exputilities[alternatives])

 # also put these probabilities on the diagonal of a matrix that only contains zeros
 p_diag=diag(p_set)

 # compute middle term P-pp'
 middle_term<-p_diag-p_set%o%p_set

 # pre- and postmultiply with the Xs from the design matrix for the alternatives in this choice set
 full_term<-t(design[alternatives,])%*%middle_term%*%design[alternatives,]

# Add contribution of this choice set to the information matrix
info_mat<-info_mat+full_term } # end of loop over choice sets

#get the inverse of the information matrix (i.e., gets the variance-covariance matrix)
sigma_beta<-solve(info_mat,diag(ncoefficients))

# Use the parameter values as effect size. Other values can be used here.
effectsize<-parameters

# formula for sample size calculation is n>[(z_(beta)+z_(1-alpha))*sqrt(S??)/delta]^2
N<-((z_one_minus_beta + z_one_minus_alpha)*sqrt(diag(sigma_beta))/abs(effectsize))^2

# Display results (required sample size for each coefficient)
 [1]    1.538494   13.294982   36.039596 8700.927118    8.023978 1252.061514   67.840047   81.279202 1085.416641 1075.452107
[11]  487.428669  105.297854 1022.750707  338.943237   16.820629 1091.652549  236.748089   24.018278
```

7.4 Appendix IV – Descriptive results Survey

The content of this appendix is mentioned in § 4.1.2 on page 58 of this thesis

Frequency Distributions:

Language		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2	42	16,4	16,4	16,4
	1	214	83,6	83,6	100,0
	Total	256	100,0	100,0	

Gender		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2	126	49,2	49,2	49,2
	1	130	50,8	50,8	100,0
	Total	256	100,0	100,0	

AgeCat		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	22	8,6	8,6	8,6
	5	4	1,6	1,6	10,2
	2	78	30,5	30,5	40,6
	3	56	21,9	21,9	62,5
	4	96	37,5	37,5	100,0
	Total	256	100,0	100,0	

Occupation		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2	142	55,5	55,5	55,5
	3	44	17,2	17,2	72,7
	5	4	1,6	1,6	74,2
	1	58	22,7	22,7	96,9
	4	8	3,1	3,1	100,0
	Total	256	100,0	100,0	

Household		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	3	134	52,3	52,3	52,3
	5	1	0,4	0,4	52,7
	4	5	2,0	2,0	54,7
	1	39	15,2	15,2	69,9
	2	77	30,1	30,1	100,0
	Total	256	100,0	100,0	

IncomeCat		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	63	24,6	24,6	24,6
	5	67	26,2	26,2	50,8
	2	26	10,2	10,2	60,9
	3	44	17,2	17,2	78,1
	4	27	10,5	10,5	88,7
	6	29	11,3	11,3	100,0
	Total	256	100,0	100,0	

The content of this appendix is mentioned in § 4.2.1 on page 60 of this thesis

Means and ANOVA tables of statements regarding data privacy

Gender		PrivacyStatement 1	PrivacyStatement 2	PrivacyStatement 3	PrivacyStatement 4	PrivacyStatement 5
1	Mean	3,50	3,56	3,55	3,43	3,89
	Std. Deviation	1,04	1,04	,957	1,06	,990
2	Mean	3,64	3,69	3,81	3,64	4,10
	Std. Deviation	,907	,881	,836	1,02	,757
Total	Mean	3,57	3,63	3,68	3,54	4,00
	Std. Deviation	,980	,966	,907	1,04	,888

ANOVA

Gender		Sum of Squares	df	Mean Square	F	Sig.
PrivacyStatement1	Between Groups	1,306	1	1,306	1,363	,244
	Within Groups	243,429	254	,958		
	Total	244,734	255			
PrivacyStatement2	Between Groups	1,064	1	1,064	1,140	,287
	Within Groups	236,936	254	,933		
	Total	238,000	255			
PrivacyStatement3	Between Groups	4,183	1	4,183	5,169	,024
	Within Groups	205,552	254	,809		
	Total	209,734	255			
PrivacyStatement4	Between Groups	2,878	1	2,878	2,660	,104
	Within Groups	274,805	254	1,082		
	Total	277,684	255			
PrivacyStatement5	Between Groups	2,845	1	2,845	3,647	,057
	Within Groups	198,151	254	,780		
	Total	200,996	255			

Age		PrivacyStatement 1	PrivacyStatement 2	PrivacyStatement 3	PrivacyStatement 4	PrivacyStatement 5
1	Mean	3,32	3,14	3,45	3,68	3,77
	Std. Deviation	,894	,941	,963	1,09	,869
2	Mean	3,37	3,44	3,62	3,29	3,87
	Std. Deviation	1,13	1,09	,996	1,20	1,04
3	Mean	3,55	3,45	3,71	3,66	4,07
	Std. Deviation	,807	,893	,929	,940	,783
4	Mean	3,77	3,97	3,77	3,61	4,10
	Std. Deviation	,934	,801	,788	,933	,827
5	Mean	4,25	4,25	3,50	3,75	4,00
	Std. Deviation	,500	,500	1,291	1,26	,000
Total	Mean	3,57	3,63	3,68	3,54	4,00
	Std. Deviation	,980	,966	,907	1,044	,888

ANOVA

Age		Sum of Squares	df	Mean Square	F	Sig.
PrivacyStatement1	Between Groups	9,314	3	3,105	3,323	,020
	Within Groups	235,420	252	,934		
	Total	244,734	255			
PrivacyStatement2	Between Groups	22,430	3	7,477	8,740	,000
	Within Groups	215,570	252	,855		
	Total	238,000	255			
PrivacyStatement3	Between Groups	2,150	3	,717	,870	,457
	Within Groups	207,585	252	,824		
	Total	209,734	255			
PrivacyStatement4	Between Groups	6,579	3	2,193	2,039	,109
	Within Groups	271,104	252	1,076		
	Total	277,684	255			
PrivacyStatement5	Between Groups	3,700	3	1,233	1,575	,196

		Sum of Squares	df	Mean Square	F	Sig.
	Within Groups	197,296	252	,783		
	Total	200,996	255			

Occupation		PrivacyStatement 1	PrivacyStatement 2	PrivacyStatement 3	PrivacyStatement 4	PrivacyStatement 5
1	Mean	3,34	3,36	3,57	3,36	3,90
	Std. Deviation	1,163	1,135	,993	1,266	,986
2	Mean	3,66	3,73	3,68	3,51	3,99
	Std. Deviation	,959	,931	,902	1,009	,922
3	Mean	3,45	3,50	3,68	3,68	4,07
	Std. Deviation	,730	,792	,829	,800	,661
4	Mean	3,88	3,88	4,00	3,88	4,25
	Std. Deviation	,991	,641	,756	,991	,707
5	Mean	4,25	4,75	4,50	4,50	4,50
	Std. Deviation	,500	,500	,577	,577	,577
Total	Mean	3,57	3,63	3,68	3,54	4,00
	Std. Deviation	,980	,966	,907	1,044	,888

ANOVA

Occupation		Sum of Squares	df	Mean Square	F	Sig.
PrivacyStatement1	Between Groups	7,322	4	1,831	1,935	,105
	Within Groups	237,412	251	,946		
	Total	244,734	255			
PrivacyStatement2	Between Groups	11,690	4	2,922	3,241	,013
	Within Groups	226,310	251	,902		
	Total	238,000	255			
PrivacyStatement3	Between Groups	4,225	4	1,056	1,290	,274
	Within Groups	205,509	251	,819		
	Total	209,734	255			
PrivacyStatement4	Between Groups	7,395	4	1,849	1,717	,147
	Within Groups	270,289	251	1,077		
	Total	277,684	255			
PrivacyStatement5	Between Groups	2,349	4	,587	,742	,564
	Within Groups	198,647	251	,791		
	Total	200,996	255			

Household		PrivacyStatement 1	PrivacyStatement 2	PrivacyStatement 3	PrivacyStatement 4	PrivacyStatement 5
1	Mean	3,46	3,69	3,44	3,31	3,90
	Std. Deviation	1,12	1,03	1,07	1,23	,995
2	Mean	3,69	3,75	3,83	3,58	4,03
	Std. Deviation	,921	,920	,923	1,080	,903
3	Mean	3,54	3,54	3,67	3,57	4,01
	Std. Deviation	,978	,978	,839	,961	,863
4	Mean	3,60	3,60	3,40	3,60	4,00
	Std. Deviation	,894	,548	,894	1,14	,707
Total	Mean	3,57	3,63	3,68	3,54	4,00
	Std. Deviation	,980	,966	,907	1,044	,888

ANOVA

Household		Sum of Squares	df	Mean Square	F	Sig.
PrivacyStatement1	Between Groups	2,009	4	,502	,519	,722
	Within Groups	242,725	251	,967		
	Total	244,734	255			
PrivacyStatement2	Between Groups	4,949	4	1,237	1,333	,258
	Within Groups	233,051	251	,928		
	Total	238,000	255			
PrivacyStatement3	Between Groups	4,587	4	1,147	1,403	,233
	Within Groups	205,147	251	,817		
	Total	209,734	255			

PrivacyStatement4	Between Groups	2,579	4	,645	,588	,671
	Within Groups	275,105	251	1,096		
	Total	277,684	255			
PrivacyStatement5	Between Groups	,466	4	,116	,146	,965
	Within Groups	200,530	251	,799		
	Total	200,996	255			

Income		PrivacyStatement 1	PrivacyStatement 2	PrivacyStatement 3	PrivacyStatement 4	PrivacyStatement 5
1	Mean	3,38	3,40	3,57	3,52	3,92
	Std. Deviation	1,084	1,071	,979	1,162	,989
2	Mean	3,35	3,42	3,54	3,42	3,96
	Std. Deviation	,936	,987	,811	,857	,774
3	Mean	3,25	3,34	3,61	3,39	3,82
	Std. Deviation	,811	,805	,841	1,04	,815
4	Mean	3,85	3,67	3,74	3,59	4,07
	Std. Deviation	,989	,877	,944	1,19	,829
5	Mean	3,72	3,93	3,76	3,57	4,01
	Std. Deviation	,950	,876	,889	,973	,913
6	Mean	4,07	4,00	3,90	3,76	4,34
	Std. Deviation	,799	,964	,939	,988	,814
Total	Mean	3,57	3,63	3,68	3,54	4,00
	Std. Deviation	,980	,966	,907	1,04	,888

ANOVA

Income		Sum of Squares	df	Mean Square	F	Sig.
PrivacyStatement1	Between Groups	18,861	5	3,772	4,175	,001
	Within Groups	225,873	250	,903		
	Total	244,734	255			
PrivacyStatement2	Between Groups	18,061	5	3,612	4,106	,001
	Within Groups	219,939	250	,880		
	Total	238,000	255			
PrivacyStatement3	Between Groups	3,359	5	,672	,814	,541
	Within Groups	206,376	250	,826		
	Total	209,734	255			
PrivacyStatement4	Between Groups	2,915	5	,583	,530	,753
	Within Groups	274,769	250	1,099		
	Total	277,684	255			
PrivacyStatement5	Between Groups	5,497	5	1,099	1,406	,223
	Within Groups	195,499	250	,782		
	Total	200,996	255			

	Education	PrivacyStatement 1	PrivacyStatement 2	PrivacyStatement 3	PrivacyStatement 4	PrivacyStatement 5
1	Mean	4,00	4,11	3,56	4,11	4,56
	Std. Deviation	,707	,601	,882	,782	,527
2	Mean	3,07	3,13	3,23	3,07	3,83
	Std. Deviation	1,11	1,01	1,07	1,34	,986
3	Mean	3,60	3,61	3,68	3,68	3,94
	Std. Deviation	,858	,894	,883	,864	,721
4	Mean	3,55	3,63	3,63	3,53	3,98
	Std. Deviation	,916	,983	,868	,979	,907
5	Mean	3,81	3,76	4,11	3,54	4,14
	Std. Deviation	,995	1,01	,774	1,169	,948
6	Mean	3,66	3,81	3,78	3,53	4,00
	Std. Deviation	1,15	,931	,870	1,04	1,02
Total	Mean	3,57	3,63	3,68	3,54	4,00
	N	256	256	256	256	256

Std. Deviation	,980	,966	,907	1,044	,888

ANOVA						
Education		Sum of Squares	df	Mean Square	F	Sig.
PrivacyStatement1	Between Groups	11,740	5	2,348	2,519	,030
	Within Groups	232,994	250	,932		
	Total	244,734	255			
PrivacyStatement2	Between Groups	11,156	5	2,231	2,459	,034
	Within Groups	226,844	250	,907		
	Total	238,000	255			
PrivacyStatement3	Between Groups	13,468	5	2,694	3,431	,005
	Within Groups	196,267	250	,785		
	Total	209,734	255			
PrivacyStatement4	Between Groups	10,826	5	2,165	2,028	,075
	Within Groups	266,857	250	1,067		
	Total	277,684	255			
PrivacyStatement5	Between Groups	4,587	5	,917	1,168	,326
	Within Groups	196,409	250	,786		
	Total	200,996	255			

The content of this appendix is mentioned in § 4.2.1 on page 60 of this thesis

Means and ANOVA tables of statements regarding Energy Consumption

Gender		EnergyStatement1	EnergyStatement2	EnergyStatement3
1	Mean	3,75	3,52	3,77
	Std. Deviation	,845	,934	,831
2	Mean	3,59	3,55	3,51
	Std. Deviation	,832	,835	,767
Total	Mean	3,67	3,53	3,64
	Std. Deviation	,841	,885	,809

ANOVA						
Gender		Sum of Squares	df	Mean Square	F	Sig.
EnergyStatement1	Between Groups	1,775	1	1,775	2,523	,113
	Within Groups	178,663	254	,703		
	Total	180,438	255			
EnergyStatement2	Between Groups	,066	1	,066	,085	,771
	Within Groups	199,684	254	,786		
	Total	199,750	255			
EnergyStatement3	Between Groups	4,369	1	4,369	6,825	,010
	Within Groups	162,569	254	,640		
	Total	166,938	255			

Age		EnergyStatement1	EnergyStatement2	EnergyStatement3
1	Mean	3,45	3,36	3,50
	Std. Deviation	,800	,790	,740
2	Mean	3,55	3,50	3,53
	Std. Deviation	,921	,936	,879
3	Mean	3,43	3,34	3,48
	Std. Deviation	,871	,996	,763
4	Mean	3,97	3,73	3,88
	Std. Deviation	,672	,747	,743
5	Mean	3,50	3,00	3,25
	Std. Deviation	1,000	1,155	,957
Total	Mean	3,67	3,53	3,64
	Std. Deviation	,841	,885	,809

Age		ANOVA Sum of Squares	df	Mean Square	F	Sig.
EnergyStatement1	Between Groups	13,224	3	4,408	6,643	,000
	Within Groups	167,214	252	,664		
	Total	180,438	255			
EnergyStatement2	Between Groups	5,606	3	1,869	2,425	,066
	Within Groups	194,144	252	,770		
	Total	199,750	255			
EnergyStatement3	Between Groups	7,257	3	2,419	3,817	,011
	Within Groups	159,681	252	,634		
	Total	166,938	255			

Occupation		EnergyStatement1	EnergyStatement2	EnergyStatement3
1	Mean	3,53	3,53	3,57
	Std. Deviation	,922	,941	,775
2	Mean	3,74	3,55	3,76
	Std. Deviation	,796	,855	,798
3	Mean	3,66	3,50	3,50
	Std. Deviation	,805	,849	,699
4	Mean	3,50	3,38	3,00
	Std. Deviation	1,069	1,302	1,309
5	Mean	3,75	3,50	3,25
	Std. Deviation	1,258	1,000	,957
Total	Mean	3,67	3,53	3,64
	Std. Deviation	,841	,885	,809

Occupation		ANOVA Sum of Squares	df	Mean Square	F	Sig.
EnergyStatement1	Between Groups	2,011	4	,503	,707	,588
	Within Groups	178,427	251	,711		
	Total	180,438	255			
EnergyStatement2	Between Groups	,289	4	,072	,091	,985
	Within Groups	199,461	251	,795		
	Total	199,750	255			
EnergyStatement3	Between Groups	7,104	4	1,776	2,789	,027
	Within Groups	159,833	251	,637		
	Total	166,938	255			

Household		EnergyStatement1	EnergyStatement2	EnergyStatement3
1	Mean	3,72	3,64	3,62
	Std. Deviation	,916	,873	,747
2	Mean	3,69	3,55	3,58
	Std. Deviation	,862	,925	,784
3	Mean	3,66	3,49	3,69
	Std. Deviation	,813	,874	,853
4	Mean	3,60	3,80	3,60
	Std. Deviation	,548	,447	,548
Total	Mean	3,67	3,53	3,64
	Std. Deviation	,841	,885	,809

Household		ANOVA Sum of Squares	df	Mean Square	F	Sig.
EnergyStatement1	Between Groups	2,904	3	,968	1,374	,251
	Within Groups	177,533	252	,704		
	Total	180,438	255			
EnergyStatement2	Between Groups	3,255	3	1,085	1,392	,246
	Within Groups	196,495	252	,780		
	Total	199,750	255			
EnergyStatement3	Between Groups	,459	3	,153	,232	,874

	Within Groups	166,478	252	,661		
	Total	166,938	255			

Income		EnergyStatement1	EnergyStatement2	EnergyStatement3
1	Mean	3,62	3,56	3,44
	Std. Deviation	,941	,996	,894
2	Mean	3,38	3,23	3,62
	Std. Deviation	,983	,951	,752
3	Mean	3,57	3,55	3,48
	Std. Deviation	,625	,663	,762
4	Mean	3,48	3,41	3,59
	Std. Deviation	,935	,844	,797
5	Mean	3,97	3,63	3,90
	Std. Deviation	,696	,902	,781
6	Mean	3,69	3,62	3,79
	Std. Deviation	,850	,862	,675
Total	Mean	3,67	3,53	3,64
	Std. Deviation	,841	,885	,809

ANOVA

Income		Sum of Squares	df	Mean Square	F	Sig.
EnergyStatement1	Between Groups	9,743	5	1,949	2,854	,016
	Within Groups	170,694	250	,683		
	Total	180,438	255			
EnergyStatement2	Between Groups	3,652	5	,730	,931	,461
	Within Groups	196,098	250	,784		
	Total	199,750	255			
EnergyStatement3	Between Groups	8,705	5	1,741	2,751	,019
	Within Groups	158,232	250	,633		
	Total	166,938	255			

Education		EnergyStatement1	EnergyStatement2	EnergyStatement3
1	Mean	4,22	4,11	4,11
	Std. Deviation	,667	,782	,782
2	Mean	3,40	3,40	3,43
	Std. Deviation	,968	1,003	,679
3	Mean	3,55	3,35	3,52
	Std. Deviation	,739	,889	,763
4	Mean	3,71	3,56	3,76
	Std. Deviation	,765	,820	,766
5	Mean	3,73	3,68	3,65
	Std. Deviation	,962	,915	,949
6	Mean	3,84	3,59	3,63
	Std. Deviation	,920	,875	,907
Total	Mean	3,67	3,53	3,64
	Std. Deviation	,841	,885	,809

ANOVA

Education		Sum of Squares	df	Mean Square	F	Sig.
EnergyStatement1	Between Groups	7,079	5	1,416	2,042	,073
	Within Groups	173,359	250	,693		
	Total	180,438	255			
EnergyStatement2	Between Groups	6,431	5	1,286	1,663	,144
	Within Groups	193,319	250	,773		
	Total	199,750	255			
EnergyStatement3	Between Groups	5,394	5	1,079	1,669	,142
	Within Groups	161,544	250	,646		
	Total	166,938	255			

7.5 Appendix V - Multinomial and Mixed Logit Models

```
Multinomial Logit Model (Page 63):
Call:
mlogit(formula = Choice ~ 0 + Constant + Type1 + Type2 + Type3 +
    Why1 + Why2 + Why3 + Act1 + Share1 + Share2 + Share3 + Remove1 +
    Remove2 + Remove3 + Trade1 + Trade2 + Trade3, data = DataLong,
    method = "nr")

Frequencies of alternatives:
      A       B      No
0.44336 0.45996 0.09668

nr method
5 iterations, 0h:0m:0s
g'(-H)^-1g = 0.000261
successive function values within tolerance limits

Coefficients :
           Estimate Std. Error  z-value  Pr(>|z|)
Constant -1.5017168  0.0817171 -18.3770 < 2.2e-16 ****
Type1     0.4007592  0.0565352   7.0887 1.354e-12 ****
Type2     0.3084363  0.0544796   5.6615 1.501e-08 ****
Type3    -0.0206293  0.0552541  -0.3734 0.7088848
Why1     -0.0557569  0.0598400  -0.9318 0.3514567
Why2     -0.2017656  0.0608105  -3.3179 0.0009068 ****
Why3      0.1996068  0.0567842   3.5152 0.0004395 ****
Act1     -0.0055069  0.0605888  -0.0909 0.9275797
Share1   -0.0344511  0.0537839  -0.6405 0.5218171
Share2    0.1836950  0.0556001   3.3039 0.0009536 ****
Share3   -0.0548566  0.0578126  -0.9489 0.3426866
Remove1   0.0863363  0.0676194   1.2768 0.2016733
Remove2   0.0538746  0.0613705   0.8779 0.3800213
Remove3  -0.0837937  0.0588409  -1.4241 0.1544252
Trade1   -0.4400710  0.0574952  -7.6540 1.954e-14 ****
Trade2   -0.0534055  0.0569039  -0.9385 0.3479771
Trade3    0.1210924  0.0578356   2.0937 0.0362837 **
---
Significance codes:  (0 = '****') (0.001 = '***') (0.01 = '**') (0.05 = '*')

Log-Likelihood: -1776.4
--------------------------------------------------------------------
```

Mixed Logit Model

```
Call:
mlogit(formula = Choice ~ 0 + Constant + Type1 + Type2 + Type3 +
    Why1 + Why2 + Why3 + Act1 + Share1 + Share2 + Share3 + Remove1 +
    Remove2 + Remove3 + Trade1 + Trade2 + Trade3, data = DataLong,
    reflevel = "No", rpar = c(Type1 = "n", Type2 = "n", Type3 = "n",
        Why1 = "n", Why2 = "n", Why3 = "n", Act1 = "n", Share1 = "n",
        Share2 = "n", Share3 = "n", Remove1 = "n", Remove2 = "n",
        Remove3 = "n", Trade1 = "n", Trade2 = "n", Trade3 = "n"),
    R = 125, Halton = NA, Panel = TRUE, Method = BFGS, Correlation = FALSE)

Frequencies of alternatives:
     No       A       B
0.09668 0.44336 0.45996

bfgs method
38 iterations, 0h:2m:8s
g'(-H)^-1g = 3.03E-07
gradient close to zero

Coefficients :
              Estimate  Std. Error  z-value  Pr(>|z|)
Constant   -1.5612e+00  1.3533e-01 -11.5356 < 2.2e-16 ****
Type1       6.0208e-01  1.4462e-01   4.1631 3.139e-05 ****
Type2       4.5856e-01  1.3264e-01   3.4573 0.0005456 ****
Type3       9.2841e-02  1.0355e-01   0.8966 0.3699257
Why1       -1.2344e-01  1.0547e-01  -1.1704 0.2418278
Why2       -2.8502e-01  1.1257e-01  -2.5320 0.0113410 **
Why3        4.5395e-01  1.3815e-01   3.2858 0.0010168 ***
Act1       -7.2612e-02  1.1007e-01  -0.6597 0.5094625
Share1     -6.5350e-02  9.4538e-02  -0.6913 0.4894026
Share2      3.3602e-01  1.1150e-01   3.0137 0.0025804 ***
Share3     -2.0248e-01  1.0369e-01  -1.9527 0.0508519 *
Remove1     2.5975e-01  1.1779e-01   2.2051 0.0274489 **
Remove2     6.1798e-02  1.0112e-01   0.6112 0.5410962
Remove3    -2.3612e-01  1.1808e-01  -1.9998 0.0455261 **
Trade1     -7.0682e-01  1.5787e-01  -4.4773 7.558e-06 ****
Trade2     -6.4856e-02  1.0327e-01  -0.6280 0.5299718
Trade3      2.5697e-01  1.0616e-01   2.4206 0.0154944 **
sd.Share1   8.7208e-02  6.7303e-01   0.1296 0.8969013
sd.Share2   4.5078e-01  6.3293e-01   0.7122 0.4763404 *
sd.Share3  -8.2538e-01  5.0938e-01  -1.6204 0.1051547
sd.Remove1 -1.2424e+00  4.1118e-01  -3.0215 0.0025151 ***
sd.Remove2  3.8441e-01  5.5245e-01   0.6958 0.4865433
sd.Remove3  1.1301e+00  4.6742e-01   2.4177 0.0156168 **
---
Significance codes:  (0 = ‘****’) (0.001 = ‘***’) (0.01 = ‘**’) (0.05 = ‘*’)

Log-Likelihood: -1770.3

random coefficients
        Min.     1st Qu.      Median        Mean      3rd Qu. Max.
Type1   -Inf  0.56349579  0.60207780  0.60207780  0.640659821  Inf
Type2   -Inf  0.40575588  0.45856340  0.45856340  0.511370931  Inf
Type3   -Inf  0.01958403  0.09284080  0.09284080  0.166097562  Inf
Why1    -Inf -0.50480380 -0.12344453 -0.12344453  0.257914731  Inf
Why2    -Inf -0.42525463 -0.28501778 -0.28501778 -0.144780937  Inf
Why3    -Inf -0.10669611  0.45395085  0.45395085  1.014597798  Inf
Act1    -Inf -0.57117660 -0.07261221 -0.07261221  0.425952171  Inf
Share1  -Inf -0.12417127 -0.06535008 -0.06535008 -0.006528902  Inf
Share2  -Inf  0.03198063  0.33602448  0.33602448  0.640068317  Inf
Share3  -Inf -0.75918605 -0.20247900 -0.20247900  0.354228049  Inf
Remove1 -Inf -0.57823601  0.25974628  0.25974628  1.097728573  Inf
Remove2 -Inf -0.19747946  0.06179824  0.06179824  0.321075944  Inf
Remove3 -Inf -0.99836400 -0.23612253 -0.23612253  0.526118936  Inf
```

```
Trade1  -Inf -0.79865548 -0.70681719 -0.70681719 -0.614978907  Inf
Trade2  -Inf -0.24323253 -0.06485619 -0.06485619  0.113520160  Inf
Trade3  -Inf  0.25691340  0.25696567  0.25696567  0.257017940  Inf
----------------------------------------------------------------
```

Mixed Logit Model + Interaction Gender + Type

```
Call: mlogit(formula = Choice ~ Constant + Type1 + Type2 + Type3 +
    Why1 + Why2 + Why3 + Act1 + Share1 + Share2 + Share3 + Remove1 +
    Remove2 + Remove3 + Trade1 + Trade2 + Trade3 + Type1:Gender +
    Type2:Gender + Type3:Gender + 0, data = DataLong, reflevel = "No",
    rpar = c(Why1 = "n", Why2 = "n", Why3 = "n", Remove1 = "n",
        Remove2 = "n", Remove3 = "n"), R = 125, Halton = NA,
    Panel = TRUE, Method = BFGS, Correlation = FALSE)

Frequencies of alternatives:
     No       A       B
0.09668 0.44336 0.45996

bfgs method
20 iterations, 0h:0m:31s
g'(-H)^-1g = 3.42E-07
gradient close to zero

Coefficients :
              Estimate Std. Error  z-value  Pr(>|z|)
Constant     -1.453414   0.096023 -15.1361 < 2.2e-16 ***
Type1         0.508284   0.102846   4.9422 7.725e-07 ***
Type2         0.376113   0.087445   4.3011 1.699e-05 ***
Type3         0.043434   0.082257   0.5280 0.5974784
Why1         -0.097959   0.090278  -1.0851 0.2778885
Why2         -0.213729   0.083219  -2.5683 0.0102211 *
Why3          0.366933   0.099913   3.6725 0.0002402 ***
Act1         -0.030123   0.045943  -0.6557 0.5120461
Share1       -0.054239   0.079510  -0.6822 0.4951338
Share2        0.270453   0.089238   3.0307 0.0024398 **
Share3       -0.167563   0.082363  -2.0344 0.0419068 *
Remove1       0.206417   0.097691   2.1130 0.0346040 *
Remove2       0.084983   0.081711   1.0401 0.2983161
Remove3      -0.214562   0.107809  -1.9902 0.0465675 *
Trade1       -0.603668   0.112901  -5.3469 8.949e-08 ***
Trade2       -0.078820   0.087973  -0.8960 0.3702749
Trade3        0.235494   0.088444   2.6626 0.0077530 **
Type1:Gender -0.323312   0.080550  -4.0138 5.975e-05 ***
Type2:Gender -0.060261   0.070153  -0.8590 0.3903460
Type3:Gender  0.026960   0.069290   0.3891 0.6972102
sd.Why1       0.174237   2.054094   0.0848 0.9324012
sd.Why2      -0.475468   0.596991  -0.7964 0.4257761
sd.Why3       0.371272   0.705723   0.5261 0.5988273
sd.Remove1    0.876283   0.400764   2.1865 0.0287766 *
sd.Remove2    0.192792   1.499444   0.1286 0.8976933
sd.Remove3    1.175928   0.363304   3.2368 0.0012089 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Log-Likelihood: -1753.1

random coefficients
        Min.     1st Qu.      Median        Mean    3rd Qu. Max.
Why1    -Inf -0.21547968 -0.09795880 -0.09795880 0.01956207  Inf
Why2    -Inf -0.53442685 -0.21372874 -0.21372874 0.10696938  Inf
Why3    -Inf  0.11651342  0.36693286  0.36693286 0.61735230  Inf
Remove1 -Inf -0.38462679  0.20641719  0.20641719 0.79746117  Inf
Remove2 -Inf -0.04505302  0.08498342  0.08498342 0.21501986  Inf
Remove3 -Inf -1.00771413 -0.21456244 -0.21456244 0.57858926  Inf
```

Mixed Logit Model including Interaction Gender + Trade

```
Call:
mlogit(formula = Choice ~ Constant + Type1 + Type2 + Type3 +
    Why1 + Why2 + Why3 + Act1 + Share1 + Share2 + Share3 + Remove1 +
    Remove2 + Remove3 + Trade1 + Trade2 + Trade3 + Trade1:Gender +
    Trade2:Gender + Trade3:Gender + 0, data = DataLong, reflevel = "No",
    rpar = c(Why1 = "n", Why2 = "n", Why3 = "n", Remove1 = "n",
        Remove2 = "n", Remove3 = "n"), R = 125, Halton = NA,
    Panel = TRUE, Method = BFGS, Correlation = FALSE)

Frequencies of alternatives:
     No       A       B
0.09668 0.44336 0.45996

bfgs method
23 iterations, 0h:0m:31s
g'(-H)^-1g = 2.43E-07
gradient close to zero

Coefficients :
               Estimate Std. Error  z-value  Pr(>|z|)
Constant      -1.444849   0.092824 -15.5655 < 2.2e-16 ***
Type1          0.488167   0.099587   4.9019 9.492e-07 ***
Type2          0.359497   0.087539   4.1067 4.014e-05 ***
Type3          0.032289   0.083487   0.3868 0.6989366
Why1          -0.094888   0.088865  -1.0678 0.2856208
Why2          -0.204333   0.081390  -2.5105 0.0120552 *
Why3           0.369435   0.099211   3.7237 0.0001963 ***
Act1          -0.034689   0.045658  -0.7598 0.4473937
Share1        -0.048001   0.079157  -0.6064 0.5442450
Share2         0.260750   0.089273   2.9208 0.0034912 **
Share3        -0.166198   0.081321  -2.0437 0.0409820 *
Remove1        0.223041   0.096796   2.3042 0.0212094 *
Remove2        0.066377   0.080480   0.8248 0.4095032
Remove3       -0.226448   0.108576  -2.0856 0.0370140 *
Trade1        -0.591040   0.110384  -5.3544 8.584e-08 ***
Trade2        -0.060665   0.088357  -0.6866 0.4923370
Trade3         0.225574   0.088959   2.5357 0.0112224 *
Trade1:Gender -0.068395   0.075664  -0.9039 0.3660303
Trade2:Gender -0.038815   0.074905  -0.5182 0.6043266
Trade3:Gender -0.026185   0.072226  -0.3625 0.7169498
sd.Why1        0.133550   2.354930   0.0567 0.9547755
sd.Why2       -0.535480   0.546061  -0.9806 0.3267793
sd.Why3        0.269131   1.103242   0.2439 0.8072732
sd.Remove1     0.734489   0.490711   1.4968 0.1344489
sd.Remove2     0.181885   1.544399   0.1178 0.9062492
sd.Remove3     1.232994   0.362949   3.3972 0.0006809 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Log-Likelihood: -1768.3

random coefficients
        Min.     1st Qu.     Median       Mean      3rd Qu. Max.
Why1    -Inf -0.18496594 -0.0948876 -0.0948876 -0.004809255  Inf
Why2    -Inf -0.56550832 -0.2043328 -0.2043328  0.156842792  Inf
Why3    -Inf  0.18790919  0.3694350  0.3694350  0.550960868  Inf
Remove1 -Inf -0.27236446  0.2230407  0.2230407  0.718445935  Inf
Remove2 -Inf -0.05630261  0.0663771  0.0663771  0.189056814  Inf
Remove3 -Inf -1.05808994 -0.2264479 -0.2264479  0.605194124  Inf
```

Mixed Logit Model including Interaction Age + Type

```
Call:
mlogit(formula = Choice ~ Constant + Type1 + Type2 + Type3 +
    Why1 + Why2 + Why3 + Act1 + Share1 + Share2 + Share3 + Remove1 +
    Remove2 + Remove3 + Trade1 + Trade2 + Trade3 + Type1:Age1 +
    Type1:Age2 + Type2:Age1 + Type2:Age2 + Type3:Age1 + Type3:Age2 +
    0, data = DataLong, reflevel = "No", rpar = c(Share1 = "n",
    Share2 = "n", Share3 = "n", Remove1 = "n", Remove2 = "n",
    Remove3 = "n"), R = 125, Halton = NA, Panel = TRUE, Method = BFGS,
    Correlation = FALSE)

Frequencies of alternatives:
     No       A       B
0.09668 0.44336 0.45996

bfgs method
20 iterations, 0h:0m:29s
g'(-H)^-1g = 8.85E-07
gradient close to zero

Coefficients :
             Estimate Std. Error  z-value  Pr(>|z|)
Constant    -1.436211   0.091585 -15.6817 < 2.2e-16 ***
Type1        0.558739   0.105499   5.2962 1.183e-07 ***
Type2        0.381992   0.095711   3.9911 6.577e-05 ***
Type3        0.036067   0.091286   0.3951 0.6927725
Why1        -0.104670   0.086799  -1.2059 0.2278630
Why2        -0.214699   0.088134  -2.4360 0.0148490 *
Why3         0.394200   0.113194   3.4825 0.0004967 ***
Act1        -0.023209   0.045638  -0.5085 0.6110757
Share1      -0.039754   0.081931  -0.4852 0.6275211
Share2       0.261111   0.090647   2.8805 0.0039699 **
Share3      -0.173155   0.092025  -1.8816 0.0598911 .
Remove1      0.225739   0.099462   2.2696 0.0232323 *
Remove2      0.065481   0.089958   0.7279 0.4666711
Remove3     -0.240825   0.111684  -2.1563 0.0310593 *
Trade1      -0.608887   0.109617  -5.5547 2.781e-08 ***
Trade2      -0.078021   0.088414  -0.8824 0.3775353
Trade3       0.240632   0.090307   2.6646 0.0077081 **
Type1:Age1   0.059225   0.097033   0.6104 0.5416217
Type1:Age2   0.326918   0.121453   2.6917 0.0071085 **
Type2:Age1  -0.011240   0.096153  -0.1169 0.9069405
Type2:Age2   0.056505   0.113431   0.4981 0.6183831
Type3:Age1   0.098603   0.098149   1.0046 0.3150789
Type3:Age2  -0.093273   0.117706  -0.7924 0.4281132
sd.Share1    0.151992   1.700601   0.0894 0.9287835
sd.Share2   -0.034193   2.090580  -0.0164 0.9869506
sd.Share3    0.492481   0.888560   0.5542 0.5794106
sd.Remove1   0.806683   0.384938   2.0956 0.0361160 *
sd.Remove2  -0.038467   1.739868  -0.0221 0.9823608
sd.Remove3   1.320971   0.339381   3.8923 9.930e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Log-Likelihood: -1756.9

random coefficients
        Min.     1st Qu.      Median        Mean    3rd Qu. Max.
Share1  -Inf -0.14227165 -0.03975448 -0.03975448 0.06276269  Inf
Share2  -Inf  0.23804822  0.26111100  0.26111100 0.28417379  Inf
Share3  -Inf -0.50532794 -0.17315461 -0.17315461 0.15901871  Inf
Remove1 -Inf -0.31836030  0.22573880  0.22573880 0.76983790  Inf
Remove2 -Inf  0.03953552  0.06548126  0.06548126 0.09142700  Inf
Remove3 -Inf -1.13180654 -0.24082490 -0.24082490 0.65015674  Inf
```

Mixed Logit Model including Interaction Age + Trade

```
Call:
mlogit(formula = Choice ~ Constant + Type1 + Type2 + Type3 +
    Why1 + Why2 + Why3 + Act1 + Share1 + Share2 + Share3 + Remove1 +
    Remove2 + Remove3 + Trade1 + Trade2 + Trade3 + Trade1:Age1 +
    Trade1:Age2 + Trade2:Age1 + Trade2:Age2 + Trade3:Age1 + Trade3:Age2 +
    0, data = DataLong, reflevel = "No", rpar = c(Share1 = "n",
    Share2 = "n", Share3 = "n", Remove1 = "n", Remove2 = "n",
    Remove3 = "n"), R = 125, Halton = NA, Panel = TRUE, Method = BFGS,
    Correlation = FALSE)

Frequencies of alternatives:
     No       A       B
0.09668 0.44336 0.45996

bfgs method
21 iterations, 0h:0m:33s
g'(-H)^-1g = 5.95E-07
gradient close to zero

Coefficients :
             Estimate Std. Error  z-value  Pr(>|z|)
Constant    -1.440738   0.098025 -14.6976 < 2.2e-16 ***
Type1        0.471546   0.093667   5.0343 4.796e-07 ***
Type2        0.358009   0.090690   3.9476 7.893e-05 ***
Type3        0.042666   0.087334   0.4885 0.6251703
Why1        -0.110342   0.087223  -1.2651 0.2058495
Why2        -0.209007   0.085225  -2.4524 0.0141903 *
Why3         0.369629   0.110334   3.3501 0.0008078 ***
Act1        -0.024689   0.043826  -0.5634 0.5731928
Share1      -0.043705   0.078979  -0.5534 0.5800051
Share2       0.254478   0.087321   2.9143 0.0035651 **
Share3      -0.152145   0.090106  -1.6885 0.0913129 .
Remove1      0.218509   0.098825   2.2111 0.0270311 *
Remove2      0.051915   0.091085   0.5700 0.5687009
Remove3     -0.227247   0.106475  -2.1343 0.0328196 *
Trade1      -0.571822   0.110710  -5.1650 2.404e-07 ***
Trade2      -0.112916   0.090287  -1.2506 0.2110697
Trade3       0.246940   0.090943   2.7153 0.0066209 **
Trade1:Age1 -0.123941   0.105489  -1.1749 0.2400279
Trade1:Age2  0.018889   0.124173   0.1521 0.8790936
Trade2:Age1  0.117004   0.106020   1.1036 0.2697652
Trade2:Age2 -0.213947   0.127883  -1.6730 0.0943289 .
Trade3:Age1  0.023671   0.098142   0.2412 0.8094101
Trade3:Age2  0.147766   0.117680   1.2557 0.2092390
sd.Share1    0.157060   1.605493   0.0978 0.9220697
sd.Share2   -0.040832   2.077507  -0.0197 0.9843192
sd.Share3   -0.337968   1.120561  -0.3016 0.7629525
sd.Remove1   0.743672   0.388432   1.9145 0.0555502 .
sd.Remove2   0.135014   1.694427   0.0797 0.9364910
sd.Remove3   1.190537   0.340516   3.4963 0.0004718 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Log-Likelihood: -1768.1

random coefficients
        Min.     1st Qu.      Median        Mean    3rd Qu. Max.
Share1  -Inf -0.14964072 -0.04370505 -0.04370505 0.06223063  Inf
Share2  -Inf  0.22693769  0.25447825  0.25447825 0.28201881  Inf
Share3  -Inf -0.38010038 -0.15214453 -0.15214453 0.07581132  Inf
Remove1 -Inf -0.28308967  0.21850915  0.21850915 0.72010797  Inf
Remove2 -Inf -0.03914987  0.05191545  0.05191545 0.14298077  Inf
Remove3 -Inf -1.03025257 -0.22724731 -0.22724731 0.57575796  Inf
```

Mixed Logit Model including Interaction Income + Type

```
Call:
mlogit(formula = Choice ~ Constant + Type1 + Type2 + Type3 +
    Why1 + Why2 + Why3 + Act1 + Share1 + Share2 + Share3 + Remove1 +
    Remove2 + Remove3 + Trade1 + Trade2 + Trade3 + Type1:Income1 +
    Type1:Income2 + Type1:Income3 + Type2:Income1 + Type2:Income2 +
    Type2:Income3 + Type3:Income1 + Type3:Income2 + Type3:Income3 +
    0, data = DataLong, reflevel = "No", rpar = c(Share1 = "n",
    Share2 = "n", Share3 = "n", Remove1 = "n", Remove2 = "n",
    Remove3 = "n"), R = 125, Halton = NA, Panel = TRUE, Method = BFGS,
    Correlation = FALSE)

Frequencies of alternatives:
     No       A       B
0.09668 0.44336 0.45996

bfgs method
30 iterations, 0h:0m:41s
g'(-H)^-1g = 2.66E-07
gradient close to zero

Coefficients :
                Estimate Std. Error  z-value  Pr(>|z|)
Constant      -1.4393262  0.0908522 -15.8425 < 2.2e-16 ***
Type1          0.4530842  0.0963203   4.7039 2.552e-06 ***
Type2          0.3295629  0.0955106   3.4505 0.0005595 ***
Type3          0.0979525  0.0908975   1.0776 0.2812057
Why1          -0.1094939  0.0839404  -1.3044 0.1920889
Why2          -0.2175585  0.0895283  -2.4301 0.0150966 *
Why3           0.3940152  0.1172012   3.3619 0.0007742 ***
Act1          -0.0334873  0.0454476  -0.7368 0.4612237
Share1        -0.0457608  0.0810477  -0.5646 0.5723354
Share2         0.2759127  0.0911002   3.0287 0.0024563 **
Share3        -0.1825109  0.0919473  -1.9850 0.0471499 *
Remove1        0.2294565  0.1008332   2.2756 0.0228696 *
Remove2        0.0694087  0.0816185   0.8504 0.3951005
Remove3       -0.2469312  0.1115844  -2.2130 0.0269009 *
Trade1        -0.6222710  0.1154255  -5.3911 7.003e-08 ***
Trade2        -0.0852507  0.0888728  -0.9592 0.3374356
Trade3         0.2552003  0.0912229   2.7975 0.0051492 **
Type1:Income1  0.1400828  0.1280213   1.0942 0.2738610
Type1:Income2  0.2648350  0.1133711   2.3360 0.0194912 *
Type1:Income3 -0.2696121  0.1205732  -2.2361 0.0253461 *
Type2:Income1  0.2422545  0.1291518   1.8757 0.0606918 .
Type2:Income2  0.0428253  0.1137584   0.3765 0.7065763
Type2:Income3  0.0067846  0.1212547   0.0560 0.9553788
Type3:Income1 -0.0171290  0.1261815  -0.1357 0.8920201
Type3:Income2 -0.0928314  0.1112317  -0.8346 0.4039560
Type3:Income3 -0.0343533  0.1208921  -0.2842 0.7762838
sd.Share1      0.0539156  2.1073251   0.0256 0.9795885
sd.Share2     -0.0235671  2.0417845  -0.0115 0.9907907
sd.Share3     -0.4442535  0.9486911  -0.4683 0.6395841
sd.Remove1     0.9405537  0.3345722   2.8112 0.0049355 **
sd.Remove2     0.1043895  1.7027896   0.0613 0.9511163
sd.Remove3     1.3090716  0.3530686   3.7077 0.0002092 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Log-Likelihood: -1758.9

random coefficients
        Min.     1st Qu.      Median        Mean      3rd Qu. Max.
Share1  -Inf -0.08212631 -0.04576077 -0.04576077 -0.009395228  Inf
Share2  -Inf  0.26001692  0.27591270  0.27591270  0.291808484  Inf
```

```
Share3  -Inf -0.48215530 -0.18251089 -0.18251089  0.117133511  Inf
Remove1 -Inf -0.40493729  0.22945652  0.22945652  0.863850327  Inf
Remove2 -Inf -0.00100094  0.06940872  0.06940872  0.139818388  Inf
Remove3 -Inf -1.12988658 -0.24693117 -0.24693117  0.636024231  Inf
```

Mixed Logit Model including Interaction Income + Trade

```
Call:
mlogit(formula = Choice ~ Constant + Type1 + Type2 + Type3 +
    Why1 + Why2 + Why3 + Act1 + Share1 + Share2 + Share3 + Remove1 +
    Remove2 + Remove3 + Trade1 + Trade2 + Trade3 + Trade1:Income1 +
    Trade1:Income2 + Trade1:Income3 + Trade2:Income1 + Trade2:Income2 +
    Trade2:Income3 + Trade3:Income1 + Trade3:Income2 + Trade3:Income3 +
    0, data = DataLong, reflevel = "No", rpar = c(Share1 = "n",
    Share2 = "n", Share3 = "n", Remove1 = "n", Remove2 = "n",
    Remove3 = "n"), R = 125, Halton = NA, Panel = TRUE, Method = BFGS,
    Correlation = FALSE)

Frequencies of alternatives:
     No       A       B
0.09668 0.44336 0.45996

bfgs method
23 iterations, 0h:0m:33s
g'(-H)^-1g = 3.93E-07
gradient close to zero

Coefficients :
                 Estimate Std. Error  z-value  Pr(>|z|)
Constant       -1.4349621  0.0919520 -15.6056 < 2.2e-16 ***
Type1           0.4779471  0.0913236   5.2336 1.663e-07 ***
Type2           0.3654161  0.0920095   3.9715 7.142e-05 ***
Type3           0.0489042  0.0861293   0.5678 0.5701710
Why1           -0.1184102  0.0867214  -1.3654 0.1721247
Why2           -0.2072841  0.0855566  -2.4228 0.0154026 *
Why3            0.3915162  0.1110640   3.5251 0.0004233 ***
Act1           -0.0274605  0.0449596  -0.6108 0.5413445
Share1         -0.0429986  0.0814220  -0.5281 0.5974334
Share2          0.2563609  0.0892899   2.8711 0.0040903 **
Share3         -0.1633058  0.0902769  -1.8089 0.0704599 .
Remove1         0.2309949  0.1000021   2.3099 0.0208937 *
Remove2         0.0579534  0.0912840   0.6349 0.5255141
Remove3        -0.2460572  0.1114887  -2.2070 0.0273130 *
Trade1         -0.6061413  0.1106235  -5.4793 4.270e-08 ***
Trade2         -0.1126754  0.0960585  -1.1730 0.2408010
Trade3          0.2384339  0.0922290   2.5852 0.0097312 **
Trade1:Income1 -0.0658780  0.1425229  -0.4622 0.6439183
Trade1:Income2  0.0795425  0.1191904   0.6674 0.5045441
Trade1:Income3  0.0015089  0.1333481   0.0113 0.9909717
Trade2:Income1  0.0847441  0.1383115   0.6127 0.5400719
Trade2:Income2  0.0090053  0.1214891   0.0741 0.9409117
Trade2:Income3  0.1239538  0.1340548   0.9247 0.3551476
Trade3:Income1  0.1042480  0.1308933   0.7964 0.4257795
Trade3:Income2 -0.0179392  0.1149756  -0.1560 0.8760126
Trade3:Income3 -0.0618633  0.1215279  -0.5090 0.6107198
sd.Share1       0.1450427  1.7873766   0.0811 0.9353239
sd.Share2      -0.0123654  2.1760717  -0.0057 0.9954661
sd.Share3       0.3362651  1.1393051   0.2951 0.7678798
sd.Remove1      0.7995331  0.3830812   2.0871 0.0368781 *
sd.Remove2     -0.0263437  1.7768685  -0.0148 0.9881710
sd.Remove3      1.3169609  0.3414791   3.8566 0.0001150 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Log-Likelihood: -1768.8
```

```
random coefficients
        Min.     1st Qu.      Median        Mean    3rd Qu. Max.
Share1  -Inf -0.14082841 -0.04299858 -0.04299858 0.05483126  Inf
Share2  -Inf  0.24802060  0.25636093  0.25636093 0.26470125  Inf
Share3  -Inf -0.39011313 -0.16330576 -0.16330576 0.06350160  Inf
Remove1 -Inf -0.30828198  0.23099493  0.23099493 0.77027184  Inf
Remove2 -Inf  0.04018478  0.05795337  0.05795337 0.07572196  Inf
Remove3 -Inf -1.13433385 -0.24605724 -0.24605724 0.64221937  Inf
```

Mixed Logit Model including Interaction Privacy + Type

```
Call:
mlogit(formula = Choice ~ 0 + Constant + Type1 + Type2 + Type3 +
    Why1 + Why2 + Why3 + Act1 + Share1 + Share2 + Share3 + Remove1 +
    Remove2 + Remove3 + Trade1 + Trade2 + Trade3 + Type1:FactorPrivacy +
    Type2:FactorPrivacy + Type3:FactorPrivacy, data = DataLong,
    reflevel = "No", rpar = c(Why1 = "n", Why2 = "n", Why3 = "n",
        Remove1 = "n", Remove2 = "n", Remove3 = "n"), R = 125,
    Halton = NA, Panel = TRUE, Method = BFGS, Correlation = FALSE)

Frequencies of alternatives:
     No       A       B
0.09668 0.44336 0.45996

bfgs method
19 iterations, 0h:0m:29s
g'(-H)^-1g = 5.11E-07
gradient close to zero

Coefficients :
                     Estimate Std. Error  z-value  Pr(>|z|)
Constant            -1.448554   0.090384 -16.0267 < 2.2e-16 ***
Type1                0.487231   0.092952   5.2417 1.591e-07 ***
Type2                0.354435   0.087533   4.0492 5.140e-05 ***
Type3                0.026907   0.081431   0.3304 0.7410724
Why1                -0.098722   0.085305  -1.1573 0.2471549
Why2                -0.202798   0.079212  -2.5602 0.0104613 *
Why3                 0.363283   0.096917   3.7484 0.0001780 ***
Act1                -0.028561   0.044770  -0.6379 0.5235111
Share1              -0.041347   0.077101  -0.5363 0.5917679
Share2               0.249418   0.087297   2.8571 0.0042748 **
Share3              -0.155795   0.080871  -1.9265 0.0540453 .
Remove1              0.210952   0.094287   2.2373 0.0252644 *
Remove2              0.070234   0.080811   0.8691 0.3847878
Remove3             -0.224979   0.107172  -2.0992 0.0357970 *
Trade1              -0.579977   0.103457  -5.6059 2.071e-08 ***
Trade2              -0.067946   0.086958  -0.7814 0.4345875
Trade3               0.215345   0.087371   2.4647 0.0137123 *
Type1:FactorPrivacy  0.177881   0.074273   2.3950 0.0166220 *
Type2:FactorPrivacy  0.029120   0.068890   0.4227 0.6725146
Type3:FactorPrivacy  0.041157   0.070700   0.5821 0.5604711
sd.Why1              0.065956   2.819296   0.0234 0.9813355
sd.Why2             -0.477120   0.574051  -0.8311 0.4058917
sd.Why3              0.292343   0.931127   0.3140 0.7535465
sd.Remove1           0.634623   0.516857   1.2279 0.2195025
sd.Remove2           0.258341   1.307352   0.1976 0.8433534
sd.Remove3           1.221342   0.350330   3.4863 0.0004898 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Log-Likelihood: -1762.8

random coefficients
        Min.    1st Qu.      Median        Mean     3rd Qu. Max.
```

```
Why1    -Inf -0.1432093 -0.09872240 -0.09872240 -0.05423548  Inf
Why2    -Inf -0.5246102 -0.20279759 -0.20279759  0.11901500  Inf
Why3    -Inf  0.1661010  0.36328321  0.36328321  0.56046537  Inf
Remove1 -Inf -0.2170948  0.21095212  0.21095212  0.63899905  Inf
Remove2 -Inf -0.1040145  0.07023358  0.07023358  0.24448168  Inf
Remove3 -Inf -1.0487617 -0.22497908 -0.22497908  0.59880348  Inf
```

<u>Mixed Logit Model including Interaction Privacy + Trade</u>

```
Call:
mlogit(formula = Choice ~ 0 + Constant + Type1 + Type2 + Type3 +
    Why1 + Why2 + Why3 + Act1 + Share1 + Share2 + Share3 + Remove1 +
    Remove2 + Remove3 + Trade1 + Trade2 + Trade3 + Trade1:FactorPrivacy +
    Trade2:FactorPrivacy + Trade3:FactorPrivacy, data = DataLong,
    reflevel = "No", rpar = c(Why1 = "n", Why2 = "n", Why3 = "n",
        Remove1 = "n", Remove2 = "n", Remove3 = "n"), R = 125,
    Halton = NA, Panel = TRUE, Method = BFGS, Correlation = FALSE)

Frequencies of alternatives:
     No       A       B
0.09668 0.44336 0.45996

bfgs method
19 iterations, 0h:0m:27s
g'(-H)^-1g = 1.86E-07
gradient close to zero

Coefficients :
                      Estimate Std. Error  z-value  Pr(>|z|)
Constant             -1.449494   0.087583 -16.5500 < 2.2e-16 ***
Type1                 0.473285   0.090470   5.2314 1.683e-07 ***
Type2                 0.354848   0.084252   4.2117 2.534e-05 ***
Type3                 0.031837   0.081699   0.3897 0.6967732
Why1                 -0.090545   0.085213  -1.0626 0.2879782
Why2                 -0.208726   0.079036  -2.6409 0.0082687 **
Why3                  0.353499   0.096287   3.6713 0.0002413 ***
Act1                 -0.036938   0.044579  -0.8286 0.4073259
Share1               -0.045240   0.076611  -0.5905 0.5548434
Share2                0.255164   0.086066   2.9647 0.0030293 **
Share3               -0.152380   0.079366  -1.9200 0.0548610 .
Remove1               0.200508   0.094100   2.1308 0.0331054 *
Remove2               0.058646   0.079714   0.7357 0.4619135
Remove3              -0.200272   0.104053  -1.9247 0.0542664 .
Trade1               -0.569609   0.101946  -5.5874 2.305e-08 ***
Trade2               -0.060744   0.085388  -0.7114 0.4768476
Trade3                0.207935   0.086516   2.4034 0.0162418 *
Trade1:FactorPrivacy -0.152838   0.076384  -2.0009 0.0454024 *
Trade2:FactorPrivacy -0.115076   0.075110  -1.5321 0.1254962
Trade3:FactorPrivacy  0.203279   0.074428   2.7312 0.0063104 **
sd.Why1               0.066308   2.830800   0.0234 0.9813123
sd.Why2              -0.544045   0.511985  -1.0626 0.2879549
sd.Why3               0.276254   1.081141   0.2555 0.7983208
sd.Remove1            0.664085   0.510955   1.2997 0.1937061
sd.Remove2            0.150302   1.605612   0.0936 0.9254188
sd.Remove3            1.115349   0.367482   3.0351 0.0024044 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Log-Likelihood: -1765.7

random coefficients
        Min.     1st Qu.      Median        Mean     3rd Qu. Max.
Why1    -Inf -0.13526901 -0.09054501 -0.09054501 -0.04582102  Inf
Why2    -Inf -0.57567838 -0.20872573 -0.20872573  0.15822692  Inf
Why3    -Inf  0.16716861  0.35349908  0.35349908  0.53982955  Inf
```

```
Remove1 -Inf -0.24741017  0.20050814  0.20050814  0.64842645  Inf
Remove2 -Inf -0.04273134  0.05864556  0.05864556  0.16002245  Inf
Remove3 -Inf -0.95256314 -0.20027197 -0.20027197  0.55201920  Inf
```

<u>Mixed Logit Model including Interaction Energy + Type</u>

```
Call:
mlogit(formula = Choice ~ 0 + Constant + Type1 + Type2 + Type3 +
    Why1 + Why2 + Why3 + Act1 + Share1 + Share2 + Share3 + Remove1 +
    Remove2 + Remove3 + Trade1 + Trade2 + Trade3 + Type1:FactorEnergy +
    Type2:FactorEnergy + Type3:FactorEnergy, data = DataLong,
    reflevel = "No", rpar = c(Why1 = "n", Why2 = "n", Why3 = "n",
        Remove1 = "n", Remove2 = "n", Remove3 = "n"), R = 125,
    Halton = NA, Panel = TRUE, Method = BFGS, Correlation = FALSE)

Frequencies of alternatives:
     No       A       B
0.09668 0.44336 0.45996

bfgs method
21 iterations, 0h:0m:29s
g'(-H)^-1g = 6.33E-07
gradient close to zero

Coefficients :
                    Estimate Std. Error  z-value  Pr(>|z|)
Constant           -1.450845   0.090299 -16.0671 < 2.2e-16 ***
Type1               0.480971   0.096879   4.9647 6.882e-07 ***
Type2               0.360273   0.085333   4.2220 2.422e-05 ***
Type3               0.036570   0.082754   0.4419 0.6585482
Why1               -0.093892   0.088273  -1.0636 0.2874893
Why2               -0.206531   0.080919  -2.5523 0.0107013 *
Why3                0.354620   0.098288   3.6080 0.0003086 ***
Act1               -0.030614   0.045604  -0.6713 0.5020309
Share1             -0.048483   0.078434  -0.6181 0.5364808
Share2              0.261009   0.087858   2.9708 0.0029702 **
Share3             -0.165926   0.080431  -2.0630 0.0391157 *
Remove1             0.213061   0.096320   2.2120 0.0269664 *
Remove2             0.068521   0.079732   0.8594 0.3901295
Remove3            -0.202693   0.106586  -1.9017 0.0572122 .
Trade1             -0.583226   0.108216  -5.3895 7.066e-08 ***
Trade2             -0.065368   0.086659  -0.7543 0.4506588
Trade3              0.220218   0.087250   2.5240 0.0116029 *
Type1:FactorEnergy  0.063434   0.068464   0.9265 0.3541727
Type2:FactorEnergy -0.064177   0.069460  -0.9239 0.3555145
Type3:FactorEnergy  0.040901   0.067314   0.6076 0.5434393
sd.Why1             0.122698   2.390184   0.0513 0.9590594
sd.Why2            -0.520344   0.565624  -0.9199 0.3576010
sd.Why3             0.365998   0.926073   0.3952 0.6926839
sd.Remove1          0.812285   0.459095   1.7693 0.0768408 .
sd.Remove2          0.106834   1.711769   0.0624 0.9502353
sd.Remove3          1.111964   0.389890   2.8520 0.0043446 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Log-Likelihood: -1769.4

random coefficients
        Min.      1st Qu.      Median        Mean     3rd Qu. Max.
Why1    -Inf -0.176649988 -0.09389167 -0.09389167 -0.01113335  Inf
Why2    -Inf -0.557497914 -0.20653126 -0.20653126  0.14443540  Inf
Why3    -Inf  0.107758325  0.35462047  0.35462047  0.60148262  Inf
Remove1 -Inf -0.334816672  0.21306112  0.21306112  0.76093891  Inf
Remove2 -Inf -0.003537719  0.06852062  0.06852062  0.14057896  Inf
Remove3 -Inf -0.952701404 -0.20269294 -0.20269294  0.54731553  Inf
```

Mixed Logit Model including Interaction Energy + Trade

```
Call:
mlogit(formula = Choice ~ 0 + Constant + Type1 + Type2 + Type3 +
    Why1 + Why2 + Why3 + Act1 + Share1 + Share2 + Share3 + Remove1 +
    Remove2 + Remove3 + Trade1 + Trade2 + Trade3 + Trade1:FactorEnergy +
    Trade2:FactorEnergy + Trade3:FactorEnergy, data = DataLong,
    reflevel = "No", rpar = c(Why1 = "n", Why2 = "n", Why3 = "n",
        Remove1 = "n", Remove2 = "n", Remove3 = "n"), R = 125,
    Halton = NA, Panel = TRUE, Method = BFGS, Correlation = FALSE)

Frequencies of alternatives:
     No       A       B
0.09668 0.44336 0.45996

bfgs method
19 iterations, 0h:0m:28s
g'(-H)^-1g = 3.88E-07
gradient close to zero

Coefficients :
                     Estimate Std. Error  z-value  Pr(>|z|)
Constant            -1.448352   0.090106 -16.0739 < 2.2e-16 ***
Type1                0.486422   0.097284   5.0000 5.733e-07 ***
Type2                0.359729   0.085446   4.2100 2.554e-05 ***
Type3                0.033200   0.083929   0.3956 0.6924185
Why1                -0.096168   0.088871  -1.0821 0.2792025
Why2                -0.202648   0.080679  -2.5118 0.0120119 *
Why3                 0.363261   0.099193   3.6622 0.0002501 ***
Act1                -0.032349   0.045610  -0.7092 0.4781743
Share1              -0.043208   0.078457  -0.5507 0.5818211
Share2               0.253317   0.088582   2.8597 0.0042405 **
Share3              -0.165944   0.080963  -2.0496 0.0404024 *
Remove1              0.222180   0.096345   2.3061 0.0211055 *
Remove2              0.070416   0.079885   0.8815 0.3780664
Remove3             -0.221089   0.107710  -2.0526 0.0401080 *
Trade1              -0.596167   0.109298  -5.4545 4.910e-08 ***
Trade2              -0.060455   0.087779  -0.6887 0.4909995
Trade3               0.226174   0.088796   2.5471 0.0108615 *
Trade1:FactorEnergy -0.128467   0.076995  -1.6685 0.0952145 .
Trade2:FactorEnergy -0.027004   0.076549  -0.3528 0.7242674
Trade3:FactorEnergy  0.120949   0.073321   1.6496 0.0990271 .
sd.Why1              0.127385   2.349315   0.0542 0.9567583
sd.Why2             -0.498532   0.585755  -0.8511 0.3947173
sd.Why3              0.374868   0.902704   0.4153 0.6779422
sd.Remove1           0.787248   0.471960   1.6680 0.0953078 .
sd.Remove2           0.081605   1.749120   0.0467 0.9627882
sd.Remove3           1.184385   0.377452   3.1378 0.0017019 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Log-Likelihood: -1768.2

random coefficients
        Min.     1st Qu.      Median        Mean     3rd Qu. Max.
Why1    -Inf -0.18208781 -0.09616819 -0.09616819 -0.01024857  Inf
Why2    -Inf -0.53890297 -0.20264823 -0.20264823  0.13360651  Inf
Why3    -Inf  0.11041642  0.36326132  0.36326132  0.61610623  Inf
Remove1 -Inf -0.30881080  0.22218015  0.22218015  0.75317110  Inf
Remove2 -Inf  0.01537363  0.07041552  0.07041552  0.12545741  Inf
Remove3 -Inf -1.01994479 -0.22108932 -0.22108932  0.57776614  Inf
```

7.6 Appendix VI – Principle Component Analysis

Principle Component Analysis Privacy Statements (Page 72)

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,813
Bartlett's Test of Sphericity	Approx. Chi-Square	411,371
	df	10
	Sig.	,000

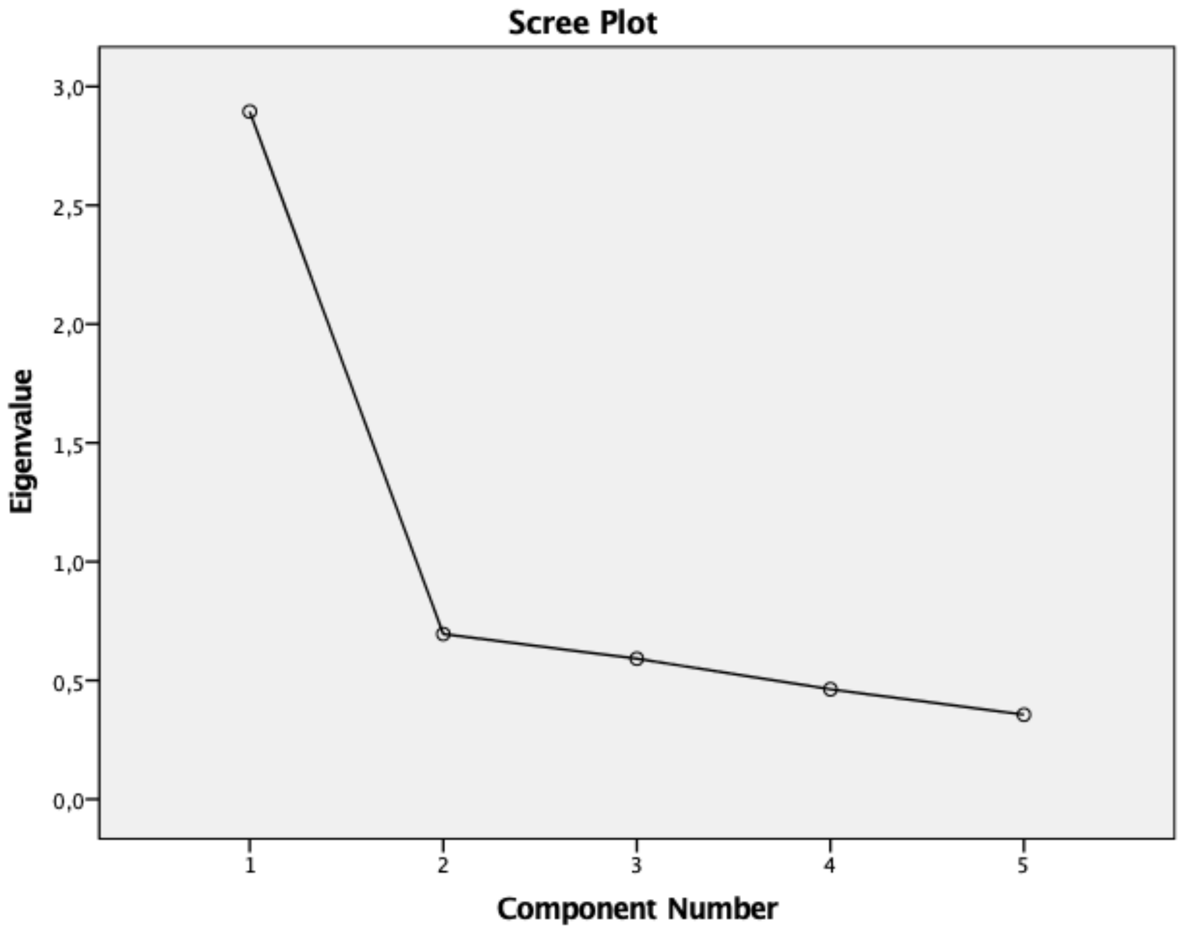
Communalities

	Initial	Extraction
Privacy1	1,000	,686
Privacy2	1,000	,541
Privacy3	1,000	,530
Privacy4	1,000	,558
Privacy5	1,000	,579

Extraction Method: Principal Component Analysis.

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	**Total**	**% of Variance**	**Cumulative %**	**Total**	**% of Variance**	**Cumulative %**
1	2,894	57,887	57,887	2,894	57,887	57,887
2	,695	13,909	71,797			
3	,592	11,830	83,627			
4	,463	9,257	92,884			
5	,356	7,116	100,000			



Principle Component Analysis Energy Statements (Page 72)

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,638
Bartlett's Test of Sphericity	Approx. Chi-Square	190,993
	df	3
	Sig.	,000

Communalities	Initial	Extraction
Energy1	1,000	,681
Energy2	1,000	,766
Energy3	1,000	,537

Extraction Method: Principal Component Analysis.

Total Variance Explained	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2,894	57,887	57,887	2,894	57,887	57,887
2	,695	13,909	71,797			
3	,592	11,830	83,627			
4	,463	9,257	92,884			
5	,356	7,116	100,000			

Extraction Method: Principal Component Analysis.