R.J. van den Boogert

Engaging the crowd in sensing for smart mobility

A discrete choice experiment investigating users' preferences in participatory sensing applications

TU Delft

AKKA

Source cover illustration: Pixabay, picture by Geralt

Engaging the crowd in sensing for smart mobility

A discrete choice experiment investigating users' preferences in participatory sensing applications

By

Ria Johanna van den Boogert

Student number: 4582535

in partial fulfilment of the requirements for the degree of

Master of Science
in Complex Systems Engineering and Management

at the Delft University of Technology,

to be defended publicly on Tuesday February 8, 2022.

Thesis committee:

Chair and second supervisor:	Prof.dr.ir. C.G. Chorus, TU Delft
First supervisor:	Dr. A.Y. Ding, TU Delft
Advisor:	T. Szép, MSc., TU Delft
External supervisor:	Dr. R. Kherrazi, AKKA Technologies

TU Delft

Preface

This thesis concludes my time at the Faculty of Technology, Policy, and Management at the Delft University of Technology. It has been a fascinating couple of years. Since my start at TPM, I have been intrigued by solving complex problems in society using an interdisciplinary point of view, and I am very grateful for all the opportunities I have been given to explore this field within my studies.

I wrote this thesis during a graduation internship at AKKA Technologies. The consultancy company was interested in understanding the motivations of users regarding data sharing, in order to improve the design of smart applications in the mobility sector. The project allowed me to combine my interest in choice behaviour with my enthusiasm for the evolving field of data and smart solutions. It is my hope that this research will be a valuable contribution to both science and society.

I could certainly not have done it alone. In these few words, I would like to take the opportunity to express my gratitude to some of the people who have surrounded and supported me throughout my studies.

During the process of writing this thesis, I had the pleasure of being guided by an amazing graduation committee. I would like to thank Caspar Chorus, my chair and second supervisor, for the excellent support and ideas you provided throughout the process. Your enthusiasm about choice modelling has been inspiring and motivated me to make the most out of this thesis. To Aaron Ding, my first supervisor, thank you for taking the time to think along with me during our weekly meetings. Thank you for all the feedback you provided and for encouraging me. To Teodóra Szép, my advisor, thank you for your availability and for providing the sharp advice I needed while setting up and conducting the choice experiment.

I want to thank the team of AKKA Netherlands for giving me the opportunity to graduate at their company and to see how the knowledge from my studies can be put into practice in a business environment. Special thanks go out to Rachid Kherrazi, who has been my external supervisor from AKKA. Thank you for being so closely involved with the project and supporting me in every step of the process. Furthermore, I want to thank everyone who helped me with the distribution of my survey and for providing me with the right contacts to interview. Also, many thanks go out to everyone who participated in the survey or in the interviews. Without them, this study would not have been a success.

Looking back at the past five and a half years, I feel blessed for all the family and friends I have been surrounded with during this period. To my parents, my grandparents, my sisters Hanna and Sara, and my boyfriend David, thank you for supporting me and believing in me. To Merel, Saskia, Katherina, Lidha, and Pauline, thank you for making my journey at TPM such a pleasant one. To Amila, thank you for providing me with tips and tricks and for proof-reading parts of my thesis. Last but not least, I want to thank my friends and roommates in Delft who gave me a wonderful time as a student and who kept my mind off the project now and then.

I hope you enjoy reading this thesis. With everything I have learned as a student, I look forward to making a positive impact in our society.

Mirjam van den Boogert
Delft, January 2022

Summary

In 2050, it is expected that 70% of the world's population will live in cities (Jin et al., 2014), leading to increasing congestion in and surrounding cities. This will raise new challenges, requiring more efficient and interactive cities. A novel paradigm contributing to these so-called smart cities is participatory sensing. Also known as mobile crowdsourcing, this solution enables both public and professional users to actively gather, analyse, and share local data about the urban environment using built-in sensors in smart devices (Truong et al., 2019). Considering that over 94% of the population has access to a mobile network, obtaining real-time data from these already existing sensors can be a low-cost solution for acquiring a huge amount of information (International Telecommunication Union, 2016). These real-time data can be used to analyse and predict mobility flows, and make public and private transport more efficient, safe, and sustainable.

However, a clear benefit is required to motivate smart device users to share data about their activities and their environment. Sharing data comes with the risk of disclosing private information, as location data can lead to the identification of living and work locations, as well as individual habits. Research on motivations of smart device users to engage in participatory sensing tasks is required in order to be able to design value-sensitive participatory sensing applications. This study aims to identify factors related to incentives and privacy that explain choice behaviour of users in participatory sensing applications. The main research question being addressed is as follows: *"How do factors relating to incentives and privacy affect the willingness among smart device users to contribute to participatory sensing systems for smart mobility?"*

A choice modelling approach was taken in order to identify the trade-offs made by users between potential benefits and costs of sharing data. This is an approach not often used before in the field of participatory sensing and provides novel insights in user behaviour in these systems. First, a literature review was conducted identifying possible factors relating to incentives and privacy, influencing the willingness of people to share data. Five factors were selected: monetary reward, effort, risk of re-identification, types of data, and data use. These factors were incorporated in a stated choice experiment distributed among smart device users through an online survey. In total, 125 valid responses were collected. The required effort of participating was regarded the most important factor influencing the willingness to share data in sensing applications for smart mobility. This provides new insights, as previous studies do not include effort in choice experiments regarding data sharing. As expected, the perceived ease of use declines if more inputs by the user are required. Moreover, respondents are reluctant to the collection of contextual and multimedia in addition to location and motion data, a finding which is confirmed by recent studies. Almost half of the respondents indicated to be highly concerned about their privacy. Therefore, a surprising finding is that the risk of re-identification was regarded the least important factor influencing the willingness to share data. However, when taking a deeper look at the data, it appears there is a group of people having extreme preferences regarding privacy and trust, who assign a higher importance to privacy related factors (risk of re-identification, types of data, data use).

The identified trade-offs were used to evaluate the implications for different use cases in the field of smart mobility. Three interviews were conducted, which each led to the definition of a use case, being crowd management in a city, safety research using car accident information, and real-time travel information in public transport. By aggregating the quantitative and qualitative parts of the research, it can be concluded that the accuracy of collected information can be improved by collecting more types of data in addition to location data. However, this will lead to a decline in the acceptance rate. A proposed solution is to provide tailor-made sensing applications, giving the user control to indicate which data they agree to share. Furthermore, the communication of the purpose of the data collection is important to users. Moreover, being transparent about the risks related to the data collection can help users to make a well-informed decision and will ensure an ethical design of sensing applications. Finally, increasing the attractiveness of the application is recommended to reduce the perceived effort, which could be done by gamification of sensing tasks.

Besides societal implications, this study provides several recommendations for future research. First, it is recommended to repeat the experiment without the provision of a financial compensation, in order to see if this leads to different choice behaviour and preferences of users. Furthermore, research on the understanding of privacy risks among users is recommended. Finally, this study can be used outside the smart mobility field, in order to analyse the willingness to share data in a broader sense.

Contents

List of figures and tables

List of figures

List of tables

1

Introduction

This chapter introduces the research problem and describes the objectives of the research. Subsection 1.1 gives a short description on the research background. In subsection 1,2, the problem statement is defined and the research objectives are described. Subsection 1.3 states the main research question, after which the research approach and the sub-research questions are defined in subsection 1.4. The subsequent subsection argues for the scientific as well as the societal relevance of the research. Subsection 1.6 explains the relation with the CoSEM programme. Finally, the structure of the thesis is presented in subsection 1.7.

1.1. Research background

As it is expected that 70% of the world's population will live in cities and surrounding regions by 2050, the need arises for more interactive and efficient cities (Jin et al., 2014). The development of these so-called smart cities is largely dependent on the availability of information and communication technologies that provide the sharing of knowledge about the city (Szabo et al., 2013). A major enabler of smart cities is the concept of Internet of Things (IoT). Interconnections between all kinds of digital devices generate large amounts and varieties of data, which are used to create applications delivering new services to citizens, companies, and public administrations (Zanella et al., 2014). Furthermore, decisions, actions and future planning can be based on this wealth of real-time information (Jin et al., 2014). IoT-enabled smart cities aim to improve the quality of life of their citizens in various ways, by creating smart solutions that contribute to sustainable, eco-friendly, and healthy cities (Kamel Boulos & Al-Shorbaji, 2014).

Participatory sensing is a recently emerged solution in the field of Internet of Things. This sensing paradigm, also known as mobile crowdsourcing, enables both public and professional users to gather, analyse, and share local data about the urban environment using built-in sensors and applications in smart mobile devices (Truong et al., 2019). Considering that over 94% of the population has access to a mobile network in 138 countries (International Telecommunication Union, 2016), obtaining real-time data from smartphones and other smart devices has a huge potential. Participatory sensing is seen as a valuable tool to engage the public in urban monitoring (Muller et al., 2015). These solutions are able to improve policy and decision-making, but also enable citizens to play an active role in their community (Rosa et al., 2020). However, the introduction of participatory sensing also poses significant challenges concerning the quality and the trustworthiness of collected data, as well as the privacy and security of users (Antoniou, 2017).

A field to which participatory sensing can be especially relevant is smart mobility. Every day, a huge number of people make use of traffic or public transport. It is predicted that congestion in and surrounding cities will increase (Rijksoverheid, 2019). There is a need to analyse traveller data that could contribute to solutions to improve transport systems. Challenges of new developments regarding smart transportation, smart charging, and Mobility as a Service (MaaS) ask for new ways to acquire large amounts of data. These data can be used to analyse and predict mobility flows and make public and private transport more efficient, safe, and sustainable. Therefore, this study specifically focuses on participatory sensing applications for smart mobility.

1.2. Problem statement and research objectives

Gathering more accurate knowledge about the urban environment increases possibilities for designing effective solutions enhancing people's wellbeing (Issarny et al., 2018). A key challenge in unlocking this potential of participatory sensing systems is to identify robust incentives that ensure participation of individuals (Ogie, 2016; Riahi et al., 2017). However, sharing data comes with the risk of disclosing private information. Sensing measurements might be tagged with location information (Du et al., 2019). Also, data might be gathered on health behaviours or symptoms, moods, eating or sleep (Shilton & Estrin, 2012). For users, there should be a clear benefit in order to encourage them to share their personal data about their activities and environment (Salim & Haque, 2015). Incentive mechanisms can motivate users to participate, but also require a quantification of privacy (Bennati, Dusparic, Shinde, & Jonker, 2018).

The best way to incentivize users remains unclear. A general study on the motivations of volunteers to engage in participatory sensing tasks and the effectiveness of participatory sensing incentives across different contexts is still lacking (Restuccia et al., 2016). Furthermore, the specific privacy concerns of users, which can be linked to their different characteristics, have to be further researched (Kong et al., 2019).

This research aims to fill these knowledge gap by exploring which factors related to privacy and incentives affect the willingness of smart device users to contribute to participatory sensing systems. Explaining choice behaviour regarding sharing data in these systems will contribute to developing more effective and value-based participatory sensing applications, which means applications are designed taking into account the values as perceived by users.

1.3. Main research question

It appears that there is a lack of understanding of which incentives stimulate individuals to contribute to participatory sensing systems. It is unclear how participants in a such systems value privacy and how they make a trade-off between privacy and incentives that motivate them to participate, such as a reward. The main research question that is addressed in this research is as follows:

How do factors relating to incentives and privacy affect the willingness among smart device users to contribute to participatory sensing systems for smart mobility?

1.4. Research approach and sub-research questions

As described, there is a lack of understanding in the preferences of potential contributors to participatory sensing systems. Therefore, a choice modelling approach is suitable to address the knowledge gap. A discrete choice model can be used to describe the choices of decision makers between different alternatives (Train, 2009). First, this approach is used to understand user behaviour and explore what factors determine choice behaviour. Second, it is suitable for identifying what individuals will choose, predicting market demand and facilitating optimal design (Chorus & Van Cranenburgh, 2020). According to Johansson et al. (2021), discrete choice modelling is a suitable method for innovative privacy research.

This research aims to determine what factors influence the privacy perception of smart device users regarding participatory sensing systems, and how this affects their choice whether or not to contribute to the system. Besides that, the choice model is used to identify the trade-off that these individuals make between potential costs and benefits of participating in sensing applications. Moreover, this research explores factors to be taken into account when designing incentive schemes and facilitates the design of value-sensitive participatory sensing systems.

The main research question will be answered by addressing a set of sub-questions (SQ's).

1. What factors potentially incentivize or disincentivize individuals to contribute to participatory sensing systems?
2. What factors relating to privacy potentially influence the decision of individuals to share data in participatory sensing systems?
3. What trade-offs do individuals make between potential costs and benefits of participation, when choosing to share data in participatory sensing systems?
4. Regarding these trade-offs, what are implications for different applications in the field of smart mobility?

The first two sub-questions are answered by conducting a literature research, in order to identify factors that potentially influence the willingness of individuals to contribute to participatory sensing systems. These factors are used as an input for sub-questions 3 and 4. The most relevant factors are included in a choice experiment design. Choice modelling is used to measure the influence of factors related to incentives and privacy on the willingness to contribute to participatory sensing systems. An elaboration on the methodology can be found in section 3.

In order to answer sub-question 4, a qualitative approach is used by conducting interviews with relevant parties in the field of smart mobility. Furthermore, this question aims to aggregate the outcomes of the quantitative and the qualitative part of the research. Since this question cannot be answered directly, 2 sub-sub-questions are used:

4a. What use cases can be defined for which participatory sensing can be relevant?
4b. What is the acceptance of smart device users for these use cases, according to the results?

Question 4a is answered by conducting interviews. Question 4b uses the results of both sub-question 3 and sub-question 4a as an input, predicting the acceptance of the use cases. Lastly, the insights from both the qualitative and quantitative research in order to describe principles that can be used for designing value-sensitive participatory sensing applications. These recommendations are described in chapter 9.

1.5. Scientific and societal relevance

Little research has been conducted on the trade-off between benefits and costs in participatory sensing applications. Also, there is limited knowledge on the user side of participatory sensing. Thus, this thesis can expand this field of research by investigating factors that influence the decision of individuals regarding data sharing. Furthermore, a choice experiment is a method not often applied to the topic of participatory sensing. Usually, this research method is used in the field of travel behaviour and health care. Therefore, this thesis can provide insights on the applicability of this method to other disciplines, such as research on digital innovations.

Besides the scientifical relevance of this research, the study will also have a societal contribution. This research will deliver insights in how users perceive privacy, and in the trade-off they make between perceived benefits and costs of sharing data. These insights can enable organisations initiating sensing campaigns to create a better design of participatory sensing applications by implementing a right balance between benefits, risks, and required effort. This will contribute to the attractiveness of such applications and might incentivize more users to participate. Eventually, this can contribute to the analysis, prediction, and optimisation of mobility flows and enhance developments in the field of smart mobility.

Furthermore, this study contributes to the societal debate regarding data sharing and privacy. Insights in user perceptions can help organisations collecting data to make ethical choices, keeping the user and their preferences in mind. This can lead to the development of useful IoT-based applications while protecting values like trust and privacy.

1.6. Relation to CoSEM program

In this study, several issues need to be considered. On one hand, the technical and data quality aspects of participatory sensing systems need to be addressed. On the other hand, incentive mechanisms to engage the public in urban monitoring require research (Ekman & Weilenmann, 2021).

This asks for a multidisciplinary approach, taking into account technical, institutional, and social perspectives. This makes this research typical for a CoSEM (Complex Systems Engineering and Management) thesis, which is aimed at designing solutions for large and complex socio-technical systems. This research focuses on identifying factors for a value-sensitive design approach towards the development of IoT-based applications, taking into account values related to trust and privacy. Looking beyond the technical design of participatory sensing applications and concentrating on what is needed to implement these systems by investigating incentives and behaviour, again stresses the socio-technical character of this research.

1.7. Thesis layout

The research includes an Experiment Design Phase, in which the choice scenarios are selected, a Choice Modelling Phase, in which the model estimation takes place and the results of the experiments are analysed, and an Application Phase, in which implications for different use cases are described.

The thesis is structured as follows. Chapter 2 provides a general background on the research field of participatory sensing. Also, it describes the scope and the concepts of the research. Chapter 3 addresses the methodologies that are used for conducting the research. In chapter 4, a literature review is conducted in order to explore factors potentially influencing the decision of individuals to share data. These factors are used as an input for the choice experiment design, which is constructed in chapter 5. Chapter 6 elaborates on the model estimation as well as the interpretation of the results. In chapter 7, interviews are conducted aiming to define use cases for which participatory sensing applications can be relevant. Chapter 8 provides a discussion of the research, addressing the limitations. Finally, conclusions and recommendations for further research are described in chapter.

The design of the research is displayed in Figure 1. On the left side, the chapters in this thesis are displayed. Sub-question 1 and sub-question 2 are answered in chapter 4, sub-question 3 in chapter 6, and sub-question 4 in chapter 7. In chapter 9, the main research question is addressed.

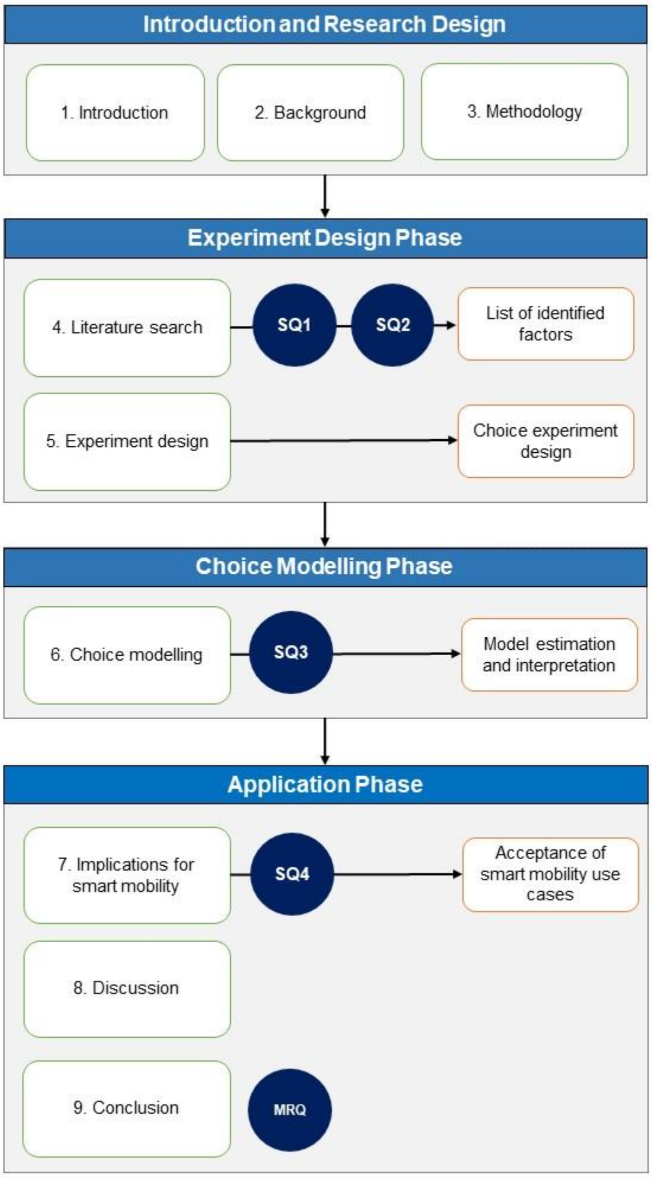


Figure 1. Thesis structure

2

Background

In this section, a general background is provided on participatory sensing and the applications for smart mobility. Subsection 2.1 introduces the concept of participatory sensing and frames it in the broader research field of Internet of Things. In subsection 2.2, the relevance of participatory sensing applications for smart mobility is discussed. Subsection 2.3 addresses the core challenges related to trust, privacy, and incentives for participatory sensing. Finally, a theoretical framework is established in subsection 2.4, based on Privacy Calculus theory.

2.1. Sensing and the smart city

2.1.1. Internet of Things

The Internet of Things (IoT) is seen as an essential aspect of the smart city of the future (Cottrill et al., 2020). The Internet of Things is defined by Patel & Patel (2016) as “a network of physical objects, interacting and cooperating with each other through wired and wireless connections in order to create new applications and services”. Devices of all types and sizes, such as vehicles, smartphones, home appliances, cameras, medical instruments can communicate and share information and make energy, transport, cities and many other areas more intelligent.

2.2.2. Smart city

A smart city is “a system of interconnected smart systems in order to express full potential of ICT and specifically IoT” (Weber, 2017). IoT-enabled smart cities aim to improve the quality of life of their citizens in various ways, by creating smart solutions that contribute to sustainable, eco-friendly, and healthy cities (Kamel Boulos & Al-Shorbaji, 2014).

According to Du et al. (2019), a smart city can be defined as an “urban area that uses the information that is collected by various types of sensors and devices to monitor and manage its infrastructures and its resources efficiently”. Cities can be seen as a cyber physical system (CPS), in which the city’s physical status is monitored continuously through cyber components like sensors and processors (Puliafito et al., 2021). Smart monitoring systems should provide better connections between citizens and services. This should be done by enabling citizens to participate in sensing, and being more open regarding data, policies, and government (Du et al., 2019).

In order to increase the efficiency of city management, there is a need for live monitoring of urban process parameters (Jin et al., 2014). Current data collection exercises often take a lot of effort and money. Thus, municipalities are looking for ways to collect the required data and analyse them in real time, by incorporating smart technologies. IoT infrastructures, including data processing and management, actuation, and analytics, can enhance this process by gathering and evaluating data in real time, extracting information, and converting it into useful knowledge. This knowledge will improve and inform the decision making of both city management and citizens (Jin et al., 2014).

Figure 2 presents the four-layer model of a smart city as proposed by Bawany & Shamsi (2015) and Puliafito et al. (2021). The foundation of the smart city is the Information and Communication Technologies (ICT) infrastructure, which include sensors in public and private devices, such as traffic lights, bus GPS, lamp posts, air pollution and weather stations, and citizen’s mobile devices. The *infrastructure* layer provides the possibilities for collecting, managing, and elaborating data (Puliafito et al., 2021). The *management* layer formulates policies, rules, and legislations regarding, among other things, privacy, credit-reward systems, and incentive mechanisms. Building upon these first two layers, a variety of services and *applications* can be offered, in contexts like mobility, waste, public safety, energy, and e-health (Bawany & Shamsi, 2015; Puliafito et al., 2021). Finally, the *stakeholder* layer includes all the parties involved in the smart city, such as the municipality, enterprises, telecommunication operators, citizens, and vehicles.

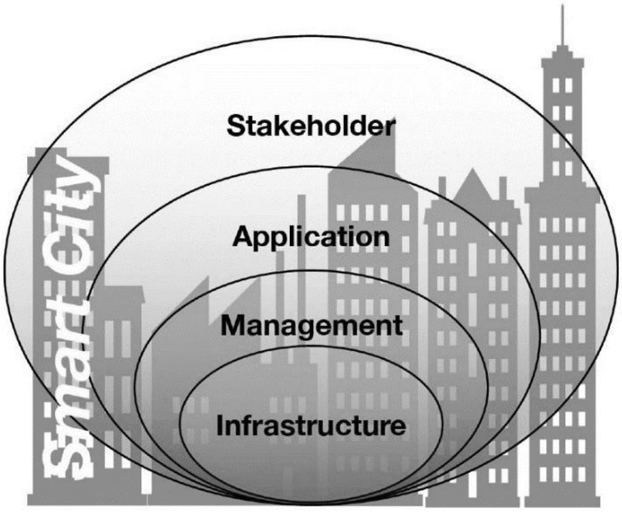


Figure 2. Four-layer model smart city

2.1.3. Sensing paradigms

Common sensing paradigms for data collection are RFID and WSN. RFID (Radio-frequency Identification) technology is mainly used in applications in retail and supply chain management. This technology automatically identifies objects to which RFID tags are attached. WSN's (Wireless Sensor Networks) are networks of sensors that collect, process, analyse, and disseminate valuable information gathered in various environments (Jin et al., 2014). WSN's can, for example, monitor and control climate conditions in smart buildings, provide drivers with information for better route planning, congestion avoidance, and safer driving, and monitor the structural health of buildings. Thus, these sensor network based systems form an essential component in building a smart city (Du et al., 2019).

Besides these two paradigms, a recently emerged sensing paradigm is *participatory sensing.* This paradigm comes from the idea of outsourcing sensing tasks to the crowd (Du et al., 2019). Instead of deployed sensors, people function as sensors, collecting and sharing sensory data. Data collected by these users is a low-cost solution and complements data from fixed infrastructures (Jin et al., 2014).

2.1.4. Participatory sensing

Burke et al. (2006) are one of the first introducing the concept of participatory sensing. Here, participatory sensing is defined as the tasking of mobile devices to form interactive, participatory sensor networks. This sensing paradigm, also known as mobile crowdsourcing, enables both public and professional users to gather, analyse, and share local data about people and the urban environment using built-in sensors and applications in smart mobile devices (Truong et al., 2019). Mobile crowd sourcing technologies makes use of distributed IoT devices in order to collect data and extract meaningful information for further use. By involving people in sensing processes, participatory sensing is a novel way to extend IoT services (Wibisono & Ahmad, 2017). Smartphones nowadays contain a lot of sensors, like GPS, camera, ambient light, accelerometers, compass and microphones. These built-in sensors can provide a lot of data in any location in a city (Du et al., 2019). Besides smartphones, other devices can potentially function as sensor nodes. Examples are wearable devices like smart watches or glasses, and autonomous driving vehicles. By using smart devices for sensing tasks in addition to existing sensing networks, smart city monitoring systems can be improved in terms of accuracy and spatial-temporal granularity (Du et al., 2019).

Different technologies can be used in the field of participatory sensing. Solutions to support large and heterogeneous networks are LTE and 5G. These technologies support a larger network size, and also improve the real time monitoring performance by enabling sensor nodes with a higher data rate. Other upcoming standards for sensor nodes are NarrowBand-Iot (NB-IoT), LoRaWAN, and IEEE 802.11ah, which allow sensor nodes to run more sustainably. Lastly, a fog computing architecture can help smart city monitoring, providing a better coverage and reducing service latency and response time (Du et al., 2019).

Participatory sensing has huge benefits compared to fixed sensor infrastructures. Since already existing, built-in sensors are used, deployment and maintenance costs are low (Heiskala et al., 2016). Also, the coverage of the sensor network can be high, since devices owned by individuals are used. Furthermore, user opinions and actions can complement sensing data gathered by fixed infrastructures (Heiskala et al., 2016).

Figure 3 provides a visualisation of participatory sensing applications.

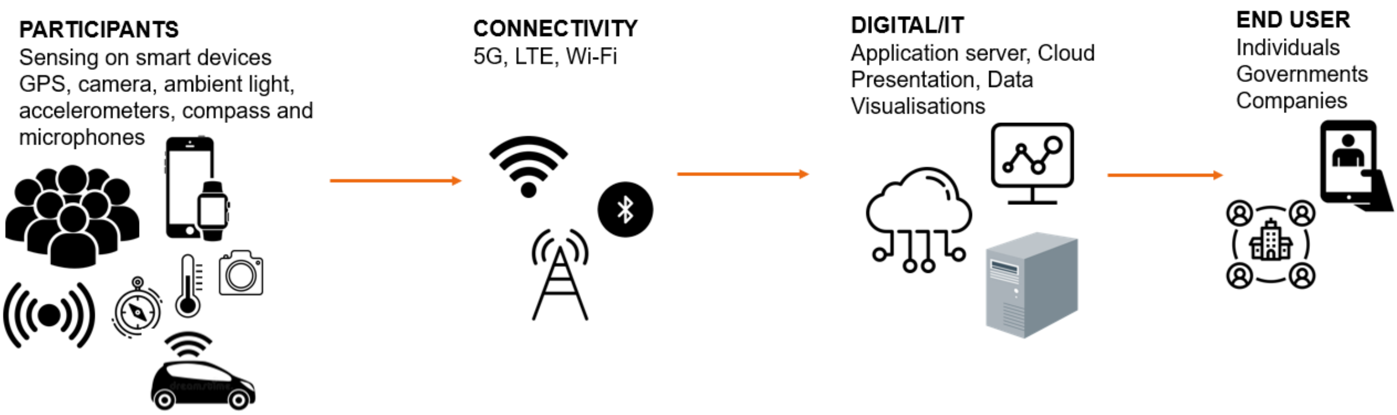


Figure 3. Visualisation of participatory sensing applications

A distinction is made in this field between *participatory sensing* and *opportunistic sensing.* In participatory sensing, users actively participate in submitting data to meet requests by the application. In opportunistic sensing, the application may run in the background and the active involvement of users is minimal (Ma et al., 2015; Salim & Haque, 2015). A benefit of opportunistic sensing, compared to participatory sensing, is that it places a lower burden or cost on the user, since there is no need for selecting and sampling data manually (Lane et al., 2010).

The fields to which participatory sensing can be of relevance are extensive. Kamilaris & Pitsillides (2016) describe applications of participatory sensing for measuring air or noise pollution, health, mobility, and public safety. This study specifically focuses on applications in the field of smart mobility, which is explained in the next subsection.

2.2. Participatory sensing for smart mobility

Considering that 50% of the world's population is currently living in cities and it is estimated that this will rise with an additional 20% by 2050, there will be new requirements for urban infrastructure and transport to become smarter (Paalosmaa & Shafie-Khah, 2021). This subsection discusses the challenges that are arising in the field of smart mobility, as well as the potential relevance of participatory sensing in this field.

2.2.1. Challenges in the field of smart mobility

In this subsection, three current topics that are challenging in the field of smart mobility are described.

Smart transportation

Current challenges related to transport include congestion, pollution, accidents, noise, and scarceness of space (Paalosmaa & Shafie-Khah, 2021; Ribeiro et al., 2021). These issues directly affect the quality of mobility services and the accessibility of a city. They can lead to increased delays, energy expenditure, and pollution (Ribeiro et al., 2021).

In order to achieve a better planning of traffic and infrastructure, urban mobility governance and the collection of real-time data is required. Besides these technical issues, social issues are to be considered, such as improving traffic safety and attractiveness, reducing environmental impacts, and enhancing information management and decision-making (Paalosmaa & Shafie-Khah, 2021).

Smart transport aims to look for new ways of transport, such as e-mobility systems, self-driving cars, continuous mobility chains and mobility services, that are efficient, user-friendly, and cost-effective (Paalosmaa & Shafie-Khah, 2021). To realise this and make transportation more automated, acquiring relevant and up-to-date data and information is essential. Thus, there is a need for development of sensing systems in transport and mobility, in order to observe all types of movements of vehicles and people. These data can support decision-making regarding mobility operation, noise, and air quality. Eventually, digitization of transport can deliver fully personalised services and commercial offers to users (Ribeiro et al., 2021).

Smart charging
Another development in the field of mobility concerns the fast rising number of electric vehicles (EV's). This increases the demand for a more reliable and comprehensive charging infrastructure. If charging is not well managed, a large number of EV's can cause severe peak loads to the power grid, since there is an increased power demand. This has significant negative impacts on the power quality and can even lead to disruption to the stability of the energy system (Paalosmaa & Shafie-Khah, 2021).

In order to handle the increasing demand for electricity, there is a need for monitoring and controlling charging. In that way, the energy consumption effect to the grid can be optimized. This is called *smart charging*, which means an EV and a charging device are connected and exchanging data (Paalosmaa & Shafie-Khah, 2021).

However, the monitor and control aspects of charging still need to be investigated. Furthermore, few knowledge is available on how to predict EV-bookings (Lopez-Carreiro et al., 2021).

Mobility as a Service (MaaS)
The concept of Mobility as a Service (MaaS) is meant to provide an attractive and more sustainable alternative for private transport. The goal is to join public and business sectors to make public transport more attractive, providing users with an unbreakable mobility chain (Paalosmaa & Shafie-Khah, 2021). This is done in several ways. First, MaaS integrates all available modes of both public and private transport into one bundle of services. It expects users not to buy the transport modes itself, but mobility services. Besides that, it aims for user-centricity by providing users with tailored and on-demand mobility solutions based on their needs, preferences, and habits. Lastly, MaaS relies on a digital platform that integrates end-to-end trip planning, booking, electronic ticketing, payment services, and real-time travel information (Lopez-Carreiro et al., 2021).

2.2.2. Relevance of participatory sensing for smart mobility

As described in the previous subsection, new developments in the field of smart mobility ask for the collection of real-time data. Related to this, Shit (2020) argues for the relevance of crowdsensing methods for the realisation of intelligent transport systems. These methods improve environmental sensing by involving anyone in the sensing process. Within the crowdsensing process, three stages can be distinguished. First, data is acquired by sensors from vehicles, devices, buildings, and human activities. Then, all the collected data is converted into a standard form. Finally, the data is analysed by using artificial intelligence techniques (Shit, 2020).

Mobility data can be described as data about individuals that include their locations at specific times (MobiDataLab, 2021). The advantages of collecting these data using participatory sensing are huge. For example, information collected by the crowd can address navigation issues in cities. Existing navigation apps provide the best route, but cannot provide information on road conditions, and weather and traffic parameters. Using the driver's feedback can improve the navigation system design. Besides showing real-time information, historical data can be used to make predictions on future situations, such as road congestion prediction (Shit, 2020).

However, also risks are involved in sharing mobility data, relating to privacy. Mobility data have the characteristic of being unique. This means that the data of different individuals is easily differentiable, since the starting and ending of their trajectories are often working and home locations. These are highly unique, and can therefore lead to re-identification (MobiDataLab, 2021). This risk of re-identification is further discussed in 2.3 and chapter 4.

2.2.3. Crowdsensing applications for smart mobility

As stated before, participatory sensing is a promising approach for applications in the field of transportation. Several applications are proposed in literature for using data sensed by the crowd in order to improve mobility. An example of this is the real-time monitoring of road surface quality (Mednis, 2013). For example, Nericell is a system for rich monitoring of road and traffic conditions, making use of sensors on smartphones (Mohan et al., 2008).

Furthermore, participatory sensing can contribute to smoother and more sustainable public transport (Heiskala et al., 2016). Since travellers carry smartphones with them everywhere, there is a high potential in collecting real-time traffic information. Using data collected from applications running on travellers' smartphones, can provide better predictions of traffic flows and of the traveller situation. Also, public transit route schedule predictions can be improved. Furthermore, the application can give user reports of incidents or crowding on a train. With this information, transportation operators get more insight in transport behaviours and are enabled to predict travel demand (Shit, 2020). This can help improving the efficiency of fleet management and operations. Operators will also be able to deliver more personalized services to travellers, which may improve travelling experiences for passengers. Another application is objectively measuring the quality of passengers' trips based on wait/travel time, distance and bumpiness of ride (Xiao et al., 2018), in order to improve services.

For transportation authorities, these data can also be highly valuable. By gaining more insight in the current traffic system, they can improve transport policies, such as optimizing congestion charges, taxation, and subsidies. This can contribute to smoother and more sustainable transportation (Heiskala et al., 2016).

Beside using smartphones carried by drivers or passengers, vehicles itself are increasingly equipped with on-board sensors and in-vehicle information systems. Thus, vehicles can be effective data collection systems for participatory sensing. An example of such an intelligent transportation application is the prediction of *Distance-to-empty* (DTE), which is the distance an electric vehicle (EV) can drive before running out of fuel (Tseng & Chau, 2017). The DTE depends on various factors including driving behaviour, road conditions, traffic, and vehicle specifications. Participatory sensing has potential to improve the accuracy of DTE prediction since it enables exploiting data from other drivers. This can help individual drivers to adjust their driving behaviour. Moreover, participatory sensing systems can be used to get a more comprehensive understanding of the nature of traffic and to explore methods of reducing congestion in urban areas and on highways (Yi et al., 2017).

Another application of participatory sensing is outdoor air quality monitoring (Weber, 2017). An example of this is P-Sense (Pollution-Sense), which is a system for air quality monitoring and control (Mendez et al., 2011). This system aims to give various parties access to pollution data and address their particular problems and needs. For government officials, it is possible to monitor and control the Air Quality Index of a city. Also, doctors can correlate respiratory problems of their patients to air quality and county officials or realtors are able to determine the best place for a new building using data. Moreover, the system delivers benefits for the user since it is possible to assess the exposure to pollution according to places visited by the user (Mendez et al., 2011). Based on these data, it can give advice to users for the healthiest route in terms of air pollution.

2.2.4. User preferences for sharing mobility data

When optimising transportation and traffic management systems, travel times of individuals can be decreased. Moreover, it can make transportation more safe and more sustainable (Shit, 2020). Yet, according to Ribeiro et al. (2021), a crucial and little addressed element regarding the digitization of transport is the impact of users and their readiness to get involved in these new opportunities. Users have to understand the process of the collection and treatment of data, as well as the production of information. When they put their trust in data governance, they will be more open to share their data (Ribeiro et al., 2021). However, in order to get users involved in participatory sensing systems for smart mobility, their preferences need to be understood. The core concepts relating to these user preferences are discussed in the next subsection.

2.3. Core concepts

Major challenges in participatory sensing systems include trust issues, privacy issues, and the provision of appropriate incentives. As recognized by Riahi et al. (2017), these three challenges are interlinked. By anonymizing a user's identity in order to protect privacy, implementing an effective trust mechanism becomes more difficult. Furthermore, users could be incentivized to sacrifice (part of) their privacy in order to improve trust mechanisms (Riahi et al., 2017). Taking into account the interdependencies between these concepts, these three challenges are discussed in the following subsections.

2.3.1. Trust in participatory sensing

An important question that is raised is how participatory sensing incentive mechanisms can encourage the collection of high quality data. In this context, Jaimes et al. (2015) describe the use of reputation schemes. Users can be ranked based on past performances, assessments of peers, or by a combination of both. This is supported by research based on pilot experiences by Kotovirta et al. (2012), that concludes that more weight should be put on reports of trustworthy users. According to Mousa et al. (2015), reputation-based trust systems can provide a guarantee towards the accountability of a user. However, an open challenge is how to manage the accountability of participants while preserving their privacy (Mousa et al., 2015), which is discussed in more detail in the next paragraph.

2.3.2. Privacy in participatory sensing

Several studies highlight the importance of studying how to incorporate privacy-preserving mechanisms into the design of participatory sensing incentives (Kotovirta et al., 2012; Ma et al., 2015; Ogie, 2016). Data captured by participatory sensing systems can reveal identity, based on location data or other data attributes (Riahi et al., 2017). Thus, a trade-off needs to be made between privacy protection and the quality of data (Jaimes et al., 2015). Mousa et al. (2015) agree that how to assure the compromise between these conflicting goals requires investigation.

Several laws and regulations aim to minimize the risk of leaking information that relates to an individual who can be directly or indirectly defined. According to GDPR principles, users should give unambiguous consent before parties are allowed to process their data. Furthermore, people have the right to be informed, the right of access, right of rectification, right to erasure, right to restrict processing, right to data portability, right to object, and rights in relation to automated decision making and profiling . Nevertheless, users perceive that they have lost control over information shared in an online context (Schomakers et al., 2020). This relates to information privacy, which means that individuals should be able to exercise a substantial degree of control over their data and its use (Clarke, 1999).

Specifically, regarding participatory sensing applications, a major concern with respect to privacy is maintaining user-level control over sensitive sensor data (Christin et al., 2011). In Christin et al. (2011), privacy in participatory sensing is defined as follows:

> "Privacy in participatory sensing is the guarantee that participants maintain control over the release of their sensitive information. This includes the protection of information that can be inferred from both the sensor readings themselves as well as from the interaction of the users with the participatory sensing system."

All parties in a sensing application can be protected from external parties by using SSL/TLS, which enables secure communications between any two parties. However, there is a risk of leakage of personal information to internal adversaries. If a Service Provider collects all data over a longer period, it might learn a great amount of sensitive information and violate their privacy in terms of movements, habits, and more (Cristofaro, 2014). Therefore, users should be able to control what data is shared exactly, and who will receive or use their data.

Clarke (1999) suggested there are different types of privacy that can be distinguished. These are privacy of the person, privacy of personal behaviour, privacy of personal communications, and privacy of personal data. However, rapid technological advances required an extension of the definition of privacy. Finn et al. (2013) added additional types of privacy and argues for seven different types of privacy, being privacy of the person, privacy of behaviour and action, privacy of personal communication, privacy of data and image, privacy of thoughts and feelings, privacy of location and space and privacy of association. Participatory sensing applications mainly affect privacy of behaviour and action, data and image, and location and space. Privacy of behaviour and action can be violated by identifying travel activities, for example (Derikx et al., 2016). The privacy of data and image includes ensuring that data of individuals is not automatically available to other individuals and organisations, as well as giving people a certain degree of control over their data (Finn et al., 2013). Privacy of location and space is mainly impacted by technologies that can identify, track, or monitor individuals while moving through public or semi-public spaces.

Something that has yet to be researched is how the way users value privacy changes with different sensing tasks or with different incentives (Ogie, 2016). Participants may trade privacy in return for some benefit. Insights are needed on effective mechanisms that allow for these privacy-trust negotiations (Riahi et al., 2017). According to work by Bennati et al. (2018), incentive mechanisms can help to increase user participation, but quantifying privacy is an issue that still needs to be addressed. Similarly, Khoi & Casteleyn (2018) recommend further research on behaviour of users when factors like length of sensing activities, type of incentive, and the public's feedback on their collected data vary.

2.3.3. Incentives for participatory sensing

A key challenge in participatory sensing systems is to identify robust incentives that ensure participation of individuals (Ogie, 2016; Riahi et al., 2017). Ogie (2016) identifies several types of incentives for participatory sensing, distinguishing monetary and non-monetary incentives, which are displayed in Figure 4.

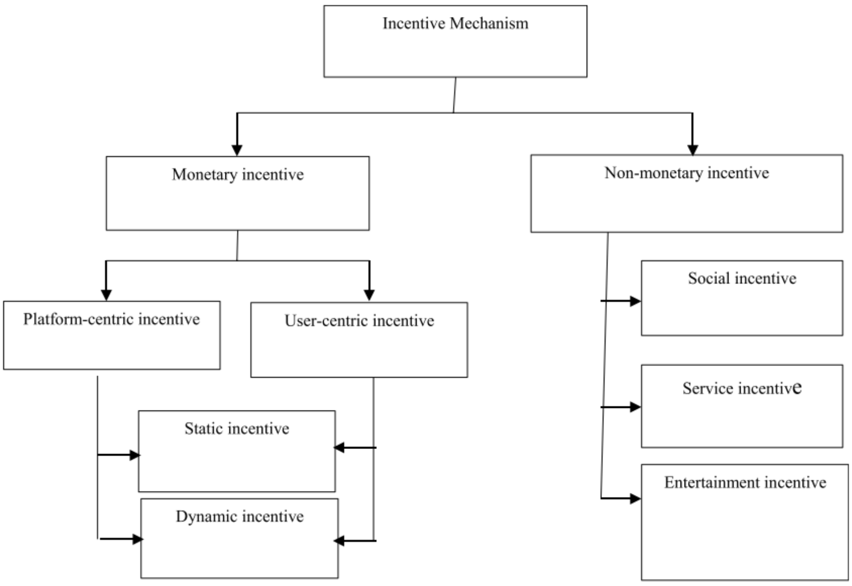


Figure 4. Categorisation of incentives (Ogie, 2016)

Money is often mentioned as a useful incentive to motivate users to engage in participatory sensing. However, how the amount of money that each participant expects to receive can be estimated is something that yet needs to be researched (Ogie, 2016). The work by Jaimes et al. (2015) supports this notion by stating that functions need to be proposed to estimate the cost for participation, based on a sum of factors such as energy, resource consumption, and privacy. In contrast, Kotovirta et al. (2012) state that incentives do not necessarily need to be monetary. According to Riahi et al. (2017), existing schemes are often not based on financial rewards, since these may clash with the participant's intrinsic motivation (Kostakos et al., 2017). Other ways to motivate users are sensing tasks for social or ethical reasons, contributing data in turn for service usage, or interestingness and enjoyment (Kotovirta et al., 2012; Ogie, 2016). An idea that is mentioned by Gao et al. (2015) is the co-existence of multiple incentive schemes. How to determine the appropriate incentive reward for each participant in order to provide flexible and personalized incentive schemes is a challenge to be researched.

Several approaches have been discussed in literature in order to design effective incentive mechanisms, mainly relying on purely game-theoretic approaches. However, users taking part in participatory sensing systems might have unequal resources and exhibit different behaviour. Therefore, according to Riahi et al. (2017), modelling approaches could complement these studies by exploring the design space of user behaviour.

2.4. Conceptual framework: Privacy Calculus Theory

From a utilitarian point of view, privacy can be seen as an interest rather than an absolute right (Clarke, 1999). This implies that privacy can be traded and users can be persuaded to participate if the benefits of a service overrun the related sacrifices (Derikx et al., 2016). Behaviours regarding privacy are a result of situational and context-specific cost-benefit analysis of information disclosure (Li et al., 2010).

A theory relevant in the context of privacy perceptions and behaviour of consumers is Privacy Calculus Theory, a model first proposed by Laufer & Wolfe (1977). According to this theory, individuals are more likely to disclose personal information if the benefits exceed the costs of data sharing (Wang et al., 2016). Before making a decision whether to provide information, individuals weigh the risks and benefits to assess the outcomes, and react accordingly (Dinev & Hart, 2006). Privacy risk is related to the expected loss of personal information to external parties or loss of control over personal information. Benefits of information disclosure include financial rewards, personalization, and social adjustment benefits (Smith et al., 2011).

Majumdar & Bose (2016) apply Privacy Calculus Theory to Internet of Things (IoT) services. The advantages of implementing such technologies are huge. However, this also means sacrificing consumer's privacy to some extent. Thus, a trade-off is made between risks and benefits in this context that provides the background to use the Privacy Calculus Theory. Perceived benefits of IoT are the opportunities it provides for real-time decision-making with collected data. On the other hand, over-tracking can lead to privacy intrusion and there is a risk of unauthorised access of data (Majumdar & Bose, 2016). In previous research, Privacy Calculus Theory has mainly been applied in the context of individuals' self-disclosure on social networks or on websites. However, Privacy Calculus Theory as the theoretical basis in the context of IoT applications has been limited (Princi & Krämer, 2019).

A conceptual model that applies the Privacy Calculus Theory to the context of participatory sensing systems is presented in Figure 5.

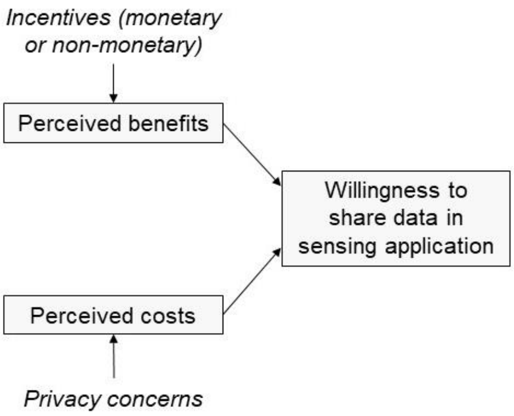


Figure 5. Conceptual Model Privacy Calculus

Perceived benefits in the benefit-risk trade-off are monetary or non-monetary incentives, that motivate individuals to engage in participatory sensing systems. Perceived costs lie in the fear that location data or other personal data will be disclosed. Individuals are assumed to weigh these incentives and privacy concerns and make a decision on whether or not to participate according to this trade-off. This trade-off is investigated in this research by conducting a discrete choice experiment. Privacy Calculus Theory is used in previous research by Potoglou et al. (2017) to inform choice experiments. The Privacy Calculus framework provides the theoretical background for the research and is used to guide the choice experiment. Chapter 3 further describes the methodology that is used to construct and analyse the choice experiment.

3
Methodology

This section addresses the methodology of the research. The first subsection gives a general introduction on choice modelling. In subsection 3.2, the Experiment Design Phase is discussed, after which the methods used in the Choice Modelling Phase are described in subsection 3.3. In subsection 3.4, the methods for applying the model to specific use cases are addressed.

3.1. Introduction to choice modelling approach

Since there is a lack of understanding in the preferences of contributors to participatory sensing systems, a choice modelling approach is used. A discrete choice model is used to describe the choices of decision-makers between different alternatives (Train, 2003). This research focuses on explaining choice behaviour in terms of underlying factors.

Choice modelling is an approach that is not usually taken in the field of participatory sensing. Therefore, this research can deliver new insights that can facilitate the improvement of participatory sensing systems. The underlying theory on which the model estimation in our research is based is Random Utility Maximization theory (Train, 2009). In this study, several Multinomial Logit (MNL) models are estimated, which is a well-known model in the field of choice modelling. In order to capture heterogeneity caused by correlation between repeated choices of individuals, an Mixed Logit (ML) panel model is estimated as well. The following subsections elaborate on these models.

3.2. Experiment Design Phase

The experiment design phase consists of two aspects. First, a literature review is conducted to identify potential factors to include in the choice experiment. Subsequently, the experiment design is constructed.

3.2.1. Identification of factors

First, a systematic literature review is conducted in order to explore factors that potentially influence participation in participatory sensing systems. These factors can be used as *attributes* in the choice experiment, which are characteristics of choice alternatives. Furthermore, suitable *attribute levels* need to be defined, which are the particular values of attributes in the choice alternatives.

A careful selection process is required in order to determine what attributes to include in the choice experiment, since the Discrete Choice Experiment method is limited to the amount of attributes included in the presented choices (Johansson et al., 2021). In order to define choice scenarios for the discrete choice experiment that include the factors most important to respondents, it is necessary to review relevant literature (Hollin et al., 2020).

Both factors related to incentives as well as factors related to privacy are considered. Since literature on privacy-related factors influencing participation in participatory sensing systems in particular is limited, articles addressing data sharing behaviour and privacy concerns in general are also considered. The databases used for this literature search are *Scopus, Webofscience,* and *GoogleScholar.* The literature search results in a list of factors related to incentives and a list of factors related to privacy concerns. The results of the literature review are presented in chapter 4. In order to determine which factors are suitable to include in the choice experiment, exclusion criteria are used to limit the number of attributes in the experiment. In chapter 5, an elaboration on this exclusion process can be found.

After conducting the literature search and evaluating the list of factors, a short list of selected factors remain that can be used as attributes in the choice experiment.

3.2.2. Choice experiment

In order to model factors influencing choice behaviour in participatory sensing systems, data collection is required. A distinction can be made between *revealed preference* (RP) and *stated preference* (SP) data. Revealed preference data mirror the actual choices made by people in real world situations. Stated preference data, on the other hand, are collected by presenting respondents with hypothetical choice situations. These data are collected through an experiment or survey. Respondents in the survey are asked to state what their preference would be among different alternatives in a hypothetical choice situation (Train, 2003).

An advantage of SP data is that these data allow for a higher degree of variation in the attributes when designing the experiment, since the values of the attributes can be varied on a wider range, beyond values of existing alternatives. However, it should be noted that a limitation of using SP data lies in the fact that what people say that they will do, is often not the same as what they actually do (Train, 2003). Specifically, the privacy paradox is relevant within the context of this research (Ioannou et al., 2020). It appears that users often state that privacy of their personal data is an important issue, but do not actively protect this data in reality (Gerber et al., 2018).

Nevertheless, participatory sensing is an emerging field, which means that historical, RP data are lacking. Therefore, SP data are used in this research. By designing choice experiments that are as realistic as possible, as well as asking respondents background questions about their current choices related to data sharing and privacy, the effect of the privacy paradox will be minimized.

The SP data is collected through an online survey that is targeted at smart device users in the Netherlands who are 18 years or older. Owning a smart phone or another smart device is a prerequisite for participating in the survey. Thus, at the beginning of the survey, respondents are asked whether they own a smart device. The survey is constructed and distributed using the online software Qualtrics.

Discrete choice models describe the choices of decision makers among alternatives (Train, 2003). This set of alternatives is called the *choice set.* The factors that are found in the literature search are used as attributes to design choice sets for the choice experiment. In the experiment, respondents are asked whether they would share data or not in each situation (“yes” or “no”). This is also known as a *binary choice task*, since respondents have to choose between two options. The choice sets are constructed using the software NGene. In chapter 5, an elaboration on the construction of the choice sets can be found.

3.3. Choice Modelling Phase

In order to analyse what factors influence the decisions of individuals regarding data sharing, different choice models are estimated. The results of the estimated models are elaborated on in chapter 6.

3.3.1. Multinomial Logit (MNL) model

The most well-known discrete choice model is the Multinomial Logit (MNL) model, which is based on Random Utility Maximization (RUM). These models are derived assuming that a decision-maker faces a choice among a certain amount of alternatives. Each alternative has a certain level of utility (or satisfaction) that the decision-maker will obtain when choosing that alternative. According to the RUM model, the decision-maker will choose the alternative that provides the greatest utility (McFadden, 2001; Train, 2003).

The utility of an alternative depends on parameters, that are estimated by the researcher. However, there are aspects of the true utility of a decision-maker that cannot be observed by the researcher. Therefore, utility is composed of a systematic part (V_i) and an unobserved part (ε_i). This means that even when the systematic utility for an alternative is the highest, the alternative may still not be chosen by a decision-maker due to other factors that influence the decision-maker’s choice, such as unobserved factors, heterogeneity in tastes, or randomness in choices.

The utility of an alternative can thus be defined as:

$$U_i = V_i + \varepsilon_i$$

Where the systematic part is defined as:

$$V_i = \sum_m \beta_m x_{im}$$

Thus, the final equation for the RUM model is as follows:

$$U_i = \sum_m \beta_m x_{im} + \varepsilon_i$$

Where:
i = alternative, e.g. scenario 1, scenario 2
m = attribute, e.g. monetary reward, type of data
U_i = utility of alternative i
ε_i = unobserved utility of alternative i (error term)
β_m = attribute weight for an attribute m in alternative i (to be estimated in the model)
x_{im} = attribute value of attribute m for alternative i, e.g. €20, €40

Because of the error term, we can only predict choices up to a probability. In other words, a higher systematic utility means there is a higher probability of the alternative being chosen. The following equation is used to determine the choice probability:

$$p_i = \frac{e^{V_i}}{1 + e^{V_i}}$$

Where:
p_i = probability that alternative i is chosen
V_i = systematic utility of alternative i

The model is estimated by using this RUM model. The model estimation is done in *Apollo*, which is a statistical tool in *R*.

3.3.2. Latent Class Model

A statistical procedure that can be used to identify different subgroups within a population is Latent Class (LC) Analysis (Weller et al., 2020). Traditionally, LC analysis assumes that each observation is a member of one and only one of *T* (unobservable) classes.

Individuals often base their choices on the perception of reality, instead of on an objectively measurable reality (Molin et al., 2017). In the experiment, two latent variables are measured concerning the perceptions of respondents regarding trust and privacy. Since there might be unobservable classes of people in the sample, a Latent Class model is estimated.

In order to estimate the Latent Class model, the package "lcpars" in Apollo is used. The choice probabilities are defined like in the MNL model. Furthermore, the privacy and distrust perceptions and the personal characteristics are added as covariates. The model estimates the parameters for every class and shows the covariates. Based on this model, it is possible to identify if there are any patterns within the data.

3.3.3. Mixed Logit model for Panel Data

A Multinomial Logit model assumes that there is no correlation between the choices made by a certain individual. This means that it assumes that multiple choices made by the same individual are not correlated. However, in reality, these choices are correlated. These are called panel effects. When ignoring the correlations in panel data, the assumption is made that every observed choice is independent of all the others. This implies that it is assumed that the dataset contains more information than it does in reality. As a result, the model will assign too much certainty to the estimated parameters. Statistically speaking, this means that the model will underestimate the standard error of parameters.

To specify the MNL model and correct for this error, a solution is the Mixed Logit (ML) model (Train, 2003). This specification is useful if there are different segments in the population, that each have its own choice behaviour or preferences. Moreover, it is able to capture utility-correlation between consecutive choices of respondents.

In this research, repeated choices are made by each decision-maker. To capture the heterogeneity in the choices, an ML model for panel data is estimated. This is done by adding an additional error component to the model. In this way, (part of) the correlation between choices that are made by the same individual is captured. Mathematically, the product of choice probabilities is integrated over the additional error term, when computing the Log-Likelihood function.

A disadvantage of the ML model is that it is less elegant than the MNL model, is more difficult to code, and requires higher computation times. However, since the ML model assumes correlations between repeated choices, which also exist in reality, this model can provide more realistic estimations of parameters and standard errors. Thus, it is expected that the ML model outperforms the MNL model.

3.4. Qualitative research

In parallel to the choice modelling phase, a qualitative research is conducted. This is done by conducting expert interviews with several relevant parties.

The interviews that are conducted are semi-structured. Some standard questions are asked on current smart mobility projects, perceived challenges of these projects, and on the potential of implementing participatory sensing. Furthermore, some questions are tailored to the specific party that was interviewed.

All interviews were recorded. Afterwards, the interviews are transcribed manually. The software *Atlas.ti* is used for coding the interviews. All relevant parts of the interviews are coded. Then, the codes are categorised using group codes. This allows for efficient comparison of the different interviews. After coding the interviews, we define use cases to which participatory sensing could be applied. Each interview leads to one specific use case. The challenges mentioned in the interviews are analysed, and the similarities and differences between the interviews are compared.

Afterwards, the results of the quantitative and qualitative research are combined. We analyse the impact of the choice experiment results on the defined use cases. This is done by describing hypothetical choice situations based on the identified use cases, and calculating the acceptance of these use cases by using the estimations of the choice modelling. Furthermore, we discuss what should be done to implement participatory sensing in these use cases, according to the results. These implications are discussed in chapter 7.

4

Literature research

This chapter aims to provide an overview of the different factors that are relevant to the choice of individuals to participate in sharing data in sensing applications. In order to identify these factors, existing literature is reviewed on this topic through a literature review. The chapter aims to answer the following research questions.

SQ1. What factors potentially incentivize or disincentivize individuals to contribute to participatory sensing systems?

SQ2. What factors relating to privacy potentially influence the decision of individuals to share data in participatory sensing systems?

First, the literature review strategy is discussed. In the second subsection, identified factors related to incentives motivating participation in sensing applications are discussed. Subsection 4.3 addresses factors related to privacy with respect to data sharing. Subsection 4.4 describes factors related to perceptions on trust and privacy, after which personal characteristics potentially influencing the willingness to share data are mentioned in subsection 4.5. Lastly, a summary of all potential factors identified in the literature review is given in subsection 4.6.

4.1. Literature review strategy

Scopus, WebOfScience and *GoogleScholar* were used as a tool to find relevant papers. Different search terms were used in order to identify factors affecting the willingness to contribute to participatory sensing systems. First, we searched for factors influencing users' behaviour in participatory sensing applications specifically. Furthermore, search terms on data sharing in general were used. Factors related to data sharing in general can also be relevant to data sharing in participatory sensing systems. Since a discrete choice experiment is conducted in this research, we also searched for factors previously used in similar choice experiments regarding data sharing. The search terms that were used to explore the literature are presented in Table 1.

Table 1. Search terms

Concepts	**Synonyms**
Participatory sensing	Mobile sensing, public sensing, citizen sensing, crowd sensing
Inventiv*	Motivation
Data sharing	Location sharing
Factor	Driver, influence
Choice modelling	Choice experiment, factor analysis
Behavio*r	Preferences, participation, contribution
Privacy calclulus	Willingness to share, willingness to disclose

These search terms were combined in order to search the databases and find relevant literature. The final search queries are presented in Table 2. Backwards and forwards snowballing was used to find all relevant papers.

Table 2. Final queries

Topic	Search terms	Results	Selected
Participatory sensing	sensing AND (participatory OR mobile OR public OR citizen OR crowd) AND (participation OR behavio*r OR preference OR contribution) AND (incentiv* OR motivat*) AND (factor OR driver OR influence)	84	10
Data sharing preferences	("data sharing" OR "location sharing") AND ("choice modelling" OR "choice experiment" OR preference* OR "factor analysis")	294	11
	("data sharing" OR "location sharing") AND ("willingness to share" OR "willingness to disclose" OR "privacy calculus")	90	

In the literature review, only scientific papers, conference papers, and journal articles submitted after 2010 were included in order to build upon the most current information. furthermore, only papers written in English were considered.

Ultimately, 21 articles were selected for the literature review. An overview of all reviewed articles can be found in Table 3.

Table 3. Selected articles for literature review

	Author	Country	Type	Topic
Related to participatory sensing specifically				
1	Anawar et al. (2017)	MY	Statistical analysis	Incentives
2	Calado & Pardal (2018)	PT	Design science research	Incentives
3	Christin et al. (2013)	DE, AT	Statistical analysis	Privacy
4	Heiskala et al. (2016)	FI	Empirical research	Incentives
5	Khoi et al. (2018)	DE, ES	Empirical research	Incentives
6	Klopfenstein et al. (2019)	IT	Design science research	Incentives
7	Mloza-Banda & Scholtz (2018)	ZA	Statistical analysis	Incentives
8	Ogie (2016)	AU	Literature review	Incentives
9	Salim & Hague (2016)	AU, UK	Literature review	Incentives, Privacy
10	Zaman et al. (2015)	BD	Literature review	Incentives
Related to data sharing in general				
11	Aitken et al. (2018)	GB	Statistical analysis	Data sharing preferences
12	Bhatnagar & Kumra (2020)	IN	Statistical analysis	Data sharing preferences
13	Heidel et al. (2021)	DE	Statistical analysis	Data sharing preferences
14	Ioannou et al. (2020)	UK	Statistical analysis	Privacy
15	Johansson et al. (2021)	SE, IS, NL	Statistical analysis	Data sharing preferences
16	Kumaraguru & Cranor (2005)	US	Literature Review	Privacy
17	Lorenzo et al. (2020)	DE	Statistical analysis	Privacy, data sharing preferences
18	Princi & Krämer (2019)	DE	Statistical analysis	Privacy, data sharing preferences
19	Schomakers et al. (2020)	DE	Empirical research	Privacy
20	Turland & Slade (2020)	CA	Statistical analysis	Data sharing preferences
21	Wang et al. (2016)	TW, US	Empirical research	Privacy, data sharing preferences

In total, 14 factors relating to incentives and privacy are identified. In addition, factors relevant to privacy perceptions are identified, as well as factors related to personal characteristics that might play a role are found. All identified factors are displayed in Table 4.

Table 4. Identified factors

Category	Factor	Found in
Relevant to incentives	Monetary reward	2,4,5,6,8,10,12,13,17
	Non-monetary benefit	1,2,6,7,8,10
	Societal benefit	4,7,8,12,18
	Required effort	4,7,8,9,10
	Resource consumption	2,4
Relevant to privacy	Level of identification	6,10,15,19
	Types of data	4,9,11,13,15,17,19
	Data recipients	3,4,11,13,15,17,19,20
	Purpose of data sharing	8,11,15,19
	App security	19
	Review of data sharing	11,15
	Consent	15,17
	Duration of data collection process	3,8
	Amount of data	18
Privacy Indexes	Distrust Index	14,16,19
	Privacy Segmentation Index	14,16,18
Personal characteristics	Age	3,5,19
	Gender	17,21
	Education level	3
	Income level	3
	Digital behaviour	3,12,19,21
	Altruism	1,3,6,12

4.2. Factors relevant to incentives for participatory sensing

First, factors that can incentivize individuals to contribute to participatory sensing applications were investigated. Perceived benefits can be a motivation for people to participate. Anawar et al. (2017) make the distinction between intrinsic and extrinsic incentives. Intrinsic incentives are motivations that are based on purely personal reasons, such as inherent interest, enjoyment, or leisure (Mloza-Banda & Scholtz, 2018). Extrinsic incentives are influenced or controlled by an external regulation or by other people (Anawar et al., 2017).

However, also factors that disincentivize smart device owners exist. These disincentives are caused by "costs" of contributing data. Users will incur direct and indirect costs, depending on the sensing tasks. These costs include time and resources consumed by performing sensing tasks, such as network bandwidth, memory, CPU, and battery usage. All identified factors related to incentives and disincentives are discussed in the following subsections.

4.2.1. Personal benefit

First, individuals can be motivated because it provides a benefit for them personally. In order to enhance participation in sensing applications, incentives are given to participants. There are two main types of incentive mechanisms that can be distinguished, which are monetary and non-monetary incentives (Anawar et al., 2017). Monetary incentives are extrinsic incentives in the form of a financial reward. Non-monetary incentives can be either extrinsic or intrinsic, by providing entertainment, providing a service, or by reputation and social interaction.

Monetary reward

When applying monetary incentives, participants are paid a given amount of money in exchange for their work (Klopfenstein et al., 2019). These can be financial rewards in the form of cash, but also in the form of discount coupons and gift vouchers (Anawar et al., 2017; Bhatnagar & Kumra, 2020). Bhatnagar & Kumra (2020) found a significant positive impact of such extrinsic, monetary rewards on the willingness to share IoT product data (*n* = 337). This is confirmed by research by Turland & Slade (2020), concluding that participation rates are significantly higher when providing even a small monetary benefit.

Several monetary incentive mechanisms have been investigated in previous research. In a study on monetary incentives, Khoi et al. (2018) conclude that a fixed micro-payment mechanism works more effectively than other incentive mechanisms in attracting participants and enticing them to complete sensing tasks. If a participant submits data to fulfil a sensing task, the participant receives a fixed amount of money, specified by the campaign author (Khoi & Casteleyn, 2018). Micro-payments are often given out in the form of coupons (Zaman et al., 2015).

In a study by Heidel et al. (2021), the effects of a monthly bonus payment for sharing self-tracked health data were investigated, using a payment ranging from €5 to €75 per month. It appeared that under certain conditions, 55% of participants would sell their self-tracked data to universities, 47% to health insurers, and 39% to pharmaceutical or medical devices companies, for €20 per month (*n* = 1114) (Heidel et al., 2021).

Heidel et al. (2021) also found that participants tend to overestimate the monetary value of their data, which can result in high prices demanded.

Non-monetary benefit

A first non-monetary benefit is entertainment. By turning the sensing tasks into a game, user participation can be enhanced (Klopfenstein et al., 2019). Therefore, some crowdsensing systems use gamification in order to motivate participants by eliciting feelings of fun and enjoyment (Mloza-Banda & Scholtz, 2018). For example, users receive points when submitting sensing requests (Calado & Pardal, 2018). Since these kind of incentives are due to inherent interest, enjoyment, or leisure, this is seen as an intrinsic motivation.

A second non-monetary benefit is service provision. This means that the sensing platform and users provide mutual benefits to each other (Klopfenstein et al., 2019). By participating in a sensing application, users get access to a useful service. This may be a service that is offered using the aggregated crowd sensed data (Ogie, 2016). An example of this is a function that allows self-monitoring by recording information. In a study by Anawar et al. (2017), the possibility for self-monitoring appeared to be an important incentive. Also, services that do not derive its value from the sensed data are possible.

Lastly, users can also be motivated because they get social recognition or improved reputation when having performed a sensing task (Ogie, 2016). This is implemented in applications by a leader board that shows a ranking or a progress percentage, or by awarding participants with points, trophies, or badges (Anawar et al., 2017). Another social incentive is the possibility for users to get feedback from peers and reinforce others information (Mloza-Banda & Scholtz, 2018).

4.2.2. Societal benefit

Besides personal benefits, the relevance of the participatory sensing application to society is a potential factor influencing the willingness to contribute to participatory sensing systems (Ogie, 2016). Research by Princi & Krämer (2019) shows that when people recognize an IoT system as advantageous, for example when it allows for faster rescue in case of emergency, their acceptance of the system will be higher.

This motivation of providing societal benefits can be related to feelings of moral obligation. These feelings can stimulate individuals to contribute and share data for a societal cause. In a study by Bhatnagar & Kumra (2020), a significant positive impact of moral obligation on the intention to share data was found (*n* = 337).

Societal benefits delivered to society can intrinsically motivate individuals to share data. Mloza-Banda & Scholtz (2018) mention collective incentives, meaning that citizens are motivated to work together for a common good. However, when the benefit of an application is that the wellbeing of the affected community is improved, the crowdsensing service offers a benefit also for people not contributing data. Collective incentives may therefore not be enough, since some people may decide to let others contribute data and wait for the results, which gives a free-riding effect (Heiskala et al., 2016).

4.2.3. Required effort

The effort required by the user to perform the sensing tasks is an important factor to take into account. Direct costs related to effort include time spent by performing sensing tasks. Indirect costs can also be incurred, for example when it is required to travel to a specific sensing location (Ogie, 2016). Effort is also related to spatial-temporal characteristics. A report from a remote area or at midnight should have a higher reward (Zaman et al., 2015). Offering incentives that are worth performing the sensing tasks is necessary in order to enhance adoption of participatory sensing systems.

Furthermore, insufficient time can be a demotivational factor to participate in sensing tasks (Mloza-Banda & Scholtz, 2018). Therefore, it is expected that a higher effort will have a negative influence on the willingness to participate.

A sensing application on a smart device collects raw sensor data. These data may be analysed and used locally before being send to a server for aggregate analytics. Also, an application can prompt users for direct observations, such as taking a photograph or answering a question (Heiskala et al., 2016). When users have to contribute to the sensing application actively, the effort they have to take is higher than with an application that only runs in the background. This might influence individuals' choices on whether to participate or not. As introduced in Chapter 2, applications can use opportunistic or participatory sensing. A combination of both forms is also possible. Thus, the effort that is required depends on the level of the level of opportunistic and participatory sensing that is used in the application. Applications that ask users to report direct observations must be careful not to overburden the users, since this might be a disincentive for them to participate (Heiskala et al., 2016).

Salim & Haque (2015) distinguish three levels regarding user engagement in crowdsensing systems. The lowest level (aware and consent) means that participants are aware of their involvement and provide consent for data being collected, however their interaction with the system remains minimal or passive. Engaged users (the second level) are more actively involved in the system and interact with the system by adding their own experiences. The collaborative level, which is the highest level of participation, means that users actively look for more data and aim for better coverage in data collection activities (Salim & Haque, 2015).

4.2.4. Resource consumption

Participating in sensing tasks can incur costs related to resource consumption, such as network bandwidth, memory, CPU, and battery usage. For example, users may worry that the application will consume too much computation power, battery power and network bandwidth in sensing activities (Calado & Pardal, 2018). When the sensing application consumes too much of these resources, the user may stop using the application (Heiskala et al., 2016). Thus, resource consumption is a disincentive for users to participate.

Respondents in a study by Calado & Pardal (2018) stated that resource consumption related to crowdsensing tasks should be at a similar level as other popular mobile applications. Also, different technical solutions are developed in order to limit the resource consumption of sensing applications (Heiskala et al., 2016). Therefore, in the experiment we assume that the resource consumption will be similar to other mobile applications.

4.3. Factors relevant to privacy in participatory sensing

Privacy concerns can have a negative effect on the willingness to participate in sensing applications. Like mentioned in chapter 2, a major element regarding privacy concerns is the control users have over data they provide. Factors regarding privacy are explored in order to determine what possibilities are desired by users to control their privacy.

4.3.1. Risk of re-identification

Requiring registration in participatory sensing applications enables recognition of users and more accurate correlation between users and observations. However, Kotovirta et al. (2012) state that each extra step that is required in the installation process reduces the amount of interested users. Users may be reluctant to share their observations with others and prefer to submit anonymous observations. Therefore, requiring no registration could attract more users. For example, the proposed platform in the study by Klopfenstein et al. (2019) is designed to require no user registration. A disadvantage in that case, however, is that it will not be possible to identify observers reporting invalid data or evaluate the past performance of users (Kotovirta et al., 2012).

The protection of user' identity is a core aspect for privacy. The study by Schomakers et al. (2020) shows that anonymization is the most important factor that influences users' decision to share data (n = 126). Even in the case of no registration, the sensing platform still stores data about contributions by users including, e.g., geolocations and timestamps. Users are not directly identifiable, but the collection of these data can still expose them to potentially being identified. These data are potentially sensitive, since they could be used to reconstruct information about individual participants, like commute patterns, routines, or private locations. Furthermore, microphone and camera data can be a threat to privacy, especially when linked with other information provided by individuals (Klopfenstein et al., 2019). Therefore, it is possible to hide or partially obfuscate information about contributions by users. This means that, for example, an approximation of the geolocation instead of the exact location is given, or a timestamp reduced to week- or month-level.

In order to provide privacy protection regarding time or location data, the concept of *k*-anonymity can be used as a privacy measure (Klopfenstein et al., 2019). This can prevent re-identification of sensitive information, which means that *k*-anonymity ensures that two categories of data cannot be connected to one another. According to Zaman et al. (2015), the degree of anonymity can be incorporated within incentive schemes, in order to give control to the user on the level of privacy protection.

In a study by Johansson et al. (2021) on user preferences for sharing health data , it appeared that the level of identification was so important to participants, that it was crucial for sharing data at all. Thus, this attribute may dominate people's decisions. Therefore, this attribute was set at a fixed level, which meant that participants were informed that their personal data would be replaced with a code, also known as "pseudonymised data". However, since our research is related to urban monitoring and health data is viewed more sensitive by participants than location-based data (Schomakers et al., 2020), we assume that the level of identification is less dominant in our research and is an attribute that can be varied in the alternatives.

In Schomakers et al. (2020), *k*-anonymity is used as an attribute in a choice experiment. In order to make sure the attribute levels were comprehensible to participants without background knowledge on anonymization techniques, the fictive level of "complete" anonymization was used, besides the levels of k=5, k=2, and no anonymization. However, complete anonymization does not exist in reality. In the choice experiment, a large gap was seen in utility between the anonymization level of k=5 and "complete" anonymization. In order to make the anonymization level easier to understand for participants, another way to frame this is the "risk of re-identification".

For using anonymized data, no consent is needed from the user, since the GDPR applies only to data relating to an identifiable natural person. However, even with anonymization, users appear to want to control what data is collected, for what purpose, and they want to be rewarded (Schomakers et al., 2020).

4.3.2. Types of data

Different types of data can be collected in participatory sensing systems. The type of data that is collected can influence that willingness to contribute to participatory sensing applications (Lorenzo et al., 2020; Schomakers et al., 2020). Personal data is not only related to data that allows to identify a person directly, like names and addresses. When users share information about their environment, personal data can also be inferred, which can be derived indirectly. This can be a threat to the privacy of a user. For example, in a participatory sensing system related to transportation, location data is usually needed. These data can lead to the common daily commutes, work and home locations of the user (Heiskala et al., 2016). Besides that, frequent visits to hospitals could provide information about the users' medical conditions (Christin et al., 2013). Furthermore, accelerometer readings can infer information about the current activity of a participant, sound samples may contain private conversations, and pictures and videos can give information about the environment of a participant. The release of these kind of data can be a threat to the privacy of the user (Christin et al., 2013).

According to Christin et al. (2013), data that can be potential threats to privacy when being shared with unauthorized parties are time and location data, sound samples, pictures and videos, acceleration, environmental data, and biometric data. The willingness of people to share data in a participatory sensing system may depend on the kind of data that is collected. For example, Schomakers et al. (2020) found that users are more hesitant to share data related to health than location data.

The type of data that is collected highly depends on the sensors used for the data collection. Different sensors that are embedded in smart devices can be distinguished (Masoud et al., 2019; Salim & Haque, 2015):

- Sensors used for localisation: GPS, Wi-Fi, and Bluetooth.
- Physical motion sensors: accelerometer, gyroscope, magnetometer
- Environmental sensors: temperature and humidity sensors, barometer
- Multimedia sensors: camera, microphone, fingerprint

4.3.3. Data recipients

The parties with whom the data is being shared, is a factor potentially influencing the willingness of individuals to participate (Heiskala et al., 2016; Johansson et al., 2021; Schomakers et al., 2020). For example, users may claim different rewards for scientific and corporate institutions, since individuals may be more reluctant to share data with corporate institutions compared to academic institutions (Christin et al., 2013). This is confirmed by research by Aitken et al. (2018), which shows that participants have greater support for data usage by the public sector compared to usage by the private sector. However, Turland & Slade (2020) find an opposite effect in a study on participatory sensing for farm management. Here, users are more concerned about sharing data with the government compared to sharing with private organizations, which echoes concerns about government surveillance. Furthermore, these users are most willing to share data with university researchers (Turland & Slade, 2020).

Something related to this is whether collected data is sold to third parties (Lorenzo et al., 2020). If raw data or statistically processed data gathered by a participatory sensing application is sold for profit, users may be less willing to participate (Heidel et al., 2021).

4.3.4. Purpose of data sharing

The purpose for which the data is used, plays a role for individuals when considering participation in sensing tasks (Aitken et al., 2018; Ogie, 2016; Schomakers et al., 2020). The study by Aitken et al. (2018) emphasizes that general public benefits are the most important purpose according to participants. Also, being transparent about the purpose of data collection can enhance perceived privacy protection (Princi & Krämer, 2019). Furthermore, when participants understand why their data are useful to the recipient, they may be more willing to provide data (Schomakers et al., 2020).

4.5.5. App security

An application installed on a smart device can be protected by, among others, a password, a double password, a fingerprint scanner, or facial recognition. Out of these security measures, the most accepted option according to previous research is double password protection, followed by password protection. Facial recognition is seen as the least secure measure (n = 126) (Schomakers et al., 2020).

4.5.6. Review of data sharing

The process of collecting and aggregating data could be monitored or overseen by an independent body, such as an ethics committee or a data access committee, or by relevant public services. Such a party can review and control whether data is used and shared appropriately (Johansson et al., 2021). In this case, individuals' willingness to participate may be higher, when compared to the scenario in which the collection and aggregation process is monitored by the organisation(s) undertaking the research (Aitken et al., 2018).

4.5.7. Provision of information and consent

Lorenzo et al. (2020) define two types of consent for collecting data. Users can be given the ability to opt in for the data collection (opt-in), which means that respondents have to provide explicit consent to collect and use their data. The other option is opt-out, which means that the respondent implicitly consents to collect his or her data by using the application. There is always a possibility to withdraw the implicit consent (Lorenzo et al., 2020).

Participants in a survey conducted by Johansson et al. (2021) indicated they prefer to actively consent. If that is not an option, they want to know what happens to their data and be able to opt out. Also, participants expressed importance of having clear information before deciding to consent.

Since individuals know beforehand that participatory sensing applications will collect data, we assume that consent will be asked beforehand (opt-in) as a fixed attribute.

4.5.8. Duration of data collection process

The duration of the data collection process may also be a factor influencing the willingness to participate in sensing applications. Data can be gathered on a shorter or longer term. However, according to research by Christin et al. (2013), among others, participants are more sensitive to the purpose of data sharing than to the duration of the data collection. A longer gathering duration does not significantly mean a higher expected reward.

4.5.9. Amount of data

According to Princi & Krämer (2019), people allow their (personal) data to be collected to a certain limit. When this limit is reached, the willing to accept the collection of data declines. Thus, it is expected that a higher amount of collected data has a negative influence on the willingness to contribute to participatory sensing applications.

4.4. Privacy Indexes

Besides characteristics of the sensing application, the willingness of individuals to contribute to participatory sensing applications depends on personal attitudes regarding data sharing, relating to privacy and trust. The researcher dr. Alan Westin conducted over 30 surveys related to privacy. In these surveys, respondents were asked about how they valued organizations and laws and regulations regarding privacy. In order to summarize the results and show trends in privacy concerns, Westin created a “Privacy Index” for most of his studies. We use the Privacy Index and the Distrust Index created by Westin in this research to analyse how attitudes and concerns regarding privacy and data sharing influence the willingness of individuals to participate in crowdsensing.

4.4.1. Westin’s Privacy Segmentation Index

Individuals have a varying degree of perceived privacy concerns and attach a different value to their privacy (Princi & Krämer, 2019). This may have an effect on their intention to participate in data sharing. People who have higher privacy concerns, may be more reluctant to contribute to participatory sensing systems.

Groups of people show different levels in privacy concerns. Kumaraguru & Cranor (2005), reviewing works on privacy indexes by Westin between 1978 and 2004, mention three categories to refer to different groups regarding privacy concerns:

1. High and Fundamentalist,
2. Medium and Pragmatist,
3. Low and Unconcerned.

Privacy Fundamentalists are generally hesitant to share information with organisations. They will probably choose privacy controls over personal benefits, in the case these compete with each other. Privacy Pragmatists weigh the benefits of sharing data against the degree of intrusiveness of personal information, and make their decision based on this trade-off. Individuals in the Unconcerned group generally trust organizations that want to collect their personal information. They are willing to forgo privacy claims in order to secure personal or societal benefits (Kumaraguru & Cranor, 2005).

In his 2001 report, Westin gives a definition of the Privacy Segmentation Index (Westin, 2001), which he created for studies conducted between 1995 and 1999. For deriving the index, he provided statements relating to privacy perceptions. For each statements, respondents have to indicate how strongly they agree or disagree. The following statements were used to derive the Privacy Index:

For each of the following statements, do you agree strongly, agree somewhat, disagree somewhat or disagree strongly?

1. Consumers have lost all control over how personal information is collected and used by companies.
2. Most businesses handle the personal information they collect about consumers in a proper and confidential way.
3. Existing laws and organizational practices provide a reasonable level of protection for consumer privacy today.

Respondents who agree with the first statement and disagree with the second and third statements, are considered Privacy Fundamentalists. Privacy Unconcerned respondents are those respondents who disagree with the first statement and agree with the second and third statements. All other respondents are categorized as Privacy Pragmatists.

According to Ioannou et al. (2020), there are some universal factors affecting privacy concerns of individuals, with individuals in different countries and context sharing the same concerns. These factors are in line with the statement in Westin's Privacy Index. Users who believe they have a high amount of control over the information that is collected or shared by an application, e.g. through privacy settings, might have less concerns over their privacy (Ioannou et al., 2020). Also, consumers having more trust in businesses and organizations handling their data in a confidential way are more likely to have fewer concerns regarding privacy. Furthermore, the perceived effectiveness of government regulations regarding privacy plays a role. Individuals being more sceptical about the effectiveness of privacy regulations may have higher privacy concerns (Ioannou et al., 2020).

4.4.2. Westin's Distrust Index

Trust in an organization means that one accepts the vulnerability of disclosing personal information and considers the provider competent to protect this information from unauthorized uses. Consumers who have more trust in online providers may be more likely to have fewer concerns regarding their privacy (Ioannou et al., 2020; Schomakers et al., 2020). Therefore, it is expected that more trust in the party handling their data will have a positive effect on the willingness to contribute to participatory sensing applications.

In Westin's study in 1994, he created the Distrust Index (Westin, 1994). The following questions were used to derive the index.

For each of the following statements, do you agree strongly, agree somewhat, disagree somewhat or disagree strongly?

1. Technology has almost gotten out of control.
2. Government can generally be trusted to look after our interests.
3. The way one votes has no effect on what the government does.
4. In general business helps us more than harms us.

To create the Distrust Index, the respondent's answers to the four questions are examined. If a respondent gives 3-4 distrustful answers, the respondent is classified as High Distrust. Two distrustful answers are classified as Medium Distrust, one distrustful answer as Low Distrust, and no distrustful answers as No Distrust.

4.5. Personal characteristics

Several personal characteristics are considered in order to see if they influence the willingness to contribute data to participatory sensing applications. Furthermore, measuring these factors can show the variety of people included in the survey, and if this is similar to the diversity in the population. Demographic factors like age, gender, education, and income level, as well as digital behaviour and involvement in altruistic activities are considered.

4.5.1. Age

The age of a participant can influence their assessment of the importance of privacy. Privacy perceptions and behaviour regarding data sharing differ between age groups (Schomakers et al., 2020). According to a study by Christin et al. (2013), young participants that already share a high amount of data online are more likely to contribute to participatory sensing campaigns. A similar effect was found by Khoi et al. (2018), stating that younger participants significantly submitted more results than older users. However, there were also studies finding no significant effect of age on the willingness to share data (Turland & Slade, 2020). In order to investigate if there is an effect in our sample, age is incorporated as a factor in the experiment.

4.5.2. Gender

Several studies include gender as a demographic variable. Wang et al. (2016) as well as Turland & Slade (2020) found that females tend to disclose more personal information via mobile apps. On the contrary, Lorenzo et al. (2020) found that women care more about privacy than the average respondent, which would indicate women being less willing to contribute data. However, other studies conclude that gender differences do not affect submission behaviour of data collectors in a participatory sensing system (Khoi et al., 2018). Since previous research is not clear on the influence of gender, we include this as a factor in the survey.

4.5.3. Education

Christin et al. (2013) include education achievements in their research and conclude that this has a limited effect on concerns regarding privacy. In order to see if an effect of education level can be observed in our sample, this factor is included in the experiment.

4.5.4. Income level

Income level is mentioned by Christin et al. (2013) as a possible factor influencing the importance of privacy. People with a higher income might expect a higher reward in turn for submitting data. Therefore, we expect that income level has a negative impact on the willingness to participate in sensing applications.

4.4.5. Digital behaviour

Participants that already share a lot of information online may be more prone to share data in participatory sensing applications (Christin et al., 2013). This can be related to “personal innovativeness”, which is the willingness of an individual to try out any new information technology. An early adopter may be more willing to share information (Bhatnagar & Kumra, 2020). Digital behaviour can be measured by the use of digital devices (Schomakers et al., 2020), for example the average time spent using mobile applications last week by users, or the number of mobile apps used weekly on average (Wang et al., 2016).

4.4.6. Altruism

Altruism is a basic human behaviour to help others. In this context, Bhatnagar & Kumra (2020) mention that enjoyment of helping can be a major motivator to share personal device data. In this study, a significant positive impact of enjoyment of helping was found on the intention to share data ($n = 337$). Altruistic behaviour of individuals can therefore affect their willingness to contribute to participatory sensing systems. Participants that are already involved in benevolent and altruistic activities may be more willing to participate when there are benefits to society and may request lower rewards (Christin et al., 2013). Thus, it is important to measure the attitude of respondents regarding altruism. This can be done by asking a question on if and to what extent they are involved in volunteering or other altruistic activities. It is anticipated that a higher level of participation in altruistic activities, will have a positive effect on the willingness to contribute to participatory sensing systems.

4.6. Conclusion

In total, 14 factors are found relating to benefits and costs of participating in sensing applications. The first sub-question of the research aims to identify potential factors incentivising or disincentivising people to share data in participatory sensing applications. According to the literature review, factors that can (dis)incentivise potential users are *monetary benefit, non-monetary benefit, societal benefit, required effort,* and *resource consumption.* A monetary, non-monetary, or societal benefit is expected to incentivise users to participate in sensing applications. The required effort and the resource consumption related to the sharing of data are expected to be a disincentive for participating.

Sub-question 2 of the research aims to identify potential factors relating to privacy that are influencing the willingness of individuals to share data. The factors identified in the research are *risk of re-identification, types of data, data recipients, purpose of data sharing, app security, review of data sharing, provision of information and consent, duration of data collection process,* and *amount of data.* Furthermore, two indexes related to privacy and trust perceptions are found, which are the Privacy Segmentation Index and the Distrust Index. People scoring higher on the Privacy Index or the Distrust Index are expected to be less willing to share data.

Finally, seven factors related to personal characteristics are found, which are *age, gender, education, income, digital behaviour, and altruism.* The characteristics age, gender, and education are mainly included in order to get insight in the representativeness of the sample. No clear expectation exists on the effect of these characteristics on the willingness to share data. It is expected that people with a higher income find small amounts of money less important when making a decision than people with a lower income. Furthermore, it is expected that people spending a lot of time using applications and people having a high amount of applications on their smartphone, will be more willing to put effort in sharing data. Lastly, people already participating in altruistic activities are expected to be more willing to spend effort on participating, without getting a personal benefit in return. In the chapter 5, most important factors identified in the literature research are further developed and refined for designing the choice experiment.

5

Experiment Design

This section addresses the design of the choice experiment. The first subsection discusses the process of selecting attributes and attribute levels to be included in the choice experiment. In subsection 5.2, the identified attributes are discussed. Subsection 5.3 describes the attribute levels, after which the refinement of the attributes leading to the final selection of attributes and levels is addressed in subsection 5.4. In subsection 5.5, the construction of the experiment design is addressed. Subsection 5.6 describes the design of the survey in which the choice experiment is conducted. Finally, subsection 5.7 explains the process of the survey distribution.

5.1. Process of experiment design construction

Johansson et al. (2021) describe guidelines for carefully selecting attributes to be included in a discrete choice experiment, using a three-step approach: Step 1) Attribute identification, Step 2) Attribute development and Step 3) Attribute refinement. This three-step process is used in our study in order to develop attributes for the choice experiment.

The first step consists of a literature review exploring all potential attributes relevant to the topic, and selecting the most important attributes to be included. In the second step, the attributes are further developed by determining appropriate attribute levels. Finally, the attributes and attribute levels are refined by implementing user feedback, in order to come to a final selection of attributes that are clear and well-defined for respondents participating in the experiment.

After selecting attributes and attribute levels, the experiment design is generated using these attributes. Then, the survey is constructed, in which the experiment will be conducted. The process from identifying attributes to the design of the survey is presented in Figure 6.

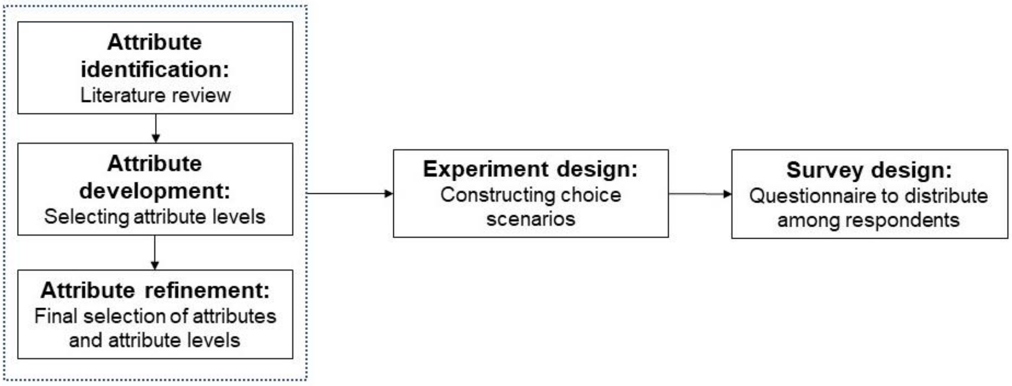


Figure 6. Experiment Design process

5.2. Attribute identification

In total, 14 factors were identified in the literature review. However, including too many attributes in the choice experiment can lead to increased choice difficulty, which may impact the validity of people's responses (Pearce et al., 2021). The number of attributes is therefore often varied from 3 to 7 (Mariel et al., 2021). Thus, the number of attributes to be included in the experiment is reduced from 14 to 5.

A few guidelines are used to determine what attributes to include in the survey. First, the most important attributes for respondents should be included. According to the study by Schomakers et al. (2020), the *data receiver* and the *types of data* were deemed factors of great importance by participants. Furthermore, *anonymization* appeared to be an important attribute. Since these three factors had higher importance and were mentioned more often in previous research compared to app security, review of data sharing, amount of data tracking and the duration of the data collection process, these were selected to be included as attributes in the survey.

The purpose of the data collection was also mentioned in several studies as an important factor affecting the decision of individuals to perform sensing tasks. The purpose of data collection is closely related to the societal benefits of the participatory sensing application. The purpose and the related benefits strongly depend on the context in which the data is collected. Therefore, the purpose of the data collection and the contribution delivered to society will be communicated at the beginning of the experiment. In this way, participants in the experiment are able to take into account the societal benefits while making their decision whether or not to share data.

Also, a second criterion is that attributes should be selected that can be influenced by policy or design. Moreover, the attribute should be directly observable. Something that can be influenced is the personal benefit that is received by the user when participating in sensing tasks. This benefit can be a monetary reward, entertainment, service provision, or reputation. Out of these factors, monetary rewards are most directly observable. Entertainment or reputation related benefits are inherent to a person and are therefore seen as less interesting for inclusion in the experiment. Furthermore, it is assumed that the extent to which the respondent is entertained by or interested in the sensing tasks depends on the purpose of the sensing and the area application, which is already included the context given in the experiment. Moreover, including monetary rewards allows us to measure the trade-off between money and privacy. Therefore, the personal benefit of receiving a *monetary reward* is included in the experiment.

Lastly, the *effort* spent by the participant is selected as an attribute. This allows us to gain insight in the trade-off between money and effort. Also, the required effort is a factor that can be highly influenced by the design of the sensing application, by making a choice on using opportunistic of participatory sensing.

The attributes selected to be included in the choice experiment are presented in Table 5.

Table 5. Selected attributes

Group	Factor
Benefits and costs	Monetary reward
	Effort spent
	Risk of re-identification
	Type of data
	Data recipients
Privacy Indexes	Privacy Segmentation Index
	Distrust Index
Personal characteristics	Age
	Gender
	Education
	Income level
	Digital behaviour
	Altruism

5.3. Attribute development

This subsection describes the levels that are used to vary the attributes in the choice experiment. First, the factors that are varied in the experiment are addressed, which are factors relating to benefits and costs of participation in sensing applications. Secondly, the factors relating to perceptions of privacy and trust are discussed, following the factors relating to personal characteristics.

5.3.1. Factors varied in experiment

Monetary reward
Several studies use monetary benefits varying between €5 and €75 per month in their experiments Another study used a fixed monetary gift of €20 (Khoi et al., 2018). For this study, we use a range from €0 to €60. An equal distance between the attribute levels is maintained, since this preserves orthogonality (ChoiceMetrics, 2018).

Effort
As described in Chapter 4, Salim & Haque (2015) distinguish three levels regarding user engagement in crowdsensing systems. Therefore, three levels are included for the effort attribute. Low effort means that the sensing application runs in the background of the smart device, which means that there is no need for users to interact with the application. Moderate effort means that the application runs in the background, but gives the user prompts for answering a short question or for providing feedback in order to collect data of higher quality. High effort means that the user is actively involved in the sensing process and has to submit sensing reports actively. Furthermore, this means that the user is willing to deviate from his or her normal route to perform sensing tasks at less "popular" locations.

Risk of re-identification
Since Schomakers et al. (2020) saw a large gap between attributes levels used in their choice experiment because of unrealistic values included, we use three levels with a slight difference in risk of re-identification. The risk of re-identification is dependent of the level of anonymization in the participatory sensing application. A higher level of anonymization means a lower risk of re-identification. In our experiment, we vary the risk of re-identification roughly based on a *k*-anonymity of 1 out of 10, 1 out of 5, and 1 out of 3. A level of "complete" anonymization is not used, since this level does not exist in reality. This leads to three attribute levels: a risk of re-identification of 10%, 20%, and 30%.

Types of data
Potential threats caused by data collected by participatory sensing systems are described in Christin et al. (2011). Time and location data are data that are gathered in most participatory sensing applications. Disclosure of these data can leak information regarding the home and workplace locations of participants, as well as their routines and habits.

Motion data includes data from accelerometers. From these data, the activities of participants can be recognized, which can give indications about a user's identity. For example, employers may verify if their employees are working by detecting anomalies in a participant's activity.

Another type of data that can used to infer identities and preferences of users is multimedia data, which includes sound samples, pictures and videos. Private conversations could be recorded, or sound patterns and pictures and videos could be used to infer the current context of a participant. Furthermore, information about the social relations of participants can be withdrawn from multimedia data (Christin et al., 2011).

These three types of data are in line with the three different of privacy that are impacted by participatory sensing, which are mentioned in chapter 4. Time and location data concerns privacy of location and space, motion data affects privacy of behaviour and action, and multimedia data regards privacy of data and image. Thus, these three types of data are appropriate to investigate the preferences of individuals regarding different types of privacy dimensions. In this study, we also choose to include a fourth type of privacy, related to environmental factors. Environmental data can also be privacy sensitive. For example, barometers can measure changes in the atmospheric pressure in the surroundings of the smart device (Masoud et al., 2019). With these data, the location of a phone inside a building can be detected with a high accuracy.

Since these four main types of data could influence the willingness of individuals to participate, these are included in the experiment. Because time and location are almost always collected in participatory sensing applications, these data are collected in all four scenarios. Furthermore, the respondents are presented with scenarios in which both time and location data and motion data is collected, as well as scenarios in which time and location, motion data, environmental, and multimedia data are collected.

Organizer of participatory sensing campaign
The nature of the party initiating the participatory sensing campaign can influence the value participants attach to their privacy (Christin et al., 2013). Possible parties organizing the sensing campaign are governmental institutions, academic institutions, and corporate institutions. Besides these parties, Wilde et al. (2020) mention societal stakeholders, like an NGO, as a potential organizer.

Thus, the four levels that are varied for this attribute are a governmental institution, an academic institution, a corporate institution, and a societal organisation. Since no clear expectation exists on what party is more accepted by people, this is a categorical attribute.

5.3.2. Factors relating to perceptions of privacy and trust

Table 6 shows the operationalisation of factors related to attitudes of respondents. For determining the Privacy Segmentation Index and the Distrust Index, the statements as defined by Westin are used. A 5-point Likert scale is used for answering the questions, where 1 represents “strongly disagree”, 2 represents “somewhat disagree”, 3 represents “undecided”, 4 represents “somewhat agree”, and 5 represents “strongly agree”. Depending on the rankings on these statements, respondents are categorized as Privacy Unconcerned, Privacy Pragmatist, or Privacy Fundamentalist, and as No distrust, Low distrust, Medium distrust, or High distrust.

Table 6. Statements for determining Privacy Segmentation Index and Distrust Index

	Statement
Privacy Segmentation Index	Consumers have lost all control over how personal information is collected and used by companies.
	Most businesses handle the personal information they collect about consumers in a proper and confidential way.
	Existing laws and organizational practices provide a reasonable level of protection for consumer privacy today.
Distrust Index	Technology has almost gotten out of control.
	Government can generally be trusted to look after our interests.
	The way one votes has no effect on what the government does.
	In general business helps us more than harms us.

5.3.3. Factors relating to personal characteristics

The factors relating to personal characteristics of respondents are age, gender, education level, income level, average time spent weekly using mobile applications, number of mobile apps used weekly, and involvement in altruistic activities.

Inspired by Ioannou et al. (2020), six age groups are used for the age attribute. For gender, respondents can identify themselves as male, female, or other. Education is measured based on the levels as used by the Centraal Bureau voor de Statistiek (CBS, 2019). Income levels are also derived from the levels used by the CBS (CBS, 2020).

Digital behaviour of respondents is measured in two ways, based on questions used in a study by Wang et al. (2016). First, the respondent has to state the average time spent using mobile applications last week. Secondly, respondents are asked to state the number of mobile apps used weekly on average.

Finally, the respondents are asked if they participate or have been participating in volunteering activities (being a sport trainer, helping in non-profit organization, donating blood, contributing to Wikipedia).

5.4. Attribute refinement

According to Johansson et al. (2021), attribute refinement is an important step in the process of determining attributes to be included in a discrete choice experiment. First, the attributes and the purpose of the data collection presented to respondents are presented. Then, a final selection of attributes and attribute levels is made.

5.4.1. Specification of attributes

The selected attributes and attribute levels were presented to a few people with varying expertise on the topic. After doing this, a few adjustments were made to refine the selected attributes and attribute levels.

First, the attribute “organizer of sensing campaign” was changed to “data use”, specifying what kind of party is collecting the data and with what aim. This was done in order to make this attribute more specific and give respondents a better idea of what will be done with their data.

Furthermore, the attribute level “environmental data” was changed to “contextual data”, since this more specifically refers to data collected about the surrounding context of an individual, such as temperature and humidity.

5.4.2. Specification of purpose of data collection

The choice situations in the experiment are specifically framed in the context of *mobility*, which is a relevant field in which participatory sensing can be of great use. Collecting data from the devices of individuals can have benefits in the context of mobility for governments, transport operators, companies, and individuals. The main benefits are the prediction of traffic flows and the improvement of trip quality. The following benefits are communicated to respondents in the choice experiment:

Predicting traffic flows
Using data collected by individuals, traffic flows can be predicted. A municipality can use these insights to manage crowds and congestion in a city, and to analyse the demand for different modes of transport. By gaining insights in demand for transportation, transport operators are able to adjust and optimize fleet management. For users of public transport, this will lead to fewer waiting times and less crowded trains, metros or buses. Furthermore, they will receive improved travel advices since the application will have insight in their personal habits and preferences regarding transport. For car drivers, better route planning can be facilitated based on predicted congestion, which will result in time savings. Also, these information can be used to find methods to reduce congestion.

Enhancing trip quality
Also, information on driving behaviour can be gathered, for example by using data on acceleration or braking. Furthermore, road conditions can be analysed by measuring vibrations, which can indicate bumpiness in the road. Government authorities and transport operators can use this information to improve infrastructure, to make roads more safe and comfortable. For car drivers, aggregated user data on driving behaviour and road and traffic conditions can be used to make predictions on e.g. the DTE (Distance-To-Empty), the distance they can travel before their (electric) vehicle runs out of fuel. Besides that, car manufacturers can use these data to get insight in car conditions, shelf lives and durability and improve their products based on these insights. Transport authorities can measure the trip quality of travellers by using data on waiting times, distance, the bumpiness of a ride, or other inputs by users. This can be used to improve their services and deliver better travel experiences to travellers.

The attributes in the choice experiment are framed in the context of these benefits. Types of data being collected are time and location data, motion data, contextual data, and multimedia data. *Location data* are collected to analyse transport flows and predict travel demand. *Motion data* are collected to analyse driving behaviour, analyse user activity (walking, cycling, driving), and to measure road conditions. *Contextual data* are collected to measure trip quality with respect to environmental conditions, such as temperature, humidity, and air pollution. This can be used to give advice to travellers on routes with more healthy environmental conditions. Lastly, *multimedia data* are collected to characterize places more easily by using location-tagged images and videos. Also, sound samples can be used to identify noisy traffic or noise pollution in public transport.

Collected data can be used by governmental institutions, academic institutions, corporate institutions, or societal organisations. Municipalities or other governmental authorities can use crowd sensed data to improve mobility in urban spaces by offering different modes of transport and good infrastructure. Academic institutions such as a university or a research institute can use data to investigate mobility or analyse opportunities for alternative modes of transport. Commercial companies aim to use data to either improve existing products or services, or to design new products or services. Lastly, sensing campaigns can be organised bottom-up by a civil organisation, aiming to address a local problem (such as parking pressure, traffic congestion, or traffic safety) and provide data to support decision making.

5.4.3. Final selection of attributes and attribute levels

The final attributes that are used in the choice experiment are presented in the following tables. Table 7 presents the factors that are used in the structural model. In Table 8, the factors related to perceptions are presented. Finally, Table 9 presents the factors related to personal characteristics.

Table 7. Operationalisation of factors in structural model

Factor	Levels	Level coding
Monetary reward	€0/month	0
	€20/month	1
	€40/month	2
	€60/month	3
Effort	Low	0
	Moderate	1
	High	2
Risk of re-identification	10% (1 out of 10)	0
	20% (1 out of 5)	1
	30% (1 out of 3)	2
Type of data	Time and location data	0
	Time and location data, Motion data	1
	Time and location data, Motion data, Contextual data	2
	Time and location data, Motion data, Contextual data, Multimedia data	3
Data use	Governmental institution aiming to improve mobility	0
	Academic institution aiming to investigate transport modes	1
	Corporate institution aiming to improve products or services	2
	Societal organisation aiming to address local issues related to mobility	3

Table 8. Operationalisation of attitude factors

Factor	Levels	Level coding
Privacy Segmentation Index	Privacy Unconcerned	0
	Privacy Pragmatist	1
	Privacy Fundamentalist	2
Distrust Index	No distrust	0
	Low distrust	1
	Medium distrust	2
	High distrust	3

Table 9. Operationalisation of factors related to personal characteristics

Factor	Levels	Level coding
Age	18-25	0
	26-35	1
	36-45	2
	46-55	3
	56-65	4
	>65	5
Gender	Male	0
	Female	1
	Other	2
Education	Basisonderwijs	0

	Vmbo, havo-, vwo-onderbouw, mbo-1	1
	Havo-, vwo-bovenbouw, mbo-2, mbo-3, mbo-4	2
	Hbo-, wo-bachelor	3
	Hbo-, wo-master, doctor	4
Income level	<€10.000	0
	€10.000-€20.000	1
	€20.000-€30.000	2
	€30.000-€40.000	3
	€40.000-€50.000	4
	€50.000-€100.000	5
	>€100.000	6
Average time spent using mobile applications last week (hours)	None	0
	1-20	1
	21-40	2
	41-60	3
	61-80	4
	81-100	5
	>100	6
Number of mobile apps used weekly on average	None	0
	1-10	1
	11-20	2
	21-30	3
	>30	4
Participation in altruistic activities	No	0
	Yes	1

Using the identified attributes, the conceptual model as described in chapter 2 is extended. This Privacy Calculus model, applied to our study, is presented in Figure 7. The solid lines represent direct effects. The dashed lines represent interaction effects.

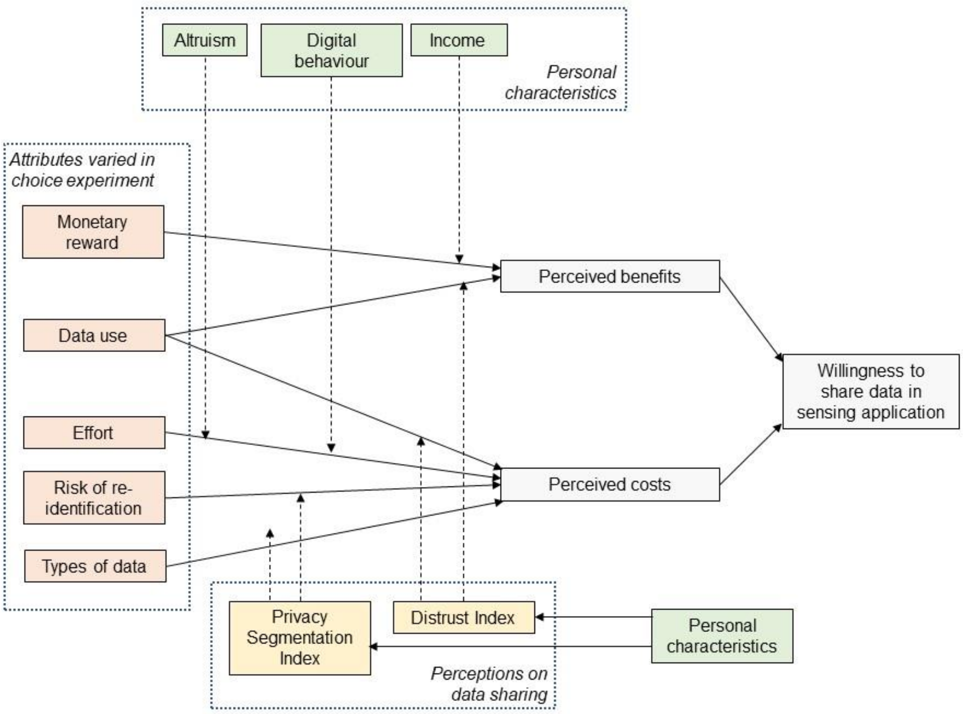


Figure 7. Extended Privacy Calculus model

The five factors displayed at the left side of the model are the attributes that are varied in the choice experiment. These factors influence the perceived benefits and costs of sharing data. Monetary rewards are assumed to incentivize participation in data sharing. The *data use* attribute can be either an incentive or a disincentive to participate. When an individual has a good relationship with the party collecting the data or is motivated to contribute to the objective of the organization, this may be a stimulation for this person to contribute by sharing data. However, when a person does not trust a certain party, this attribute may be a disincentive to participate. The attributes *effort*, *risk of re-identification*, and *types of data* are factors related to perceived costs of sharing data, since a certain amount of time or privacy is sacrificed when sharing data.

5.5. Experiment design generation

For constructing the experiment, different design types can be considered. A first option is a full factorial design, which includes all possible choice situations and thus allows for estimating all possible effects. However, this leads to a too large number of choice sets for a practical study. An alternative is a fractional factorial design, which uses only a set of choice sets from the full factorial design, and therefore has a smaller number of alternatives (ChoiceMetrics, 2018).

The *orthogonal* design is the most well-known fractional factorial design type. Orthogonal designs ensure that the attribute levels have zero correlation, which means they are independent of each other. Furthermore, all pairs of attribute levels occur equally often across all pairs of alternatives in orthogonal designs, and attribute levels are balanced, which means that each level occurs an equal amount of times for each alternative (Mariel et al., 2021). However, orthogonal designs also have some limitations. These designs cannot prevent choice situations in which one alternative is obviously more preferred than the others (ChoiceMetrics, 2018), which are so-called dominant alternatives.

Another type of fractional factorial designs, is the *efficient* design, which aim to be efficient in terms of trying to maximize the information from each choice situation. These efficient designs are able to perform better compared to orthogonal designs, however prior estimates need to be available. This means that efficient designs rely on the accuracy of the prior values (ChoiceMetrics, 2018).

In this research, a fractional factorial design is used since it allows for a smaller number of alternatives compared to a full factorial design. Since no priors are available, an orthogonal design is used.

The software NGene is used to construct the experiment design. The syntax used for the construction can be found in Appendix A. Sequential construction of alternatives is applied, since the experiment includes unlabelled alternatives, which have the same attributes and levels (ChoiceMetrics, 2018). NGene found 36 choice sets as the minimum number of choice scenario allowing for an orthogonal design. Because this number is still quite large and can be perceived too exhausting to complete for a single respondent, the choice sets are divided into three smaller blocks of choice sets by using *blocking*. Thus, there are 3 blocks with 12 choice sets each. The blocks are not orthogonal by itself, but the full design is still orthogonal (ChoiceMetrics, 2018). Respondents are randomly assigned to one of the blocks. The final experiment design is presented in Appendix B.

5.6. Survey design

The survey is constructed using the online software Qualtrics[1]. On the first page of the survey respondents are asked to give their consent for participating in the research. Then, the question is asked if the respondent owns a smartphone or another smart device. If the respondent does not consent or does not own a smart device, the respondent will automatically be directed to the end of the survey. The survey is structured as follows:

Introduction to the topic

First, a general introduction is given on participatory sensing. Then, an explanation is given on the benefits of participatory sensing in the context of mobility. The communication of the benefits is supported by visualisations. Finally, the set-up of the experiment is presented to the participant. The participant will see an example of a hypothetical choice situation, as well as an explanation of the attributes and attribute levels

[1] Qualtrics is an online survey software, licensed by TU Delft.

varied in the experiment. The rest of the survey consists of three main parts, which are the choice experiment, questions on perceptions, and questions on personal characteristics.

Part 1: Choice experiment
First, the respondents are presented with the choice scenarios. Respondents are randomly assigned to one of the three blocks of choice situations. Then, respondents are presented with 12 choice situations, in which they have to make a choice whether they would participate in sharing data in that situation or not. For each situation, they have to choose “Yes” or “No”.

Part 2: Questions regarding perceptions on privacy and trust
In the second part, respondents receive questions regarding their perceptions on privacy and trust regarding data sharing. They are asked to rate respectively 3 and 4 statements on a Likert-scale from 1 to 5 (“Strongly disagree”, “Somewhat disagree”, “Undecided”, “Somewhat agree”, “Strongly agree”).

Part 3: Questions regarding personal characteristics
Finally, respondents receive 7 questions on demographics and personal characteristics.

The survey conducted is completely anonymous. No personal information potentially leading to the identification of respondents is collected. Before distributing the survey, the survey has been approved by the Human Research Ethics Committee of the TU Delft, in order to ensure the ethical protection of respondents.

5.7. Survey distribution

Before distribution of the survey, the survey was pre-tested with a few people having varying expertise on the topic in order to make sure the questions are clear and to determine the time needed to complete the survey. From this pre-investigation, it appeared that participants need approximately 10 to 15 minutes to complete the survey. The final survey can be found in Appendix C. The survey was made available in both English and Dutch.

The survey was distributed among people owning a smart phone or a smart device. Among others, several residents’ associations were asked by e-mail to distribute the survey among their members (Amsterdam, Delft, Den Haag, Eindhoven, Rotterdam). An e-mail was sent to consultants within AKKA as well. Furthermore, the survey was distributed through social media. This included the internal social media platform Yammer within AKKA Technologies.

Several methods are proposed in literature to determine the sample size required for the model estimation of a discrete choice experiment. The rule of thumb that was most commonly cited uses the following formula in order to determine the sample size (Orme, 2010; Rose & Bliemer, 2013):

$$N \geq 500 \cdot \frac{L^{max}}{J \cdot S}$$

Here, N is the minimum number of respondents, L^{max} is the largest number of levels for any of the attributes, J is the number of alternatives and S is the number of choice sets presented to each respondent. Using these formula, the minimum amount of respondents needed for this experiment is 500·4/(2·12) = 83. In chapter 6, the results of the choice experiment are discussed.

6
Choice modelling

This chapter discusses the data analysis phase, in which the choice models are estimated. The following research question is addressed in this chapter.

SQ3. What trade-offs do individuals make between potential costs and benefits of participation, when choosing to share data in participatory sensing systems?

In 6.1, the data collected through the online survey is described. Subsection 6.2 addresses the results of the Multinomial Logit model estimation, which is considered the “base model”. In subsection 6.3, the Latent Class analysis is discussed. Subsection 6.4 further examines the people who chose the same alternative in each situation, after which the preferences of the group of people with an extreme score on the Privacy Index and the Distrust Index are compared to the preferences of the whole sample in 6.5. In subsection 6.6, the factors varied in the experiment are tested for non-linearity between levels. Subsequently, the results of the estimated ML panel models are analysed in 6.7. Subsection 6.8 provides an overview of all estimated models and argues for the final model to be selected. In subsection 6.9, the model is validated. Finally, the model is interpreted in 6.10, after which a conclusion of the chapter is drawn in 6.11.

6.1. Description of collected data

The online survey was published on October 29th 2021 and was available until November 12th 2021. After this period, the data was downloaded from Qualtrics. First, the data was cleaned and treated in SPSS to make it suitable for model estimation. Also, a closer look was taken at the collected data by describing the data.

6.1.1. Data cleaning

A total amount of 196 respondents participated in the survey. However, 63 responses appeared to be unfinished, and were therefore deemed useless. Also, in five cases the respondent indicated not owning a smart phone or a smart device, which meant they were directed to the end of the survey. Furthermore, three respondents did not give consent to participate in the research. Thus, these records were deleted from the dataset. Five people did not answer one to three questions that were related to personal characteristics, such as their income level. However, despite the missing data, these records are kept in the dataset since their responses to the choice situations are still useful. This leaves us with a number of 125 appropriate records for further analysis. Of these 125 respondents, 38 participated in block 1 (30,4%), 40 in block 2 (32,0%), and 47 in block 3 (37,6%).

6.1.2. Representativeness of sample

An important question that should be asked is if the sample is representative for the population. Therefore, the frequencies and percentages per category are presented in Table 10. From this table, it appears that the group of people between 18 and 35 years old is overrepresented in the sample, while, the group of people older than 55 is underrepresented. Furthermore, there were relatively more men who participated in the survey than women. When looking at the percentages for education levels, it appears that the people in the sample are relatively high-educated. Lastly, the income groups of <€10.000 and >€50.000 are overrepresented in the sample, while the group of people earning €10.000 to €30.000 is rather underrepresented. When interpreting the data, it should be taken into account that the sample is not fully representative.

Table 10. Representativeness of sample

Factor	Level	Frequency	Percentage in sample	Percentage in population
Age[a]	18-25	37	29,6%	12%
	26-35	36	28,8%	16%
	36-45	15	12,0%	15%
	46-55	20	16,0%	17%
	56-65	6	4,8%	17%
	>65	11	8,8%	23%
Gender[a]	Male	81	64,8%	49%
	Female	44	35,2%	51%
	Other	-	-	-
Education[b]	Basisonderwijs	1	0,8%	9%
	Vmbo, havo-, vwo-onderbouw, mbo-1	17	13,6%	20%
	Havo-, vwo-bovenbouw, mbo-2, mbo-3, mbo-4	48	38,4%	36%
	Hbo-, wo-bachelor	48	38,4%	21%
	Hbo-, wo-master, doctor	59	47,2%	13%
Income level[c]	<€10.000	31	24,8%	14%
	€10.000-€20.000	6	4,8%	23%
	€20.000-€30.000	11	8,8%	18%
	€30.000-€40.000	14	11,2%	15%
	€40.000-€50.000	15	12,0%	11%
	€50.000-€100.000	37	29,6%	16%
	>€100.000	6	4,8%	3%
	Missing	5	4,0%	-

Sources: a. CBS (2021a); b. CBS (2021b); c. CBS (2020)

6.1.2. Other descriptive statistics

Besides data on demographics, other data related to personal characteristics and behaviour were gathered. Table 11 presents the results for these questions.

Table 11. Descriptive statistics on personal characteristics

Factor	Level	Frequency	Percentage in sample
Average time spent using mobile applications last week (hours)	1-20	49	39,2%
	21-40	48	38,4%
	41-60	20	16,0%
	61-80	4	3,2%
	81-100	1	0,8%
	>100	2	1,6%
	Missing	1	0,8%
Number of mobile apps used weekly on average	None	0	0,0%
	1-10	77	61,6%
	11-20	40	32,0%
	21-30	5	4,0%
	>30	2	1,6%
	Missing	1	0,8%
Participation in altruistic activities	Yes	94	75,2%
	No	31	24,8%

6.1.3. Responses per choice set.

Figure 8 shows that the “yes” option was chosen in 39% of the cases, and the “no” option was chosen in 61% of the cases.

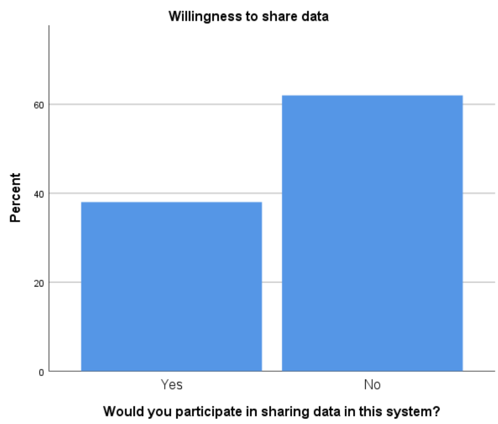


Figure 8. Willingness to share data

In Appendix D, a table is included that presents the amount of times that the “yes” and the “no” option were chosen per choice set. From this table, it appears that the choice scenario that was most accepted by respondents is choice set 5. 34 of the 38 participants (89%) that responded to this choice set indicated that they would share data in this situation. The characteristics of this choice set are presented in Figure 9.

Monetary reward	€60 per month
Effort	Low
Risk of re-identification	10%
Types of data	Time and location
Data use	Academic institution aiming to investigate transport modes

Figure 9. Choice set 5

The choice sets that were seen as least acceptable by participants were choice set 7 and choice set 12. In both of these choice situations, only 2 out of 38 participants (5%) indicated they were willing to share their data. The characteristics of these choice sets are presented in Figure 10 and Figure 11.

Monetary reward	None
Effort	High
Risk of re-identification	10%
Types of data	Time and location, Motion, Contextual, Multimedia
Data use	Corporate institution aiming to improve products or services

Figure 10. Choice set 7

Monetary reward	€20 per month
Effort	High
Risk of re-identification	30%
Types of data	Time and location, Motion
Data use	Corporate institution aiming to improve products or services

Figure 11. Choice set 12

6.1.4. Recoding into Westin's Indexes

First, the perception statements 2 and 3 for the Privacy Segmentation Index had to be inversed, as well as the statements 2 and 4 for the Distrust Index. These variables were re-coded in SPSS so that a high value (5) means a high concern, and a low value (1) means a low concern.

The scores of the Likert scales were summed in order to assign respondents to a category of Westin's Privacy and Distrust Indexes. In this way, the score for the Privacy Index ranges from a minimum of 3 to a maximum of 15 and the Distrust Index ranges from 4 to 20. These "Privacy scores" and "Distrust scores" are used in the estimations of the model.

In order to see how respondents perceive privacy and trust, they are categorized using the coding in Westin's research. The people who gave privacy concerned answers to all statements, are considered Privacy Fundamentalists. Privacy Unconcerned respondents are those respondents who give only privacy unconcerned answers. All other respondents are categorized as Privacy Pragmatists. Using these categories, we found that 6% of the respondents belonged to the Privacy Unconcerned category, 47% to the Privacy Pragmatists, and 47% to the Privacy Fundamentalists. This is also visualized in Figure 12.

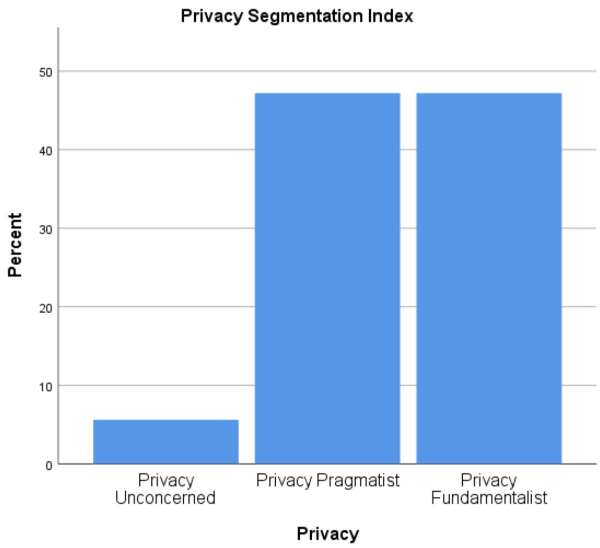


Figure 12. Privacy Segmentation Index

There might be selection bias in the survey, which means that people who do not want to share data also did not want to participate in the survey. For example, there were 3 respondents that did not give consent to participating in the survey. This means that their preferences are not included in the dataset. People that are reluctant to share any data mainly belong to the Privacy Fundamentalist category. Thus, one might expect that if there is a selection bias, the Privacy Fundamentalist category may be underrepresented in the survey compared to the population. Therefore, it is important to check if the categories in the sample are similar to those within the population.

In order to see if the percentages found in the sample are similar to the population, we compared them to percentages found in Westin's research, within the American population. These are reported in research by Kumaraguru & Cranor (2005). More recent surveys on the Privacy Index are conducted by Google Consumer Surveys (GCS) and Amazon's Mechanical Turk, reported by Woodruff et al. (2014). In Table 12, the percentages found in these researches are presented and compared to the percentages in the sample.

Table 12. Percentages per Privacy Index category

Classification	Sample	Westin (2001)	GSC1 (2014)	GSC2 (2014)	MTurk (2014)
Privacy unconcerned	6%	8%	6%	5%	10%
Privacy pragmatists	47%	58%	57%	58%	40%
Privacy fundamentalists	47%	34%	38%	37%	49%

It should be noted that the research by Westin is from 2001 and focused at the American population. However, it appears that there are no less Privacy Fundamentalists in our sample than in previous research. Thus, it is not likely that this group is underrepresented compared to the population.

For the Distrust Index, the same process was repeated. The frequencies of the categories are reported in Figure 13. A percentage of 8% belongs to the No Distrust category, 42% to the Low Distrust category, 38% to the Medium Distrust category, and 12% to the High Distrust category.

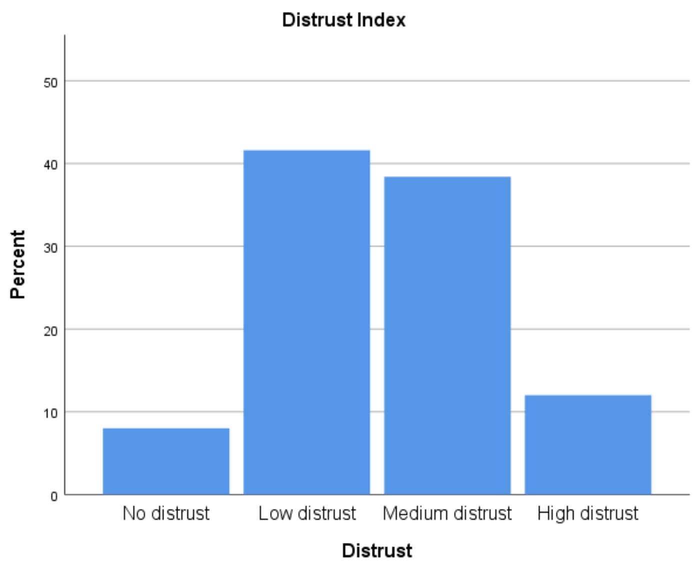


Figure 13. Distrust Index

The percentages found in the sample are again compared to the percentages found in Westin's research, reported by Kumaraguru & Cranor (2005). The percentages are presented in Table 13.

Table 13. Percentages per Distrust Index category

Classification	Sample	Westin (1994)
No distrust	8%	5%
Low distrust	42%	26%
Medium distrust	38%	38%
High distrust	12%	31%

The main difference that can be noted is that the Low Distrust category has a higher frequency in our sample, and the High Distrust category has a relatively low frequency in the sample. This could mean that people from the High Distrust category also did not want to fill in the survey, and are therefore underrepresented in the sample. This is something that should be taken into account when interpreting the results relating to the influence of the Distrust Index. However, it should again be noted that this research dates back to 1994, and was conducted within the American population. Current percentages in the population could be different.

6.1.5. Influence of demographic characteristics on perceptions

Several One-way ANOVA tests were conducted using SPSS in order to see if there were variations between different demographic groups regarding perceptions on data sharing. The influences of age, gender, and education level were tested.

For age, there were two statements that were perceived significantly different between age groups:

- "Technology has almost gotten out of control" (p=0.040): this statement was ranked highest by people of 56 years and older.
- "The way one votes has no effect on what the government does" (0.000): on average, this statement was ranked highest by people of 65 years and older, lowest by people between 46 and 65 years old, and moderately by people 45 or younger.

Furthermore, the privacy scores and distrust scores of different age groups were analysed. No significant differences between age groups were found for privacy scores. For the distrust scores, a significant difference was found between age groups. People of 65 years or older scored highest on the Distrust Index (average score of 15), while people between 46 and 55 years old appeared to be least distrustful (average score of 10).

No significant differences were found between male and female. Also, no significant differences were found between different education levels.

6.2. MNL model

Different MNL models as well as ML models were estimated, with and without interaction effects. Of these estimated models, only the 14 most interesting models are reported in this chapter. The motivations for including these models are discussed further in this subsection. The programming code used to estimate the models is presented in Appendices E, F, and G.

To prepare the data for estimating choice models, the dataset was restructured first in SPSS from "wide" to "long" format, so that every observation is one row. Since each of the 125 respondents had to make 12 choices, there are 12 observations for each respondent, resulting in a dataset with a total of 1500 records. First, the base model was estimated, which is the MNL model as described in chapter 3. Then, several variations on this base model were modelled.

6.2.1. Model 1: MNL model with only main effects

The first model estimated is the most simple model and includes the main effects of the factors mon, eff, rid, tod, and use. However, the factor *data use* is dummy coded, since this this factor has categorical levels. Table 14 shows how this factor is coded. The value of the first level, which is considered the reference category, is set to zero. The parameters use_{aca}, use_{cor}, and use_{soc} are estimated compared to the reference category, which is the governmental institution.

Table 14. Dummy coding of data use factor

	use_{aca}	**use_{cor}**	**use_{soc}**
Governmental institution aiming to improve mobility	0	0	0
Academic institution aiming to investigate transport modes	1	0	0
Corporate institution aiming to improve products or services	0	1	0
Societal organisation aiming to address local issues related to mobility	0	0	1

The systematic utility function is defined as follows:

$$V_{yes} = \beta_{yes} + \beta_{mon} \cdot mon + \beta_{eff} \cdot eff + \beta_{rid} \cdot rid + \beta_{tod} \cdot tod + \beta_{use_{aca}} \cdot (use == 1) + \beta_{use_{cor}} \cdot (use == 2) + \beta_{use_{soc}} \cdot (use == 3)$$
$$V_{no} = 0$$

where

V_{yes} = the systematic utility of sharing data
V_{no} = the systematic utility
β_{yes} = the base utility (constant) of choosing the "yes" option
β_{mon} = the marginal utility of the factor monetary reward
β_{eff} = the marginal utility of the factor effort
β_{rid} = the marginal utility of the factor risk of re-identification
β_{tod} = the marginal utility of the factor type of data

$\beta_{use_{aca}}$ = the marginal utility of the factor data use by an academic institution
$\beta_{use_{cor}}$ = the marginal utility of the factor data use by a corporate institution
$\beta_{use_{soc}}$ = the marginal utility of the factor data use by a societal organisation

The code that was used to estimate the base MNL model can be found in Appendix E. The estimated parameters of model 1 are presented in Table 15. The first column shows the parameters estimated in the model. The second column (*estimate*) presents the estimated weight of the factors. This can be seen as the utils gained or lost by 1 unit increase of the attribute. In the third column, the standard errors associated with the parameter estimates are displayed, which illustrate the variation of the estimate across the sample. Finally, the t-ratios are used to determine if the attributes have an effect on choices in the population. Factors with an indicated p-value that is higher than 0.05 are considered statistically insignificant, which means that no effect can be observed in the population.

Table 15. Estimates Model 1

Factor	Estimate	s.e.	Rob.t.rat.(0)	p(1-sided)
$\boldsymbol{\beta_{mon}}$	0.3154	0.0522	4.957	0.0000
$\boldsymbol{\beta_{eff}}$	-0.6982	0.0783	-8.404	0.0000
$\boldsymbol{\beta_{rid}}$	-0.4459	0.0770	-4.280	0.0000
$\boldsymbol{\beta_{tod}}$	-0.3685	0.0534	-6.351	0.0000
$\boldsymbol{\beta_{use_{aca}}}$	0.3157	0.1754	1.752	0.0399
$\boldsymbol{\beta_{use_{cor}}}$	-0.6582	0.1810	-3.446	0.0000
$\boldsymbol{\beta_{use_{soc}}}$	-0.1905	0.1584	-1.397	0.0812
$\boldsymbol{\beta_{yes}}$	0.7923	0.1835	3.649	0.0000

From the estimation results, it appears that the parameter for data use by societal organisations is insignificant. This means that this parameter is not different from the reference level, which is the governmental organisation. Although the estimate for this parameter is not significant, this estimate still provides more information than “knowing nothing”. Since this research is aimed at improving the design of participatory sensing applications, the best guess is still that $\beta_{use_{soc}}$ is equal to -0.1905. Therefore, we choose to keep this parameter in the model. Furthermore, it should be noted that the parameter for data use by an academic institution is significant on a 1-sided p-value, but not on a 2-sided level. Since no clear expectation exists of the sign of this parameter (e.g. that an data use by an academic institution is more acceptable than by a governmental institution), the 2-sided p-value should be considered for this parameter. Thus, this is something that should be taken into account when interpreting the differences between data use by different organisations.

Based on the estimates for the *data use* parameter, the relative importance of the different data users can be derived. It can be concluded that in this sample, respondents are relatively positive about sharing data with academic institutions. Participants are most willing to share data with academic institutions, secondly with governmental institutions, thirdly with societal organisations, and least willing to share with corporate institutions.

The final Log-Likelihood of model 1 is -878.24. In order to evaluate if this model performs better than the model that determines choices by “throwing a dice”, a Likelihood Ratio Test is performed. The formula to calculate the Likelihood Ratio Score (LRS) is as follows:

$$LRS = -2 \cdot (LL_A - LL_B)$$

where
LL_A = Null model
LL_B = Estimated model

The Likelihood Ratio Test resulted in a value of 322.96. This is higher than the Chi-Square value with 8 degrees of freedom, which is equal to 15.507. Thus, we can conclude that the estimated model fits the data better than the model of throwing a dice.

The Rho-square is equal to 0.1553. This means that the estimated model explains 15.53% of the initial uncertainty. The BIC (Bayesian Information Criterion) value equals 1814.98. This is a criterion for scoring and comparing models, based on their Log-Likelihood and their complexity. A lower BIC value is considered better than a higher BIC value.

By using the utility ranges of the attributes, the relative importance of each attribute can be calculated. The relative importance is obtained by calculating the utility contribution of an attribute as a percentage of the sum of utility contributions. The utility contribution is calculated by multiplying the estimate with the maximum value of the attribute. For example, the utility contribution of the *monetary reward* attribute in model 1 is equal to $0.3154 \cdot 3 = 0.9462$. When dividing this by the sum of maximum utility contributions, which is equal to 5.3138, we obtain a percentage of 18%.

From Figure 14, it can be concluded that the required *effort* is the most important to respondents (26%). The *risk of re-identification* (17%) shows to be of least importance. However, there is only a slight difference with the *monetary reward* attribute and the *data use* attribute (18%).

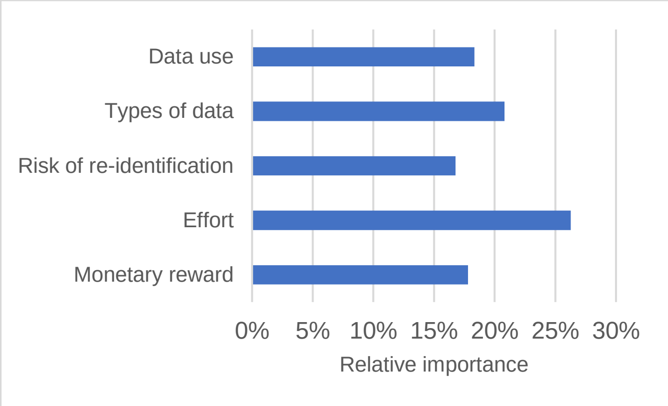


Figure 14. Relative importance of attributes (Model 1)

6.2.2. Model 2: MNL model with interaction effects

According to the literature review, the way in which participants perceive privacy and trust can influence their willingness to share data. Therefore, interaction effects are added to the model estimated previously. The privacy score and the distrust score are defined as the sum of the scores given by respondents on the statements regarding sharing data. Since the statements are scored on a Likert scale from 1 to 5, the range of the Privacy Index is from 3 to 15, and the Distrust Index ranges from 4 to 20.

The first effect that is estimated is the effect of distrust on the importance of the data use attribute. If people are distrustful about how organisations handle their data, or about technology in general, they may find the party that is collecting their data and for what purpose more important. Furthermore, the effects of privacy perceptions on the types of data and risk of re-identification are estimated. If a person is more concerned about privacy, these attributes might be more important.

In order to estimate the interaction effects, 5 parameters are added to the model. The systematic utility function for model 2 is defined as follows:

$$V_{yes} = \beta_{yes} + \beta_{mon} \cdot mon + \beta_{eff} \cdot eff + (\beta_{rid} + \beta_{rid_{privscore}} \cdot privscore) \cdot rid + (\beta_{tod} + \beta_{tod_{privscore}} \cdot privscore) \cdot tod + (\beta_{use_{aca}} + \beta_{use_{aca_{distscore}}} \cdot distscore) \cdot (use == 1) + (\beta_{use_{cor}} + \beta_{use_{cor_{distscore}}} \cdot distscore) \cdot (use == 2) + (\beta_{use_{soc}} + \beta_{use_{soc_{distscore}}} \cdot distscore) \cdot (use == 3)$$

$$V_{no} = 0$$

The code that was used to estimate model 2 can be found in Appendix E. The results of the estimated model are presented in Table 16.

Table 16. Estimates Model 2

	Estimate	s.e.	Rob.t.rat.(0)	p(1-sided)
β_{mon}	0.3366	0.0538	4.936	0.0000
β_{eff}	-0.7174	0.0804	-8.489	0.0000
β_{rid}	0.0033	0.2842	0.008	0.4968
β_{tod}	0.0653	0.1969	0.254	0.3999
$\beta_{use_{aca}}$	1.3232	0.4977	2.046	0.0204
$\beta_{use_{cor}}$	0.9659	0.5262	1.597	0.0551
$\beta_{use_{soc}}$	1.5354	0.4655	2.571	0.0051
β_{yes}	0.8249	0.1858	3.768	0.0000
$\beta_{distscore\_useaca}$	-0.0843	0.0406	-1.475	0.0701
$\beta_{distscore\_usecor}$	-0.1424	0.0448	-2.666	0.0038
$\beta_{distscore\_usesoc}$	-0.1513	0.0391	-2.823	0.0024
$\beta_{privscore\_tod}$	-0.0455	0.0258	-1.259	0.1039
$\beta_{privscore\_rid}$	-0.0422	0.0178	-1.813	0.0349

From the estimations, it appears that the interaction effects have a negative sign. This is as expected, since this means that the attributes related to trust and privacy (*risk of re-identification, types of data,* and *data use*) are most important to respondents scoring high on the Privacy and Distrust Index (Privacy Fundamentalists or High Concern), and less to respondents having a low score on Westin's indexes. However, the effect of the privacy score on the types of data is not significant (p=0.1039). Also, the interaction effect of the distrust score on data use by an academic institution is not significant.

Because of the correlation with the interaction effects, the main effects of the *risk of re-identification* and *types of data* parameters become insignificant. Also, we note that the data use by a corporate institution becomes insignificant, while data use by academic institutions and societal organisations becomes significant, contrary to model 1.

From this model, it can be concluded that people scoring higher on the Distrust Index are less willing to share data with corporate institutions and with societal organisations than people with a lower distrust score. Besides that, people with a higher score on the Privacy Index find the *risk of re-identification* factor more important and are less willing to share data if the risk of re-identification increases.

The final Log-Likelihood for this model is -842.21, the Rho-square is 0.1900, and the BIC is equal to 1779.48. This means model 2 performs better than model 1, which is the base model.

6.3. Latent Class Model

A goal of traditional Latent Class Analysis (LCA) is to determine the smallest number of latent classes *T* that is sufficient to explain away the associations (relationships) observed among the manifest variables (Magidson & Vermunt, 2004). In order to check for patterns in the data, a Latent Class Analysis is conducted. Classes are expected to have different preferences regarding factors (betas), and different personal characteristics. The Latent Class model is operationalised according to the conceptual model described in section 5.4. Since personal characteristics are assumed to influence the willingness to share data indirectly and not directly according to this model, the β_{yes} (the general inclination to participate) is set fixed among classes.

First, a 2-class model was estimated. This model appeared to perform better on the data then the 3-class model, that was estimated afterwards. Furthermore, the privacy score, distrust score, and factors related to personal characteristics were added as covariates, in order to see if there is a significant difference between classes. Since there were 5 respondents who did not answer all the questions on personal characteristics (such as income), these were deleted from the dataset for this exercise.

The Latent Class model was estimated in *R*Studio. The code that was used for estimation is presented in Appendix F.

6.3.1. Model 3: LC model with covariates

The 2-class model with covariates is presented in Table 17. In this table, the betas for class A and class B are presented. Also, the gammas are presented per class, which are the covariates. The utility for class B is set to zero, which means this is the reference level to which class A is compared.

Table 17. Estimates Model 3

	Estimate	s.e.	Rob.t.rat.(0)	p(1-sided)
β_{mon_a}	0.7296	0.1172	4.029	0.0000
β_{mon_b}	0.1859	0.0847	1.301	0.0966
β_{eff_a}	-0.6543	0.1424	-3.128	0.0000
β_{eff_b}	-1.0144	0.1170	-7.453	0.0000
β_{rid_a}	0.0797	0.1693	0.315	0.3763
β_{rid_b}	-0.8942	0.1257	-6.116	0.0000
β_{tod_a}	-0.3080	0.1057	-2.361	0.0091
β_{tod_b}	-0.5088	0.0787	-6.050	0.0000
β_{useaca_a}	0.2653	0.3403	0.895	0.1855
β_{useaca_b}	0.3562	0.2736	1.004	0.1576
β_{usecor_a}	-1.3369	0.3257	-2.609	0.0045
β_{usecor_b}	-0.3594	0.2822	-1.345	0.0894
β_{usesoc_a}	-0.1625	0.3092	-0.505	0.3068
β_{usesoc_b}	0.0997	0.2393	0.443	0.3287
β_{yes}	1.0753	0.2062	4.532	0.0000
delta_a	6.0805	2.5122	2.251	0.0122
delta_b	0	NA	NA	NA
gamma_age_a	-0.6748	0.3475	-1.813	0.0349
gamma_age_b	0	NA	NA	NA
gamma_gen_a	-1.0542	0.7317	-1.235	0.1085
gamma_gen_b	0	NA	NA	NA
gamma_edu_a	0.7285	0.5568	1.173	0.1203
gamma_edu_b	0	NA	NA	NA
gamma_inc_a	0.4065	0.2356	1.324	0.0928
gamma_inc_b	0	NA	NA	NA
gamma_tim_a	0.9793	0.3807	1.868	0.0309
gamma_tim_b	0	NA	NA	NA
gamma_app_a	-1.6924	0.8147	-1.362	0.0866
gamma_app_b	0	NA	NA	NA
gamma_alt_a	-1.9946	0.8990	-1.590	0.0560
gamma_alt_b	0	NA	NA	NA
gamma_S1_1_a	-0.3097	0.3600	-0.848	0.1981

gamma_S1_1_b	0	NA	NA	NA
gamma_S1_2_a	0.8320	0.4402	1.625	0.0521
gamma_S1_2_b	0	NA	NA	NA
gamma_S1_3_a	-1.5380	0.5312	-2.280	0.0113
gamma_S1_3_b	0	NA	NA	NA
gamma_S2_1_a	0.2676	0.3276	0.720	0.2358
gamma_S2_1_b	0	NA	NA	NA
gamma_S2_2_a	-0.3912	0.3321	-1.077	0.1407
gamma_S2_2_b	0	NA	NA	NA
gamma_S2_3_a	-0.5280	0.3360	-1.554	0.0601
gamma_S2_3_b	0	NA	NA	NA
gamma_S2_4_a	-1.1640	0.3869	-3.327	0.0000
gamma_S2_4_b	0	NA	NA	NA

The model assigned 33% of the choices to class A and 67% to class B. In order to draw conclusions at the population level regarding the differences between classes, the significance of the differences needs to be determined. This is done by taking the difference in parameter sizes, and dividing this by the standard error associated with this difference. The resulting value is the associated t-ratio. If this value is higher than 1.96, it can be concluded that the difference is statistically different from 0 at a 5% level of significance. The standard errors and t-ratios associated with the differences between parameters are displayed in Table 18.

Table 18. Significance of differences

	Diff. in parameter	**s.e. (diff.)**	**t-ratio (diff.)**
$\boldsymbol{\beta_{mon}}$	0.5437	0.0383	14.20
$\boldsymbol{\beta_{eff}}$	0.3600	0.0731	4.93
$\boldsymbol{\beta_{rid}}$	0.9739	0.1065	9.15
$\boldsymbol{\beta_{tod}}$	0.2008	0.0464	4.33
$\boldsymbol{\beta_{useaca}}$	0.0909	0.0581	1.56
$\boldsymbol{\beta_{usecor}}$	0.9775	0.2451	3.99
$\boldsymbol{\beta_{usesoc}}$	0.2622	0.0969	2.70

From these results, it can be concluded that there is a difference between class A and class B in the population for most of the factors. Only for the data use by an academic institution, the t-ratio is not higher than 1.96. For the covariates, the gammas for class B are set to zero. Thus, for determining if there are significant differences between the gammas of class A and class B, the robust t-ratios from Table 17 can directly be compared to a value of 1.96. By doing this, the conclusion can be drawn that the preferences of people regarding the third Privacy Index statement as well as the fourth Distrust statements, are significantly different between class A and class B.

Based on the significant differences, the following conclusions can be drawn about people in class A and people in class B.

People in class B can be characterised as people that are concerned about their privacy and are not sure that current privacy laws are able to protect their privacy sufficiently. Also, they are relatively distrustful of corporate institutions. This is in line with model 2, which shows that people with a higher distrust score are less willing to share their data with corporate institutions. If the effort of participating in sensing applications increases, people in class B become reluctant to share their data. Furthermore, if the amount of different types of data increases, they are less willing to share data as well. Another factor that is highly important to the people in class B is the *risk of re-identification* factor. If risks are involved, they are less willing to share data. Since the monetary reward factor is of relatively low importance for class B, these people cannot be easily convinced to participate by offering a financial compensation for their contributions.

People in class A, on the other hand, are less sensitive to sensing applications requiring more effort than people in class B. The willingness of class A people to share data decreases when more types of data are being collected, but the decrease in willingness to share is lower than for people in class B. Furthermore, people in class A are less concerned about their privacy and are more trustful of corporate institutions, compared to people in class B. As a result, the risk of re-identification does not play an important role in their decision to share data or not. The amount of money they receive for participating, however, is a factor which is relatively important for people in class A. This means that offering a financial compensation can be successful in increasing the willingness to share data.

The final Log-Likelihood of the Latent Class model is -733.24, the Rho-square is 0.2654, and the BIC is 1684.65. These values indicate a better performance than the base model. Of these three tests, however, the final Log-Likelihood cannot be compared directly with the base model, since model 3 includes less observations.

6.4. Non-participation

When taking a closer look at the data, it appears that there are multiple respondents that did not want to participate in any situation. Also, some participants chose to share their data in every choice situation. The amount of people per block that chose the same alternative ("yes" or "no") in every situation is displayed in Table 19.

Table 19. Full participation and non-participation

	Full participation	**Percentage of block**	**No participation**	**Percentage of block**
Block 1	2	5.3%	2	5.3%
Block 2	0	0.0%	8	20.0%
Block 3	1	2.1%	8	20.0%
Total	3	2.4%	18	14.4%

The situation where a respondents always chooses the same alternative across choice sets, is referred to as *non-trading* (Hess et al., 2010). Several explanations exist for this kind of choice behaviour. First, non-trading may reflect an extreme preference for a certain alternative. If a respondent has a strong preference for sharing no data at all, this will be reflected in a choice for the "no" alternative in all choice scenarios. Another explanation is that the non-trading behaviour arises from misunderstanding, boredom, or fatigue during the stated choice task. Lastly, the non-trading behaviour may reflect a form of political or strategic behaviour. This expresses itself mainly in the case of controversial topics (Hess et al., 2010). People may be so opposed to parties collecting their data that they will never choose the "yes" option.

The behaviour of respondents choosing the same alternative in each situation could also be related to a phenomenon called *lexicographic answering.* This means that some respondents always choose an option based on one attribute, notwithstanding the levels of the other attributes (Killi, 2007). An explanation for this behaviour is that only one attribute in the attribute set matters to the respondent, which is known as a *lexicographic preference.* Another explanation could be that there are steep indifference curves. This means that the interval of an attribute that is strongly preferred may be too small or too wide. Moreover, it could be that the range of the attribute levels is right, but the values are chosen too high or too low. As a consequence, trade-offs with other attributes are irrelevant. For example, in our experiment, it could be that a monetary reward is highly important to some respondents, and that these respondents require a financial compensation that is much higher than the maximum of €60 per month. As a result, they will not accept sharing data in any choice situation, irrespective of the other attributes.

In order to further examine the personal characteristics of people giving lexicographic answers, several descriptive analyses were conducted using SPSS. In order to do this, a dummy variable was used. This variable indicates a respondent choosing the "no" alternative for every choice set. Only the personal characteristics of the people answering "no" in each situation are analysed. Since only 3 people chose the "yes" alternative in each situation, analysing their characteristics did not provide useful insights. The results of the analyses in SPSS are presented in Appendix H.

From these analyses, it can be concluded that a relatively high percentage of people giving lexicographic answers belong to the “Privacy Concerned” category of Westin’s Privacy Index (72% vs. 43% in the rest of the sample). For the Distrust Index, people in the “High Distrust” category are relatively overrepresented in the subset of people answering lexicographically. Lastly, we found that the yearly income of people giving lexicographic answers is relatively high (€50.000-€100.000). This could indicate that in their opinion, the range of the monetary reward is too low. Also, it could mean that the monetary reward does not matter to them, since they cannot be incentivised by obtaining a small amount of money in return. Out of the 18 people that chose the “no” alternative in each situation, 4 did not provide an answer on the question on income (vs. 1 out of 107 for the rest of the sample). This could mean that those people chose to provide minimum information out of boredom or caused by fatigue. However, this could also mean that the people that indicated that they were not willing to share data in any choice situation, are more privacy concerned and reluctant to share information on their income in a survey.

6.4.1. Model 4: Non-participation

In order to analyse the effect of respondents choosing the same alternative in every scenario, model 4 is estimated. In this model, the participants that chose the same alternative in every choice scenario were excluded from the dataset. This led to the exclusion of 21 respondents in total. The estimates for the model with the exclusion of these respondents are presented in Table 20.

Table 20. Estimates Model 4

	Estimate	**s.e.**	**Rob.t.rat.(0)**	**p(1-sided)**
β_{mon}	0.3938	0.0584	5.238	0.0000
β_{eff}	-0.9141	0.0903	-10.724	0.0000
β_{rid}	-0.4623	0.0858	-3.881	0.0000
β_{tod}	-0.4189	0.0601	-6.135	0.0000
$\beta_{use_{aca}}$	0.2396	0.1958	1.212	0.1127
$\beta_{use_{cor}}$	-0.9665	0.2013	-4.751	0.0000
$\beta_{use_{soc}}$	-0.2612	0.1753	-1.548	0.0608
β_{yes}	1.3099	0.2154	5.255	0.0000

From these estimates, it appears that there is a slight difference with the parameter estimates of model 1. When removing the “non-trading” people from the dataset, it appears that the *effort* attribute becomes relatively more important (29%) and the *risk of re-identification* attribute becomes less important (14%). The other attributes do not differ much from the values of model 1. This can also be seen in Figure 15.

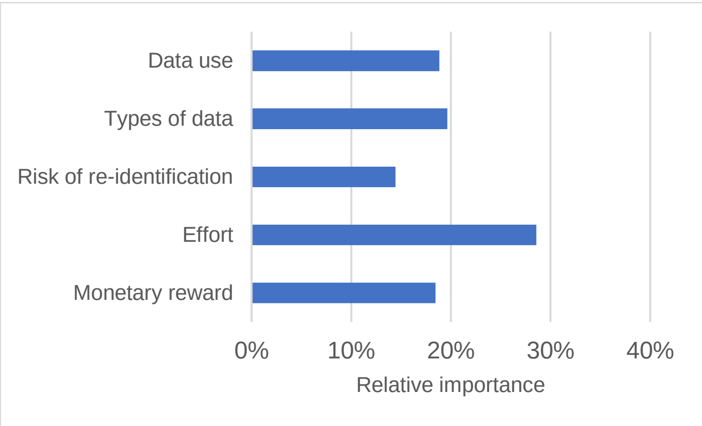


Figure 15. Relative importance of attributes (Model 4)

The relative importance of the risk of re-identification attribute decreases from 17% to 14% when excluding the people giving lexicographic answers. This is in line with our findings from the descriptive analyses. People choosing the “no” alternative in every situation are generally more concerned about their privacy. It is important to include these people in the dataset, since they represent a group of people that have such a high concern about their privacy that they do not want to share their data in any situation.

In terms of performance, the final Log-Likelihood is -716.03, the BIC is 1489.10, and the Rho-square is 0.1723.

6.5. Extreme preferences

It appears that a percentage of 47% of respondents can be categorised as “Privacy Fundamentalists”, according to Westin’s Privacy Index. In order to further examine the preferences of these Privacy Fundamentalists, a model was estimated using only the choices of respondents from this group. Besides that, a model was estimated on the subset of people who were categorised in the “High Distrust” group.

6.5.1. Model 5: Preferences of Privacy Fundamentalists

For the estimation of the model for the Privacy Fundamentalist group, a simple MNL model was used. The results of the estimated model are presented in Table 21.

Table 21. Estimates Model 5

Factor	Estimate	s.e.	Rob.t.rat.(0)	p(1-sided)
β_{mon}	0.2921	0.0790	3.295	0.0000
β_{eff}	-0.6034	0.1188	-5.004	0.0000
β_{rid}	-0.5521	0.1208	-3.584	0.0000
β_{tod}	-0.4121	0.0825	-4.284	0.0000
$\beta_{use_{aca}}$	0.3739	0.2644	1.294	0.0979
$\beta_{use_{cor}}$	-1.0035	0.2945	-2.955	0.0016
$\beta_{use_{soc}}$	-0.0995	0.2362	-0.462	0.3219
β_{yes}	0.6011	0.2671	1.835	0.0333

The subset with the Privacy Fundamentalist people consisted of 708 observations. Out of these 708 observations, the “yes” alternative was chosen 227 times (32% compared to 39% in the whole sample) and the “no” alternative was chosen 481 times (68% compared to 61% in the whole sample). This indicates that people being highly concerned about their privacy are slightly more reluctant to share data than people in the whole sample. Out of the 59 individuals categorised as Privacy Fundamentalists, 13 chose the “no” alternative in every choice situation.

Most estimated parameters are significant. Only the data use parameter is insignificant. When comparing the estimated model to the base model (Model 1), it can already be observed that there is a difference between the estimated parameters. A comparison between the relative importance of attributes of model 5 and model 1 is presented in Figure 16. The blue bars represent the importance for the Privacy Fundamentalist group, and the orange bars represent the importance for the whole sample.

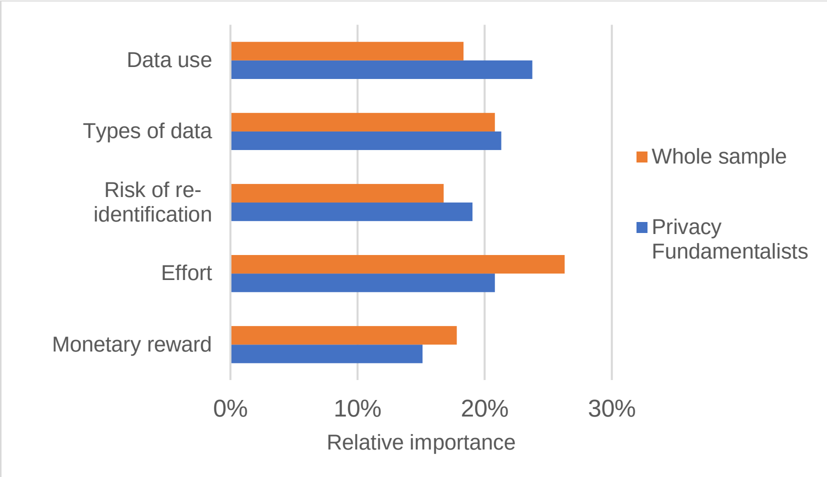


Figure 16. Relative importance of attributes (Model 5 vs. Model 1)

From this figure, it can be concluded that to people who are concerned about their privacy, the privacy related factors (*data use*, *risk of re-identification*, and *types of data*) are more important than for people who are less concerned about their privacy. Also, compared to the base model, the *effort* and the *monetary reward* parameters become less important. The biggest change is observed for the *data use* parameter, which has a relative importance of 26% for people who are categorised as Privacy Fundamentalists.

In Chapter 5, the statements that are used for creating the Privacy Segmentation Index were described. From these statements, we can conclude that to Privacy Fundamentalists are people who think consumers have no control over how personal information is collected and used by companies, who think that business do not handle personal information in a proper and confidential way, and who think that laws do not provide a sufficient level of privacy protection. For this group of people, the party collecting the data and the purpose of the data collection highly influence their decision on data sharing. Especially when the data is collected by a corporate institution, these people become reluctant to share their data.

6.5.2. Model 6: Preferences of High Distrust group

The same process was followed in order to investigate the preferences of people in the High Distrust category. The results are presented in Table 22.

Table 22. Estimates Model 6

Factor	Estimate	s.e.	Rob.t.rat.(0)	p(1-sided)
$\boldsymbol{\beta_{mon}}$	0.2271	0.2005	0.866	0.1934
$\boldsymbol{\beta_{eff}}$	-0.5193	0.3157	-1.988	0.0234
$\boldsymbol{\beta_{rid}}$	-0.9742	0.3459	-4.246	0.0000
$\boldsymbol{\beta_{tod}}$	-0.4681	0.2240	-3.943	0.0000
$\boldsymbol{\beta_{use_{aca}}}$	1.1923	0.7216	1.464	0.0716
$\boldsymbol{\beta_{use_{cor}}}$	0.4713	0.8492	0.635	0.2628
$\boldsymbol{\beta_{use_{soc}}}$	1.0076	0.7036	2.089	0.0184
$\boldsymbol{\beta_{yes}}$	-0.9811	0.6576	-1.663	0.0482

The subset with the respondents in the High Distrust group consisted of 180 observations. Because of this relatively low amount of observations, some parameters (monetary reward, data use) are not significant. Out of all observations, the “yes” alternative was chosen 27 times (15% compared to 39% in the whole sample), and the “no” alternative was chosen 153 times (85% compared to 61% in the whole sample). This indicates that people in the High Distrust group are more reluctant than people scoring lower on the Distrust Index. Out of the 15 people in the High Distrust subset, 6 peoples indicated they did not want to share data in any situation.

The relative importance of the attributes in model 6, compared to the relative importance of attributes in model 5, is presented in Figure 17. The blue bars represent the importance of factors for people in the High Distrust group, and the orange bars represent the importance of factors for the whole sample.

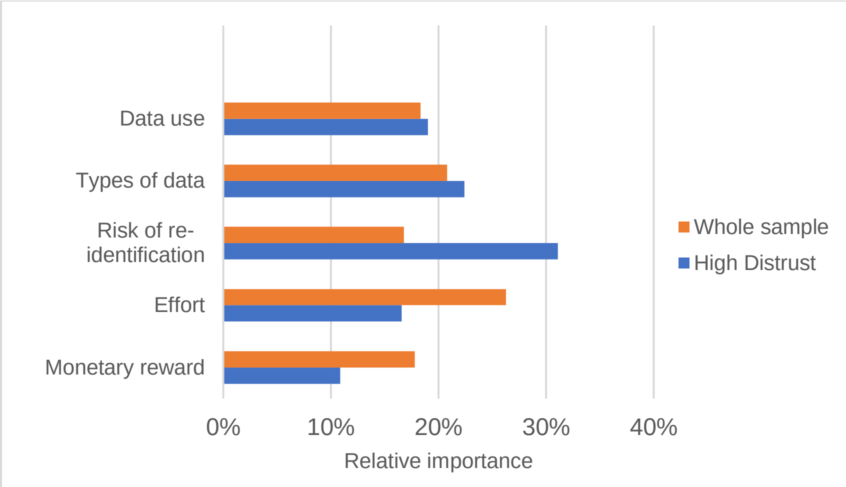


Figure 17. Relative importance of attributes (Model 6 vs. Model 1)

From this figure, it is immediately observable that there is a large difference in importance of the risk of re-identification parameter. The relative importance of this parameter for people in the High Distrust category is 14% higher than for people in the whole sample. Furthermore, the data use parameter and the types of data parameter are seen as somewhat more important by people who have a high distrust, which is something that was also observed for the privacy concerned people. The effort factor and the monetary reward factor are considered as less important.

From the statements used to derive the Distrust Index, we can conclude that people who are distrustful do not trust technology, do not trust the government to look after their interests, and think business do more harm than good to consumers. The distrustful attitude against the government can also be observed in the parameter estimates of model 6. To these people, a governmental institution collecting their data is least accepted, compared to other parties.

6.6. Testing for non-linearity

Non-linearity means that there is an unequal utility contribution from each attribute level of a factor. If this is the case, parameter estimates could be over- or underestimated. Therefore, it is important to evaluate if there is non-linearity in the attribute levels. In order to test if the attributes in the model show non-linear effects, the factors are re-coded to dummy variables. The first levels are the reference levels, which means their utility contribution is set to zero. By doing this, the utility contribution of each attribute level can be estimated. From these utility contributions, conclusions can be drawn on the linearity of the factors.

6.6.1. Non-linearity of monetary reward parameter

First, the *monetary reward* parameter is tested for non-linearity effects. The results are presented in Table 23.

Table 23. Estimates Model 7

	Estimate	s.e.	Rob.t.rat.(0)	p(1-sided)
β_{mon20}	0.4269	0.1895	2.102	0.0178
β_{mon40}	0.5754	0.1789	2.983	0.0014
β_{mon60}	0.9935	0.1640	5.049	0.0000
β_{eff}	-0.7029	0.0786	-8.421	0.0000
β_{rid}	-0.4372	0.0819	-4.321	0.0000
β_{tod}	-0.3711	0.0535	-6.422	0.0000
$\beta_{use_{aca}}$	0.2773	0.1889	1.490	0.0680
$\beta_{use_{cor}}$	-0.6885	0.1854	-3.682	0.0000
$\beta_{use_{soc}}$	-0.1934	0.1582	-1.439	0.0751
β_{yes}	0.7827	0.1845	3.547	0.0000

From the results, it can be concluded that an increase from no reward to €20 per month and an increase from €40 to €60 per month lead to a higher increase in utility than an increase from €20 to €40 per month. This is also displayed in Figure 18, in which the utility contributions per level are presented. The figure shows that the *monetary reward* parameter is non-linear.

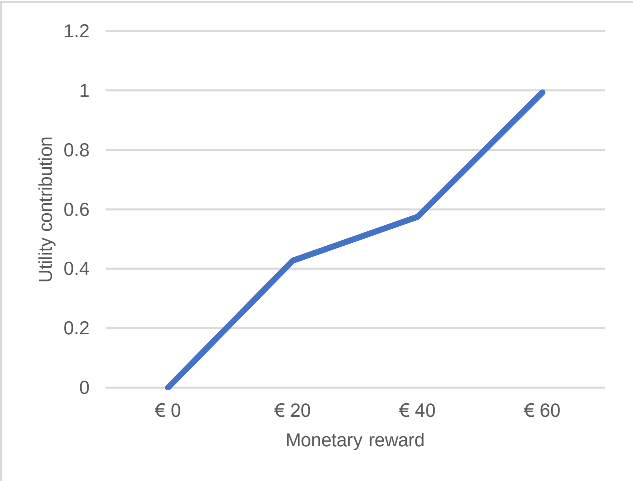


Figure 18. Utility contribution of mon attribute levels

6.6.2. Non-linearity of effort parameter

Furthermore, the *effort* attribute is dummy coded in order to test for non-linearity. The results are presented in Table 24.

Table 24. Estimates Model 8

	Estimate	s.e.	Rob.t.rat.(0)	p(1-sided)
β_{mon}	0.3210	0.0525	4.989	0.0000
β_{effmod}	-0.5105	0.1371	-4.422	0.0000
β_{effhig}	-1.4188	0.1587	-8.218	0.0000
β_{rid}	-0.4477	0.0769	-4.305	0.0000
β_{tod}	-0.3709	0.0534	-6.376	0.0000
$\beta_{use_{aca}}$	0.2969	0.1748	1.660	0.0484
$\beta_{use_{cor}}$	-0.6756	0.1807	-3.562	0.0000
$\beta_{use_{soc}}$	-0.1965	0.1587	-1.442	0.0747
β_{yes}	0.7328	0.1863	3.369	0.0000

It appears that an increase from moderate effort to high effort leads to a slightly higher increase in disutility than an increase from low to moderate effort. This is also displayed in Figure 19. A small effect of non-linearity can be observed.

Figure 19. Utility contribution of eff attribute levels

6.6.3. Non-linearity of risk of re-identification parameter

In model 7, the *risk of re-identification* attribute is tested for non-linearity. The results are displayed in Table 25.

Table 25. Estimates Model 9

	Estimate	s.e.	Rob.t.rat.(0)	p(1-sided)
β_{mon}	0.3147	0.0519	4.993	0.0000
β_{eff}	-0.7116	0.0790	-8.540	0.0000
β_{rid20}	-0.2128	0.1418	-1.789	0.0368
β_{rid30}	-0.8974	0.1558	-4.178	0.0000
β_{tod}	-0.3698	0.0535	-6.295	0.0000
$\beta_{use_{aca}}$	0.2997	0.1752	1.689	0.0456
$\beta_{use_{cor}}$	-0.6928	0.1830	-3.566	0.0000
$\beta_{use_{soc}}$	-0.1811	0.1576	-1.354	0.0879
β_{yes}	0.7311	0.1853	3.535	0.0000

According to the results, the increase from a re-identification risk of 20% to 30% leads to a much higher increase in disutility compared to the increase from 10% to 20%. This can also be noted in the visualisation of the utility contributions per attribute level in Figure 20. A clear effect of non-linearity is observed.

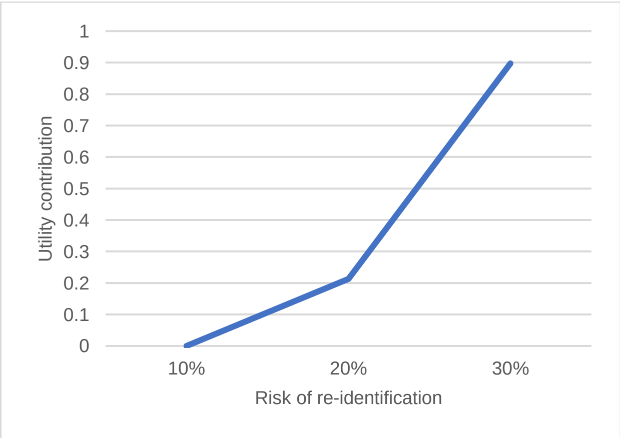


Figure 20. Utility contribution of rid attribute levels

6.6.4. Non-linearity of types of data parameter

Lastly, the *types of data* factor is tested for non-linearity. In Table 26, the results are presented.

Table 26. Estimates Model 10

	Estimate	s.e.	Rob.t.rat.(0)	p(1-sided)
β_{mon}	0.2861	0.0532	4.481	0.0000
β_{eff}	-0.7790	0.0826	-8.796	0.0000
β_{rid}	-0.4384	0.0784	-4.084	0.0000
β_{tod2}	-0.0213	0.1686	-0.163	0.4355
β_{tod3}	-0.2642	0.1657	-1.866	0.0310
β_{tod4}	-1.2023	0.1746	-6.159	0.0000
$\beta_{use_{aca}}$	0.3529	0.1761	1.944	0.0260
$\beta_{use_{cor}}$	-0.6515	0.1850	-3.284	0.0000
$\beta_{use_{soc}}$	-0.0696	0.1619	-0.503	0.3075
β_{yes}	0.6760	0.1883	3.159	0.0000

From the visualisation in Figure 21, it is clearly notable that the attribute levels do not have an equal utility contribution. Sharing motion data in addition to location only causes a slight increase in disutility (which is also not significant), while sharing context data in addition to location and motion data leads to a higher increase in disutility. A large increase in disutility is observed when moving from level 3 to level 4, which means that multimedia data is also shared, is even higher. This means that people are more sensitive to the types of data attribute when more types of data are being shared. Thus, non-linearity can be clearly observed for this parameter.

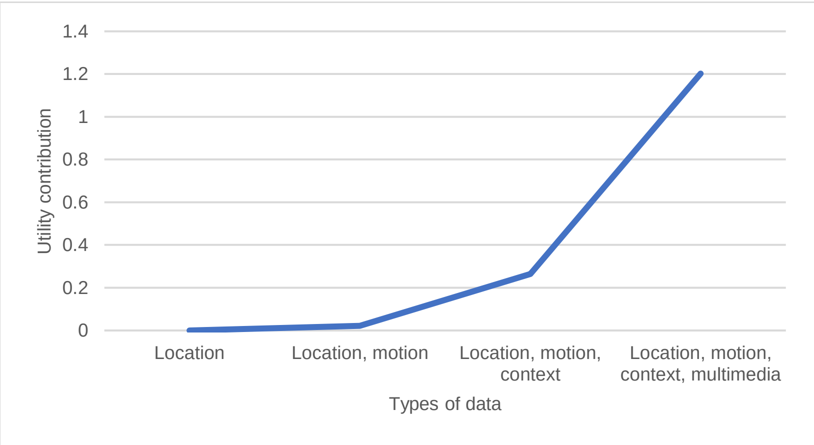


Figure 21. Utility contribution of tod attribute levels

6.6.5. Model 11: MNL model with non-linear parameters

From the previous paragraphs, it appears that all the attributes show an unequal utility contribution of attribute levels. Therefore, a model is estimated that incorporates the non-linearity of these factors. The results are presented in Table 27.

Table 27. Estimates Model 11

	Estimate	s.e.	Rob.t.rat.(0)	p(1-sided)
β_{mon20}	0.3889	0.2102	1.885	0.0297
β_{mon40}	0.4749	0.2011	2.505	0.0061
β_{mon60}	0.9431	0.1701	4.697	0.0000
β_{effmod}	-0.6036	0.1420	-5.064	0.0000
β_{effhig}	-1.5977	0.1690	-8.488	0.0000
β_{rid20}	-0.4080	0.1635	-2.828	0.0023
β_{rid30}	-0.8367	0.1716	-3.801	0.0000
β_{tod2}	-0.0458	0.1835	-0.280	0.3897
β_{tod3}	-0.2698	0.1916	-1.449	0.0737
β_{tod4}	-1.2325	0.1758	-6.230	0.0000
$\beta_{use_{aca}}$	0.3021	0.1961	1.706	0.0440
$\beta_{use_{cor}}$	-0.7063	0.1948	-3.573	0.0000
$\beta_{use_{soc}}$	-0.0790	0.1632	-0.575	0.2827
β_{yes}	0.6137	0.1935	2.882	0.0020

The final Log-Likelihood of model 11 is equal to -869.08, the BIC is 1840.55, and the Rho-square is 0.1621. In Figure 22, the percentages that indicate the relative importance of the attributes are displayed. When comparing these percentages to those of model 1 and model 2, the ranking in relative importance does not change. When looking at the BIC value, it can be concluded that model 11 does not perform better on the data than model 1. However, the differences in utility contributions of the attribute levels can deliver valuable insights on non-linearity of attributes.

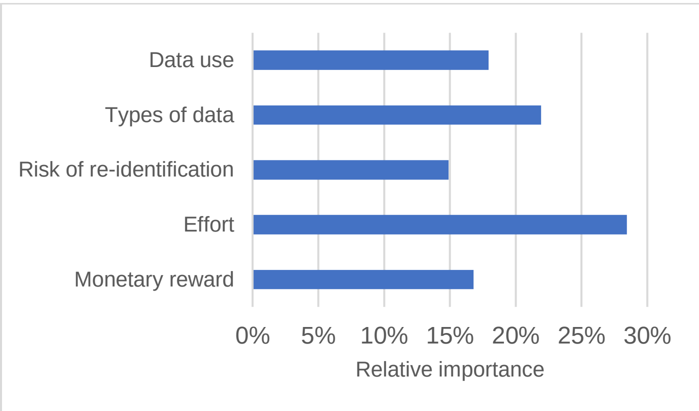


Figure 22. Relative importance of attributes (Model 11)

6.7. ML Panel model

In order to capture heterogeneity in choices, a Mixed Logit (ML) model is estimated. This model assumes that if people make multiple choices, their choices are correlated, in contrast to the MNL model. These are called panel effects. Ignoring these correlations between multiple choices of individuals will lead to biased standard errors. Therefore, an ML panel model is estimated. The code that was used to estimate the ML models in *R*Studio is presented in Appendix G.

6.7.1. Model 12: ML model with error term

First, a basic ML model is estimated. In this model, an error term is added in order to capture panel effects. The results are presented in Table 28.

Table 28. Estimates Model 12

	Estimate	s.e.	Rob.t.rat.(0)	p(1-sided)
β_{mon}	0.4669	0.0657	4.908	0.0000
β_{eff}	-1.0517	0.1046	-10.032	0.0000
β_{rid}	-0.6707	0.1081	-5.320	0.0000
β_{tod}	-0.5519	0.0703	-7.026	0.0000
$\beta_{use_{aca}}$	0.4088	0.2275	1.657	0.0488
$\beta_{use_{cor}}$	-1.0491	0.2319	-3.824	0.0000
$\beta_{use_{soc}}$	-0.2938	0.2043	-1.301	0.0966
β_{yes}	1.1947	0.2890	3.805	0.0000
Sigma_yes	1.8533	0.1830	8.916	0.0000

From these results, it can be seen that the standard error of β_{yes} is equal to 0.2890. This is relatively high when compared to the other betas. From this, we can conclude that there is a relatively large variation in the intrinsic motivation to share data. In Model 3, this inclination was assumed not to vary among classes. However, the ML model provides a new insight by showing that in reality, this intrinsic inclination does vary among different groups of people.

In Figure 23, the percentages that indicate the relative importance of the attributes are displayed. From the results, it appears that the ML model fits the data better than the previously estimated MNL model. The final Log-Likelihood is -735.61, the BIC is 1537.03, and the Rho-square is 0.2925. These values show a significant improvement in model fit when adding a random error term. This means that there is heterogeneity in the choices. Therefore, this model is further investigated.

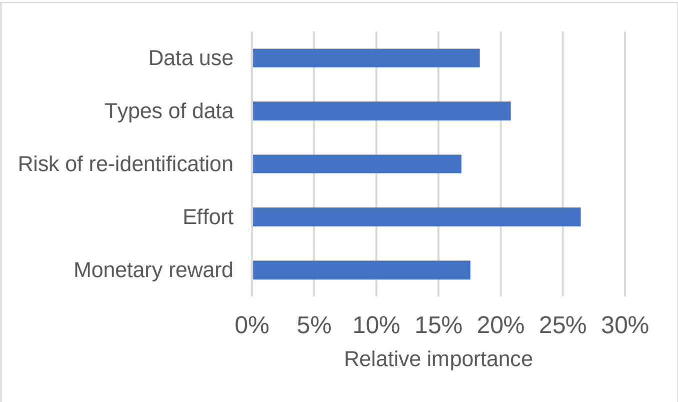


Figure 23. Relative importance of attributes (Model 12)

6.7.2. Model 13: ML model with all parameters random

It is also possible to add an additional error term for all parameters. This allows for capturing more heterogeneity in preferences. This is done in model 13, which is presented in Table 29.

Table 29. Estimates Model 13

	Estimate	s.e.	Rob.t.rat.(0)	p(1-sided)
β_{mon}	0.9956	0.1898	4.682	0.0000
β_{eff}	-1.9692	0.2885	-6.225	0.0000
β_{rid}	-1.1258	0.2402	-4.845	0.0000
β_{tod}	-0.9455	0.1583	-5.382	0.0000
$\beta_{use_{aca}}$	0.9819	0.3910	2.353	0.0093
$\beta_{use_{cor}}$	-2.2735	0.5754	-3.184	0.0000
$\beta_{use_{soc}}$	-0.2708	0.3287	-0.713	0.2379
β_{yes}	2.1037	0.4666	4.491	0.0000
Sigma_yes	2.8825	0.4840	5.475	0.0000
Sigma_mon	1.4645	0.2493	4.949	0.0000
Sigma_eff	1.2758	0.2346	5.346	0.0000
Sigma_rid	1.6647	0.3079	-4.832	0.0000
Sigma_tod	0.6276	0.1465	4.611	0.0000
Sigma_useaca	1.6863	0.4099	3.925	0.0000
Sigma_usecor	3.7651	0.7929	3.807	0.0000
Sigma_usesoc	1.4526	0.4781	2.523	0.0058

When evaluating the results of model 13, it can be seen that the model performs better than the ML model with only one error term. Also, all sigma's are significant. Thus, more heterogeneity is captured in this model with the extra error terms for all parameters. From the estimates, it can be seen that the sigma's vary for different attributes. The estimated sigma's for each parameter are presented in Figure 24.

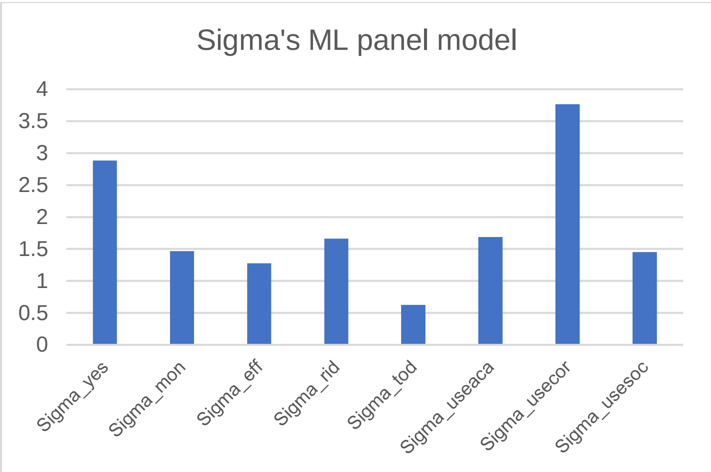


Figure 24. Sigma's Model 13

For example, the sigma for data use by a corporate institution is relatively high. This means that there is a relatively high degree of unobserved heterogeneity between people. When people base their decision on whether data is collected by a corporate institution at t = 1, then they are likely to base their decision on this attribute at t = 2 as well. In contrast, the sigma for the types of data attribute is relatively small. This means that for this attribute, repeated choices made by the same individual are less correlated.

In Figure 25, the relative importance of the parameters in this ML model is visualised.

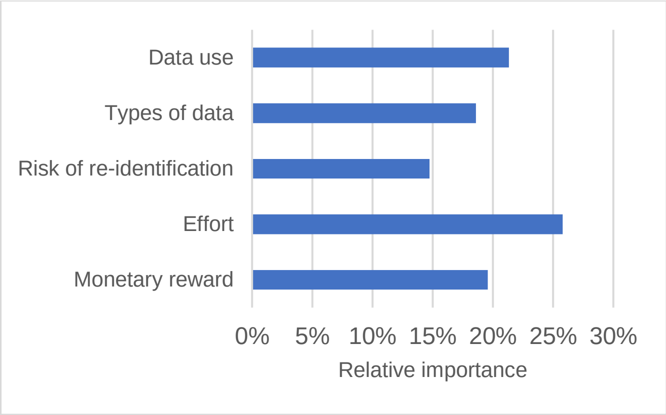


Figure 25. Relative importance of attributes (Model 13)

When compared to the previously estimated models, it appears that model 13 differs slightly in terms of relative importance of attributes. *Effort* is still seen as the most important attribute influencing the decision to participate in sensing applications. In contrast to, e.g., model 12, *data use* is the second most important attribute, closely followed by the *monetary reward* and the *types of* data attributes. The *risk of re-identification* is still the least important attribute.

The final Log-Likelihood of this model is -675.93, the BIC is 1468.87, and the Rho-square is 0.3499. Again, a large improvement in model fit can be observed, compared to model 1. Also, the model performs better on the data than model 12, in which only one random error term was added.

6.7.3. Model 14: ML model with non-linear parameters

From paragraph 6.6, it appears that the attributes included in the model show non-linearity among attribute levels. Therefore, the ML model is estimated again with non-linear parameters. The results are presented in Table 30.

Table 30. Estimates Model 14

	Estimate	**s.e.**	**Rob.t.rat.(0)**	**p(1-sided)**
$\boldsymbol{\beta_{mon20}}$	0.5216	0.2563	1.865	0.0311
$\boldsymbol{\beta_{mon40}}$	0.7232	0.2450	2.639	0.0042
$\boldsymbol{\beta_{mon60}}$	1.4086	0.2099	4.685	0.0000
$\boldsymbol{\beta_{effmod}}$	-0.9123	0.1796	-5.059	0.0000
$\boldsymbol{\beta_{effhig}}$	-2.4280	0.2208	-1.083	0.0000
$\boldsymbol{\beta_{rid20}}$	-0.6601	0.1914	-3.382	0.0000
$\boldsymbol{\beta_{rid30}}$	-1.2950	0.2324	-4.506	0.0000
$\boldsymbol{\beta_{tod2}}$	-0.0502	NaN	-0.267	0.3948
$\boldsymbol{\beta_{tod3}}$	-0.3945	0.1360	-1.830	0.0337
$\boldsymbol{\beta_{tod4}}$	-1.8467	0.1811	-6.818	0.0000
$\boldsymbol{\beta_{use_{aca}}}$	0.4425	0.2270	2.026	0.0214
$\boldsymbol{\beta_{use_{cor}}}$	-1.0803	0.2390	-4.012	0.0000
$\boldsymbol{\beta_{use_{soc}}}$	-0.1140	0.1835	-0.677	0.2492
$\boldsymbol{\beta_{yes}}$	0.9260	0.2781	3.103	0.0000
Sigma_yes	1.9058	0.1882	8.894	0.0000

Judging from the results, this ML model with non-linear parameters performs better in term of goodness-of-fit than the ML model which assumes linear parameters. The final Log-Likelihood is -722.72, the BIC is 1555.14, and the Rho-square is 0.3049. However, the model with a random error added to all parameters still performs better than model 14.

In Figure 26, the percentages that indicate the relative importance of the attributes are displayed.

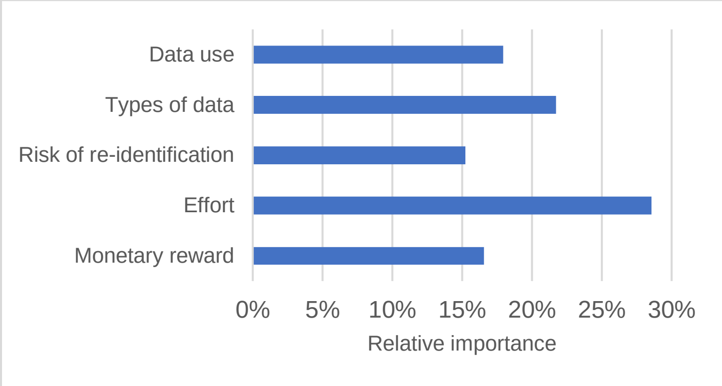


Figure 26. Relative importance of attributes (Model 14)

The relative importance of attributes is quite similar to the percentages of model 10. The *effort* attribute is relatively more important in model 14 than in model 12, and the *risk of re-identification* attribute is relatively less important.

6.8. Selection of final model

In Table 31, an overview is provided of all the models included in the research, and their model fit in terms of Log Likelihood, BIC, and Rho-square.

Table 31. Results of estimated models

Model	Description	Parameters	Observations	LL(final)	LL(0)	BIC	Rho-square
1	MNL with main effects	8	1500	-878.24	-1039.72	1814.98	0.1553
2	MNL with interactions	13	1500	-842.21	-1039.72	1779.48	0.1900
3	LC with covariates	30	1440	-733.24	-998.13	1684.65	0.2654
4	Non-participation	8	1248	-716.03	-865.05	1489.10	0.1723
5	Privacy Fundamentalists	8	708	-490.75	-383.31	819.12	0.2189
6	High Distrust group	8	180	-64.56	-124.77	170.66	0.4184
7	Non-linearity of monetary reward factor	10	1500	-877.73	-1039.72	1828.60	0.1558
8	Non-linearity of effort factor	10	1500	-876.85	-1039.72	1819.52	0.1566
9	Non-linearity of risk of re-identification factor	10	1500	-876.32	-1039.72	1818.46	0.1572
10	Non-linearity of types of data	10	1500	-871.19	-1039.72	1815.51	0.1621
11	Non-linearity of all factors	14	1500	-869.08	-1039.72	1840.55	0.1641
12	ML model with error term	9	1500	-735.61	-1039.72	1537.03	0.2925
13	ML model with all parameters random	16	1500	-675.93	-1039.72	1468.87	0.3499
14	ML model with non-linear parameters	15	1500	-722.72	-1039.72	1555.14	0.3049

From the results of the estimated models, it appears that the Mixed Logit model with all parameters random (Model 13) has the best model fit, when looking at the final Log-Likelihood, the BIC value, and the Rho-square. The model has a Rho-square of 0.3499, meaning that it explains 35% of the initial uncertainty. Therefore, this model is selected as the final model for explaining the data. However, insights from other models are used as well for interpretations.

6.9. Model validation

The choice model was validated in order to check how successful the model is in predicting choices of individuals. This was done by conducting an out-of-sample test. For this purpose, the dataset was split into two parts. Two thirds of the data was used as the estimation set, and one third of the data was used as the validation set.

First, the base MNL model (Model 1) was estimated again on two thirds of the data. The values resulting from this model estimation are presented in Table 32.

Table 32. Estimates for validation

Factor	Estimate	s.e.	Rob.t.rat.(0)	p(1-sided)
$\boldsymbol{\beta_{mon}}$	0.2921	0.0790	3.295	0.0000
$\boldsymbol{\beta_{eff}}$	-0.6034	0.1188	-5.004	0.0000
$\boldsymbol{\beta_{rid}}$	-0.5521	0.1208	-3.584	0.0000
$\boldsymbol{\beta_{tod}}$	-0.4121	0.0825	-4.284	0.0000
$\boldsymbol{\beta_{use_{aca}}}$	0.3739	0.2644	1.294	0.0979
$\boldsymbol{\beta_{use_{cor}}}$	-1.0035	0.2945	-2.955	0.0016
$\boldsymbol{\beta_{use_{soc}}}$	-0.0995	0.2362	-0.462	0.3219
$\boldsymbol{\beta_{yes}}$	0.6011	0.2671	1.835	0.0333

These parameter estimates were then used to make predictions on the other one third of the data. The *hit rate* was used as a measure for the predictability of the model. This is a measure that indicates for what percentage of observations the model succeeds in predicting the alternative that is chosen by an individual.

For calculating the Log-Likelihood and the hit rate, *R*Studio was used. The Out-of-sample Log-Likelihood of the model is -350.14. The hit rate is equal to 64.9%. This means that in 64.9% of the choices, the model succeeds in predicting the alternative ("yes" or "no") that will be chosen. This percentage is higher than when taking a random draw, which would lead to a percentage of 50%. Thus, the model performs better than "throwing a dice".

Something that should be noted in this context is that the trade-offs included in the choice experiment were designed to be difficult trade-offs. Therefore, in most choice sets there was not one alternative that was clearly more "popular" than the other alternative. As a consequence, most percentages lie somewhere around 50%. Taking this into account, a hit rate of 64.9% indicates that the model is relatively successful in making predictions.

As an additional measure, the *mean absolute deviation* of the choice probabilities is calculated. This is done by comparing the predicted percentage with the observed percentage of the "yes" alternative being chosen, and calculating the mean difference in percentage points. In Appendix J, the respective percentages are reported. The corresponding mean absolute standard deviation is equal to 7 percentage points. This means that the model predicts the choice probabilities with an accuracy of 7 percentage points. If the model, e.g., predicts a percentage of 50%, the "real" percentage will on average lie somewhere between 43% and 57%.

6.10. Model interpretation

The attributes in the model are discussed in the following subsections, in order of importance (according to model 13).

6.10.1. Effort

According to the estimated model, the effort that is required is most important for individuals when choosing to share data or not. As expected, sensing applications requiring a higher degree of user involvement have a negative influence on the willingness of users to contribute to the application. If applications ask the user to report observations, users might feel overburdened, as noted by Heiskala et al. (2016). No previous research was found that included both required effort and privacy-related attributes in a choice experiment. However, from the results, it can be concluded that effort does play an important role in consumers' decisions regarding data sharing, and is even regarded as more important than the types of data being collected, the data use, or the risk of re-identification.

6.10.2. Data use

Participants are most willing to share data with academic institutions. The least acceptable party collecting data is a corporate institution aiming to improve products or services. This is in line with research by Christin et al. (2013). In this study, it was concluded that participants claim a higher reward when sharing data with a corporate institution, compared to an academic institution. Public institutions are also more accepted as data collectors than companies in previous research (Schomakers et al., 2020), which is confirmed in our research.

According to the results, the different parties collecting data from most accepted to least accepted are 1) Academic institution aiming to investigate transport modes, 2) Governmental institution aiming to improve mobility, 3) Societal organisation aiming to address local issues regarding mobility, and 4) Corporate institution aiming to improve products or services. However, the difference between a governmental institution and a societal organisation collecting data is not significant. A reason for this could be that the purpose of the data is quite similar for these two parties; both aim to improve mobility in the local area.

6.10.3. Monetary reward

In our sample, we found that people are more likely to share data when they get a higher financial reward in return. This is in line with research by Derikx et al. (2016) and Schomakers et al. (2020), who also found that individuals are willing to share data in return for a financial compensation.

K. E. Train (2003) describes a function to calculate the value of time, which is defined as the extra cost that a person would be willing to incur in order to save time. This is done by using the estimated coefficients of cost and various time components. Based on this definition, we can specify a function that calculates the **Value of Privacy (VoP).** This function is defined as follows:

$$VoP = \frac{\frac{\delta V}{\delta RID}}{\frac{\delta V}{\delta MON}} = \frac{\beta_{rid}}{\beta_{mon}}$$

According to model 11, the Value of Privacy is equal to 1.13 €/%. This means that people want to receive an amount of €11.30 per month if the risk of re-identification is increased by 10 percentage points.

To see if there is a difference between the estimated models, the Value of Privacy is also estimated for other models. The results are presented in Table 33.

Table 33. Value of Privacy

Model	Value of Privacy (VoP)
Model 1	1.41 €/%
Model 5	1.89 €/%
Model 6	4.29 €/%
Model 12	1.44 €/%
Model 13	1.13 €/%

According to these calculations, the group of people categorised as Privacy Fundamentalists (model 5) assign a higher value to privacy, compared to people scoring lower on the Privacy Index. Also, the Value of Privacy calculated for model 6, which are the people who are highly distrustful according to Westin's Distrust Index, is significantly higher than the VoP for the other models. This is caused by the high importance this group of people assigns to the *risk of re-identification* factor. The calculation for this group is only based on a subset of 15 respondents. However, it indicates that there is a group of people for which a higher incentive in terms of a financial compensation is needed, in order for them to be willing to share data.

It should be noted that the Value of Privacy function assumes a linear function of money as well as risk. However, model 7 and model 9 showed that these functions are not linear. Therefore, the calculated VoP does not fully reflect the Value of Privacy. Nevertheless, it gives an idea of how much money individuals approximately wish to receive when the risk of re-identification increases, and can thus deliver a valuable insight. However, in reality, the amount of money people want to receive when the risk is increased from 20% to 30% is higher than when the risk increases from 10% to 20%.

6.10.4. Types of data

In a study by Schomakers et al. (2021) on privacy and trust in smart home applications, it appeared that users have no problem sharing their motion data. Also, the collection of position data was accepted my most of the respondents, while audio data was least accepted, followed by image data. This can explain the large gap that was found between the utility contribution of the different attribute levels related to the *types of data* attribute. Derikx et al. (2016) also found that users derive a higher disutility from attributes related to privacy of behaviour and action than from attributes related to privacy of location and space. However, in our sample we only observe a slight difference between the utility contribution of sharing only location data and sharing both location and motion data.

Besides the effort required from the user, the types of data attribute is relatively important (18% in model 13) to users when making a decision regarding data sharing. This is quite in line with research on health data by Aitken et al. (2018), who found the type of information to be the most important attribute.

6.10.5. Risk of re-identification

The risk of re-identification was least important for the participants in our study. From the statements categorizing individuals in Westin's Privacy Index, it appeared that only a small amount of people were categorized as Privacy Unconcerned, while 47% of the participants were in the Privacy Concerned group. Thus, people indicate to be concerned about their privacy and about how their data is handled. However, this is not reflected in the choices they make, since the *risk of re-identification* factor has a relatively low importance. This phenomenon can be explained by a concept known as the "Privacy Paradox". This describes that on one hand, people express their concerns about the handling of their personal data, but at the same time, they often choose to share their data voluntarily and rarely make an effort to actively protect their data (Gerber et al., 2018).

The finding that the *risk of re-identification* is the least important factor influencing the decisions of individuals regarding data sharing is not in line with research by Schomakers et al. (2020). In this choice experiment, the level of anonymization appeared to be the attribute with the highest relative importance. However, in this study the level of anonymization ranged from "no anonymization" to "complete anonymization", a level that does not exist in reality. In our experiment, we chose to have a smaller range with levels from 10% to 30%. This could be the reason that the risk of re-identification was less important to participants when making the decision whether or not to share data.

Furthermore, Schomakers et al. (2020) found that users with high privacy concerns assign more importance to the anonymization level than users with lower privacy concerns. This is in line with our research. Users with a higher score on the Privacy Index, have a significant higher disutility for the 30% level of the *risk of re-identification* attribute than users with a lower score on the Privacy Index.

The fact that the *risk of re-identification* is the least important factor influencing the decision of individuals on sharing data does not mean that this is not an important aspect when designing applications for participatory sensing. This is further addressed in Chapter 7.

6.10.6. Privacy perceptions

It was expected that people scoring higher on Westin's Privacy Index as well as the Distrust Index, would be more reluctant to share data in participatory sensing applications. According to model 5 and model 6, a difference can indeed be observed in the relative importance Privacy Fundamentalists and people in the High Distrust group attribute to the factors in the choice experiment, when compared to people in the whole sample. This subset of people indicating being more concerned about how their data is being handled consider the privacy related factors (*risk of re-identification, types of data,* and *data use*) as highly important, compared to the other factors. People who are categorised as Privacy Fundamentalists attach a relatively high importance to the use of the data that is being collected. People in the High Distrust category regard the risk of re-identification as the most important factor when deciding whether or not to share their data. For some of the people in this group, even the lowest rate of risk (10%) is not acceptable when sharing data.

6.10.7. Personal characteristics

It was expected that income would have an influence on the *monetary reward* factor, that digital behaviour would have an influence on the effort that people are willing to take, and that people participating in altruistic activities require a lower monetary reward and are willing to put more effort in participating. Therefore, models with interaction effects were estimated. However, these interaction effects appeared to be insignificant. Furthermore, no influences of age, gender, or education were observed. Thus, these personal characteristics seem to have no influence on the willingness to share data.

It appears that personal characteristics do not have a significant effect on the willingness of individuals to share data. Since these factors do not significantly moderate the decision to share data, we can conclude that the non-representativeness we found in the sample is not an issue.

6.11. Conclusion

From the choice modelling results, several conclusions can be drawn. According to the estimated models, *effort* is the most important factor influencing the benefit-cost trade-off by individuals (26%). Furthermore, users take into account the party collecting the data and for what purpose, when making their decision. Data collection by academic institutions is most accepted, while data collection by corporate institutions for improving products and services is least accepted. This finding is in line with previous studies on data sharing.

Also, conclusions can be drawn on the trade-off people make between money and privacy. According to model 13, people want to receive an additional amount of €11 per month when the risk of re-identification increases with 10 percentage points.

Furthermore, our finding that there is a large gap in acceptance between the collection of multimedia data and other types of data confirms previous research on sharing data from smart home applications. The collection of motion data in addition to location data is highly accepted by users, while the collection of multimedia causes a large decrease in acceptance.

Lastly, the *risk of re-identification* appeared to be the least important factor influencing the willingness to share data (15%). However, this does not mean that privacy is not influencing the benefit-cost trade-off by smart device users. A high amount of respondents (47%) indicated they are concerned about their privacy. Also, a group of these privacy-concerned respondents indicated that they are not willing to share data in any situation. For these respondents, privacy is so important that they do not accept the minimum risk of re-identification of 10%. When looking at the group of respondents that score extremely on the Privacy Index and the Distrust Index, it appears that these groups of people assign a higher importance to the privacy related factors (risk of re-identification, types of data, and data use), compared to the rest of the sample. Thus, it can be concluded that privacy concernedness has an impact on the preferences of users regarding data sharing.

7

Implications for smart mobility

This chapter addresses the implications of the experiment results for different areas of application. The research question that is addressed in this section is as follows.

SQ4. Regarding these trade-offs, what are implications for different applications in the field of smart mobility?

The choices made by individuals are often influenced by a set of factors that form the context of a choice situation (Goulias & Pendyala, 2014). In the choice experiment, scenarios were specifically focused on the context of mobility. This chapter discusses what the implications of users' preferences will be for different applications in the field of smart mobility.

The research question on implications for different applications in the field of smart mobility cannot be answered directly. Therefore, 2 sub-sub questions are defined.

1. What use cases can be defined for which participatory sensing can be relevant?
2. What is the acceptance of smart device users for these use cases, according to the results?

The first question is answered by conducting interviews with relevant parties in the field of smart mobility. From each interview, a use case is identified describing a potential application of participatory sensing in the field of smart mobility. The results of these interviews are analysed and compared. These outcomes are used as an input for question 2. In the latter question, the results of the quantitative and qualitative research are aggregated. The model estimations from Chapter 6 are used to predict the acceptance in the identified use cases.

In paragraph 7.1, the background of the conducted interviews is described. Paragraph 7.2 discusses the three use cases that are derived from these interviews. In 7.3, the most important findings from the qualitative research are presented. Also, this paragraph addresses the differences and commonalities between the conducted interviews. In 7.4, the implications of the quantitative results for the identified use cases are discussed, after which the acceptance of users in each of these use cases is predicted in 7.5. Finally, subsection 7.6 provides a conclusion on the insights gained in this chapter.

7.1. Interviews

In order to determine the relevance of participatory sensing for different applications in the field of smart mobility, several semi-structured interviews are conducted with experts from different parties for which participatory sensing could be a relevant application. We follow the process for the interviews as described in Chapter 3.

In total, 3 interviews were conducted. The interviewees were people from a municipality (Interview A), from a research group on connected cars (Interview B), and from a transport operator (Interview C). Table 34 presents an overview of the interviews that were conducted, and provides some information about the background of the interviewees.

Table 34. Conducted interviews

Interview	Party	Role of interviewee	Use case derived from interview
Interview A	Municipality	Innovation team Smart Mobility	Crowd management in a city
Interview B	Research group on connected cars	Knowledge management	Safety research using car accident information
Interview C	Transport operator in a city	Business analyst	Real-time travel information in public transport

The transcripts of the interviews can be found in Appendix I. The interviews are coded using the software *Atlas.ti*. This resulted in a total amount of 39 codes. After coding all the interviews, codes were categorised into 5 group codes, which are *Current smart mobility projects, Data collection practices, Challenges of smart mobility, Technical aspects of implementing crowd sensing,* and *Ethical aspects of implementing crowd sensing.* An overview of all codes and the grouping of these codes is provided in Appendix I.

The use cases identified from the interviews are discussed in the following paragraph.

7.2. Potential applications

The results of these interviews are used to define use cases in the field of smart mobility. In the following subsections, the use cases *Crowd management in a city, Safety research using car accident information,* and *Real-time travel information in public transport* are discussed subsequently. These use cases are derived from, respectively, Interview A, Interview B, and Interview C. For each interview, the insights gained on current smart mobility projects, the challenges related to these projects, and the opinions of the interviewees on the implementation of crowd sensing are described.

7.2.1. Crowd management in a city

This use case is derived from Interview A, which is the interview with the municipality. A first project that was mentioned by the interviewee from the municipality is crowd management. Currently, cameras and Wi-Fi sensors are used to count the amount of people and a certain moment in time. Data gathered by these cameras and sensors can be used to prepare for expected crowdedness or congestion. For example, this is done by giving people advice on alternative routes. By doing this, overcrowding can be prevented and traffic can be regulated more efficiently. Especially during the Covid-19 crisis, this kind of information was relevant.

The municipality compares real-time data with historical data to prepare, monitor, and control traffic flows. Besides camera data, data from transport and parking providers are used. For example, public transport operators know how much people checked in within a certain time range. Also, parking providers know the number of cars in their garage. This kind of information can be used to give travellers insight in crowdedness in trains or in parking capacity, and give them advice on alternative routes. Another application is monitoring and analysing the use of public transport. The interviewed municipality cooperates with transport operators in order to get insight in the development of public transport use.

Mobility data that is collected can be used in several ways. It can be used get insight in the current situation in real-time. These can be compared with historical data. Also, data can be used to predict future situations and prepare for these situations. Besides that, data can be used to simulate scenarios. In the crowd management project, the municipality is already looking at how future situations can be predicted. A goal is to answer questions such as, “is this location more crowded when it is raining?”.

Challenges

The municipality is looking at how crowd management can be digitized. Instead of only using data from cameras, there is a wish for more data in order to get a more detailed view on crowdedness. For example, they would like to know in real-time if it’s crowded at a certain location, how much traffic is coming that way, if a train is arriving soon at that location, and the number of passengers on that train.

Currently, the interviewed municipality does not receive a lot of information about mobility flows. Eventually, the municipality wants to look at how this can be done differently and how digital mobility management can be realized. Before realizing this, several questions should be asked, regarding what kind of data is collected, how these data are being collected, the ethical issues, and the protection of privacy

With respect to laws and regulations, the GDPR determines what data can be collected and under what conditions. Also, data minimization is important: no more data should be collected than actually needed. However, from an ethical point of view, there can be somewhat of a grey area. As a guideline, the following principles are used to ensure that ethics are a fixed part of the innovation process: *inclusive, control, tailored to the people, legitimate and monitored, open and transparent*, and *from everyone, to everyone*. For all innovations, these principles are taken into account. An important aspect is that the goal of the data collection should be clear and justified before collecting the data. This is even more important when it involves human subjects. Also, an important question that should be asked before collecting data, is when data should be collected, and in what situations this is not desirable. For crowd management, it is not necessary to collect data at places that are not crowded. When only one person is in a certain area, it is not desirable to collect data, since this individual will not be anonymous anymore. Lastly, it was mentioned that it is important that individuals consciously choose to share data.

Implementation of crowd sensing

If individuals want to share data which can help the municipality get insight in traffic flows or crowdedness, this can be a valuable data source. This can help to make data-driven policy, to get insight in phenomenon in the city, and to determine if measures are successful.

In order to implement crowdsensing, the municipality needs to be able to receive the data. Therefore, a technical infrastructure is needed. Standards may be needed in order to be able to exchange data efficiently between different parties. Also, a platform needs to be developed on which the data can be collected, aggregated and analysed.

Furthermore, legal aspects need to be in place and privacy protection should be ensured. This can be different when people actively make the choice to share their data. However, these users can also set conditions under which they want to share data. Moreover, if different transport providers and municipalities work together on realising smart mobility, they need to agree on what the cooperation will look like and how the parties will deal with privacy issues. Realising this kind of cooperation and the necessary contracts can require a lot of organisation, according to the interviewee.

7.2.2. Safety research using car accident information

In Interview B, another application was mentioned related to traffic information. Cars can be connected using sensors, to let them communicate with each other. This is also called *swarm intelligence.* If one car gets in an accident, it can then send a signal to all other cars nearing the accident. The cars that get the signal will know that and where the accident happened and that they have to be careful. Based on this information, they can change their route in order to prevent getting in a traffic jam and therefore reduce their delay.

Furthermore, data on accidents can be used for safety research. If car dealers are allowed to use drivers' information when they get in an accident, they can use it for research on the safety of the car. For example, employees from a car dealer often go to accidents immediately. They have an agreement with the authorities that enables them to go to the accident and do research on how the safety of the car can be improved. This can be a hard job to do, since seeing an accident can be traumatizing. If a car would be able to directly send this information to the car dealer, the safety analysis can be performed more easily.

Challenges

Sharing data has benefits and risks. Before sharing their data, people need to know who is receiving the data, as well as how the data is used. Moreover, some people want to know what they are going to get in return for sharing data. When downloading an app, people usually make a trade-off between the benefits and the risks. In this case, the benefit is often receiving a useful service. In the case of having a car that remembers the user's preferences and automatically adjusts the seats, for example, the benefit is that it makes driving the car more easily and comfortable. These benefits should outweigh the risks that come with using the application. The perceived risk depends on the kind of data that is collected. Users will ask themselves the question what will happen if their information is being leaked, and how severe it will be if that happens. If people do not really care about a certain kind of data being leaked, they will be more willing to share their data.

Implementation of crowd sensing
It was mentioned that it is important that the party collecting the data is transparent to the user about the reason for data collection, since people will want to know how their data is used. Furthermore, the benefit-cost trade-off has to be considered. If sharing information can increase the safety of a car or if it can lead to a decrease in traffic delays, there would be a clear benefit for users. Also, people like it if a car automatically adjusts to their preferences. However, another aspect that should be taken into account is the user-friendliness of a car. In the case where a lot of input is required from the user, the application can become less attractive.

7.2.3. Real-time travel information in public transport

From Interview C, a use case on travel information in public transport is derived. Currently, the transport operator is working on a crowdedness indicator, in order to provide travellers with information on crowdedness. Information is mainly collected from OV chipcard data. Via OV chipcards, data is collected on all travel flows, including check-ins, check-outs, and time stamps and locations of these check-ins and check-outs. Also, data on travel distances and kilometre prices are collected. These data are used to make transport analyses, to develop time tables, and to determine the required capacity of vehicles at specific locations. Furthermore, data can be used to determine where to start marketing campaigns. Moreover, data can be used to understand travel behaviour of people. This information can then be used to improve products.

Data that is collected from OV chipcards is not real-time yet. Therefore, crowdedness is only predicted based on historical data. Besides this, bus drivers communicate when a bus is full, or when there is a delay. This information is shown on the website and in the app in order to inform travellers. Another way in which the transport operator is currently collecting data is by using weights in vehicles. Vehicles can provide information on the weight on all locations in the vehicle. From this, information can be derived on crowdedness and insight can be gained on locations that are more crowded than others. A system is currently being developed in order to show the capacity in a vehicle. In a metro, people often tend to crowd in one place, while other places in the vehicle are not crowded at all. The aim is to show the capacity information on a screen within the vehicle, in order to stimulate travellers to spread through the metro.

Information on crowdedness is also shared with other parties, such as the municipality or other transport operators. These data is uploaded to a public platform which can be accessed by everyone. Thus, the information from the crowdedness indicator can also be used by other companies.

Challenges
The transport operator is looking for ways to get insight in fare dodging. Before, infrared sensors were placed on vehicles in order to count the amount of people entering and exiting the vehicle. However, the sensors were only placed on 13% of the vehicles. Also, the system could not handle detours well. Since the usefulness of the system was low, the system was shut down. Crowdsensing could be a way to draw conclusions on fare dodging. However, the challenge is that a sufficient amount of people should participate in order to ensure the usefulness of the system. Collecting data via smartphones of travellers was proposed before, however, it appeared that data from only 40% of travellers would be collected. In order to get more accurate insights, a higher usage would be required.

A challenge is to motivate people to participate in such a system, especially when it aims to detect fare dodging. People may be participating by accident, but in that case, ethical questions can be raised. Transparency is an important value. However, when being fully transparent, it would probably still not be possible to track fare dodgers.

Another issue is the accuracy of location data. For trains, this would not be an issue, since there is a lot of distance between stations. However, for city transport like trams and buses, it is difficult to determine exactly in what vehicle a person is located. Also, people on a bike may be hard to distinguish from people being in a vehicle.

Implementation of crowd sensing
Crowd management is an important topic for the transport operator. In the future, the transport operator aims to find solutions to collect more data, and to collect data in real-time. Crowdedness in vehicles is often caused by delays or cancellations of other vehicles. These are things that are difficult to predict with current models. Thus, there is a need for collecting and sharing real-time travel information. In the future, this information can help to provide more personalised services to travellers. Crowdedness in public transport is something that is perceived differently by different kind of people. Gaining insights in behaviour and preferences of travellers can contribute to making transport more personalised.

Crowd sensing could be a valuable data source in addition to data that is currently being collected. From a technological point of view, implementation of crowd sensing would be a relatively cheap and simple solution. The transport provider should be able to receive the data, but the integration platform that is needed to collect the data for analysis is already in place. The main issues concern the legal and aspects of implementing crowd sensing. Legal requirements need to be in place, and a transparent communication to travellers about the data collection is necessary. When implementing new applications in the field of smart mobility, communication is one of the most important aspects. Furthermore, people need to be incentivised to participate in collecting data. Whether people receive a reward and what such a reward would look like, is something that needs to be determined.

7.3. Implementing participatory sensing for smart mobility

In the field of smart mobility, crowd sensing can be a valuable data source in addition to existing sources of data. From the conducted interviews, it appears that parties are already collecting data. However, this data is often not real-time. Involving users is required to collect more data. These data can help to gain understanding of traffic flows and travel behaviour and respond to incidents more quickly. Furthermore, the interaction with the user can lead to more personalised services. Implementing crowd sensing raises some challenges, according to the interviewees. Also, the interviewees mentioned other aspect that should be taken into account when designing sensing applications. An overview of the aspects related to the implementation of participatory sensing that were described in the interviews are presented in Table 35.

Table 35. Aspects mentioned in interviews

Aspects	**Interview A: Municipality**	**Interview B: Research group connected cars**	**Interview C: Transport operator**
Incentivising individuals		X	X
Transparency about purpose and data use	X	X	X
Transparency about risks	X		X
Perceived severity of types of data being leaked		X	
Required effort by the user		X	
Accuracy of location data			X
Percentage of participants			X

First, users need to be incentivised to participate in sharing data. People make a trade-off between the benefits of using a certain application, and the risks that come with it. In order to motivate people to use an application and share data, the benefits should outweigh the risks. All interviews address parts of this benefit-cost trade-off. Interview A mainly highlights the aspects related to risk, by addressing the ethical aspects that come with collecting data from individuals. Transparency about the purpose of data collection and the related risks are highly important when designing applications for participatory sensing. This is confirmed in Interview C, in which it was emphasised that the party collecting data needs to be transparent about the data that is collected, how it is being used, and for what purpose. Therefore, the goal of the data collection has to be specified clearly before starting with the data collection. This goal should justify the collection of potentially sensitive data from individuals.

According to Interview B, the perceived risk also depends on the types of data that are collected, as well as the perceived severity of these types of data being leaked. Furthermore, the required input from the user was addressed in this interview. If more effort is required from the users, the application can be regarded less attractive.

Lastly, in Interview C it was mentioned that the accuracy of data collected by individuals needs to be ensured. This was something not addressed in the other interviews. According to the interviewee from the transport operator, only GPS data might not be sufficient for determining the precise location of individuals. This finding is confirmed by research by Wilde et al. (2020). It appears that when using GPS data on smartphones, the accuracy of travel mode identification lies somewhere between 38% and 56%. Thus, other information may be needed in addition to location data. The application may ask users to validate their travel modes through prompts. Since walking appears to be the easiest mode to detect (Wilde et al., 2020), this might not be a problem for the use case defined in Interview A, in comparison. At the moment, crowd management is mainly used to identify flows of people walking in a crowded area.

Also, the transport operator mentioned that a high percentage of people would need to participate in order for the data to be useful. Therefore, determining how to motivate travellers to share their data is something that needs to be done before implementation.

The challenges mentioned in the interviews relate to the outcomes of the choice experiment. The implications of these outcomes for the challenges mentioned in the interviews are discussed in the next subsection.

7.4. Implications of quantitative results

From the choice experiment, it can be concluded that people are willing to share data in sensing applications if the benefits outweigh the risks. Several implications for the applications in the field of smart mobility can be derived from the results of the choice experiment.

Effort: Ease of use

First, the level of user participation is an important factor to users of sensing applications. Therefore, applications that require minimal effort from users are expected to be more successful. Effort is closely related to the perceived ease of use of an application. The influence of this factor on the willingness to share data using a participatory sensing application can be explained by the Technology Acceptance Model (TAM), presented in Figure 27. This model, originated by Davis (1989), is an instrument used to predict the likelihood that a new technology will be adopted within a group.

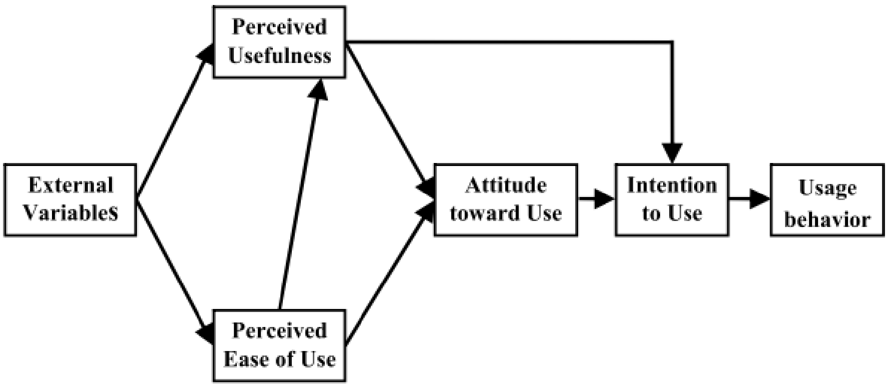


Figure 27. Technology Acceptance Model

In this model, the perceived ease of use is defined as the degree to which the user expects the proposed system or application to be free of effort (Tang & Chen, 2011). If the perceived ease of use is high, it is more likely that the application will be adopted by the user. This is a relation that is also observed in our research. Furthermore, the TAM model describes that if an application is easy to use, users will also believe the application to be more useful in general. Thus, requiring minimal effort from the user is critical when designing effective applications for participatory sensing.

Using the insights from both the choice experiment and the interviews, the Technology Acceptance Model can be extended, specifying the external variables that play a role. First, the perceived ease of use is influenced by the *required level of user participation*. As stated in Interview B, users will find an application less attractive when a lot of input by the user is required. Furthermore, in Interview A and Interview C the importance of a *transparent communication of the purpose* of the data collection was discussed. Here, it was observed that if people are informed about the use of their data, this can influence their attitude towards use.

For the crowd management application, minimal effort for the user can be realised by designing an application that only has to be turned on and off by the user. For the application related to safety research, more inputs might be required from the user. Also, in order to determine the location of travellers in public transport in more detail, other information may be needed, according to Interview C. To provide this information, a sensing application could ask for feedback from the user, in order to confirm if a traveller is located in a certain vehicle. However, increasing the level of user participation also requires more effort from the user side. When designing such applications, it should be taken into account that the application needs to be easy to use for the user. Also, an additional benefit could be considered, such as a reward. This reward could be presented as a financial compensation, but could also be a service or a discount voucher.

Data use: Purpose of data collection

Users are least willing to share data with corporate institutions. Thus, a higher reward may be needed to incentivize people to participate in sensing applications by companies. For example, a car manufacturer asking car drivers to share data in order to be able to do safety research aimed to improve car safety, might have to offer them some benefit in return. Governmental institutions are seen as relatively acceptable by users. Thus, this might benefit municipalities or other governmental institutions when collecting data with the aim of crowd management.

Besides the party that is receiving the data, the goal of the data collection needs to be clear. This is highlighted by Interview A, in which it was mentioned that the purpose of collection needs to be defined beforehand. Also, Interview B confirms that parties need to be transparent about the purpose for which they want to use the collected data.

From the choice experiment results, it appears that data collection by corporate institutions is least accepted by individuals. In the choice experiment, the goal of the corporate institutions was described as “improving products and services”, which is a goal that is still quite abstract. If a corporate institution would formulate clearly what kind of product or service is being improved or realised by the collected data, users might be more willing to share their data. Also, if they would personally benefit from the designed product or service, their willingness to share might increase even more.

Monetary reward: Incentives for sharing mobility data

According to Interview C, only 40% of travellers would participate in crowd sensing in public transport. Individuals need to be willing to share their data, and turn on the required sensors on their smart phone for collecting the data. In order to make sure that a higher percentage of people use the sensing application, it is clear that there needs to be a certain incentive that motivates them to participate. This is also confirmed by Interview B, in which it was mentioned that people need to see the benefit of sharing their data.

From the choice experiment, it can indeed be concluded that people are more willing to share data if they receive a higher amount of money. Thus, this can be an incentive for participating in sensing applications for smart mobility. A reward could be offered to users of public transport if they agree to share their data with transport operators or with a municipality. This could be location and motion data to get insight in capacity of vehicles. In addition the application could ask the user for feedback using short surveys to confirm their location and gather information on the quality of their trip. In return for these data, a small financial compensation could be given based on the amount of surveys that were completed, or the total amount of travel time during which the traveller had the application turned on. Another option is to provide discounts on travel costs. For the safety research case, car manufacturers can ask car drivers to give consent for sharing data from their car. In return for these data, users could be offered a monthly compensation. Car manufacturers can also ask car users if they are willing to share data with third parties, which could lead to a higher financial compensation. Also, insurance companies might be interested in data on driving behaviour. These companies could offer a discount if users consent to data sharing.

Types of data: Flexibility in sharing

According to Interview B, the willingness of people to share data also depends on the types of data and the perceived severity of these data being leaked. From the choice experiment, it can be concluded that the collection of location and motion data is significantly more accepted by users than the collection of context and multimedia data in addition. This is an interesting insight for applications in crowd management. If sensing applications only collect data on location and motion, users are more willing to participate. These data can be valuable for the analysis and prediction of traffic flows.

Interviewee C indicated that a barrier for implementing participatory sensing is that location data often does not provide a sufficient level of accuracy. As discussed in chapter 4, according to Masoud et al. (2019), collecting contextual data in addition to location data can provide a higher level of accuracy. Thus, collecting contextual data could be a solution to obtain accurate information about the precise location of an individual.

However, when, e.g., sound samples are collected in order to get insight in noise pollution, a lot of users might be less willing to participate in a sensing application. This issue could be addressed by giving users the option to voluntarily turn on the function sharing sound samples or other multimedia in addition to location and motion data. In this way, flexibility is provided to users and the users can decide which data they want to share according to their preferences.

Risk of re-identification: Privacy practices

According to Interview A, legal aspects need to be in place when starting to collect data. From the Latent Class Model (Model 3), it appears that people in class B are relatively concerned about the level of privacy that is provided by existing laws and privacy practices. Since the majority of people in our sample (67%) are in class B, attention should be paid to this concern. Additional privacy practices may be needed to ensure a sufficient level of privacy protection for users of sensing applications, and increase their willingness to share data.

From the choice experiment, it appeared that the risk of re-identification was the least important factor influencing the willingness of individuals to share data. However, this does not mean that this is something that should not be taken into account when designing sensing applications. From the interviews, it can be concluded that parties collecting data should be careful when involving the user in the data collection process. The decision whether or not to collect certain data needs to be based on ethical considerations. First of all, in Interview A it was indicated that one of the most important things is that people choose to share their data consciously. Therefore, communication to users is an important aspect (Interview C). Parties collecting data need to be transparent about the data being collected, and about the goal of the data collection. Also, from an ethical point of view, parties need to determine beforehand what kind of data they want to collect from the user, and for what purpose. The goal for which the data is being collected should justify the data collection.

In our choice experiment, people were presented with information on the types of data that were being shared, the risk of re-identification and the use of the data. When developing sensing applications, these are things that should be communicated to the users. In this way, users are enabled to make a well-informed decision whether or not to share data. Besides that, when cooperating with other parties in the field of mobility, parties need to agree on how to minimise the risk of re-identification and how to deal with other privacy-related issues (Interview A).

7.5. Prediction of acceptance in use cases

In this paragraph, the insights from both the interviews and the choice experiment are aggregated. The use cases defined from the interviews are used for the purpose of defining choice situations, with attributes as varied in our choice experiment. By doing this, the acceptance of this specific use cases can be calculated, using the estimations from chapter 6. Based on the acceptance rates of the use cases, different parties in the field of smart mobility (municipalities, car manufacturers, transport operators) can gain insight in which factors they could influence in order to increase the acceptance of their applications.

As described in Chapter 3, the probability that an alternative will be chosen can be calculated using the estimated utility. For all choice scenarios in the choice experiment, we calculated the probability that the "yes" alternative will be chosen. For this calculation, the estimated betas from model 1 (the simple MNL model) were used. The probabilities are presented in Appendix J.

Choice set 5 has the highest probability of individuals choosing to share data (0.89). In this scenario, a monetary reward of €60 per month is offered, a low effort is required, the risk of re-identification is 10%, and only location data is collected. Since the "benefits" are at the highest level and the "costs" at the lowest level, it was already to be expected that this scenario has a high probability of being chosen. In this situation, the data is collected by an academic institution.

Using the betas from model 1, the probabilities for combinations of attribute levels can be calculated that were not included in the choice experiment. This means we can also design hypothetical choice situations for the use cases that were defined using the interviews, and predict the acceptance of these use cases. For every use case, the attribute levels that were used to calculate the acceptance rate for one base scenario are presented in a table. In Appendix K, a full table can be found with acceptance rates for all calculated scenarios.

Crowd management in a city (Interview A)
In this situation, data would be collected by the municipality, which is a governmental institution. For crowd management, location and motion data should be sufficient in order to get insight in the movements of people. Also, low participation would be needed from users, since they would only have to give permission to share their location and motion data. Thus, effort is set at 0, types of data at 1, and data use at 0. For crowd management by a municipality, these would be the minimal requirements. The monetary reward is set at 3 and the risk of re-identification at 0, in order to calculate the maximum probability of this use case. The results of this "base scenario" for the crowd management application are displayed in Table 36.

Table 36. Base acceptance rate for crowd management in a city

Factor	Base scenario
Monetary reward	€60 per month
Effort	Low
Risk of re-identification	10%
Types of data	Location data, motion data
Data use	Governmental institution aiming to improve mobility
Acceptance	**80%**

Using these attribute level values, the probability of the "yes" option being chosen for this use case is 80%. In this situation, the municipality would have to provide a financial compensation of €60 per month, and ensure that the risk of re-identification is at 10%. If the municipality would choose to give no compensation (monetary reward = 0), the acceptance of this use case would decrease to 60%.

If the municipality would, for some reason, only be able to provide a risk of re-identification that is equal to 30%, the acceptance would decrease to 62%. If the risk of re-identification equals 20%, the acceptance is 72%.

According to Interview A, the municipality is also working with academic institutions. The municipality could choose to collaborate with an academic institution and give them the responsibility for the data collection. In that case, the willingness of people to share data would increase to 84%. The municipality would be less involved in the data collection and the data analysis process, which could be a disadvantage. However, the municipality could agree with the academic institution on which data they can use. Since a higher coverage is achieved, the collected data will be more accurate, and thus more useful.

Safety research using car accident information (Interview B)
When applying participatory sensing to safety research, contextual and multimedia data would be useful to get complete information about an accident. These data would be collected by a car manufacturer, which can be categorised as a corporate institution. When setting the monetary reward to €60 per month, the level of effort to low and the risk of re-identification to 10%, the acceptance in this use case is equal to 49%, as

presented in Table 37. This is significantly lower than the crowd management use case. However, for this use case, a lower coverage could be sufficient. A small amount of car drivers willing to share data regarding safety, can already give valuable insights for the kind of safety research that car manufacturers wish to conduct.

Table 37. Base acceptance rate for safety research using car accident information

Factor	**Base scenario**
Monetary reward	€60 per month
Effort	Low
Risk of re-identification	10%
Types of data	Location data, motion data, contextual data, multimedia data
Data use	Corporate institution aiming to improve products or services
Acceptance	**49%**

If the risk of re-identification would be equal to 30%, instead of 10%, the acceptance of sharing data would decrease to a percentage of 29%. This is a decline of 20 percentage points. Thus, if the car manufacturer would succeed in ensuring a low risk for the users, the acceptance can be improved significantly.

If the car manufacturer aims for a higher coverage, an option could be to lower the amount of data that are collected. For example, if location data, motion data, and contextual data would be collected, the acceptance would increase to 58%, which is an increase of 9 percentage points. However, it should be noted that this will provide less information. Another option is to increase the financial compensation received by the users, but this falls outside the range of our choice experiment. Also, the car manufacturer could ask for more inputs by the users. When requiring a moderate level of participation, the acceptance becomes 41%. When the effort by the user is high, the acceptance further declines to 26%. Requiring more inputs from the users decreases the attractiveness of the application, as also indicated in Interview B. However, the data that is collected could also become more valuable for the safety research, since more insights are gained on the preferences, the behaviour, and the observations of the car driver. Thus, a trade-off should be made between the amount of information that is gained, and the attractiveness of the application.

Real-time traffic information in public transport (Interview C)
As described in Interview C, transport operators have the wish to collect more real-time data. These transport operators do not have a specific category in the experiment. They could be categorised as either a governmental institution or a societal organisation. For acquiring accurate information, location and motion data would be required as a minimum. However, it appears that by only using GPS, the precise location of individuals can often not be identified in sufficient detail. Therefore, the transport operator could choose to ask feedback from the users, which can improve the reliability of the data. This would increase the required effort by participants. When setting the level of effort to medium, the risk of re-identification to 10%, and the monetary reward to none, the acceptance is 39% for a societal organisation and 43% for a governmental organisation. The attribute levels used to calculate this base acceptance rate are presented in Table 38. This finding is quite in line with the percentage of 40% expected by the transport operator. As stated in the interview, this percentage does not provide sufficient information.

Table 38. Base acceptance rate for real-time travel information in public transport

Factor	**Base scenario**
Monetary reward	None
Effort	Moderate
Risk of re-identification	10%
Types of data	Location data, motion data
Data use	Governmental institution aiming to improve mobility
Acceptance	**43%**

The data that is collected by the transport operator already provides a benefit for travellers, since more accurate travel information can be delivered. However, in order to achieve a higher percentage of acceptance, the transport operator could choose to give travellers a financial compensation for participating in the application. Providing a reward of €20 per month would lead to an acceptance of 46% (societal organisation) or 51% (governmental institution). Increasing the compensation to €60 per month leads to an acceptance of 62% (societal organisation) or 66% (governmental institution). This is an increase of 23 percentage points, compared to the situation without a monetary reward.

Instead of asking feedback from users, transport operators could also additionally choose to collect contextual data and multimedia data. These data could provide more accurate information about the context and the mode of transport of the individual at a certain moment. Furthermore, this could give information on the quality of a trip, such as noise pollution, air quality, or bumpiness. In this case, when keeping the monetary reward at €60 per month, the acceptance would be equal to 61% (societal organisation) and 65% (governmental institution). This percentage is slightly lower than the situation in which the effort is set to medium and only location and motion data is collected. If the transport operator would choose not to collect multimedia data, the acceptance would increase to 69% (societal organisation) or 73% (governmental organisation). This is an increase of 30 percentage points when compared to the “base case” of this use case. Since the research by Masoud et al. (2019) argued that the accuracy of location information can be improved by collecting environmental data, this could be an effective way to gain accurate insights, while keeping the acceptance rate at a relatively high level.

An option for the transport operator to increase the acceptance could be to require a low level of participation, and only collect location and motion data. This would lead to an acceptance of 56% (societal organisation) and 60% (governmental institution). When also adding a monetary reward of €60 per month, the acceptance will increase to respectively 76% or 80%. However, accuracy will be lost in this case. This shows a trade-off has to be made by the transport operator between accuracy on one hand, and required effort or privacy on the other hand.

7.6. Conclusion

In conclusion, several implications can be derived for applications in the field of smart mobility, based on the results from chapter 6. Examples of applications of participatory sensing are crowd management in a city, safety research using car accident information, and providing real-time travel information in public transport. Regarding the *required effort* factor, parties implementing sensing applications have to make a decision on the inputs and feedback they want to ask from the users. However, ensuring ease of use means requiring minimal effort for the user. Additional user feedback can provide richer information, but will also have a negative influence on the acceptance of users, since the application could be perceived less attractive. Especially for applications in public transport, a high acceptance, which can lead to a higher coverage, is important. Furthermore, the types of data being collected is an important aspect to be considered. Here, a trade-off between richness of information and attractiveness of the application needs to be made as well.

Also, an implication for parties in the field of smart mobility is that the purpose of the data collection and the use of the collected data has to be clearly communicated to the users, in order to increase the acceptance. According to insights from the interviews, this purpose needs to justify the collection of data that can be potentially sensitive. Ethical considerations, and thus keeping the risk of re-identification at a minimum, are important when implementing participatory sensing applications.

Based on these implications, practical recommendations can be proposed for developing applications for participatory sensing that take into account values like trust and privacy. Section 9 elaborates on these recommendations.

8 Discussion

This section reflects on the research and addresses its limitations. Subsection 8.1 provides a reflection on the conducted research. In 8.2, the limitations of this study are discussed.

8.1. Reflection

The conducted research provides a new understanding of user preferences in participatory sensing systems. We learned that the required effort is a factor that is highly important for users when making a decision whether or not to participate in sensing applications. Also, we gained new insights in the preferences of people that are highly concerned about their privacy and the handling of their personal data, by using the indexes as defined by Westin. Besides understanding the user side of participatory sensing applications, we were able to get insight in the challenges experienced by parties in the field of smart mobility as well, when implementing such applications. By conducting interviews with relevant parties, we gained a broad understanding of current developments in the smart mobility sector. Combining both quantitative and qualitative approaches provides us with a unique insight in the acceptance of users in scenarios derived from realistic use cases.

Furthermore, this study can be interpreted in a broader sense, by getting an understanding of the motivations of individuals behind sharing data in general. In this research, the focus was on applications in the field of smart mobility. However, the factors included in our choice experiment are generic and can be used as an inspiration for conducting similar experiments in other fields of research, such as health- or environment-related applications. This delivers valuable insights for implementing smart city applications outside of the smart mobility sector.

Reflecting on the research process, there were also some difficulties we experienced. Combining the quantitative results of a choice experiment with qualitative results from interviews is an approach which is not taken before in this field, according to our knowledge. Therefore, we could not rely on previous studies for applying the choice modelling results to realistic use cases. Looking back, we would have integrated the interviews with relevant parties in an earlier stadium of the research, involving them in the selection of factors for the choice experiment as well. In that way, the experiment design would have been founded on knowledge of experts in the field, besides knowledge gained from previous studies. Also, the results could have been of higher relevance for these parties. Furthermore, because of the relatively short time frame of the research, we were only able to interview three different parties. If more parties would have been interviewed, use cases could have been built upon views and experiences from multiple parties, addressing a broader part of the smart mobility field. However, we think that the three parties mentioned in this study cover a large variety of potential applications for participatory sensing, making them useful for applying the results of the choice experiment.

The decision for using a discrete choice experiment as the main method in this research requires some reflection. An alternative method that could have been applied in order to address the knowledge gap is a gaming approach. By studying the actions of individuals in a simulated environment, their preferences can be elicited. This approach could be extended with Q methodology, which is used to investigate different perspectives among a group of participants. The influence of these perspectives on the choices made by individuals in the serious game can then be investigated. This alternative would also be a suitable method for exploring motivations behind data sharing choices. However, the advantage of a discrete choice experiment compared to alternative methods, is that the method allows for an explication of the trade-offs that play a role in the motivation of individuals regarding data sharing. Based on these trade-offs, predictions can be made on choices for other combinations not included in the choice experiment. The DCE indirectly recovers the values behind people's choices, which provides insights on ethical aspects to incorporate in application design. A unique characteristic of the choice modelling approach is that is assigns a numerical value to the weighing of factors underlying people's motivations. Moreover, using a survey allows for reaching a larger group of people in a relatively short amount of time. However, a disadvantage of the discrete choice experiment is that there is no possibility to ask open-ended questions or follow-up questions regarding users' motivations, which could enrich the explanation of the observed choice behaviour. Alternative methods, like a gaming approach, do have this possibility. Therefore, alternative methods could complement and enrich the conducted research. Recommendations for alternative methods to be used in future studies are provided in section 9.2.

Although the research provides a valuable contribution to previous research, there are several limitations that should be noted when interpreting the results of the study. These limitations are discussed in the next subsection.

8.2. Limitations

First, there were some limitations related to the data collection. Via an online platform, the responses to the survey were collected. However, some respondents indicated that the survey was difficult and took quite some time to complete. This could have led to people not finishing the survey. This is reflected in the relatively high number of unfinished surveys (63). Thus, the preferences of these people are not included in the results.

Furthermore, it appeared that the academic institution was regarded as most acceptable by respondents. However, this could be due to a selection bias in the sample. Since a relatively high amount of students responded to the survey, it could be that they are more positive in general about academic institutions and sharing data with universities or for research. The willingness to share data with academic institutions could therefore be overestimated in our research.

Besides the relatively high number of academics in the sample, there could have been a selection bias regarding the general willingness to share data. It appeared that 3 respondents did not consent with participating in the research. People that do not want to share data in a survey for research might also be reluctant to participate in sharing data in sensing applications. Thus, this group of people not sharing data in general may be underrepresented in our sample.

Another limitation regarding the data collection is that we do not have information on the location or the living situation of the respondents. However, preferences could be different per area. In certain areas, the willingness to share data could be different. Furthermore, people living in cities or people using public transport regularly may be more willing to participate in sensing applications for smart mobilities, since they are directly affected by the benefits of the data collection. This is an effect that was not included in our survey.

Furthermore, some limitations regarding the setup of the experiment can be noted. First, in order to limit the length of the survey, the amount of attributes was reduced to only 5 factors, while a total amount of 14 factors was identified in the literature research. Due to this simplification, other factors potentially influencing the willingness to share data are unaddressed in this study. However, these factors may have led to a more richer explanation of the choice behaviour of smart device users regarding data sharing.

The second limitation that is related to the setup of the experiment relates to the monetary reward attribute. When participants know that several alternatives provide a financial compensation, this might influence their decisions on alternatives without a reward. Adding the monetary reward may have led to "crowding out" of intrinsic motivations of participants.

Also, in our experiment we chose to vary the attribute levels of the *monetary reward* attribute between €0 and €60. If a lower or a higher upper limit would be chosen, the relative importance of this attribute might be different. Besides that, the range of the monetary reward attribute can influence the calculated Value of Privacy. If the *monetary reward* attribute would range, e.g. from €0 to €100 instead, the Value of Privacy might be higher.

A final limitation in the experiment setup concerns the *risk of re-identification* attribute. This attribute gives only a limited indication of the degree of privacy protection. The risk of re-identification was based on the principle of *k*-anonymisation. In our study, the risk of re-identification was the least important attribute influencing the decision of individuals to share data. This could be due to two limitations. First, we chose to let the attribute vary between 10% and 30%, which is only a limited range. Although "full" anonymisation is not really possible (which would be the 0% level), the range could be larger in reality. Secondly, the question is whether participants really understood what the risk of re-identification means and what effect it can have on their privacy. In the survey, an explanation of the attribute was provided. However, the risk, described in percentages, is still a somewhat intangible attribute, especially when compared to a more tangible attribute like a financial compensation in euros.

Finally, some limitations regarding the research method should be noted. First, the use of a choice experiment, distributed via an online survey, is close-ended and does not give room to ask participants about their underlying motivations when making a choice. Also, stated choice experiments assume a utilitarian view. In this study it was assumed that values related to privacy can be traded off against monetary or societal benefits. However this method gives useful and quantitative insights in user preferences, ethical aspects regarding privacy and trust also play an important role, that should not be overlooked. Even if consumers indicate they want to sacrifice a part of their privacy in turn for some benefit, the question should always be asked if certain data should be collected (Shilton & Estrin, 2012). This is underlined by Interview A, in which was mentioned that ethical aspects have to be taken into account before collecting data. The party collecting data should always keep in mind that participants may not fully understand the risks related to sharing data and that the privacy of participants should not be brought in any danger.

The limitations mentioned should be taken into account carefully when interpreting the results of the study. Also, the limitations provide a basis for conducting further research. Recommendations for future studies are described in chapter 9.

9
Conclusion

This final section aims to answer the research questions addressed in this research. The first subsection reflects on the sub-research questions and the main research question of the study. Subsection 9.2 provides recommendations for further research, after subsection 9.3 addresses practical recommendations on how the insights from this study can be incorporated in a value-sensitive design of applications for participatory sensing.

9.1. Reflection on research questions

This subsection reflects on the research questions addressed in this study.

SQ1. What factors potentially incentivize or disincentivize individuals to contribute to participatory sensing systems?

First, smart device users can be incentivised to participate in sensing applications by offering them a personal benefit. Several types of benefits are found in literature. A first way to motivate individuals to share data is by giving them a monetary reward for submitting data. According to the literature, providing even a small amount of money can significantly increase participation rates. These monetary rewards can be provided in the form of cash, coupons, or discounts. Also, non-monetary benefits could be offered, by turning sensing tasks into a game, by giving participants access to a useful service, or by implementing reputation mechanisms. A final incentive that was found is the societal benefit of the data collection campaign. When people recognise the relevance of the application to society, this can increase their willingness to share data. Societal benefits can intrinsically motivate individuals, or arouse feelings of moral obligation. However, only implementing societal benefits can cause free-riding effects.

Besides incentives that can increase the willingness to share data, factors are found in literature that potentially disincentivise smart device users to participate in sensing applications. The required level of effort by the user of the application could influence individuals' choices. Three levels of effort are distinguished. In the lowest level, participants are aware of the data being collected and provide consent, but are minimally or passively involved in the system. Applications requiring moderate effort engage users by requiring them to respond to prompted questions. On the highest level of participation, users are actively involved in the sensing process by looking for more data and submitting sensing reports. It was expected that a higher effort decreases the willingness of individuals to participate. Furthermore, sensing applications can consume computation power, battery power, and network bandwidth. This can be another disincentive for people to participate in sensing applications.

Based on the importance of factors according to previous research and the possibility for factors to be influenced by policy or design, the factors considered most suitable for inclusion in the choice experiment are a *monetary reward* and the *effort* required from the user.

SQ2. What factors relating to privacy potentially influence the decision of individuals to share data in participatory sensing systems?

In literature, the risk of re-identification is considered one of the most important factors influencing users' willingness to share data. It was expected that a lower level of anonymisation, i.e. a higher risk of re-identification, significantly decreases people's willingness to share data. Different types of data can be collected in sensing applications. The types of data considered in this study are location data, motion data, contextual data, and multimedia data. Prior to the choice experiment, the expectation was that collecting more types of data decreases the willingness to share data. Furthermore, it was expected that data collection by academic institutions is more accepted by users than data collection by corporate institutions. Existing literature does not agree on the acceptance of data collection by governmental institutions. Other privacy-related factors influencing the willingness to share data, according to literature, are the purpose of the data collection, the security of the application, the review of the data sharing by an independent body, the extent to which information if provided, the duration of the data collection process, and the amount of data being collected. The factors that were developed to be included in the choice experiment are the *risk of re-identification,* the *types of data,* and the *data use.* The choice was made to include these factor because of their prevalence in literature and the expected importance of these factors to users' privacy perception.

Furthermore, factors indirectly influencing the decision to share data are found in literature. People scoring high on the Privacy Index and the Distrust Index as defined by Westin, are considered as more concerned about their privacy, and more distrustful about party's handling their data. Therefore, this could influence the importance they assign to privacy-related factors.

Also, personal characteristics are taken into account, due to the influence these factors can have on the perception of people on technology, privacy, and data sharing, according to literature. These are *age, gender, education, income level, digital behaviour,* and *altruism.*

SQ3. What trade-off do individuals make between potential costs and benefits of participation, when choosing to share data in participatory sensing systems?

Using the factors identified from the literature review, a choice experiment was constructed. When making a trade-off between potential costs and benefits of participating in a sensing application, the required effort appeared to be the most important factor for users. This factor determines the choice of smart device users to participate in sensing applications for a percentage of 26%. The use of the collected data was the factor that was the second most important factor influencing the willingness to share data. Data collection by academic institutions is regarded most acceptable by users, followed by governmental institutions, societal organisations, and corporate institutions, in order of acceptance. Furthermore, people are more likely to share data if they get a monetary benefit. According to the calculated "Value of Privacy", people want to receive an amount of €11.30 per month if the risk of re-identification is increased by 10 percentage points. Besides that, the willingness to share data depends on the types of data that are being collected. The collection of both location and motion data is accepted by most of the respondents, while a lot of respondents are reluctant to share multimedia in addition. This finding is in line with previous studies conducted recently. Lastly, the risk of re-identification is considered the factor of least importance when deciding whether or not to share data, which is a surprising result when compared to previous studies.

However, this does not mean that privacy is not important to smart device users. A group of people does not trust their data being handled carefully by parties. For this group, the risk of re-identification is the most important factor influencing their willingness to share data. Also, it appears that there is a group of people who are highly concerned about their privacy, also called "Privacy Fundamentalists", who attach high importance to the use of the data being collected. Especially when data is collected by a corporate institution, they are more concerned and less willing to share their data. To these groups of people, the monetary reward and the required effort are less important when making a trade-off between benefits and costs of sharing data.

SQ4. Regarding these trade-offs, what are implications for different applications in the field of smart mobility?

Based on the estimated models indicating the trade-offs made by individuals, the acceptance of different use cases in the field of smart mobility was calculated. Three interviews were conducted, with each interview leading to the definition of a use case in the field of smart mobility. The first use case to which the results were applied is the case of crowd management in a city. By using participatory sensing applications collecting mobility data, a municipality can gain insights in crowdedness or congestion in a city. When gathering location data and motion data for this purpose, requiring a low effort from users, providing a monetary reward of €60 per month, and keeping the risk of re-identification equal to 10%, the acceptance rate of this use case is equal to 80%. Collaborating with academic institutions could provide a higher coverage.

The second use case describes the use of data gathered on car accidents, for the purpose of allowing car manufacturers to improve the safety of their cars. For this aim, location data, motion data, contextual data, and multimedia data can be collected with sensing applications. In combination with a monetary reward of €60 per month, a low effort required, and a risk of re-identification equal to 10%, the acceptance rate for this use case will be equal to 49%. The car manufacturer could also choose to require more inputs by the user, in order to add more value to the data. However, a trade-off needs to be made between the richness of information and the attractiveness of the application.

The last interview led to the definition of a use case on public transport. The transport operator that was interviewed wishes to collect real-time data, in order to respond to incidents more quickly and provide personalised travel information to travellers. When achieving a coverage that is sufficiently high, participatory sensing could make this possible. In order to ensure the accuracy of location information, feedback from the user can be asked. In this case, with a moderate effort, a risk of re-identification of 10%, and collecting location and motion data, the acceptance rate is 43% when no compensation is provided, which is a percentage that is in line with the expectations of the transport provider. However, when providing a monetary reward, the acceptance rate increases up to a rate of 66% in the case of a compensation of €60 per month. Alternatively, the transport operator can collect additional data in order to improve accuracy, instead of requiring more user feedback, which will lead to a higher acceptance. Also, the acceptance can be improved by requiring both low effort and collecting less types of data. However, accuracy will partly be lost in that case.

In conclusion, the four sub-research questions that were addressed answer the main research question of the research, which is:

How do factors relating to incentives and privacy affect the willingness among smart device users to contribute to participatory sensing systems for smart mobility?

The results of this study provide new insights in user preferences in applications for participatory sensing, and on the implications of these preferences for the design of such applications. The following subsection addresses the value of the insights gained in this study to the research field.

9.2. Scientific contribution and recommendations for further research

The conducted research contributes to science in several ways. First, our study uses a stated choice experiment to explain user behaviour in participatory sensing applications. Since little research has been conducted before in this area, this study provides new insights regarding user behaviour and incentives for participatory sensing.

Studies conducted before used a choice modelling approach to investigate the Benefit-Cost trade-off made by users when sharing data, based on Privacy Calculus theory. However, in this study we chose to include the effort attribute besides other attributes related to privacy and incentives, in contrast to previous studies. The effort required by users of sensing applications appeared to be highly important when deciding whether or not to participate. This is a valuable insight that adds to Privacy Calculus research. Furthermore, a strength of this research is that a novel approach was used by combining a choice experiment with qualitative interviews. In this way, the quantitative results could directly be applied to specific use cases and connected to current practices. The insights gained in this research raise new questions that could be addressed in further research. Building upon this research, several future studies are recommended.

First, the choice experiment can be repeated with a larger amount of respondents that is more representative for the population. This can increase the reliability of the results. Furthermore, approached individuals that do not want to participate in the experiment could be asked to give a reason for not participating. In that way, it can be seen if there is a selection bias in the sample related to the general willingness to share data.

Regarding the setup of the experiment, it is recommended to repeat the experiment without a monetary reward, in order to see if the crowding out process influences the decision of participating in sensing applications. Also, discounts, vouchers, or access to a useful service could be proposed as a personal benefit instead of a financial compensation, in order to investigate if this changes the choice behaviour of users. Another incentive that was not addressed in this research is attracting users through gamification. Future research could investigate the effects of including gaming aspects in participatory sensing applications.

Further research is also needed on the range of the monetary reward attribute. Pre-investigations could be done on the amount of money that participants would want to receive in turn for sharing their data. In this way, a more realistic range can be used.

Derikx et al. (2016), in a study on sharing data with insurance companies, found that people want to sell data for third party advertisements, but want *pay* money when they receive relevant personalised promotions from the insurance company itself. In future research, it could be interesting to investigate if people would want to pay for participating in a sensing application, if it provides them with a useful service.

Besides the monetary reward attribute, the risk of re-identification requires further investigation. In the conducted research, this factor appeared to be the least important factor, which contradicts previous studies. As described in section 8, this finding could be due to insufficient understanding of the *risk of re-identification* factor. Further research is required in order to investigate if the risk of re-identification is indeed of lower importance to individuals when compared to other factors. For example, future studies could take a more technical approach to investigate the risks that are related to participating in sensing applications, and in what way they can be quantified. These insights could then be used to inform a new choice experiment, which uses a more realistic range for the risk of re-identification attribute.

Something that was mentioned in the conducted interviews is the understanding of privacy risk by the user. If users do not know about the potential risks when sharing data, this might influence their decision. In our study, this was partly addressed by asking questions derived from Westin's Privacy Index. However, a recommendation for further research is to take a deeper look at the knowledge respondents have on the risks that come with sharing data, and how they perceive the severity of these risks. This can give more insights in risk perceptions of users and conclusions could be drawn on how knowledge on privacy risks influences the trade-offs made by consumers.

In our research, it appeared that the required effort plays an important role in decisions regarding data sharing. Thus, this factor requires further investigation. Wider ranges of attribute levels can be used in a new choice experiment to look at how much time people are willing to spend on a specific application and under what conditions. Also, it would be interesting to include a question in the survey on the interestedness of the respondent in the application. The answers to this question can be used to investigate if a higher interest in a particular application leads to an increase in the willingness to put more effort in contributing to the application. Furthermore, future research can look at the effort aspect from a design science point of view. For example, the question how to increase the user-friendliness of sensing applications can be addressed.

The conducted choice experiment had a general focus on smart mobility. It would be interesting to conduct an experiment that is specifically focused on one application, such as crowd management in a specific city. This could make the benefits of participatory sensing more relatable for respondents.

Also, it is advised to repeat our experiment for applications in different fields, such as health, in order to see if preferences regarding such applications differ from applications for smart mobility.

In Interview A, it was mentioned that it is important that in the future, data can be standardised so it can be exchanged easily between parties. Further research can take a technical approach and investigate how data collected from different sources can be aggregated in an efficient way.

The final recommendations concern the approach of the research. As mentioned in the previous subsection, the use of discrete choice modelling as a method has some limitations. The research field of participatory sensing for smart mobility can be enriched by conducting research with a different methodology. Using a more open-ended way of asking questions to respondents, such as an interview, can give more insight in underlying motivations and behaviour. Another idea is to conduct an empirical research by performing a gaming study with a sensing application and monitor users and their behaviour. Investigating behaviour of users in real situations instead could give different insights compared to studies using hypothetical situations.

9.3. Societal contribution and recommendations for design

Besides the contribution to science, this research also provides new insights that can be incorporated when designing applications for participatory sensing. In this study, implications for a municipality, a car manufacturer, and a transport operator are discussed. The use cases described for these parties can be used as an example for other parties aiming to implement participatory sensing as a solution for collecting real-time data. Taking into account the findings on user preferences, more effective sensing applications can be designed. Data collected by these applications can provide a better understanding of traffic flows and contribute to future-proof mobility by improving safety and efficiency.

From the conducted interviews and the results of the choice experiment, some practical recommendations can be derived for designing participatory sensing applications. The recommendations are structured by describing principles to take into account in the design process. Specific examples are given that could be used for implementation.

Tailor-made sensing applications

From the choice experiment and the interviews, it appears that the personal preferences of individuals regarding data sharing differ. For example, some people indicate to be comfortable with sharing multimedia data, while others are only willing to share location and motion data. In order to meet preferences of different kinds of users, the participatory sensing application should be dynamic. This can be accomplished by providing more flexibility for users. Thus, a first recommendation is to give users a high amount of control. An example of such an application can be found in a study by Heiskala et al. (2016). In this application, users were given control over how their personal data was used and sold. Users could receive a discount if they opted-in for location-guided advertisement. Also, they could pay a subscription fee to be able to opt-out of sharing personal data within the sensing system. Inspired by this example, parties designing participatory sensing applications could give users the option to indicate in the application which types of data they are willing to share (location data, motion data, contextual data, multimedia data). Also, if users are willing to share a higher amount of information, which could increase their risk of re-identification, they could be offered a higher reward for participating. This could be a higher financial compensation, or another kind of benefit, such as access to a service. Designing tailor-made applications for participatory sensing could attract a higher variety of users.

Transparency by design

From the choice experiment, it appears that the risk of re-identification was least important in the decision of individuals on sharing data. Nevertheless, a high amount of people indicated to be concerned about their privacy. As clarified in the interviews, it is highly important to take into account this privacy-related factor from an ethical point of view, when collecting data from individuals. A reason that individuals did not find this factor more important in the choice experiment could be that the risk of re-identification is a somewhat intangible attribute to respondents. Users might not fully understand the risks that come with sharing data. However, when collecting data from individuals, parties should communicate the purpose of data collection as well as the related risks clearly (Interview C). Especially to people who are highly concerned about their privacy, the data use factor is highly important. Therefore, being transparent about the use of the collected data is important when aiming to address this group of Privacy Fundamentalists as well.

An idea for doing this is by designing systems that interact with users to help them understand privacy risks (Shilton & Estrin, 2012). Furthermore, a strong form of giving consent could be implemented. In order to do this, systems might go beyond requiring passive consent by encouraging people to engage with the system. This could mean that users are asked to make decisions about system use, or even about the design. Autonomy can be given to users by empowering them to decide how data collection, data analysis, and research results will be handled. For example, they could be involved in discussions about what parties should be able to get access to what types of data. A practical way to involve users in making decisions, is to let them participate in prototype development from the beginning. In this way, they can also engage in discussions of how privacy should be incorporated in the application (Shilton & Estrin, 2012). Being transparent to users from the beginning of the design process ensures that people are aware of the risks that come with sharing their data. This can help them to make a well-informed decision and can increase the ethical value of the design.

Ensure attractiveness of applications

As described previously, the required level of participation is highly important to users. Therefore, this factor is essential when designing effective sensing applications. Minimising the user-burden can increase the acceptance of the application. A way to realise this is to design applications that run in the background and have an easy-to-use interface. However, for some applications, such as for the collection of travel data, a higher level of user participation might be required in order to ensure the reliability of data. Therefore, another solution may be needed. An idea to do this is to make the application more attractive and enjoyable for users by including gamification in the sensing applications. For example, outstanding contributions can be identified and given social incentives, like awards. If the sensing task can be translated in into an enjoyable game action, this may be a motivation for users to participate (Ogie, 2016). According to Ogie (2016), this is an appropriate solution when the sensing tasks do not have time constraints, i.e. when the application runs in the background. When a lot of manual inputs are required by the user, monetary incentives are considered to be more appropriate. In this case, instead of awarding users with a fixed amount of money per month, like in the conducted choice experiment, the financial compensation could be based on the effort spent by a particular user. The idea behind this is that some sensing tasks are more difficult or require more time. Therefore, higher amounts of incentive are necessary for these tasks.

Bibliography

Aitken, M., McAteer, G., Davidson, S., Frostick, C., & Cunningham-Burley, S. (2018). Public Preferences Regarding Data Linkage for Health Research: A Discrete Choice Experiment. *International Journal of Population Data Science*, *3*(11), 1–13. https://doi.org/10.23889/ijpds.v3i1.429

Anawar, S., Adilah, W., Adnan, W., & Ahmad, R. (2017). *A Design Guideline for Non-Monetary Incentive Mechanics in Mobile Health Participatory Sensing System*. *12*(21), 11039–11049.

Antoniou, V. (2017). Mapping and the Citizen Sensor. In *Mapping and the Citizen Sensor*. https://doi.org/10.5334/bbf.a

Bawany, N. Z., & Shamsi, J. A. (2015). Smart City Architecture: Vision and Challenges. *International Journal of Advanced Computer Science and Applications*, *6*(11), 246–255.

Bennati, S., Dusparic, I., Shinde, R., & Jonker, C. M. (2018). Volunteers in the smart city: Comparison of contribution strategies on human-centered measures. *Sensors (Switzerland)*, *18*(11). https://doi.org/10.3390/s18113707

Bhatnagar, S., & Kumra, R. (2020). Understanding consumer motivation to share IoT products data. *Journal of Indian Business Research*, *12*(1), 5–22. https://doi.org/10.1108/JIBR-09-2019-0268

Burke, J., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S., & Srivastava, M. B. (2006). Participatory sensing. *The 4th ACM Conference on Embedded Networked Sensor Systems*, 5. https://escholarship.org/uc/item/19h777qd

Calado, D., & Pardal, M. L. (2018). Tamper-proof incentive scheme for mobile crowdsensing systems. *NCA 2018 - 2018 IEEE 17th International Symposium on Network Computing and Applications*, 0–7. https://doi.org/10.1109/NCA.2018.8548093

CBS. (2019). *Opleidingsniveau*. https://www.cbs.nl/nl-nl/nieuws/2019/33/verschil-levensverwachting-hoog-en-laagopgeleid-groeit/opleidingsniveau

CBS. (2020). *Inkomen van personen; inkomensklassen, persoonskenmerken*. https://opendata.cbs.nl/statline/?dl=D4D1#/CBS/nl/dataset/83931NED/table

CBS. (2021a). *Bevolking; geslacht, leeftijd en burgerlijke staat, 1 januari*. StatLine. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/7461BEV/table?fromstatweb

CBS. (2021b). *Bevolking; onderwijsniveau; geslacht, leeftijd en migratieachtergrond*. StatLine. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/82275NED/table?fromstatweb

ChoiceMetrics. (2018). *Ngene 1.2 User Manual & Reference Guide*.

Chorus, C., & Van Cranenburgh, S. (2020). Choice Models. In *TPM Modelling Book*.

Christin, D., Büchner, C., & Leibecke, N. (2013). What's the Value of Your Privacy? Exploring Factors That Influence Privacy-sensitive Contributions to Participatory Sensing Applications. *2013 Workshop on Privacy and Anonymity for the Digital Economy*, 918–923.

Christin, D., Reinhardt, A., Kanhere, S. S., & Hollick, M. (2011). A survey on privacy in mobile participatory sensing applications. *The Journal of Systems & Software*, *84*(11), 1928–1946. https://doi.org/10.1016/j.jss.2011.06.073

Clarke, R. (1999). Internet Privacy Concerns Confirm the Case for Intervention. *Communications of the ACM*, *42*(2), 60–67.

Cottrill, C. D., Jacobs, N., Markovic, M., & Edwards, P. (2020). Sensing the City: Designing for Privacy and Trust in the Internet of Things. *Sustainable Cities and Society*, *63*(July), 1–9. https://doi.org/10.1016/j.scs.2020.102453

Cristofaro, E. De. (2014). *Participatory Privacy : Enabling Privacy in Participatory Sensing Participatory Privacy: Enabling Privacy in Participatory Sensing ∗. June*. https://doi.org/10.1109/MNET.2013.6423189

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*, 123–130.

Derikx, S., de Reuver, M., & Kroesen, M. (2016). Can privacy concerns for insurance of connected cars be compensated? *Electronic Markets*, *26*(1), 73–81. https://doi.org/10.1007/s12525-015-0211-0

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.l060.0080

Du, R., Member, S., Santi, P., Xiao, M., Member, S., Vasilakos, A. V, & Fischione, C. (2019). *The Sensable City : A Survey on the Deployment and Management for Smart City Monitoring*. *21*(2), 1533–1560.

Ekman, K., & Weilenmann, A. (2021). Behind the scenes of planning for public participation: planning for air-quality monitoring with low-cost sensors. *Journal of Environmental Planning and Management*, *64*(5), 865–882. https://doi.org/10.1080/09640568.2020.1787129
Finn, R. L., Wright, D., & Friedewald, M. (2013). Seven Types of Privacy. *European Data Protection: Coming of Age*. https://doi.org/10.1007/978-94-007-5170-5
Gao, H., Liu, C. H., Wang, W., Zhao, J., Song, Z., Su, X., Crowcroft, J., & Leung, K. K. (2015). A survey of incentive mechanisms for participatory sensing. *IEEE Communications Surveys and Tutorials*, *17*(2), 918–943. https://doi.org/10.1109/COMST.2014.2387836
Gemeente Amsterdam. (2021). *Innovatie*. Gemeente Amsterdam. amsterdam.nl/innovatie/mobiliteit
Gerber, N., Gerber, P., & Volkamer, M. (2018). Explaining the privacy paradox: A systematic review of literature investigating privacy attitude and behavior. *Computers and Security*, *77*, 226–261. https://doi.org/10.1016/j.cose.2018.04.002
Goulias, K. G., & Pendyala, R. M. (2014). Choice context. In *Handbook of Choice Modelling* (pp. 101–130). https://doi.org/https://doi.org/10.4337/9781781003152.00011
Heidel, A., Hagist, C., & Schlereth, C. (2021). Pricing through health apps generated data- Digital dividend as a game changer: Discrete choice experiment. *PLoS ONE*, *16*(7 July), 1–14. https://doi.org/10.1371/journal.pone.0254786
Heiskala, M., Jokinen, J. P., & Tinnilä, M. (2016). Crowdsensing-based transportation services - An analysis from business model and sustainability viewpoints. *Research in Transportation Business and Management*, *18*, 38–48. https://doi.org/10.1016/j.rtbm.2016.03.006
Hess, S., Rose, J. M., & Polak, J. (2010). Non-trading, lexicographic and inconsistent behaviour in stated choice data. *Transportation Research Part D: Transport and Environment*, *15*(7), 405–417. https://doi.org/10.1016/j.trd.2010.04.008
Hollin, I. L., Craig, B. M., Coast, J., Beusterien, K., Vass, C., DiSantostefano, R., & Peay, H. (2020). Reporting Formative Qualitative Research to Support the Development of Quantitative Preference Study Protocols and Corresponding Survey Instruments: Guidelines for Authors and Reviewers. *Patient*, *13*(1), 121–136. https://doi.org/10.1007/s40271-019-00401-x
International Telecommunication Union. (2016). *Mobile network coverage by country*. https://www.theglobaleconomy.com/rankings/Mobile_network_coverage/
Ioannou, A., Tussyadiah, I., & Lu, Y. (2020). Privacy concerns and disclosure of biometric and behavioral data for travel. *International Journal of Information Management*, *54*(January), 102122. https://doi.org/10.1016/j.ijinfomgt.2020.102122
Issarny, B., Bouloukakis, G., & Georgantas, N. (2018). When Service-Oriented Computing Meets the IoT: A Use Case in the Context of Urban Mobile Crowdsensing Invited Paper. *ESOCC*, 1–16. https://doi.org/10.1007/978-3-319-99819-0
Jaimes, L. G., Vergara-Laurens, I. J., & Raij, A. (2015). A Survey of Incentive Techniques for Mobile Crowd Sensing. *IEEE Internet of Things Journal*, *2*(5), 370–380. https://doi.org/10.1109/JIOT.2015.2409151
Jin, J., Gubbi, J., Marusic, S., & Palaniswami, M. (2014). An information framework for creating a smart city through internet of things. *IEEE Internet of Things Journal*, *1*(2), 112–121. https://doi.org/10.1109/JIOT.2013.2296516
Johansson, J. V., Shah, N., Haraldsdóttir, E., Bentzen, H. B., Coy, S., Kaye, J., Mascalzoni, D., & Veldwijk, J. (2021). Governance mechanisms for sharing of health data: An approach towards selecting attributes for complex discrete choice experiment studies. *Technology in Society*, *66*(March). https://doi.org/10.1016/j.techsoc.2021.101625
Kamel Boulos, M. N., & Al-Shorbaji, N. M. (2014). On the Internet of Things, smart cities and the WHO Healthy Cities. *International Journal of Health Geographics*, *13*, 1–6. https://doi.org/10.1186/1476-072X-13-10
Kamilaris, A., & Pitsillides, A. (2016). Mobile Phone Computing and the Internet of Things: A Survey. *IEEE Internet of Things Journal*, *3*(6), 885–898. https://doi.org/10.1109/JIOT.2016.2600569
Khoi, N. M., & Casteleyn, S. (2018). Analyzing spatial and temporal user behavior in participatory sensing. *ISPRS International Journal of Geo-Information*, *7*(9). https://doi.org/10.3390/ijgi7090344
Khoi, N. M., Casteleyn, S., Mehdi Moradi, M., & Pebesma, E. (2018). Do monetary incentives influence users' behavior in participatory sensing? *Sensors (Switzerland)*, *18*(5), 1–29. https://doi.org/10.3390/s18051426
Killi, M. (2007). Lexicographic answering in travel choice: Insufficient scale extensions and steep indifference curves? *European Journal of Transport and Infrastructure Research*, *7*(1). https://doi.org/10.18757/ejtir.2007.7.1.3372

Klopfenstein, L. C., Delpriori, S., Aldini, A., & Bogliolo, A. (2019). “Worth one minute”: An anonymous rewarding platform for crowd-sensing systems. *Journal of Communications and Networks*, *21*(5), 509–520. https://doi.org/10.1109/JCN.2019.000051

Kong, X., Liu, X., Jedari, B., Li, M., Wan, L., & Xia, F. (2019). Mobile Crowdsourcing in Smart Cities: Technologies, Applications, and Future Challenges. *IEEE Internet of Things Journal*, *6*(5), 8095–8113. https://doi.org/10.1109/JIOT.2019.2921879

Kostakos, V., Rogstadius, J., Ferreira, D., Hosio, S., & Goncalves, J. (2017). Human sensors. *Understanding Complex Systems*, *9783319256566*, 69–92. https://doi.org/10.1007/978-3-319-25658-0_4

Kotovirta, V., Toivanen, T., Tergujeff, R., & Huttunen, M. (2012). Participatory Sensing in Environmental Monitoring - Experiences. *Proceedings - 6th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, IMIS 2012*, 155–162. https://doi.org/10.1109/IMIS.2012.70

Kumaraguru, P., & Cranor, L. F. (2005). *Privacy Indexes: A Survey of Westin’s Studies*. https://doi.org/10.1001/archpedi.1986.02140220052031

Lane, N. D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., Campbell, A. T., & College, D. (2010). A Survey of Mobile Phone Sensing. *IEEE Communications Magazine*, *September*, 140–150.

Laufer, R., & Wolfe, M. (1977). Privacy as a Concept and a Social Issue: A Multidimensional Developmental Theory. *Journal of Social Issues*, *33*, 22–42.

Li, H., Sarathy, R., & Xu, H. (2010). Understanding situational online information disclosure as a privacy calculus. *Journal of Computer Information Systems*, *51*(1), 62–71. https://doi.org/10.1080/08874417.2010.11645450

Lopez-Carreiro, I., Monzon, A., & Lopez-Lambas, M. E. (2021). Comparison of the willingness to adopt MaaS in Madrid (Spain) and Randstad (The Netherlands) metropolitan areas. *Transportation Research Part A: Policy and Practice*, *152*(August), 275–294. https://doi.org/10.1016/j.tra.2021.08.015

Lorenzo, P., Padilla, J., & Requejo, A. (2020). *Consumer Preferences For Personal Data Protection in Social Networks: A Choice Modelling Exercise*. https://doi.org/10.2139/ssrn.3716206

Ma, H., Zhao, D., & Yuan, P. (2015). Opportunities in mobile crowd sensing. *Infocommunications Journal*, *7*(2), 32–38.

Magidson, J., & Vermunt, J. K. (2004). Latent Class Models. In *The Sage Handbook of Quantitative Methodology for the Social Sciences* (pp. 175–198).

Majumdar, A., & Bose, I. (2016). Privacy calculus theory and its applicability for emerging technologies. *Lecture Notes in Business Information Processing*, *258*, 191–195. https://doi.org/10.1007/978-3-319-45408-5_20

Mariel, P., Hoyos, D., Meyerhoff, J., Czajkowski, M., Dekker, T., Glenk, K., Bredahl Jacobsen, J., Liebe, U., Bøye Olsen, S., Sagebiel, J., & Thiene, M. (2021). *Environmental Valuation with Discrete Choice Experiments Guidance on Design, Implementation and Data Analysis*.

Masoud, M., Jaradat, Y., Manasrah, A., & Jannoud, I. (2019). Sensors of Smart Devices in the Internet of Everything (IoE) Era: Big Opportunities and Massive Doubts. *Journal of Sensors*. https://doi.org/10.1155/2019/6514520

McFadden, D. (2001). Economic choices. *American Economic Review*, *91*(3), 351–378. https://doi.org/10.1257/aer.91.3.351

Mednis, A. (2013). Implementation of participatory sensing approach in mobile vehicle based sensor networks. *Baltic J. Modern Computing*, *1*(1–2), 1–8.

Mendez, D., Pérez, A. J., Labrador, M. A., & Marron, J. J. (2011). P-Sense: A participatory sensing system for air pollution monitoring and control. *2011 IEEE International Conference on Pervasive Computing and Communications Workshops, PERCOM Workshops 2011*, 344–347. https://doi.org/10.1109/PERCOMW.2011.5766902

Mloza-Banda, C., & Scholtz, B. (2018). Crowdsensing for Successful Water Resource Monitoring An Analysis of Citizens ’ Intentions and Motivations. *SAICSIT ’18*, 55–64. 10.1145/3278681.3278688

MobiDataLab. (2021). *Security and Privacy*. https://mobidatalab.eu/knowledge-base/security-and-privacy/

Mohan, P., Padmanabhan, V. N., & Ramjee, R. (2008). Nericell: Rich Monitoring of Road and Traffic Conditions using Mobile Smartphones. *SenSys 2008*.

Molin, E., Blangé, J., Cats, O., & Chorus, C. (2017). Willingness to pay for safety improvements in passenger air travel. *Journal of Air Transport Management*, *62*, 165–175. https://doi.org/10.1016/j.jairtraman.2017.04.002

Mousa, H., Mokhtar, S. Ben, Hasan, O., Younes, O., Hadhoud, M., & Brunie, L. (2015). Trust management and reputation systems in mobile participatory sensing applications: A survey. *Computer Networks*, *90*, 49–73. https://doi.org/10.1016/j.comnet.2015.07.011
Muller, C. L., Chapman, L., Johnston, S., Kidd, C., Illingworth, S., Foody, G., Overeem, A., & Leigh, R. R. (2015). Crowdsourcing for climate and atmospheric sciences: Current status and future potential. *International Journal of Climatology*, *35*(11), 3185–3203. https://doi.org/10.1002/joc.4210
Ogie, R. I. (2016). Adopting incentive mechanisms for large-scale participation in mobile crowdsensing: from literature review to a conceptual framework. In *Human-centric Computing and Information Sciences* (Vol. 6, Issue 1). Springer Berlin Heidelberg. https://doi.org/10.1186/s13673-016-0080-3
Orme, B. K. (2010). Sample Size Issues for Conjoint Analysis. *Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research*, 57–66. https://www.sawtoothsoftware.com/download/techpap/samplesz.pdf
Paalosmaa, T., & Shafie-Khah, M. (2021). Feasibility of Innovative Smart Mobility Solutions: A Case Study for Vaasa. In *World Electric Vehicle Journal* (Vol. 12, Issue 4, p. 188). https://doi.org/10.3390/wevj12040188
Patel, K. K., & Patel, S. M. (2016). Internet of Things-IOT: Definition, Characteristics, Architecture, Enabling Technologies, Application & Future Challenges. *International Journal of Engineering Science and Computing*, *6*(5), 6122–6131. https://doi.org/10.4010/2016.1482
Pearce, A., Harrison, M., Watson, V., Street, D. J., Howard, K., Bansback, N., & Bryan, S. (2021). Respondent Understanding in Discrete Choice Experiments: A Scoping Review. In *The Patient - Patient-Centered Outcomes Research* (Vol. 14, Issue 1). Springer International Publishing. https://doi.org/10.1007/s40271-020-00467-y
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
Princi, E., & Krämer, N. C. (2019). Acceptance of smart electronic monitoring atwork as a result of a privacy calculus decision. *Informatics*, *6*(3). https://doi.org/10.3390/informatics6030040
Puliafito, A., Tricomi, G., Zafeiropoulos, A., & Papavassiliou, S. (2021). Smart Cities of the Future as Cyber Physical Systems : Challenges and Enabling Technologies. *Sensors (Switzerland)*, *21*(3349), 1–25. https://doi.org/10.3390/s21103349
Restuccia, F., Das, S. K., & Payton, J. (2016). *Incentive Mechanisms for Participatory Sensing: Survey and Research Challenges*. *12*(2), 1–40. https://doi.org/10.1145/2888398
Riahi, M., Rahman, R., & Aberer, K. (2017). Privacy, trust and incentives in participatory sensing. *Understanding Complex Systems*, *9783319256566*, 93–114. https://doi.org/10.1007/978-3-319-25658-0_5
Ribeiro, P., Dias, G., & Pereira, P. (2021). Transport systems and mobility for smart cities. *Applied System Innovation*, *4*(3). https://doi.org/10.3390/asi4030061
Rijksoverheid. (2019). *Mobility as a Service (MaaS): multimodaal reisadvies op maat*. Rijksoverheid. https://www.rijksoverheid.nl/onderwerpen/mobiliteit-nu-en-in-de-toekomst/mobility-as-a-service-maas
Rosa, L., Silva, F., & Analide, C. (2020). TMSA: Participatory Sensing Based on Mobile Phones in Urban Spaces. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12489 LNCS*, 257–267. https://doi.org/10.1007/978-3-030-62362-3_23
Rose, J. M., & Bliemer, M. C. J. (2013). Sample size requirements for stated choice experiments. *Transportation*, *40*(5), 1021–1041. https://doi.org/10.1007/s11116-013-9451-z
Salim, F., & Haque, U. (2015). Urban computing in the wild: A survey on large scale participation and citizen engagement with ubiquitous computing, cyber physical systems, and Internet of Things. *International Journal of Human Computer Studies*, *81*, 31–48. https://doi.org/10.1016/j.ijhcs.2015.03.003
Schomakers, E. M., Biermann, H., & Ziefle, M. (2021). Users' Preferences for Smart Home Automation – Investigating Aspects of Privacy and Trust. *Telematics and Informatics*, *64*(March), 101689. https://doi.org/10.1016/j.tele.2021.101689
Schomakers, E. M., Lidynia, C., & Ziefle, M. (2020). All of me? Users' preferences for privacy-preserving data markets and the importance of anonymity. *Electronic Markets*, *30*(3), 649–665. https://doi.org/10.1007/s12525-020-00404-9
Shilton, K., & Estrin, D. (2012). Ethical Issues in Participatory Sensing. *CORE Issues in Professional and Research Ethics*, *1*(5), 2160–8784. http://nationalethicscenter.org/content/article/177

Shit, R. C. (2020). Crowd intelligence for sustainable futuristic intelligent transportation system: A review. *IET Intelligent Transport Systems*, *14*(6), 480–494. https://doi.org/10.1049/iet-its.2019.0321

Smith, H. J., Dinev, T., & Xu, H. (2011). Information privacy research: An interdisciplinary review. *MIS Quarterly: Management Information Systems*, *35*(4), 989–1015. https://doi.org/10.2307/41409970

Szabo, R., Farkas, K., Ispany, M., Benczur, A. A., Batfai, N., Jeszenszky, P., Laki, S., Vagner, A., Kollar, L., Sidlo, C., Besenczi, R., Smajda, M., Kover, G., Szincsak, T., Kadek, T., Kosa, M., Adamko, A., Lendak, I., Wiandt, B., … Feher, G. (2013). Framework for smart city applications based on participatory sensing. *4th IEEE International Conference on Cognitive Infocommunications, CogInfoCom 2013 - Proceedings*, 295–300. https://doi.org/10.1109/CogInfoCom.2013.6719260

Tang, D., & Chen, L. (2011). A review of the evolution of research on information Technology Acceptance Model. *BMEI 2011 - Proceedings 2011 International Conference on Business Management and Electronic Information*, *2*, 588–591. https://doi.org/10.1109/ICBMEI.2011.5917980

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

Train, K. (2009). Discrete Choice Models with Simulation. In *Cambridge University Press* (pp. 11–33). https://doi.org/10.1017/CBO9780511805271

Truong, N. B., Lee, G. M., Um, T. W., & MacKay, M. (2019). Trust Evaluation Mechanism for User Recruitment in Mobile Crowd-Sensing in the Internet of Things. *IEEE Transactions on Information Forensics and Security*, *14*(10), 2705–2719. https://doi.org/10.1109/TIFS.2019.2903659

Tseng, C. M., & Chau, C. K. (2017). Personalized Prediction of Vehicle Energy Consumption Based on Participatory Sensing. *IEEE Transactions on Intelligent Transportation Systems*, *18*(11), 3103–3113. https://doi.org/10.1109/TITS.2017.2672880

Turland, M., & Slade, P. (2020). Farmers' willingness to participate in a big data platform. *Agribusiness*, *36*(1), 20–36. https://doi.org/10.1002/agr.21627

Wang, T., Duong, T. D., & Chen, C. C. (2016). Intention to disclose personal information via mobile applications: A privacy calculus perspective. *International Journal of Information Management*, *36*(4), 531–542. https://doi.org/10.1016/j.ijinfomgt.2016.03.003

Weber, M. (2017). *Internet of Things Context of the Smart City*. 187–193.

Weller, B. E., Bowen, N. K., & Faubert, S. J. (2020). Latent Class Analysis: A Guide to Best Practice. *Journal of Black Psychology*, *46*(4), 287–311. https://doi.org/10.1177/0095798420930932

Westin, A. F. (1994). *Equifax-Harris Consumer Privacy Survey*.

Westin, A. F. (2001). *Privacy On & Off the Internet: What Consumers Want*.

Wibisono, W., & Ahmad, T. (2017). *A Mobile Crowdsensing Framework for Integrating Smartphone and IoT Devices to Cloud Computing Services*.

Wilde, L. de, Macharis, C., & Keseru, I. (2020). Technical requirements for organising campaigns in citizen observatories. *Transportation Research Procedia*, *48*(2019), 1418–1429. https://doi.org/10.1016/j.trpro.2020.08.172

Woodruff, A., Pihur, V., Consolvo, S., Brandimarte, L., & Acquisti, A. (2014). Would a privacy fundamentalist sell their DNA for $1000... if nothing bad happened as a result? The Westin categories, behavioral intentions, and consequences. *SOUPS '14: Proceedings of the Tenth Symposium On Usable Privacy and Security*, 1–18. https://www.usenix.org/conference/soups2014/proceedings/presentation/woodruff

Xiao, Z., Lim, H., & Ponnambalam, L. (2018). *Participatory Sensing for Smart Cities : A Case Study on Transport Trip Quality Measurement*. *April 2017*. https://doi.org/10.1109/TII.2017.2678522

Yi, K., Du, R., Liu, L., Chen, Q., & Gao, K. (2017). Fast participant recruitment algorithm for large-scale Vehicle-based Mobile Crowd Sensing. *Pervasive and Mobile Computing*, *38*, 188–199. https://doi.org/10.1016/j.pmcj.2017.02.009

Zaman, S., Abrar, N., & Iqbal, A. (2015). Incentive model design for participatory sensing: Technologies and challenges. *Proceedings of 2015 International Conference on Networking Systems and Security, NSysS 2015*. https://doi.org/10.1109/NSysS.2015.7043526

Zanella, A., Bui, N., Castellani, A., Vangelista, L., & Zorzi, M. (2014). Internet of things for smart cities. *IEEE Internet of Things Journal*, *1*(1), 22–32. https://doi.org/10.1109/JIOT.2014.2306328

Appendix A. NGene syntax

This appendix shows the syntax that was used in NGene to construct the orthogonal design for the choice experiment.

```
design
;alts = yes, no
;rows = 36
;orth = seq
;block = 3
;model:
U(yes) = b1 * mon [0,1,2,3]+
         b2 * eff [0,1,2]+
         b3 * rid [0,1,2]+
         b4 * tod [0,1,2,3]+
         b5 * use [0,1,2,3]
$
```

Appendix B. Choice sets

The final experiment design that was constructed is presented in Table 39.

Table 39. Final experiment design

Choice set	yes.mon	yes.eff	yes.rid	yes.tod	yes.use	Block
1	3	0	1	2	2	1
2	2	1	0	3	2	1
3	1	2	0	2	3	1
4	0	2	1	1	3	1
5	3	0	0	0	1	1
6	1	1	1	2	1	1
7	0	2	0	3	2	1
8	3	2	1	1	0	1
9	1	0	1	1	1	1
10	2	0	2	0	1	1
11	0	1	0	0	0	1
12	1	2	2	1	2	1
13	1	1	2	0	2	2
14	2	2	2	2	1	2
15	2	0	0	1	0	2
16	2	0	1	2	2	2
17	1	0	2	0	2	2
18	3	1	0	2	3	2
19	2	2	2	3	1	2
20	0	0	1	3	3	2
21	1	0	2	3	1	2
22	2	1	2	1	3	2
23	1	2	2	2	0	2
24	3	1	2	3	0	2
25	0	0	0	0	0	3
26	0	0	2	3	2	3
27	0	1	1	2	0	3
28	0	1	1	1	1	3
29	2	1	2	0	3	3
30	0	2	0	0	3	3
31	3	2	1	0	3	3
32	1	1	0	3	1	3
33	3	0	0	3	3	3
34	3	1	1	1	2	3
35	3	2	0	1	0	3
36	2	2	1	2	0	3

Appendix C. Survey

In this appendix, the final survey is presented. After being presented with the explanation of the experiment, respondents receive one out of the three blocks of choice situations. Block 1 consists of the questions 1 to 12, block 2 of questions 13 to 24, and block 3 of questions 25 to 36.

Welcome!

You are being invited to participate in a research study into factors possibly influencing data sharing in Participatory Sensing Systems. This study is being done by Mirjam van den Boogert from Delft University of Technology.

The purpose of this research study is to gain understanding of user's preferences and behaviour in participatory sensing systems and will take you approximately 10-15 minutes to complete. The data collected in this survey will be used for statistical analysis of choice behaviour.

All participants will stay entirely anonymous and data will be treated confidentially.

Your participation in this study is entirely voluntary and you can withdraw at any time.

We believe there are no known risks associated with this research study; however, as with any online related activity the risk of a breach is always possible. To the best of our ability your answers in this study will remain confidential. We will minimize any risks by not asking any personal data and by storing the data safely in Qualtrics (temporarily) and OneDrive at TU Delft.

The results of this research will be published in a report that will be submitted to the TU Delft Repository. Furthermore, the data collected in this survey will be made public in the TU Delft Survey Database for future research. No data will be made public that could identify you in any way.

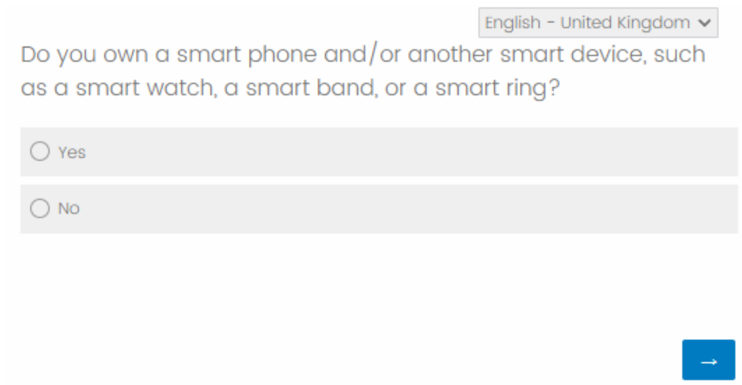
If you have any questions, please feel free to contact: R.J.vandenBoogert@student.tudelft.nl

Please state if you consent with participating in this research.

- ○ I consent for participating in this research and for data processing
- ○ I do not consent

››

Figure 28. Opening statement



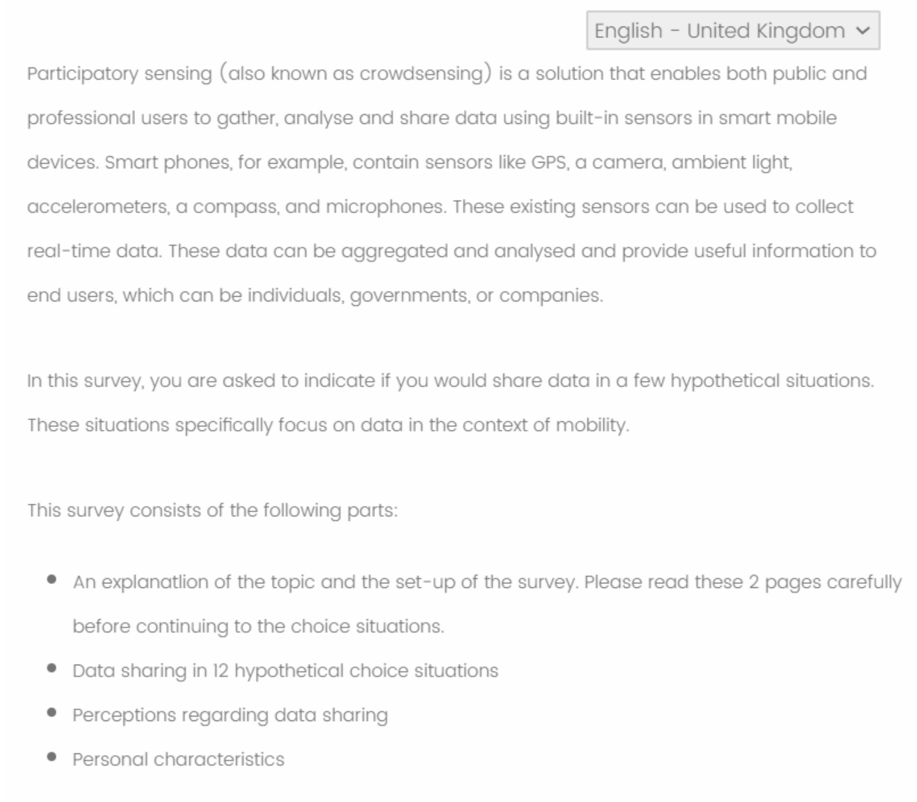
English - United Kingdom

Do you own a smart phone and/or another smart device, such as a smart watch, a smart band, or a smart ring?

- Yes
- No

→

Figure 29. Smartphone ownership



English - United Kingdom

Participatory sensing (also known as crowdsensing) is a solution that enables both public and professional users to gather, analyse and share data using built-in sensors in smart mobile devices. Smart phones, for example, contain sensors like GPS, a camera, ambient light, accelerometers, a compass, and microphones. These existing sensors can be used to collect real-time data. These data can be aggregated and analysed and provide useful information to end users, which can be individuals, governments, or companies.

In this survey, you are asked to indicate if you would share data in a few hypothetical situations. These situations specifically focus on data in the context of mobility.

This survey consists of the following parts:

- An explanatlion of the topic and the set-up of the survey. Please read these 2 pages carefully before continuing to the choice situations.
- Data sharing in 12 hypothetical choice situations
- Perceptions regarding data sharing
- Personal characteristics

Figure 30. General explanation of survey

English - United Kingdom

Sharing data in the context of mobility

This study assumes data will be collected via crowdsensing applications in the context of **mobility**. A huge number of people daily make use of traffic or public transport. It is predicted that congestion in and surrounding cities will increase. Using data collected by consumers can make transport systems more efficient and sustainable, more safe, and more flexible.

Collected data sensed by the crowd has major benefits in this context for society, transport operators, companies, and for individuals:

Predicting traffic flows

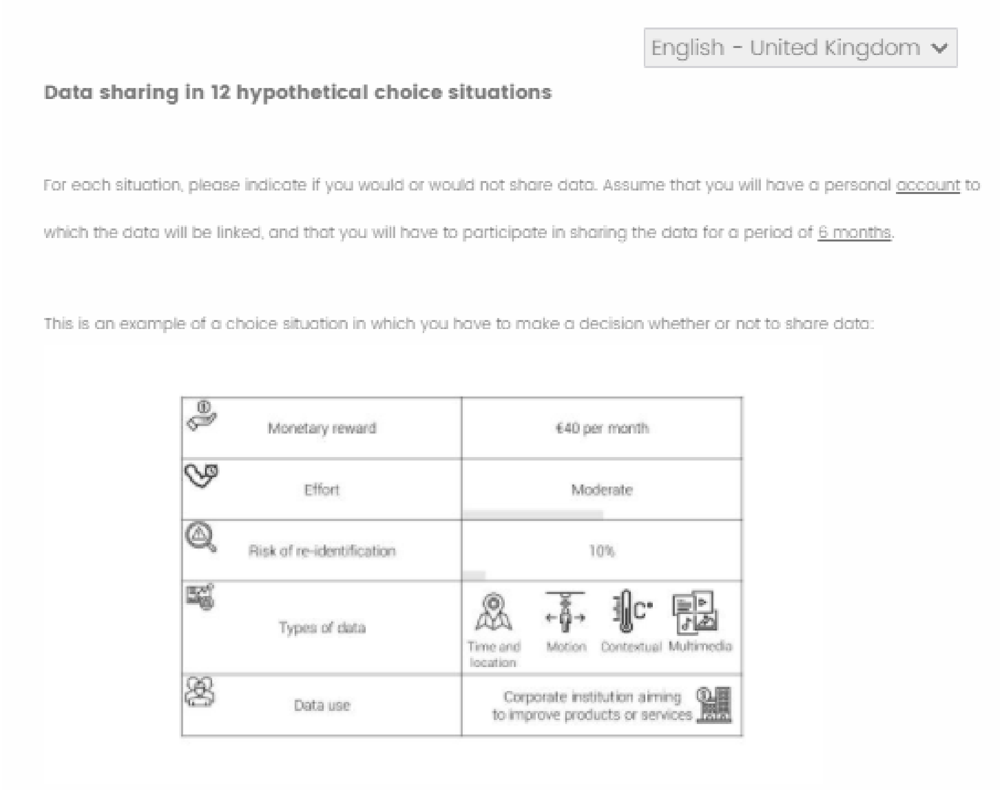
Using data collected by individuals, traffic flows can be predicted. A municipality can use these insights to manage crowds and congestion in a city, and to analyse the demand for different modes of transport. By gaining insights in demand for transportation, transport operators are able to adjust and optimize fleet management. For users of public transport, this will lead to fewer waiting times and less crowded trains, metros or buses. Furthermore, they will receive improved travel advices since the application will have insight in their personal habits and preferences regarding transport. For car drivers, better route planning can be facilitated based on predicted congestion, which will result in time savings. Also, these information can be used to find methods to reduce congestion.

Enhancing trip quality

Also, information on driving behaviour can be gathered, for example by using data on acceleration or braking. Furthermore, road conditions can be analysed by measuring vibrations, which can indicate bumpiness in the road. Government authorities and transport operators can use this information to improve infrastructure, to make roads more safe and comfortable. For car drivers, aggregated user data on driving behaviour and road and traffic conditions can be used to make predictions on e.g. the DTE (Distance-To-Empty), the distance they can travel before their (electric) vehicle runs out of fuel. Besides that, car manufacturers can use these data to get insight in car conditions, shelf lives and durability and improve their products based on these insights. Transport authorities can measure the trip quality of travellers by using data on waiting times, distance, the bumpiness of a ride, or other inputs by users. This can be used to improve their services and deliver better travel experiences to travellers.

← →

Figure 31. Explanation of benefits



English - United Kingdom

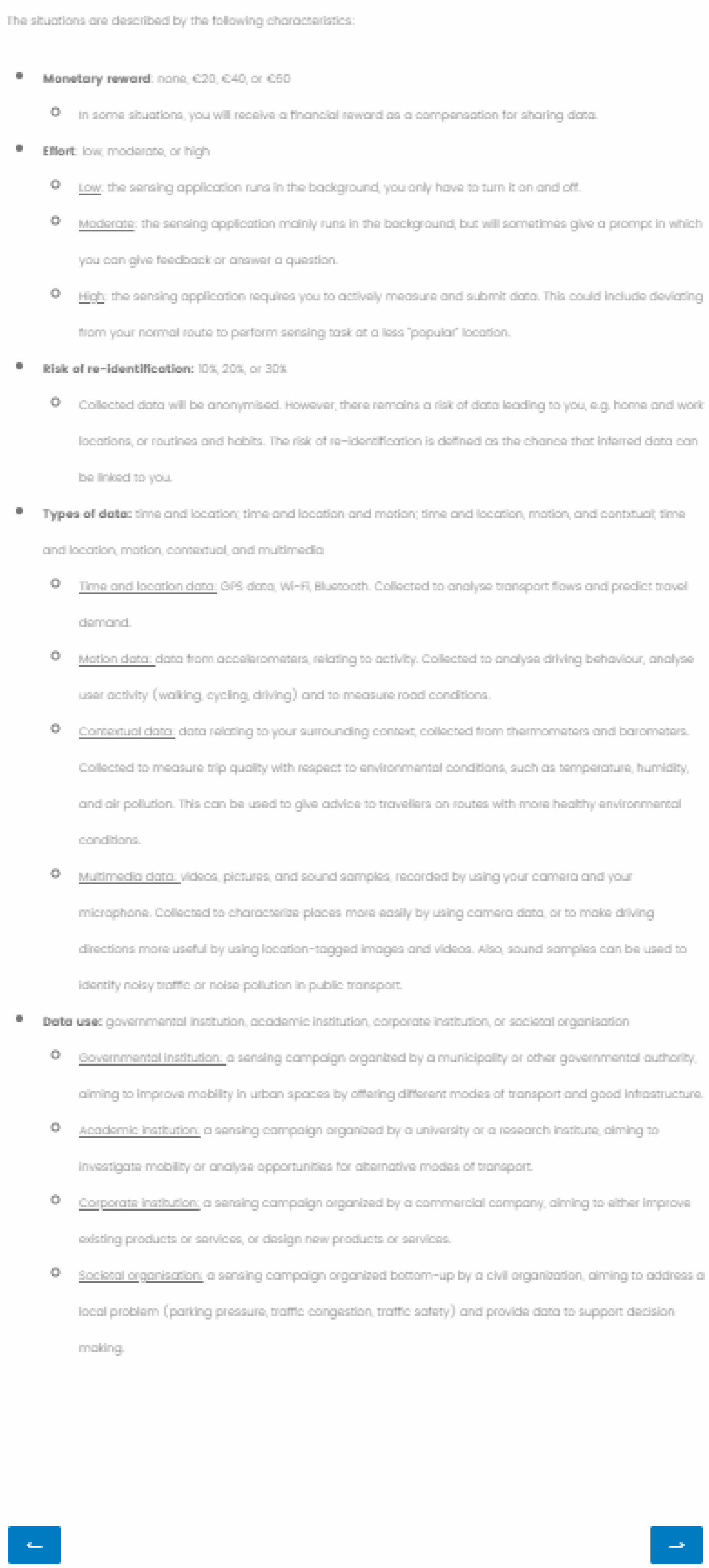
Data sharing in 12 hypothetical choice situations

For each situation, please indicate if you would or would not share data. Assume that you will have a personal account to which the data will be linked, and that you will have to participate in sharing the data for a period of 6 months.

This is an example of a choice situation in which you have to make a decision whether or not to share data:

Monetary reward	€40 per month
Effort	Moderate
Risk of re-identification	10%
Types of data	Time and location, Motion, Contextual, Multimedia
Data use	Corporate institution aiming to improve products or services

Figure 32. Explanation of experiment set-up (1)

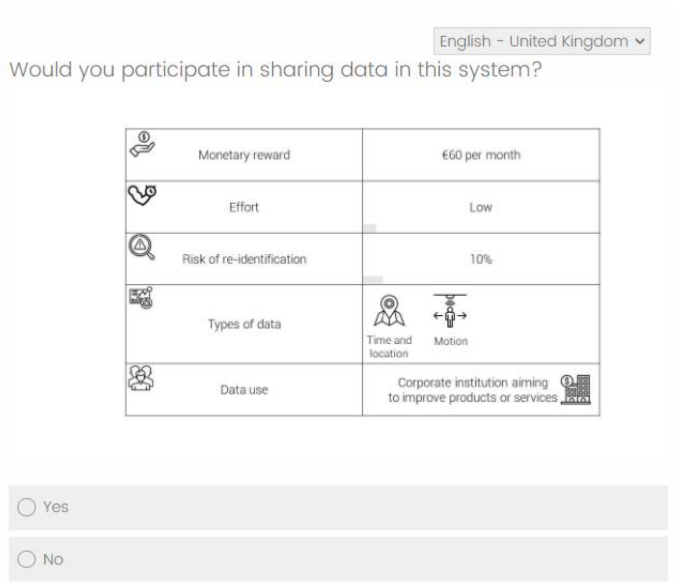


The situations are described by the following characteristics:

- **Monetary reward**: none, €20, €40, or €60
    - In some situations, you will receive a financial reward as a compensation for sharing data.
- **Effort**: low, moderate, or high
    - Low: the sensing application runs in the background, you only have to turn it on and off.
    - Moderate: the sensing application mainly runs in the background, but will sometimes give a prompt in which you can give feedback or answer a question.
    - High: the sensing application requires you to actively measure and submit data. This could include deviating from your normal route to perform sensing task at a less "popular" location.
- **Risk of re-identification:** 10%, 20%, or 30%
    - Collected data will be anonymised. However, there remains a risk of data leading to you, e.g. home and work locations, or routines and habits. The risk of re-identification is defined as the chance that inferred data can be linked to you.
- **Types of data:** time and location; time and location and motion; time and location, motion, and contxtual; time and location, motion, contextual, and multimedia
    - Time and location data: GPS data, Wi-Fi, Bluetooth. Collected to analyse transport flows and predict travel demand.
    - Motion data: data from accelerometers, relating to activity. Collected to analyse driving behaviour, analyse user activity (walking, cycling, driving) and to measure road conditions.
    - Contextual data: data relating to your surrounding context, collected from thermometers and barometers. Collected to measure trip quality with respect to environmental conditions, such as temperature, humidity, and air pollution. This can be used to give advice to travellers on routes with more healthy environmental conditions.
    - Multimedia data: videos, pictures, and sound samples, recorded by using your camera and your microphone. Collected to characterize places more easily by using camera data, or to make driving directions more useful by using location-tagged images and videos. Also, sound samples can be used to identify noisy traffic or noise pollution in public transport.
- **Data use:** governmental institution, academic institution, corporate institution, or societal organisation
    - Governmental institution: a sensing campaign organized by a municipality or other governmental authority, aiming to improve mobility in urban spaces by offering different modes of transport and good infrastructure.
    - Academic institution: a sensing campaign organized by a university or a research institute, aiming to investigate mobility or analyse opportunities for alternative modes of transport.
    - Corporate institution: a sensing campaign organized by a commercial company, aiming to either improve existing products or services, or design new products or services.
    - Societal organisation: a sensing campaign organized bottom-up by a civil organization, aiming to address a local problem (parking pressure, traffic congestion, traffic safety) and provide data to support decision making.

← →

Figure 33. Explanation of experiment set-up (2)


English - United Kingdom

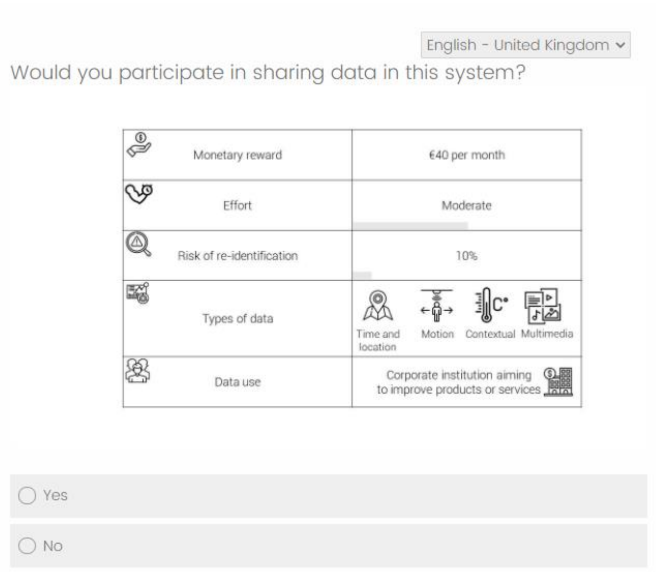
Would you participate in sharing data in this system?

Monetary reward	€60 per month
Effort	Low
Risk of re-identification	10%
Types of data	Time and location, Motion
Data use	Corporate institution aiming to improve products or services

Yes

No

Figure 34. Question 1


English - United Kingdom

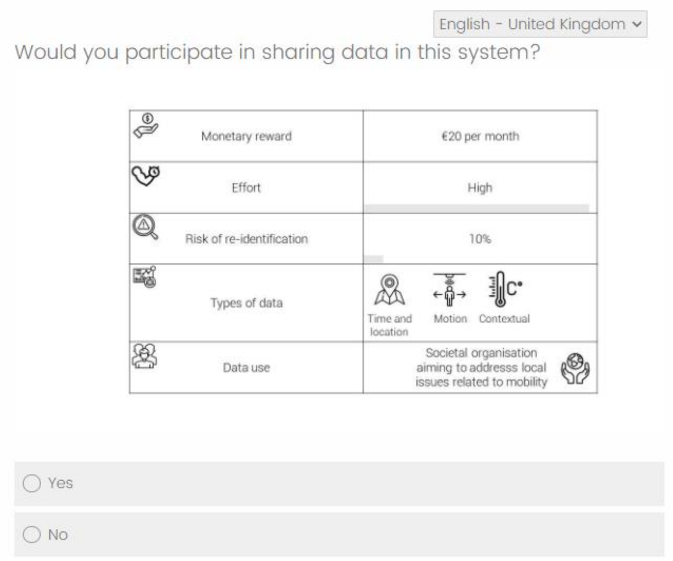
Would you participate in sharing data in this system?

Monetary reward	€40 per month
Effort	Moderate
Risk of re-identification	10%
Types of data	Time and location, Motion, Contextual, Multimedia
Data use	Corporate institution aiming to improve products or services

Yes

No

Figure 35. Question 2


English - United Kingdom

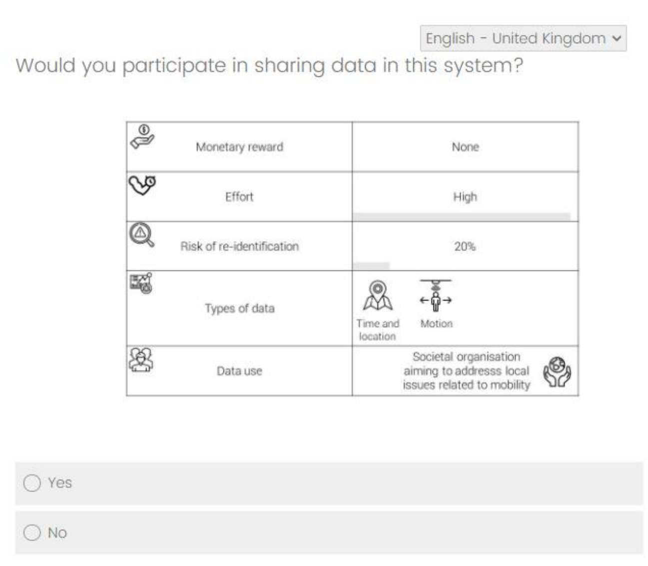
Would you participate in sharing data in this system?

Monetary reward	€20 per month
Effort	High
Risk of re-identification	10%
Types of data	Time and location, Motion, Contextual
Data use	Societal organisation aiming to addresss local issues related to mobility

Yes

No

Figure 36. Question 3



English - United Kingdom

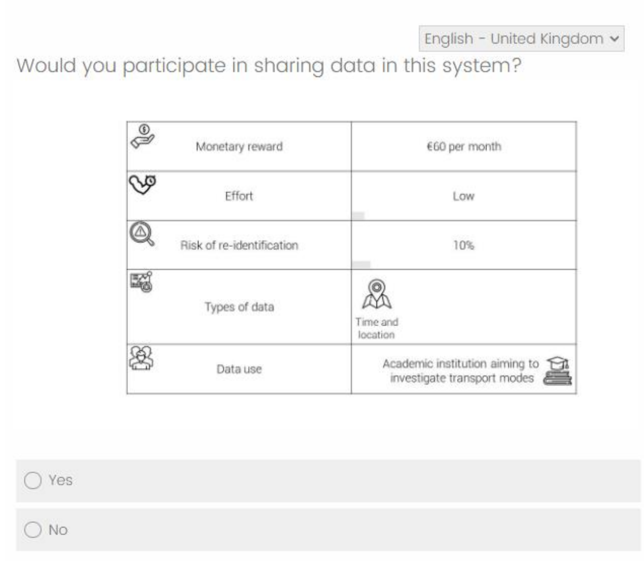
Would you participate in sharing data in this system?

Monetary reward	None
Effort	High
Risk of re-identification	20%
Types of data	Time and location; Motion
Data use	Societal organisation aiming to addresss local issues related to mobility

Yes

No

Figure 37. Question 4



English - United Kingdom

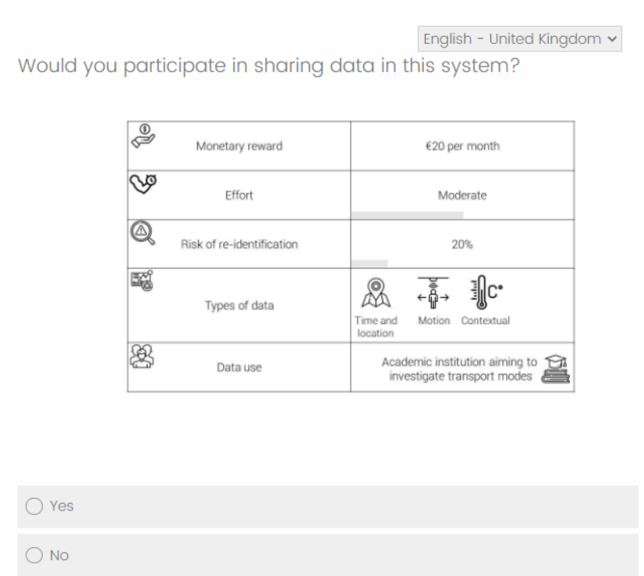
Would you participate in sharing data in this system?

Monetary reward	€60 per month
Effort	Low
Risk of re-identification	10%
Types of data	Time and location
Data use	Academic institution aiming to investigate transport modes

Yes

No

Figure 38. Question 5



English - United Kingdom

Would you participate in sharing data in this system?

Monetary reward	€20 per month
Effort	Moderate
Risk of re-identification	20%
Types of data	Time and location; Motion; Contextual
Data use	Academic institution aiming to investigate transport modes

Yes

No

Figure 39. Question 6

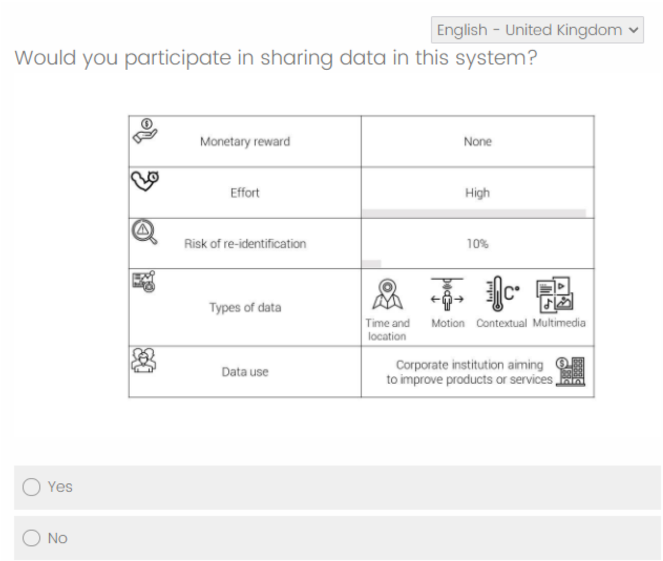


Figure 40. Question 7

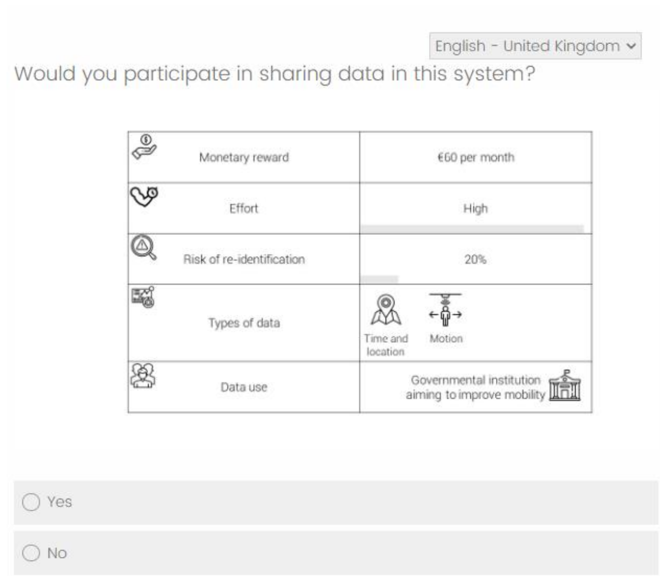


Figure 41. Question 8

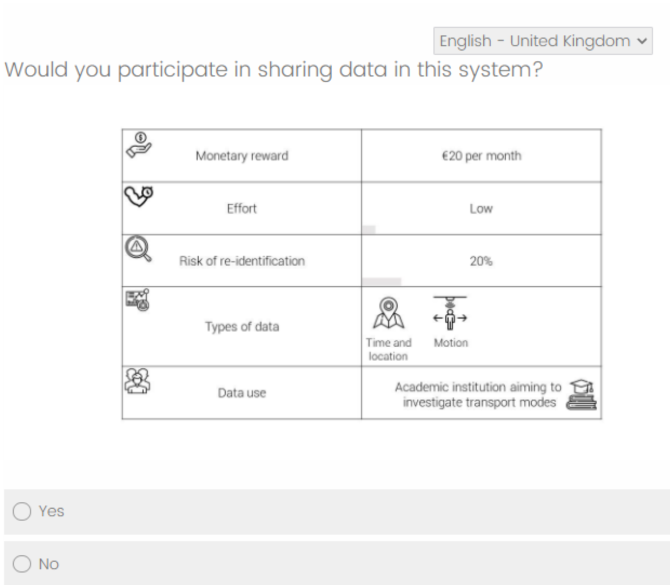
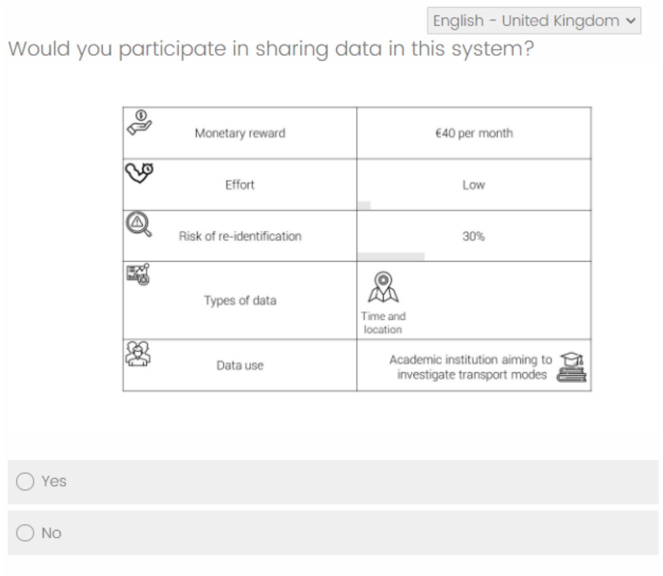


Figure 42. Question 9



English - United Kingdom

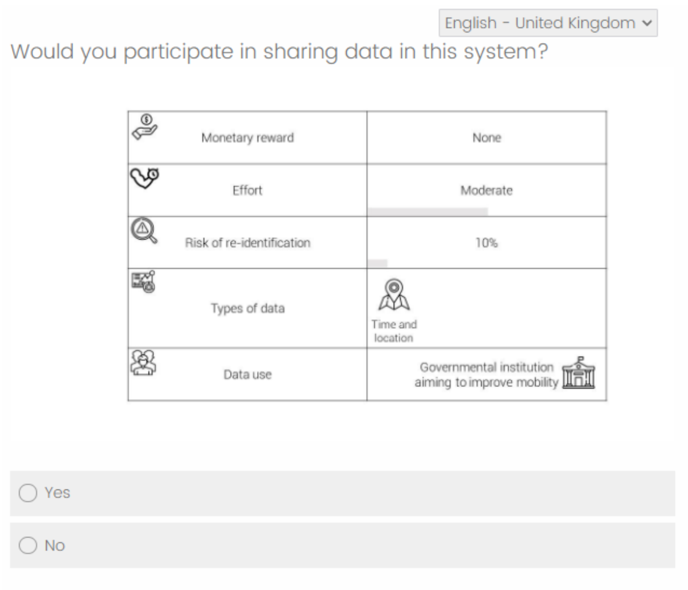
Would you participate in sharing data in this system?

Monetary reward	€40 per month
Effort	Low
Risk of re-identification	30%
Types of data	Time and location
Data use	Academic institution aiming to investigate transport modes

Yes

No

Figure 43. Question 10



English - United Kingdom

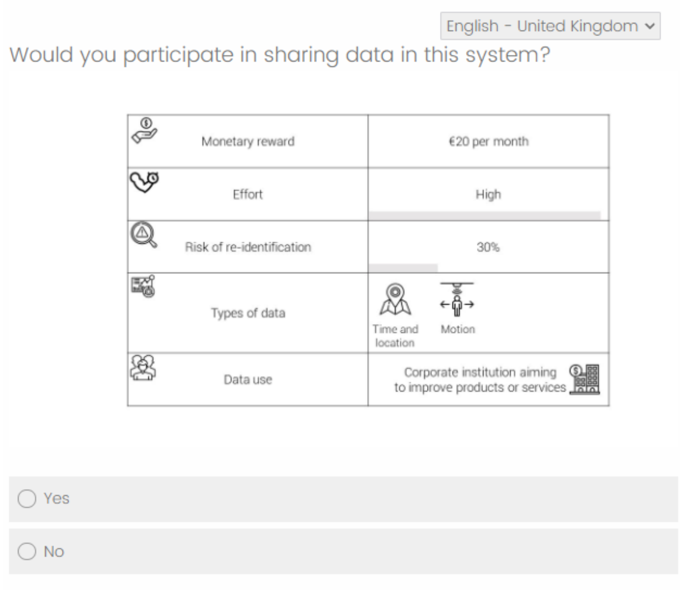
Would you participate in sharing data in this system?

Monetary reward	None
Effort	Moderate
Risk of re-identification	10%
Types of data	Time and location
Data use	Governmental institution aiming to improve mobility

Yes

No

Figure 44. Question 11



English - United Kingdom

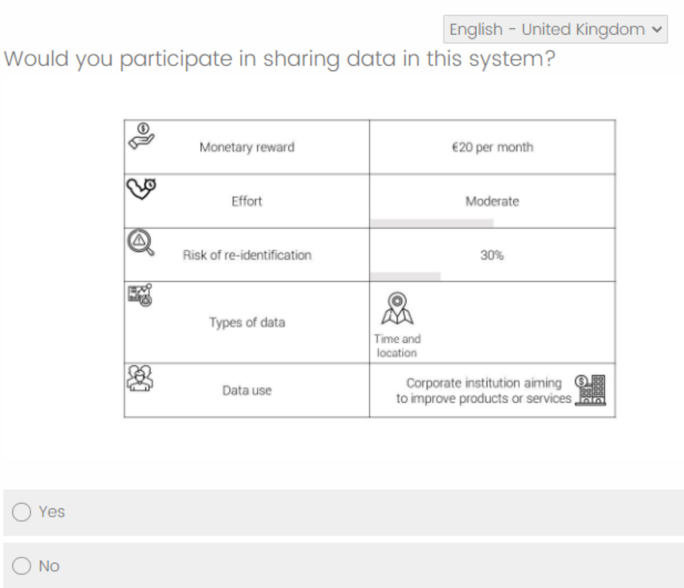
Would you participate in sharing data in this system?

Monetary reward	€20 per month
Effort	High
Risk of re-identification	30%
Types of data	Time and location; Motion
Data use	Corporate institution aiming to improve products or services

Yes

No

Figure 45. Question 12



English - United Kingdom

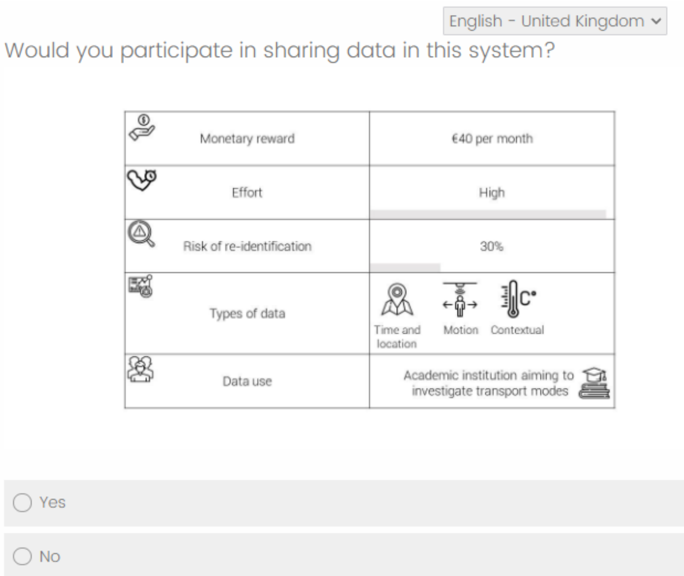
Would you participate in sharing data in this system?

Monetary reward	€20 per month
Effort	Moderate
Risk of re-identification	30%
Types of data	Time and location
Data use	Corporate institution aiming to improve products or services

Yes

No

Figure 46. Question 13



English - United Kingdom

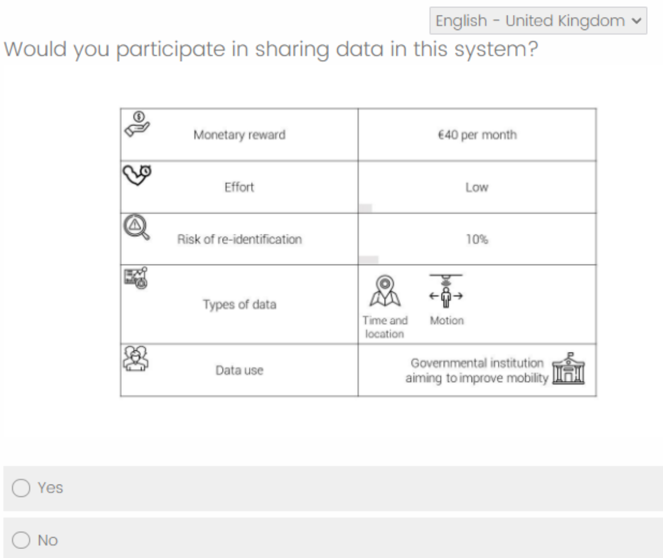
Would you participate in sharing data in this system?

Monetary reward	€40 per month
Effort	High
Risk of re-identification	30%
Types of data	Time and location, Motion, Contextual
Data use	Academic institution aiming to investigate transport modes

Yes

No

Figure 47. Question 14



English - United Kingdom

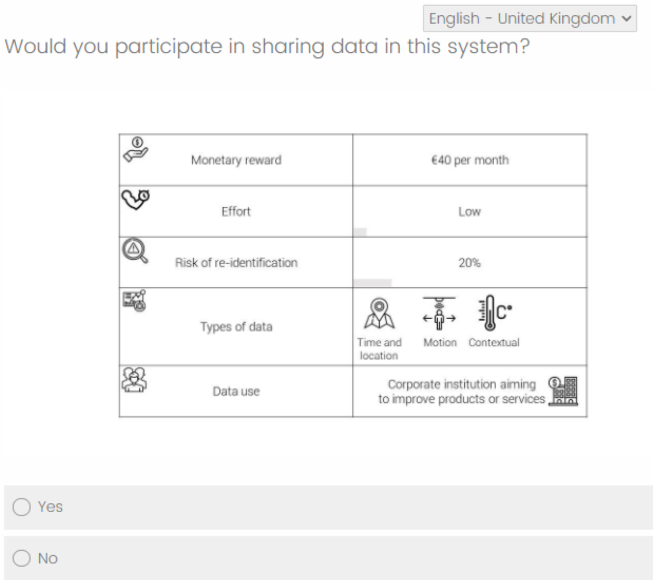
Would you participate in sharing data in this system?

Monetary reward	€40 per month
Effort	Low
Risk of re-identification	10%
Types of data	Time and location, Motion
Data use	Governmental institution aiming to improve mobility

Yes

No

Figure 48. Question 15



English - United Kingdom

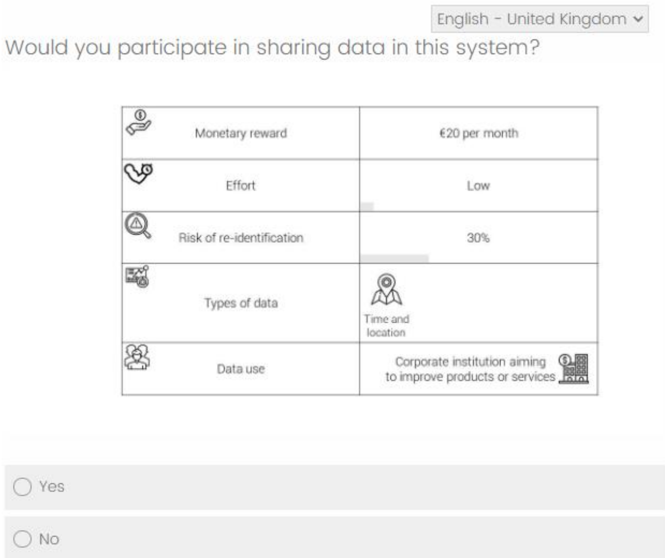
Would you participate in sharing data in this system?

Monetary reward	€40 per month
Effort	Low
Risk of re-identification	20%
Types of data	Time and location, Motion, Contextual
Data use	Corporate institution aiming to improve products or services

Yes

No

Figure 49. Question 16



English - United Kingdom

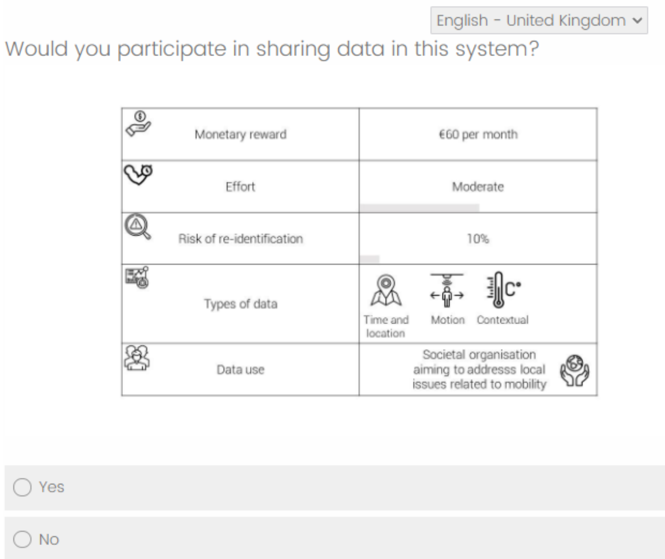
Would you participate in sharing data in this system?

Monetary reward	€20 per month
Effort	Low
Risk of re-identification	30%
Types of data	Time and location
Data use	Corporate institution aiming to improve products or services

Yes

No

Figure 50. Question 17



English - United Kingdom

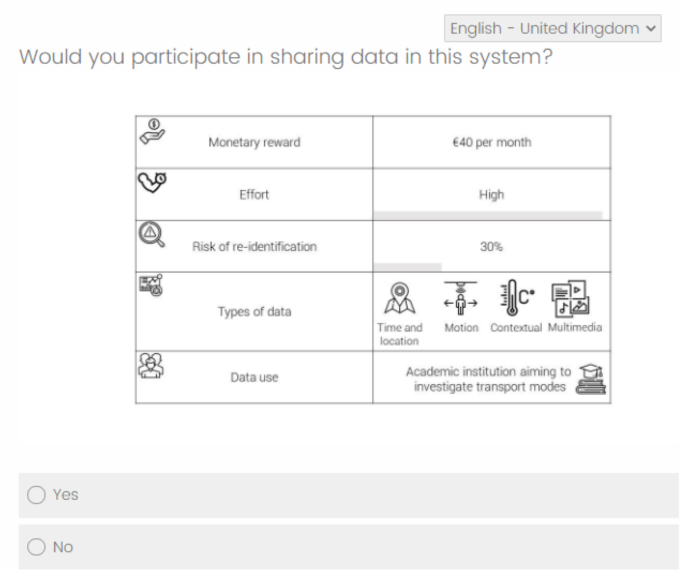
Would you participate in sharing data in this system?

Monetary reward	€60 per month
Effort	Moderate
Risk of re-identification	10%
Types of data	Time and location, Motion, Contextual
Data use	Societal organisation aiming to address local issues related to mobility

Yes

No

Figure 51. Question 18



English - United Kingdom

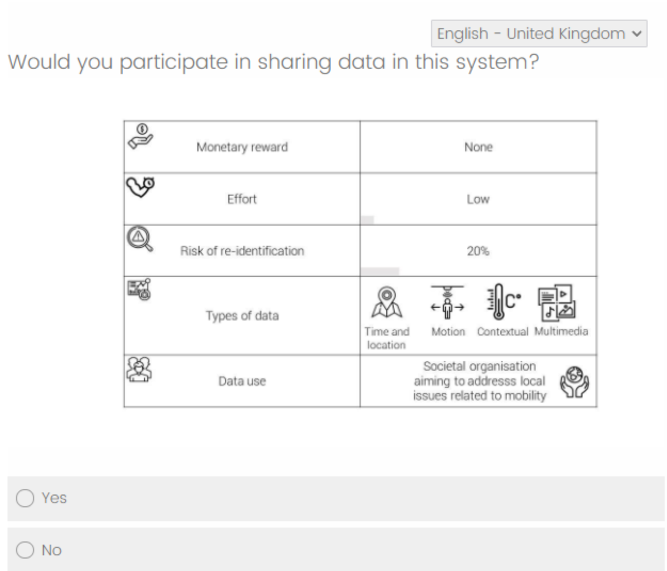
Would you participate in sharing data in this system?

Monetary reward	€40 per month
Effort	High
Risk of re-identification	30%
Types of data	Time and location, Motion, Contextual, Multimedia
Data use	Academic institution aiming to investigate transport modes

Yes

No

Figure 52. Question 19



English - United Kingdom

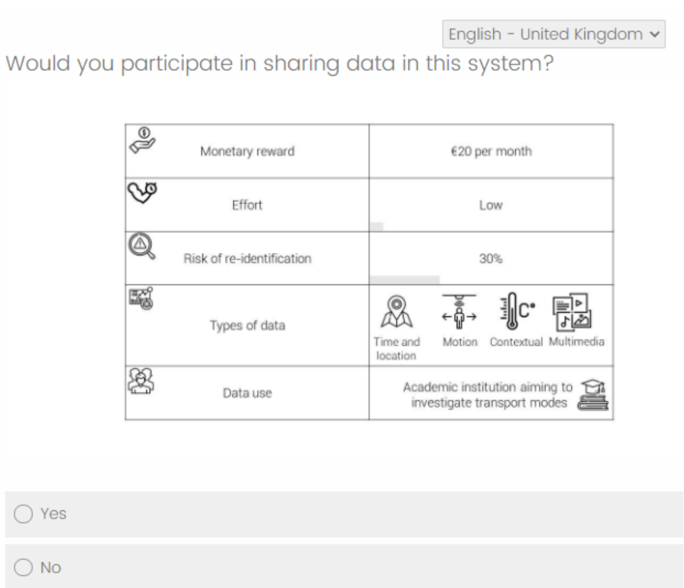
Would you participate in sharing data in this system?

Monetary reward	None
Effort	Low
Risk of re-identification	20%
Types of data	Time and location, Motion, Contextual, Multimedia
Data use	Societal organisation aiming to addresss local issues related to mobility

Yes

No

Figure 53. Question 20



English - United Kingdom

Would you participate in sharing data in this system?

Monetary reward	€20 per month
Effort	Low
Risk of re-identification	30%
Types of data	Time and location, Motion, Contextual, Multimedia
Data use	Academic institution aiming to investigate transport modes

Yes

No

Figure 54. Question 21

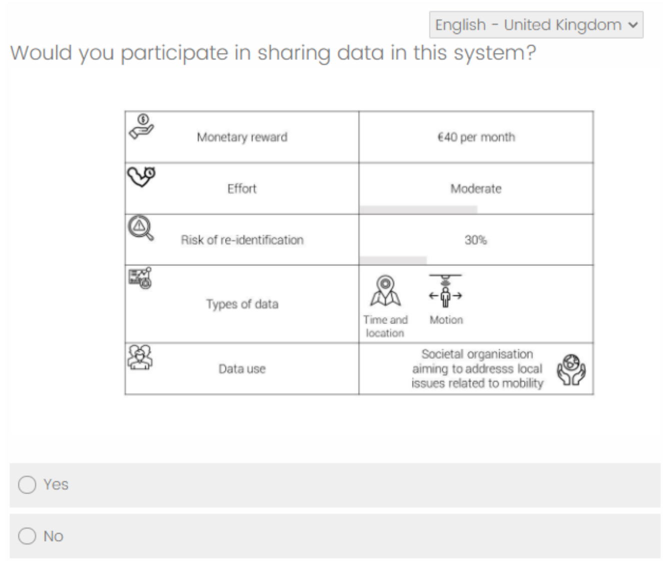


Figure 55. Question 22

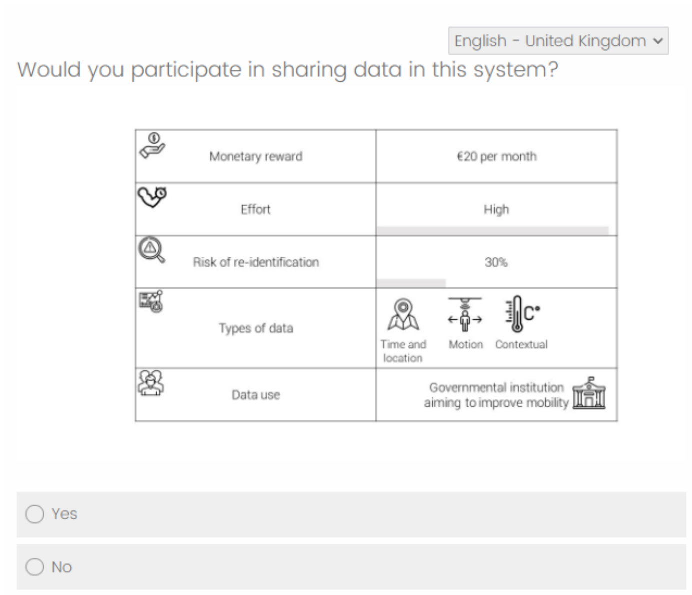


Figure 56. Question 23

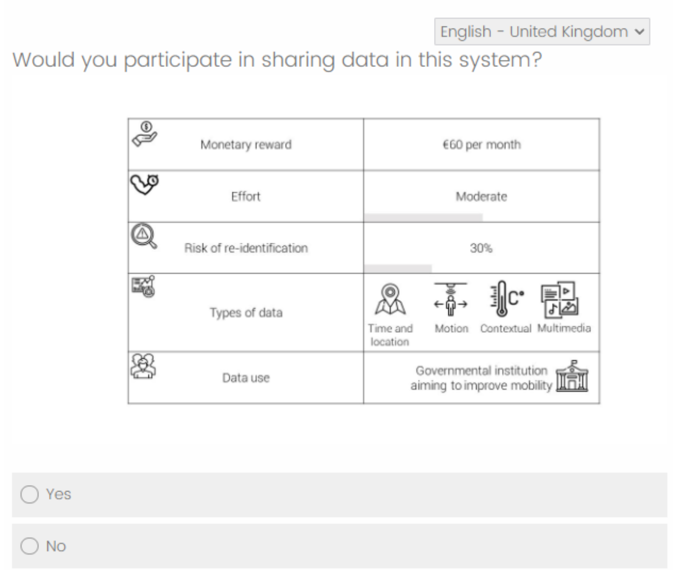


Figure 57. Question 24

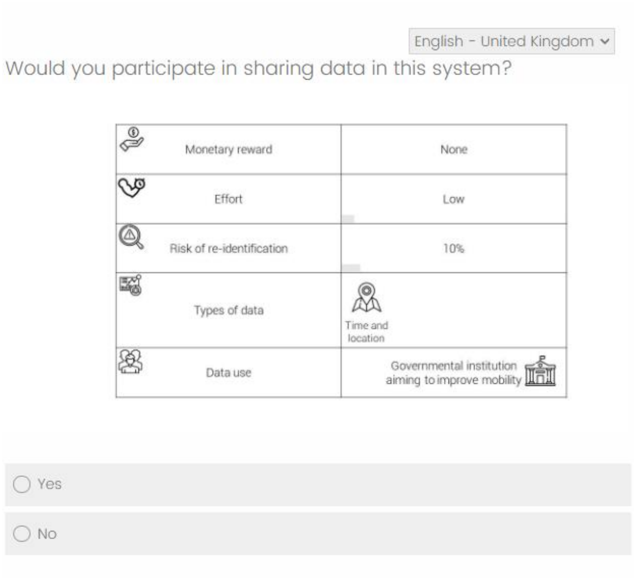


Figure 58. Question 25

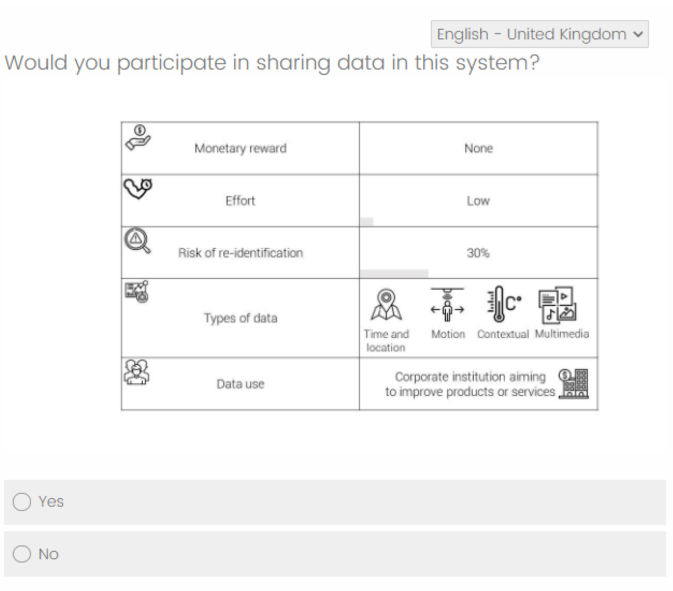


Figure 59. Question 26

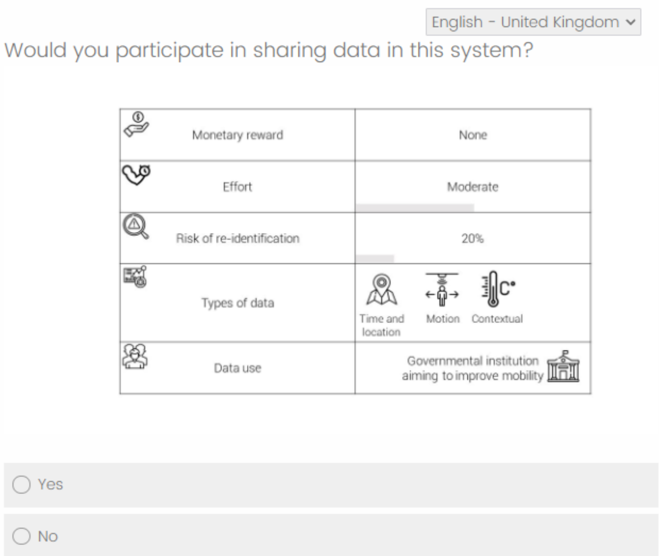


Figure 60. Question 27

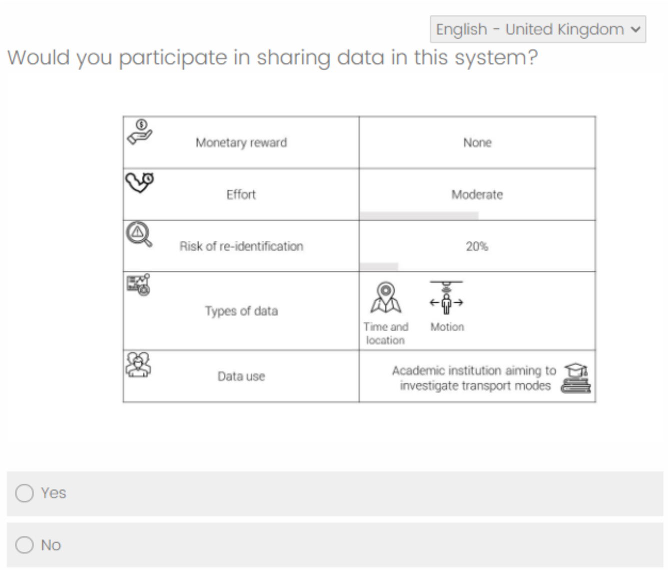


Figure 61. Question 28

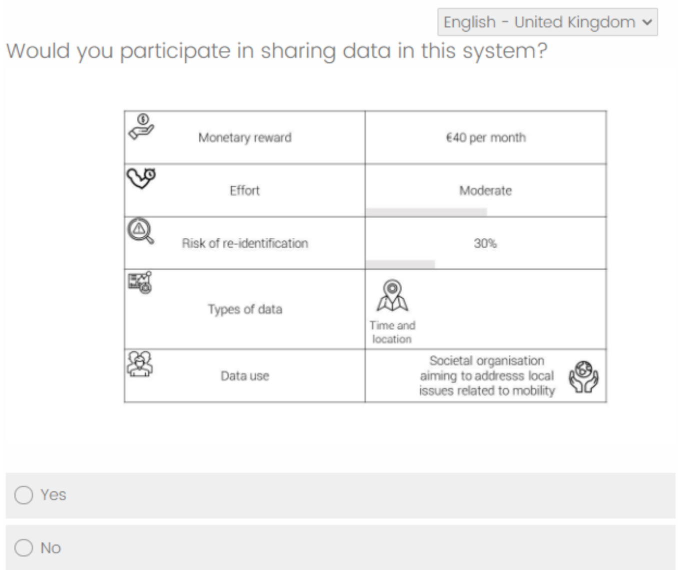


Figure 62. Question 29

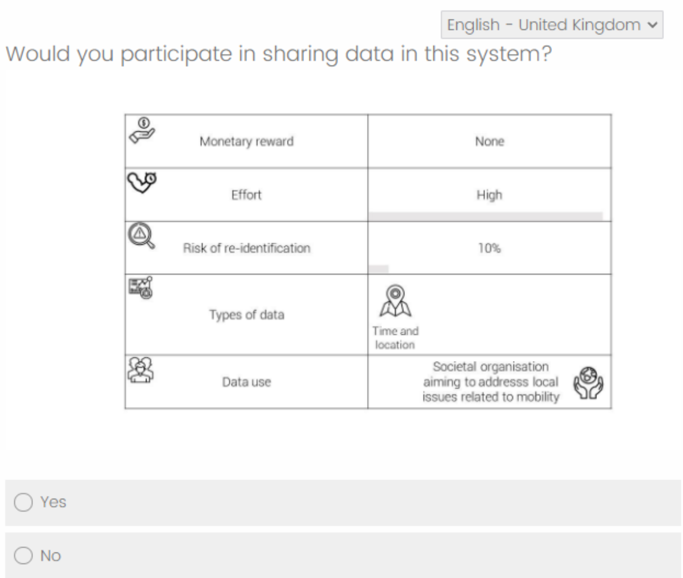


Figure 63. Question 30

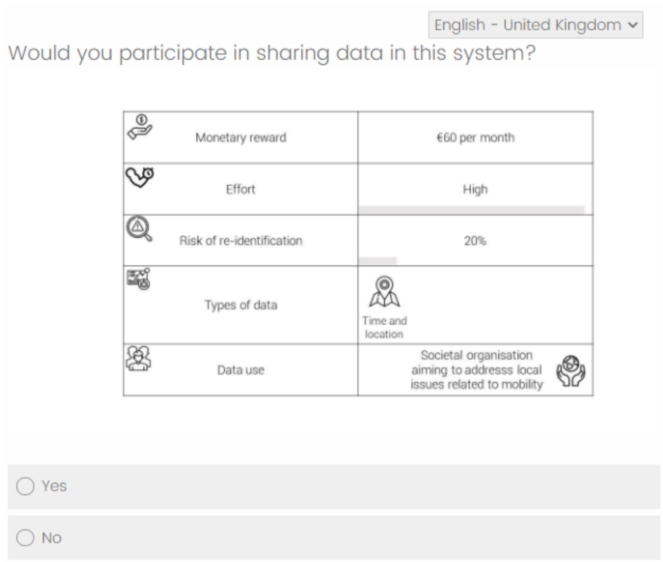


Figure 64. Question 31

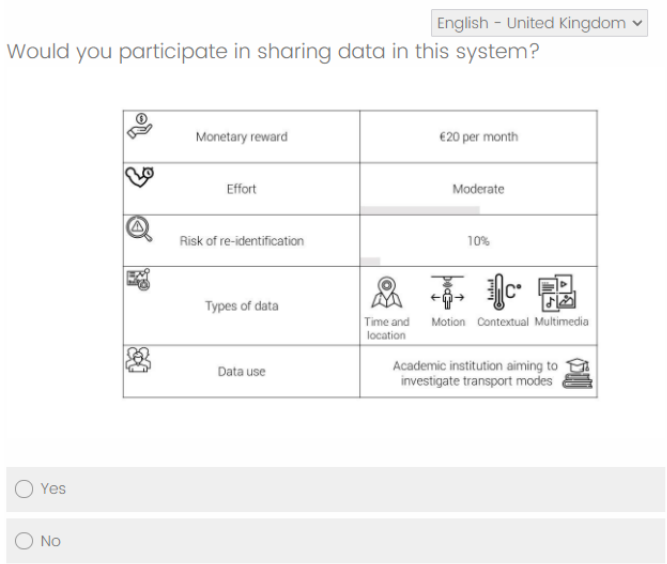


Figure 65. Question 32

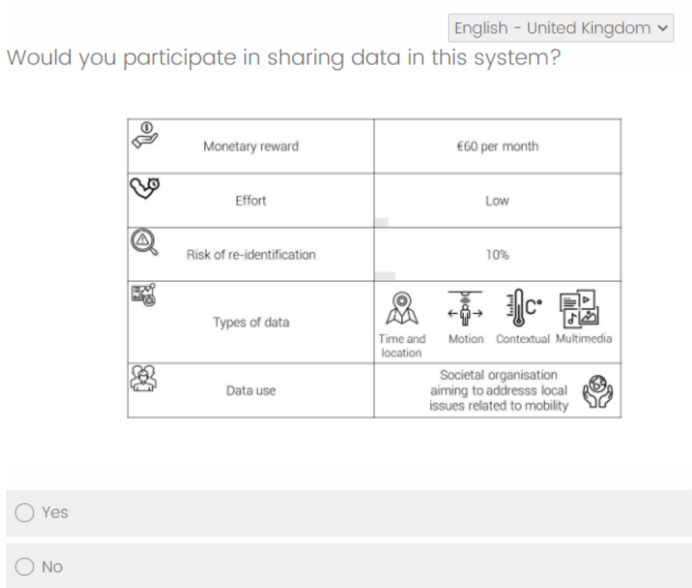


Figure 66. Question 33

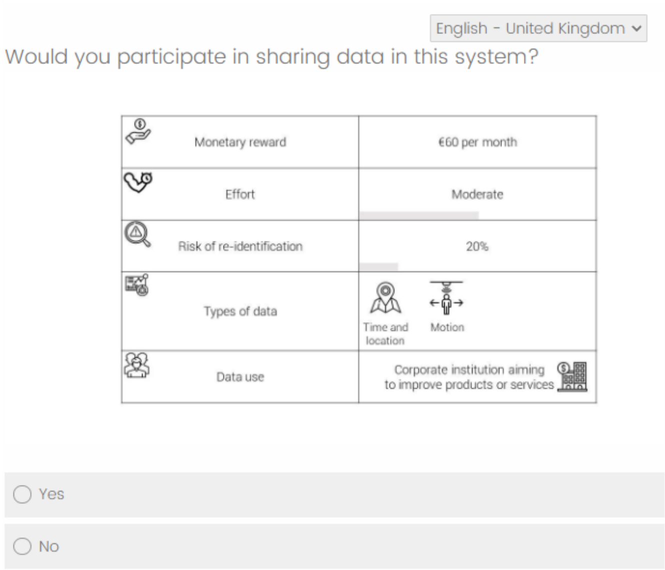


Figure 67. Question 34

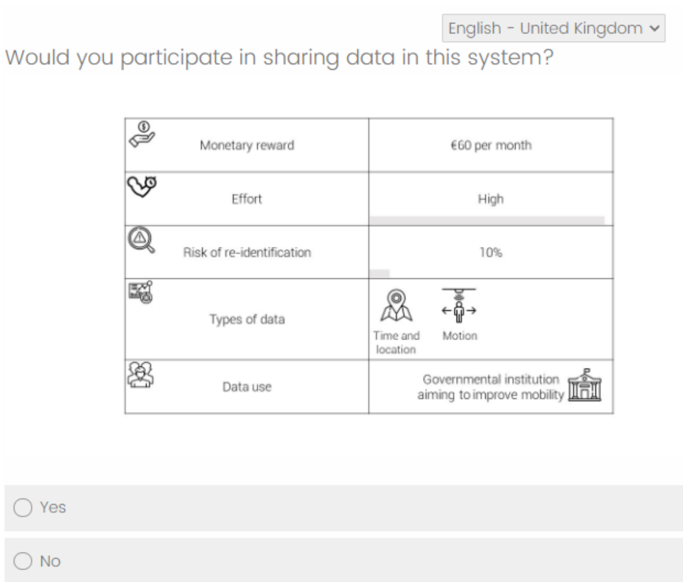


Figure 68. Question 35

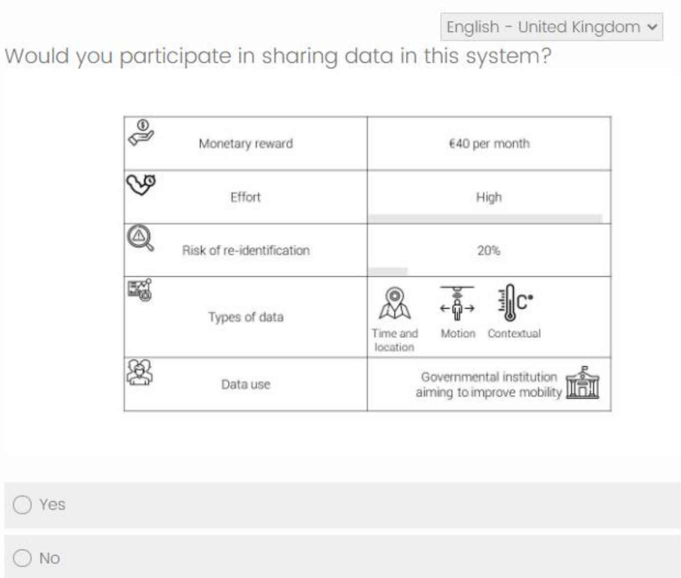
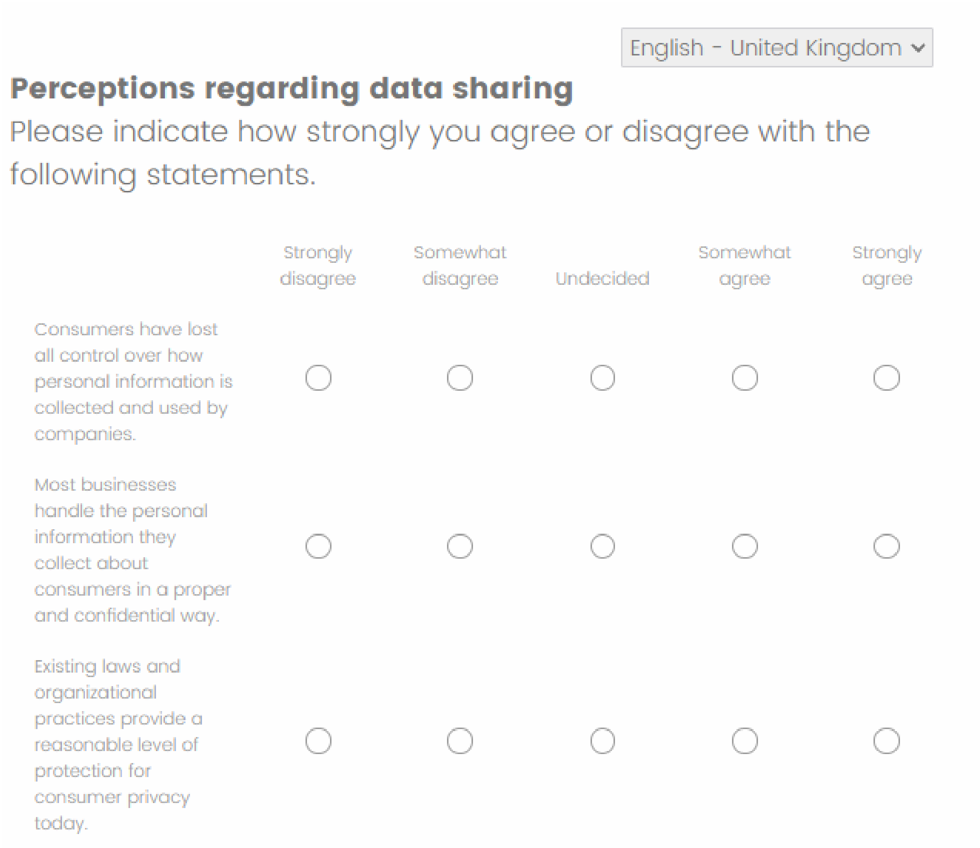


Figure 69. Question 36



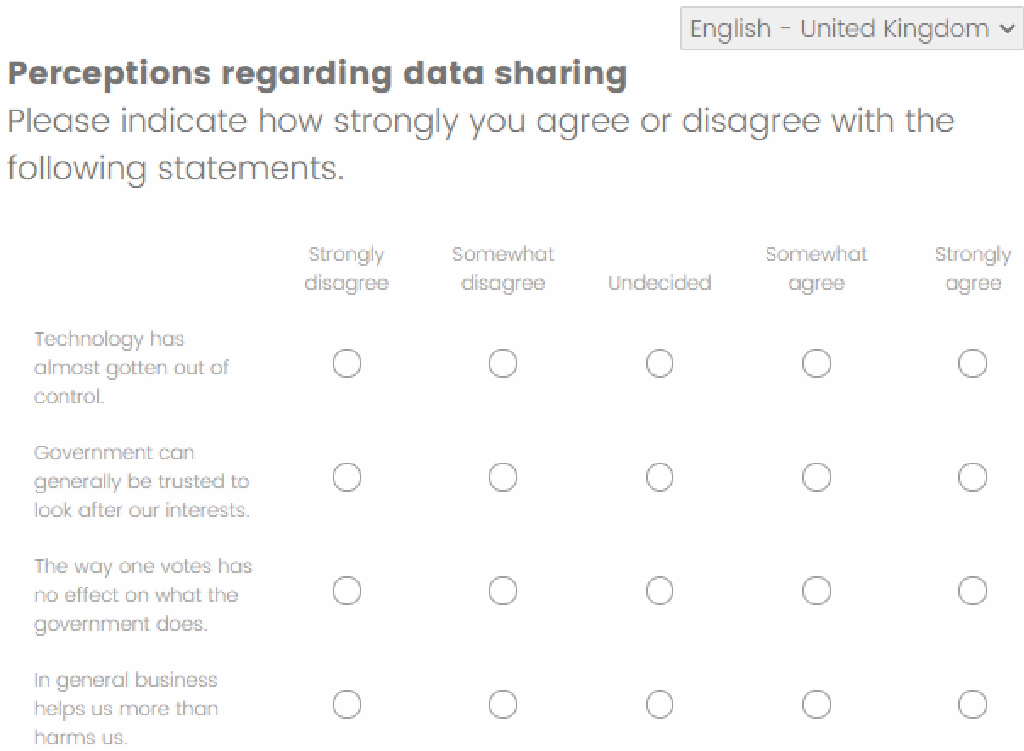
English - United Kingdom

Perceptions regarding data sharing

Please indicate how strongly you agree or disagree with the following statements.

	Strongly disagree	Somewhat disagree	Undecided	Somewhat agree	Strongly agree
Consumers have lost all control over how personal information is collected and used by companies.	○	○	○	○	○
Most businesses handle the personal information they collect about consumers in a proper and confidential way.	○	○	○	○	○
Existing laws and organizational practices provide a reasonable level of protection for consumer privacy today.	○	○	○	○	○

Figure 70. Perceptions regarding privacy



English - United Kingdom

Perceptions regarding data sharing

Please indicate how strongly you agree or disagree with the following statements.

	Strongly disagree	Somewhat disagree	Undecided	Somewhat agree	Strongly agree
Technology has almost gotten out of control.	○	○	○	○	○
Government can generally be trusted to look after our interests.	○	○	○	○	○
The way one votes has no effect on what the government does.	○	○	○	○	○
In general business helps us more than harms us.	○	○	○	○	○

Figure 71. Perceptions regarding trust

Figure 72. Personal characteristics 1

Figure 73. Personal characteristics 2

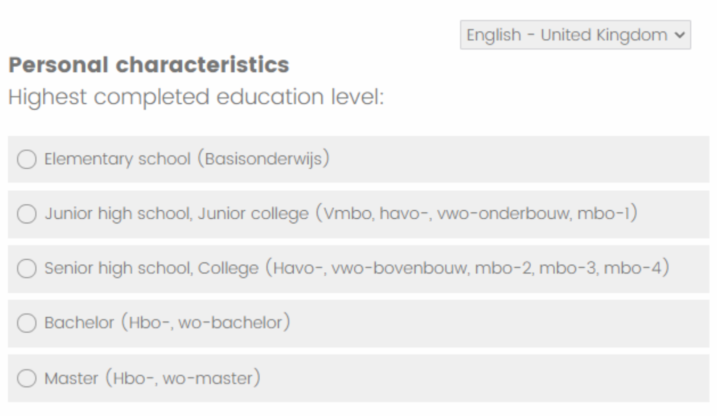
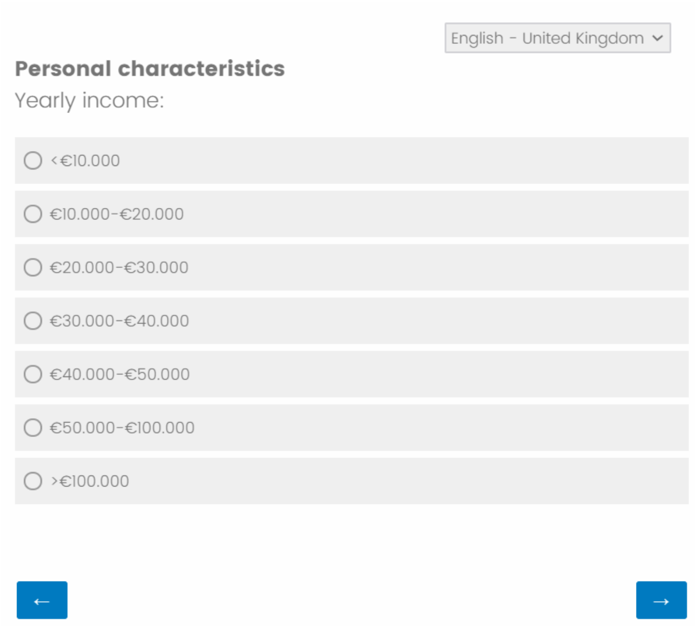


Figure 74. Personal characteristics 3



English - United Kingdom

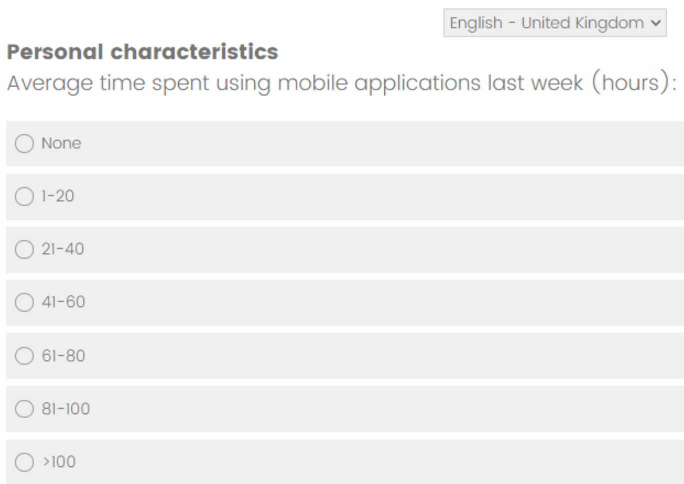
Personal characteristics

Yearly income:

- ‹€10.000
- €10.000-€20.000
- €20.000-€30.000
- €30.000-€40.000
- €40.000-€50.000
- €50.000-€100.000
- ›€100.000

← →

Figure 75. Personal characteristics 4



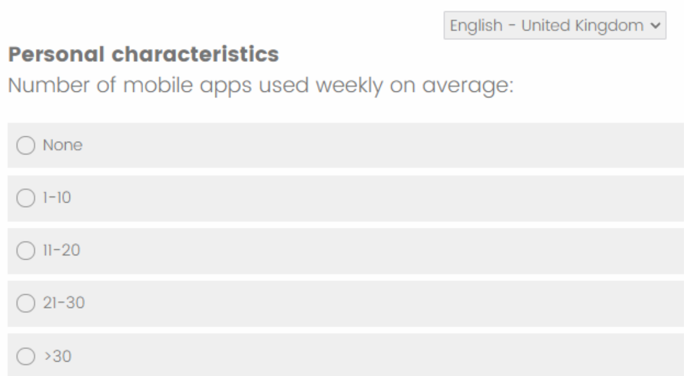
English - United Kingdom

Personal characteristics

Average time spent using mobile applications last week (hours):

- None
- 1-20
- 21-40
- 41-60
- 61-80
- 81-100
- ›100

Figure 76. Personal characteristics 5



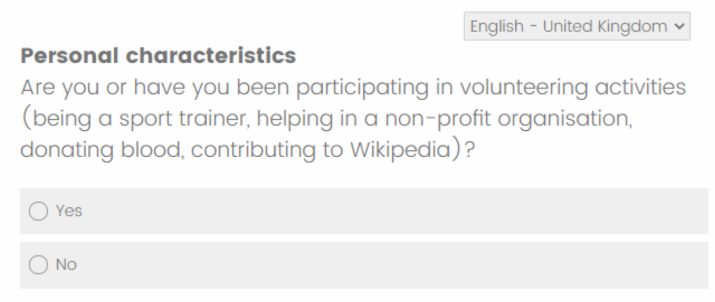
English - United Kingdom

Personal characteristics

Number of mobile apps used weekly on average:

- None
- 1-10
- 11-20
- 21-30
- ›30

Figure 77. Personal characteristics 6


English - United Kingdom

Personal characteristics

Are you or have you been participating in volunteering activities (being a sport trainer, helping in a non-profit organisation, donating blood, contributing to Wikipedia)?

- Yes
- No

Figure 78. Personal characteristics 7

Appendix D. Frequencies per choice set

Table 40 presents the frequencies of the “yes” and “no” option being chosen per choice set.

Table 40. Frequencies per choice set

Choice set	Yes (freq.)	Yes (%)	No (freq.)	No (%)	Total
1	24	63%	14	37%	38
2	9	24%	29	76%	38
3	13	34%	25	66%	38
4	8	21%	30	79%	38
5	34	89%	4	11%	38
6	18	47%	20	53%	38
7	2	5%	36	95%	38
8	17	45%	21	55%	38
9	26	68%	12	32%	38
10	24	63%	14	37%	38
11	17	45%	21	55%	38
12	2	5%	36	95%	38
13	10	25%	30	75%	40
14	5	13%	35	88%	40
15	29	73%	11	28%	40
16	17	43%	23	58%	40
17	15	38%	25	63%	40
18	27	68%	13	33%	40
19	4	10%	36	90%	40
20	8	20%	32	80%	40
21	12	30%	28	70%	40
22	18	45%	22	55%	40
23	6	15%	34	85%	40
24	11	28%	29	73%	40
25	32	68%	15	32%	47
26	7	15%	40	85%	47
27	14	30%	33	70%	47
28	25	53%	22	47%	47
29	18	38%	29	62%	47
30	10	21%	37	79%	47
31	19	40%	28	60%	47
32	19	40%	28	60%	47
33	22	47%	25	53%	47
34	15	32%	32	68%	47
35	22	47%	25	53%	47
36	11	23%	36	77%	47

Appendix E. MNL models

In this appendix, the *R* code that was used to estimate the Multinomial Logit models is displayed. For the base MNL model (model 1), the full code is included. Also, the model parameters and the utility functions for model 2 are shown. Model 4, 5, 6, 7, 8, 9, 10, and 11 are variations on these models.

Model 1: Base MNL model

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

### Initialise code
apollo_initialise()

### Set core controls
apollo_control = list(
  modelName  ="Model 1",
  modelDescr ="Model 1",
  indivID    ="ID"
)

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

### Model parameters
apollo_beta=c(B_mon   = 0,
              B_eff   = 0,
              B_rid   = 0,
              B_tod   = 0,
              B_useaca   = 0,
              B_usecor = 0,
              B_usesoc = 0,
              B_yes = 0)
apollo_fixed = c()

### VALIDATING AND PREPARING INPUTS
apollo_inputs = apollo_validateInputs()

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

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

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

  ### Create coefficients using interactions with socio-demographics

  ### List of utilites
  V = list()
  V[['yes']] = mon*B_mon + eff*B_eff + rid*B_rid + tod*B_tod + B_useaca*(use==1) +
B_usecor*(use==2) + B_usesoc*(use==3) + B_yes
  V[['no']] =  0
```

```
  ### Define settings for MNL model component
  mnl_settings = list(
  alternatives = c(yes=1, no=2),
  avail = list(yes=1, no=1),
  choiceVar = CHOICE,
  V = V
)

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

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

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

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

### Reporting and saving results
apollo_modelOutput(model,modelOutput_settings=list(printPVal=TRUE))
apollo_saveOutput(model)
```

Model 2: MNL model with interaction effects

```
### Model parameters
apollo_beta=c(B_mon   = 0,
              B_eff   = 0,
              B_rid   = 0,
              B_tod   = 0,
              B_useaca   = 0,
              B_usecor = 0,
              B_usesoc = 0,
              B_yes = 0,
              B_distscore_useaca = 0,
              B_distscore_usecor = 0,
              B_distscore_usesoc = 0,
              B_privscore_rid = 0,
              B_privscore_tod = 0)

V = list()
V[['yes']] = mon*B_mon + eff*B_eff + rid*(B_rid + B_privscore_rid*privscore) +
tod*(B_tod + B_privscore_tod*privscore) +
(B_useaca+B_distscore_useaca*distscore)*(use==1) +
(B_usecor+B_distscore_usecor*distscore)*(use==2) + (B_usesoc +
B_distscore_usesoc*distscore)*(use==3) + B_yes
V[['no']] =  0
```

Appendix F. LC model

This appendix shows the code that was used to estimate the Latent Class model (model 3) in *R*Studio.

Model 3: LC model

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

### Initialise code
apollo_initialise()

### Set core controls
apollo_control = list(
  modelName  ="Model 3",
  modelDescr ="Model 3",
  indivID    ="ID",
  nCores     = 3
)

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

### Model parameters
apollo_beta=c(B_mon_a   = 0,
              B_mon_b = 0,
              B_eff_a = 0,
              B_eff_b   = 0,
              B_rid_a   = 0,
              B_rid_b = 0,
              B_tod_a = 0,
              B_tod_b = 0,
              B_useaca_a = 0,
              B_useaca_b   = 0,
              B_usecor_a = 0,
              B_usecor_b = 0,
              B_usesoc_a = 0,
              B_usesoc_b = 0,
              B_yes = 0,
              delta_a = 0,
              delta_b = 0)
apollo_fixed = c("delta_b")

### Define latent class components
apollo_lcPars=function(apollo_beta, apollo_inputs){
  lcpars = list()

  lcpars[["B_mon"]] = list(B_mon_a, B_mon_b)
  lcpars[["B_eff"]] = list(B_eff_a, B_eff_b)
  lcpars[["B_rid"]] = list(B_rid_a, B_rid_b)
  lcpars[["B_tod"]] = list(B_tod_a, B_tod_b)
  lcpars[["B_useaca"]] = list(B_useaca_a, B_useaca_b)
  lcpars[["B_usecor"]] = list(B_usecor_a, B_usecor_b)
  lcpars[["B_usesoc"]] = list(B_usesoc_a, B_usesoc_b)

  V=list()
```

```
  V[["class_a"]] = delta_a
  V[["class_b"]] = delta_b

  mnl_settings = list(
    alternatives = c(class_a=1, class_b=2),
    avail        = 1,
    choiceVar    = NA,
    V            = V
  )

  lcpars[["pi_values"]] = apollo_mnl(mnl_settings, functionality="raw")
  lcpars[["pi_values"]] = apollo_firstRow(lcpars[["pi_values"]], apollo_inputs)

  return(lcpars)
}

### VALIDATING AND PREPARING INPUTS
apollo_inputs = apollo_validateInputs()

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

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

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

  ### Define settings for MNL model component
  mnl_settings = list(
    alternatives = c(yes=1, no=2),
    avail = list(yes=1, no=1),
    choiceVar = CHOICE
  )

  ### Loop over classes

  for(s in 1:2){

    ### Compute class-specific utilities
    V = list()
    V[['yes']] = mon*B_mon[[s]] + eff*B_eff[[s]] + rid*B_rid[[s]] + tod*B_tod[[s]] +
B_useaca[[s]]*(use==1) + B_usecor[[s]]*(use==2) + B_usesoc[[s]]*(use==3) + B_yes
    V[['no']] =  0

    mnl_settings$V = V
    mnl_settings$componentName = paste0("Class_",s)

    ### Compute within-class choice probabilities using MNL model
    P[[paste0("Class_",s)]] = apollo_mnl(mnl_settings, functionality)

    ### Take product across observation for same individual
    P[[paste0("Class_",s)]] = apollo_panelProd(P[[paste0("Class_",s)]], apollo_inputs,
functionality)
  }
```

```
  ### Compute latent class model probabilities
  lc_settings = list(inClassProb = P, classProb=pi_values)
  P[["model"]] = apollo_lc(lc_settings, apollo_inputs, functionality)

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

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

### Reporting and saving results
apollo_modelOutput(model,modelOutput_settings=list(printPVal=TRUE))
apollo_saveOutput(model)
```

Appendix G. ML models

This appendix shows the code that was used to estimate the Mixed Logit models. The code for model 12 and model 13 is displayed. Model 14 (ML model with non-linear parameters) is a variation on model 12.

Model 12: ML model with error term

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

### Initialise code
apollo_initialise()

### Set core controls
apollo_control = list(
  modelName  ="Model 12",
  modelDescr ="Model 12",
  indivID    ="ID",
  mixing     = TRUE
)

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

### Model parameters
apollo_beta=c(B_mon   = 0,
              B_eff   = 0,
              B_rid   = 0,
              B_tod   = 0,
              B_useaca   = 0,
              B_usecor = 0,
              B_usesoc = 0,
              B_yes   = 0,
              Sigma_yes = 1)
apollo_fixed = c()

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

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

  randcoeff[["EC_yes_RND"]] =  Sigma_yes * draws


  return(randcoeff)
```

```
}

### VALIDATING AND PREPARING INPUTS
apollo_inputs = apollo_validateInputs()

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

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

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

  ### Create coefficients using interactions with socio-demographics

  ### List of utilites
  V = list()
  V[['yes']] = mon*B_mon + eff*B_eff + rid*B_rid + tod*B_tod + B_useaca*(use==1) +
B_usecor*(use==2) + B_usesoc*(use==3) + B_yes + EC_yes_RND
  V[['no']] =  0

  ### Define settings for MNL model component
  mnl_settings = list(
  alternatives = c(yes=1, no=2),
  avail = list(yes=1, no=1),
  choiceVar = CHOICE,
  V = V
)

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

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

  ### Average across inter-individual draws
  P = apollo_avgInterDraws(P, apollo_inputs, functionality)

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

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

### Reporting and saving results
apollo_modelOutput(model,modelOutput_settings=list(printPVal=TRUE))
apollo_saveOutput(model)
```

Model 13: ML model with all parameters random

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

### Initialise code
apollo_initialise()

### Set core controls
apollo_control = list(
  modelName  ="Model 13",
  modelDescr ="Model 13",
  indivID    ="ID",
  mixing    = TRUE
)

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

### Model parameters
apollo_beta=c(B_mon   = 0,
              B_eff   = 0,
              B_rid   = 0,
              B_tod   = 0,
              B_useaca   = 0,
              B_usecor = 0,
              B_usesoc = 0,
              B_yes   = 0,
              Sigma_yes = 1,
              Sigma_mon = 1,
              Sigma_eff = 1,
              Sigma_rid = 1,
              Sigma_tod = 1,
              Sigma_useaca = 1,
              Sigma_usecor = 1,
              Sigma_usesoc = 1)
apollo_fixed = c()

### Set parameters for generating draws
apollo_draws = list(
  interDrawsType = "halton",
  interNDraws    = 500,
  interUnifDraws = c(),
  interNormDraws = c("draws_mon", "draws_eff", "draws_rid", "draws_tod",
"draws_useaca", "draws_usecor", "draws_usesoc", "draws_yes"),
  intraDrawsType = "halton",
  intraNDraws    = 0,
  intraUnifDraws = c(),
  intraNormDraws = c()
)

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

  randcoeff[["EC_yes_RND"]] = B_yes + Sigma_yes * draws_yes
  randcoeff[["EC_mon_RND"]] = B_mon + Sigma_mon * draws_mon
```

```
  randcoeff[["EC_eff_RND"]] = B_eff + Sigma_eff * draws_eff
  randcoeff[["EC_rid_RND"]] = B_rid + Sigma_rid * draws_rid
  randcoeff[["EC_tod_RND"]] = B_tod + Sigma_tod * draws_tod
  randcoeff[["EC_useaca_RND"]] =  B_useaca + Sigma_useaca * draws_useaca
  randcoeff[["EC_usecor_RND"]] =  B_usecor + Sigma_usecor * draws_usecor
  randcoeff[["EC_usesoc_RND"]] =  B_usesoc + Sigma_usesoc * draws_usesoc

  return(randcoeff)
}

### VALIDATING AND PREPARING INPUTS
apollo_inputs = apollo_validateInputs()

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

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

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

  ### Create coefficients using interactions with socio-demographics

  ### List of utilites
  V = list()
  V[['yes']] = mon*EC_mon_RND + eff*EC_eff_RND  + rid*EC_rid_RND  + tod*EC_tod_RND  +
EC_useaca_RND *(use==1) + EC_usecor_RND*(use==2) + EC_usesoc_RND *(use==3) +
EC_yes_RND
  V[['no']] =  0

  ### Define settings for MNL model component
  mnl_settings = list(
  alternatives = c(yes=1, no=2),
  avail = list(yes=1, no=1),
  choiceVar = CHOICE,
  V = V
)

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

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

  ### Average across inter-individual draws
  P = apollo_avgInterDraws(P, apollo_inputs, functionality)

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

#### MODEL ESTIMATION
model = apollo_estimate(apollo_beta, apollo_fixed,
```

```
                        apollo_probabilities, apollo_inputs,
estimate_settings=list(hessianRoutine="maxLik"))

### Reporting and saving results
apollo_modelOutput(model,modelOutput_settings=list(printPVal=TRUE))
apollo_saveOutput(model)
```

Appendix H. Lexicographic answering

In this appendix, the characteristics of people that answered lexicographically are examined. In Figure 79, the privacy scores of people that chose the "no" alternative in all choice situations are presented. It can be seen that these people score relatively high on the Privacy Index.

Lexino * Privscore Crosstabulation

			Privscore											
			5,00	6,00	7,00	8,00	9,00	10,00	11,00	12,00	13,00	14,00	15,00	Total
Lexino	,00	Count	1	6	3	10	16	6	19	23	14	7	2	107
		% within Lexino	0,9%	5,6%	2,8%	9,3%	15,0%	5,6%	17,8%	21,5%	13,1%	6,5%	1,9%	100,0%
	1,00	Count	0	0	1	0	1	0	3	3	4	3	3	18
		% within Lexino	0,0%	0,0%	5,6%	0,0%	5,6%	0,0%	16,7%	16,7%	22,2%	16,7%	16,7%	100,0%
Total		Count	1	6	4	10	17	6	22	26	18	10	5	125
		% within Lexino	0,8%	4,8%	3,2%	8,0%	13,6%	4,8%	17,6%	20,8%	14,4%	8,0%	4,0%	100,0%

Figure 79. Lexicographic answers with Privacy score

Figure 80 presents this effect more clearly by presenting the Privacy Index categories of the lexicographically answering people. The percentage of people belonging to the "Privacy Concerned" category is higher when compared to the rest of the sample.

Lexino * PrivIndex Crosstabulation

			PrivIndex			
			,00	1,00	2,00	Total
Lexino	,00	Count	1	60	46	107
		% within Lexino	0,9%	56,1%	43,0%	100,0%
	1,00	Count	0	5	13	18
		% within Lexino	0,0%	27,8%	72,2%	100,0%
Total		Count	1	65	59	125
		% within Lexino	0,8%	52,0%	47,2%	100,0%

Figure 80. Lexicographic answers with Privacy Index categorisation

In Figure 81, the distrust scores of people that chose the "no" alternative in all choice situations are presented. No clear conclusion can be drawn relating to the distrust scores of people giving lexicographic answers.

Lexino * Distscore Crosstabulation

			Distscore																	
			4,00	5,00	6,00	7,00	8,00	9,00	10,00	11,00	12,00	13,00	14,00	15,00	16,00	17,00	18,00	19,00	20,00	Total
Lexino	,00	Count	1	1	2	6	8	15	11	10	18	10	11	5	3	4	1	1	0	107
		% within Lexino	0,9%	0,9%	1,9%	5,6%	7,5%	14,0%	10,3%	9,3%	16,8%	9,3%	10,3%	4,7%	2,8%	3,7%	0,9%	0,9%	0,0%	100,0%
	1,00	Count	0	0	0	0	1	2	2	3	1	0	0	3	3	1	1	0	1	18
		% within Lexino	0,0%	0,0%	0,0%	0,0%	5,6%	11,1%	11,1%	16,7%	5,6%	0,0%	0,0%	16,7%	16,7%	5,6%	5,6%	0,0%	5,6%	100,0%
Total		Count	1	1	2	6	9	17	13	13	19	10	11	8	6	5	2	1	1	125
		% within Lexino	0,8%	0,8%	1,6%	4,8%	7,2%	13,6%	10,4%	10,4%	15,2%	8,0%	8,8%	6,4%	4,8%	4,0%	1,6%	0,8%	0,8%	100,0%

Figure 81. Lexicographic answers with Distrust score

In Figure 82, the Distrust Index categories of people answering lexicographically are presented. It appears that, compared to the rest of the sample, slightly more people belong to the "Low Distrust" category, and much more people belong to the "High Distrust" category.

Lexino * DistIndex Crosstabulation

			DistIndex				Total
			,00	1,00	2,00	3,00	
Lexino	,00	Count	18	36	44	9	107
		% within Lexino	16,8%	33,6%	41,1%	8,4%	100,0%
	1,00	Count	1	7	4	6	18
		% within Lexino	5,6%	38,9%	22,2%	33,3%	100,0%
Total		Count	19	43	48	15	125
		% within Lexino	15,2%	34,4%	38,4%	12,0%	100,0%

Figure 82. Lexicographic answers with Distrust Index categorisation

In Figure 83, the yearly incomes of people that chose the “no” alternative in all choice situations are presented. It can be seen that a relatively high percentage of the people giving lexicographic answers have a yearly income of €50.000-€100.000. Also, 4 people in the subset of lexicographic answers did not answer the question on income.

Lexino * Personal characteristics Yearly income: Crosstabulatio

			Personal characteristics Yearly income							Total
			<€10.000	€10.000-€20.000	€20.000-€30.000	€30.000-€40.000	€40.000-€50.000	€50.000-€100.000	>€100.000	
Lexino	,00	Count	30	5	10	12	15	29	5	106
		% within Lexino	28,3%	4,7%	9,4%	11,3%	14,2%	27,4%	4,7%	100,0%
	1,00	Count	1	1	1	2	0	8	1	14
		% within Lexino	7,1%	7,1%	7,1%	14,3%	0,0%	57,1%	7,1%	100,0%
Total		Count	31	6	11	14	15	37	6	120
		% within Lexino	25,8%	5,0%	9,2%	11,7%	12,5%	30,8%	5,0%	100,0%

Figure 83. Lexicographic answers with Yearly income

Appendix I. Interviews

In this appendix, the transcripts of the conducted interviews can be found. First, the transcripts of interview A, B, and C are presented. Then, the a table is displayed providing an overview of the codes and the grouping of the codes.

I.1. Interview A, Municipality
2 November 2021

Kun je een introductie geven over wat je doet bij de gemeente en waar je mee bezig bent op het gebied van Smart Mobility?

Ja. Ik werk dus bij het programma Smart Mobility en houd me daar met twee dingen bezig: een project waarin we onderzoeken wat voor voorwaarden we als gemeente zouden willen stellen aan mobiliteitsaanbieders om de digitalisering te helpen en om daar als gemeente ook gewoon meer grip op te hebben. In het andere project houd ik me bezig met ethiek. Ik heb zelf filosofie gestudeerd. Het terrein filosofie en technologie is nog niet helemaal bekend maar ik ben nu aan het kijken hoe kunnen we ervoor zorgen dat ethiek een vast onderdeel wordt van ons innovatieproces. Wat is daarvoor nodig en wat voor tools zijn er die ons daarbij kunnen helpen. Dus daar ben ik op aan het oriënteren. Dat zijn denk ik de belangrijkste twee dingen.

Zou je ook iets meer kunnen vertellen over de Smart Mobility projecten binnen Amsterdam?

Er zijn een aantal projecten. Bijvoorbeeld Roboat was de laatste tijd in het nieuws, zelfvarende boten. Zelfrijdend vervoer, worden pilots mee gedaan maar is nog niet zover dat er ook echt al iets gaat rijden. Mobility as a Service, dat is dat multimodaal reizen. Dat is ook iets wat we graag zouden willen dat dat kan, maar er komt nog wat organisatie bij kijken voordat je een aanbieder zover hebt dat ze multimodaal reizen met elkaar gaan organiseren. Dus dat is iets waar wij dan als gemeente aan werken om te kijken of we die innovatie kunnen versnellen. Buurthubs, dat gaat over hubs, dus plekken in een wijk, waar deelvervoer wordt aangeboden. En collega's van mij zijn er dus mee bezig om te kijken, hoe kan je dat organiseren en hoe kan je de buurt er ook in betrekken om dat te organiseren? Zodat het ook echt iets wordt van de buurt en waar de buurt vindt wat ze nodig hebben. Smart Mobility Lab Stadionplein is ook een hub maar dan een wat grotere. In een oud tankstation wordt dat georganiseerd. En hier willen ze ook wat meer functie aan toevoegen.

Is dat om te testen hoe het in de praktijk zou werken?

Ja. Dus om het te realiseren om dan te leren van, inderdaad, hoe werkt dat, wat werkt, wat heb je ervoor nodig. Dat soort dingen.
Dat is dus onder de lijn Slim en schoon reizen. Dan heb je data en mobiliteit. Public Eye, dat gaat over crowd management. Dus dat sluit denk ik het beste aan bij waar jij mee bezig bent. CDS-M, dus City Data Standaard Mobiliteit. Dat gaat over het realiseren van een datastandaard waarmee mobiliteitsaanbieders data kunnen gaan uitwisselen met de gemeente. En met andere overheden, want het liefst doen we dat natuurlijk op een gestandaardiseerde manier. Dat aanbieders niet voor elke stad weer op een andere manier hun data moeten gaan uitwisselen. Maar dat het voor hen ook overzichtelijk is. En de lessen die wij leren over hoe dat moet dat data delen, dat ze dat ook kunnen delen met andere steden die wat minder innovatiecapaciteit hebben. Drukte in de openbare ruimte, denk ik ook al over gehoord, gaat over het druktebeeld. Slimme mobiliteit met MobiLab. Dit gaat vooral over projecten dat we eigenlijk zelf in staat zijn om data te ontvangen en te analyseren. Daar heb je ook gewoon een platform voor nodig. Dus dat wordt ontwikkeld.

Wordt dat ontwikkeld door de gemeente zelf, of werken jullie daarin samen met een andere partij?

Dit is eigenlijk echt nog een lab om te testen van hoe, wat hebben we ervoor nodig en hoe gaat dat, dat data uitwisselen, en analyseren vervolgens. Dus het wordt voornamelijk zelf ontwikkeld. Uiteindelijk zouden we

dat natuurlijk gewoon zelf willen kunnen als gemeente, zeker omdat we zo groot zijn. Dit zijn voorbeelden van pilots[2] , waar ze dan mee proberen zeg maar, oefenen in data verzamelen en analyseren.
En dan heb je hier OMC. Dat is rond het Arena gebied. Daar zijn ze bezig om crowd management en het monitoren ook te digitaliseren. Eerder hadden ze eigenlijk alleen maar informatie vanuit camera's, dus gewoon beelden. En nu willen ze ook veel meer data gedreven, dus dat ze ook willen zien van daar is het zo druk, of er komt zoveel verkeer aan of er komt straks een trein aan. Met zoveel mensen erin, liefst ook nog dat soort informatie.

Van wat voor soort sensoren maak je dan gebruik? Als je geen camera's meer gebruikt, op welke manier verzamel je die data dan?
Ja die andere partners hebben die data dan, bijvoorbeeld OV weet natuurlijk hoeveel in-checks er waren. Parkeeraanbieders weten ook gewoon hoeveel auto's er hun garage in zijn gereden. Dus dat zijn de belangrijkste databronnen. We gebruiken nog niet veel sensoren en als we die gebruiken dan is het denk ik vooral dat wat onder Public Eye valt, crowd monitoring system.

Zijn dit allemaal projecten die nog in de testfase zijn of wordt dit al toegepast op dit moment?

Die hubs bijvoorbeeld worden wel echt gemaakt ook, maar wel met het doel van dit is voor het eerst dat we iets op straat hebben, hoe werkt dat dan. Dus alles wat we doen zit nog heel erg in de test- en leerfase. Omdat we natuurlijk innovatieteam zijn. Je kan wel zien, we leveren geen "af product" af zeg maar, van nu is het klaar, nu kun je het gebruiken, maar leren vooral lessen. Nu zit ik te denken over die hubs hoe dat dan verder gaat, maar dat is eigenlijk ook gewoon nog een vraagstuk dat beantwoord moet worden in de loopt van zo'n project. Als het werkt, wat doen we er dan mee, hoe kunnen we het opschalen. Dat zijn ook leervragen dan.

Er wordt best wat data verzameld. Wat is dan uiteindelijk het doel van die data, wat willen jullie daarmee doen?

Ik denk dat het sowieso wel meevalt vaak, hoeveel data er wordt verzameld. Bijvoorbeeld bij buurt hubs daar gaan ze samenwerken ook met universiteiten, ook met de TU Delft, voor onderzoek over hoe het werkt. Dus dan wordt er wel data verzameld. Ik zit niet zo heel nauw in die projecten. Sowieso hebben we als uitgangspunt bij de gemeente de Tada-principes.
Tada.city, ook een hele interessante website, zeker als je meer wilt weten over data en ethiek. Tada manifest, is een manifest dat niet door de gemeente is gemaakt maar waarvan we wel mede-ondertekenaar zijn, waarin eigenlijk verschillende burgers en andere partijen, die staan ook hieronder hebben verwoord, wat vinden we nou belangrijk voor de stad op het gebied van data. Er zijn een aantal principes uitgekomen, bijvoorbeeld we willen dat digitalisering inclusief is, dat er zeggenschap is, menselijke maat, legitiem gecontroleerd, open transparant, en van iedereen voor iedereen. En de gemeente heeft gezegd ja daar staan wij ook achter. Dus je zou kunnen zeggen dat dit wel een soort uitgangspunten zijn voor de innovaties die we doen. En daarbij heb je sowieso de AVG, die zegt al heel veel over data, wat wel en niet mag, en data minimalisatie is natuurlijk ook belangrijk. Dus we mogen niet zomaar data verzamelen om het data verzamelen, om het zo te zeggen. Dat kan niet, of dat je denkt we willen gewoon data en dan kijken we daarna wel wat we ermee kunnen of wat makkelijk is. Dat werkt natuurlijk andersom. Je hebt een doel en dat doel moet zo belangrijk zijn, rechtvaardig, dat je er data voor gaat verzamelen. Zeker als het dus over mensen gaat, persoonsgegevens. En dan pas kan je dat gaan verzamelen wat daarvoor nodig is als dat juridisch klopt en ethisch het liefst ook natuurlijk. Dus dat is een beetje de volgorde waarin dat gaat. Was dat een beetje een antwoord op je vraag?

Misschien kun je ook wat meer vertellen over hoe jullie de data willen gebruiken? Bijv. dingen voorspellen, om met verschillende partijen proberen om mobiliteitsstromen aan te passen?

We zijn wel aan het nadenken over hoe zouden we als gemeente inderdaad meer op die manier kunnen gaan werken. Maar dat zit vooral nog in de fase dat we vooral visie daarop aan het ontwikkelen zijn. Hoe kijken we er tegenaan en wat zouden we in de toekomst kunnen doen op dat gebied? Binnen het crowd management project gebeurt denk ik het meeste op het gebied van voorspellen, daar kijken ze inderdaad of we al iets zeggen als het regent of het dan druk wordt of niet, dan soort dingen. Dus daar zit denk het meest

[2] Gemeente Amsterdam (2021)

voorspellende. Maar ook voor mobiliteitsstromen zitten we meer in de fase dat we nog de data moeten gaan verzamelen. Dat wat jij onderzoekt is bijvoorbeeld heel interessant is, want het zou natuurlijk heel mooi zijn, als ik goed begrijp wat jij onderzoekt, dat als mensen zelf data leveren, terwijl ze weten dat ze het leveren, en daar ook achter staan. Dat zou eigenlijk wel het ideale geval zijn, ook voor hoe we als gemeente werken. Misschien niet per se ideaal, daar zouden we goed over na moeten denken, want je bent heel erg afhankelijk bent van elk individu en of ze mee willen werken. Voor sommige toepassingen wil je misschien wat zekerder zijn van je databronnen, maar het is wel waar we achter staan, dat we willen dat mensen bewuste keuzes daarin maken.

Je had het over crowd management, waar je een camera hebt die ook data verzamelt. Hoe gaan jullie dan om met de ethische kant daarvan?

Goede vraag. En wat is dan vooral je vraag als je het hebt over de ethische kant daarvan?

Hoe vliegen jullie zo'n project in, hoe bepaal je of iets wel of niet kan, want juridisch ligt het meeste natuurlijk wel vast, maar qua privacy heb je best wel een grijs gebied van wat je wel en niet wilt verzamelen. Dus hoe bepalen jullie hoe je daarmee omgaat?

Verschillende manieren. Ik weet dat ze bij Public Eye bijvoorbeeld een workshop hebben gedaan waarbij ze die Tada-principes als uitgangspunt hebben genomen en toen hun project daarlangs hebben gereviewd om het zo maar te zeggen. Dus dat Tada biedt wel veel houvast. Ook omdat we weten dat de gemeenteraad dat heeft vastgesteld, dus dat het niet zomaar een individueel idee is, van ik vind inclusiviteit belangrijk, maar dat we weten dat we dat als stad belangrijk vinden. Dus dat helpt, dat geeft richting, van welke keuzes moet je dan maken. Ik denk dat dat de belangrijkste tool is die we op dit moment hebben. Verder hebben we bijvoorbeeld in de gemeente zoiets als moreel beraad, maar ik weet niet of dat hier ook is ingezet. Maar dat is ook een methode voor als je dus ethisch ingewikkelde keuzes tegenkomt om jezelf te helpen van hoe ga je hier verstandig mee om. En verder is natuurlijk de gemeenteraad wel de partij die controleert wat de gemeente doet, dus zij hebben daar ook een mening over en een stem in, en een controlerende functie natuurlijk.

En als je data gebruikt van andere partijen, dus bijvoorbeeld van die vervoersbedrijven, hoe gaat dat dan in z'n werk? Ga je dan samen om de tafel om daar over na te denken?

Daar worden sowieso contracten over afgesloten. En dat er weer wordt nagedacht van hoe richten we dat in.

Wat voor projecten hebben jullie in de toekomst, zijn daar ideeën over? Bijvoorbeeld data die je graag nog zou willen gebruiken om processen in de stad te verbeteren, waar nu nog niks mee gedaan wordt. Hebben jullie projecten voor ogen die je in de toekomst nog zou willen doen op dat gebied?

Een heel groot project wat echt nog in die helemaal beginfase zit, is dat we willen nadenken over hoe je mobiliteitsmanagement meer digitaal zou kunnen doen. Dus op dit moment ontvangt de gemeente niet echt veel informatie over mobiliteitsstromen. Dus dat klinkt heel ouderwets, dat is ook heel ouderwets. Dus we willen uiteindelijk kijken hoe we dat anders kunnen gaan aanpakken. Dat het wel meer digitaal wordt. Maar dat staat echt nog helemaal in de beginfase, waarin al die vragen, van wat mag je verzamelen, wat wil je verzamelen, hoe ga je dat doen, wat is ethisch, hoe waarborg je privacy. Zit nog zelfs die fase ervoor. Dat je dus nog gaat bedenken oké waar zetten we die stip aan de horizon, wat wordt het pad ernaartoe, en hoe gaan we leren onderweg wat de antwoorden zijn op al die vragen die ik dan net noemde. Dus dat klinkt misschien niet heel spannend. Ondertussen doen we natuurlijk allemaal kleine pilots, zoals dat wat je gelezen hebt op de website. Waar we eigenlijk dingen leren over hoe ga je dan toe naar digitaal mobiliteitsmanagement. Dus bijvoorbeeld zo'n MobiLab waarin dan geoefend wordt met het ontvangen van data en het analyseren ervan. Daar gaan natuurlijk heel veel lessen uitkomen voor als we het hele mobiliteitsmanagement ander in willen richten, wat hebben we daarvoor nodig.

Misschien een stukje toepassing, de link met waar ik mee bezig ben in m'n project. Op wat voor manier denk jij dat crowd sensing zou kunnen bijdragen aan het inzicht krijgen in de mobiliteit binnen de gemeente?

Ik denk als Amsterdammers of bezoekers inderdaad via crowd sensing data willen delen en wij daardoor inzicht krijgen in verkeersstromen, of drukte in de stad, dat dat zeker heel nuttig is. En op het eerste oog klinkt het als een mooie manier om die data te krijgen. Van mensen die er heel bewust voor kiezen om die data te delen. Ik vraag me wel af, delen ze alleen data over zichzelf, want je zei ook iets over de omgeving?

Vraag ik ook in de enquête. Je hebt verschillende soorten data die je zou kunnen verzamelen, locatie data, bewegingsdata, acceleratie, contextual data, dus luchtkwaliteit bijvoorbeeld.

Maar niet bijvoorbeeld over mensen die naast hun lopen?

Nee het gaat echt meer over de omgeving. De laatste is multimedia, dat ze bijvoorbeeld een foto maken ergens van, of een video, of een geluidsopname, om te kijken hoe het geluidsoverlast is op een bepaalde plek bijvoorbeeld.

Klinkt wel echt alsof dat heel veel mogelijkheden zou bieden. En dat dat wel kan gaan helpen om eigenlijk alles wat jij noemt, als we luchtkwaliteit gaan meten kun je daar meer inzicht in krijgen. Er wordt natuurlijk ook wel luchtkwaliteit gemeten in de stad, maar dat soort data helpt dan om je beleid data-gedreven te maken, om te kijken wat er gebeurt in de stad, en om te kijken: helpen maatregelen, dat soort dingen. Over die drukte, dat kan dus helpen op het moment dat daar op gemonitord moet worden. Bijvoorbeeld in de corona-tijd was dat heel actueel. In andere tijden kun je je afvragen, in sommige delen van de stad waar het niet druk is het natuurlijk niet relevant om daar iets over te weten, hoeveel mensen daar lopen. En soms wil je het niet weten, bijvoorbeeld als er één iemand loopt, dan is dat niet anoniem zeg maar. Dus dan wil je niet binnenkrijgen: er loopt nu één iemand. Dus er is ook een soort drempel in van wanneer wil je het en wanneer mag je zulk soort dingen weten. Maar het kan zeker helpen en het klinkt ook als een mooie manier om die data in te zamelen, zo op het eerste oog.

Wat voor stappen zouden er denk je nodig zijn om zoiets te implementeren?

Wij moeten dus de data kunnen ontvangen. Dus de technische infrastructuur moet er zijn. En ik denk dat er juridisch best wel wat werk verzet zou moeten worden over hoe je dat in elkaar zet, dat er ook goed gekeken moet worden naar de privacy, hoe die gewaarborgd wordt. Al is het natuurlijk anders als mensen er zelf voor kiezen data te delen, maar ik kan me voorstellen dat zij daar wel bepaalde voorwaarden ook zelf aan willen stellen. Dus dat lijkt me belangrijk om goed geregeld te hebben.

Heb je zelf nog vragen?

Het leuke, en ook het moeilijke aan dit terrein is, het is natuurlijk allemaal nog speculatie. We zijn aan het leren, we weten niet hoe we het willen en hoe we het gaan doen. Maar ik zou zeker die Tada-website aanraden om daar eens naar te kijken. Dat geeft al mooi richting aan hoe we er in staan als stad, wat we belangrijk vinden.

I.2. Interview B, Research group Connected Cars
4 November 2021

Can you give an introduction on the research group?
So basically what we're talking about is the group that was started a long time ago, I think it was 2016. The group is called Connect2 and as I mentioned it started with people who were interested in requirements engineering and shortly after it evolved into a group that was just interested in connected car topics. It started in Wolfsburg, Germany, at our company Aircrony, which was purchased by AKKA, and it still continues today. We try to meet every Monday to discuss what's going on in the connected car world. This is a presentation from last year. At the end of the year we usually do a review of what we did that year and talk about it. Before corona, we would have a little Christmas party. This is an older presentation, there will be a new one soon, just to give you an idea.

So this is the kick-off, and what it looked like back then. Like I said, it started with engineering requirements, requirements management, and evolved in connected car topics. The nice thing about this group is that it is voluntary, nobody is forced to come do it. The topic connected car connects all of us, people who are interested and they just come. This is very characteristic of community practice. When people come together it's just this sharing of knowledge and connecting with each other. Basically what we're doing is capping in the knowledge as a group and profiting from it. Of course our knowledge expands and is continually developed as we exchange it with each other. Most of the people involved are here in Wolfsburg but we are open for everybody in the company.

We identify these people who have a really interesting project or something that they learned recently, or we gather all the news with what's been going on with connected car, and we have a session about the news, or somebody just does a presentation or tells us, or we have a topic that we want to talk about. We learned a lot about Tesla, and Starling, the sun lights to provide Internet in places where Internet isn't available, and all the ramifications about it.

We have some things that are more complex. For example, if we have projects where everyone has non-disclosure agreement, sometimes we talk about things that are very customer-oriented, that are part of our work, how we can help each other, and about understanding complex issues more.

This one is a little bit more about mobile online services. And this is also another one about connectivity and electric vehicles. It's also a topic that we had in 2020. Basically about how a vehicle can connect to a house and supply that home with electricity, bi-direction of charging, and also how can we charge from solar panels on the roof and how all that plays a mix in the whole energy needs of the home, and how can we actually improve through the use of an electric vehicle that is connected to that home.

What's the goal of all the knowledge you're gaining? Do you implement it in your daily work or are you involved in the development of connected cars?

One of the interesting things is that because we're basically doing this voluntary and don't have to supply any numbers to managers. People have come back though and have said to us, "this group is fantastic". Because they had no information about connected car, and just nearly by taking part in the group they have shortened their time it takes to be trained in for the project they're working on by 2 to 3 weeks. There are also things that are very difficult to measure, if people are motivated by the group, and stay in the company because of it.

Could you give an example of issues you talked about relating to connected cars?

Some of the issues I did mention, for example how an electric vehicle connects to a house. Insight that house there is some form of intelligence, usually it is called home energy management system. It talks to the car, it talks to other devices within the house, and it knows what the customer's preferences are. So you were talking about people with their cell phone and having apps, where these apps collectively, like Google, tell you "here are the high traffic times for the library", or for this restaurant.

Similar to that, in a home it looks at what are the patterns of the user of the energy in the house, so the house owner comes home, and they own a car, and maybe they put in a profile in the car, like program A is "I want my car charged with as much as energy from our solar panels as possible, but it does not have to be right now, I leave for work tomorrow at 9 in the morning”. So this home energy management system insight the house will think I know when the sun really starts to shine, I will fill up this car with energy from the sun. And then at the same time, we know that we want to have the battery charged at 80% or so, so I know that it’s really important otherwise I won’t be able to get to work. So if something comes up, where somebody is turning on the washing machine, or the heater to the pool, and it takes a lot of energy, and they don’t really need it, I am going to shut that off, and turn it on later to make sure he gets his charge before going to work. So these are new things that people are discussing. While electric vehicles get connected to a house, and how that actually plays a part in using energy smartly and wisely. Such as charging when electricity rates are low, so this intelligence becomes this home energy management system that gets dynamic pricing from the electricity distributor provider. If they know, right now it’s really expensive, then maybe they just charge at night. Or if it’s really expensive during the day, and the technology is there, because right now it’s there but it hurts the battery, you can actually take energy from the battery and supply your home with it at times when the electricity is really expensive.

Also what I mentioned earlier, connecting cars so that they can communicate with each other. If you’re on the highway, and one car gets in an accident, it sends a signal to all the other cars out there and says an accident has happened and you should be careful in 2 kilometres or something.

And do you also think about the privacy of those applications?

It is a huge barrier. You have to respect data privacy, but at the same time, it makes the knowledge flow more difficult. And there’s other things too. You have the data privacy thing, people have to be willing to take part in this. People will have questions like, what does this mean, what kind of data is my car giving to other people? You have cultural differences. In Germany it’s more sensitive, providing data is a bigger thing. That is my personal opinion, how I view the culture here as an American. In America they’re more relaxed. Here, people want to know, where is this data going, who’s getting it, how’s it being used? These are good questions.

You also have the aspect when people have valuable knowledge, they want to know what they’re going to get out of it. We learned at school not to have people look at our test, because sharing knowledge does not really help them in that case. Nonetheless we learn these rules about sharing, but when you come into a company, the best thing you can do is share in a proper way. Because if you’re not sharing and you have those silos in a company, you will not be as effective or fast or competitive as another company who’s doing that. If you’re working as one big team, and you have a culture that shares, and take the time to share with somebody else, this type of mentality needs to be there. You will always have the data privacy and a lot of questions to answer about that, people need to feel comfortable before sharing their knowledge, they need to go where it goes. And some people need to know what they are going to get out of it. They need to trust that a person will return the favour. And some people are intrinsically motivated and just want to help, so they just give their knowledge and don’t think about it.

Specifically to the connected cars, for example if you have such a smart energy system and the system knows your preferences, do you think there are some security issues, or people might not want to have such a system because they are concerned about their privacy?

I am not an IT guru, but there are certain things that I think most people know when they think about connecting the things. Anytime you’re connected to a network that isn’t secure, let’s say you go to a hotel somewhere, and they don’t have a secure network, it’s a risk. Some of these people have taken the necessary precautions. Still, they are having issues. So when you know about people’s preferences, especially those people are very concerned about their privacy, they want to make sure that the security is in place.

I’ll give you an example. I have Adobe as a wonderful app for my smartphone. You can make PDF files just by taking a picture of a document. And it saves it for you in the cloud. In Europe, everyone is concerned about the cloud. Nonetheless, nobody would think of not purchasing Amazon products, but they are also huge in cloud services. Why are they huge in cloud services? They could not have made that happen if they weren’t developing their own cloud. I’m not an expert about Amazon and their cloud, but they are very far ahead. All these companies, even Google and everything. There are some security standards that they have to meet, and they would never want to have a breach, because they would lose trust with other people. You have to think like, do I think Adobe's going to have a secure cloud, you take a risk. Because the benefit is there, I have all that stuff digitalized. If my house burns down, I have everything in the cloud. You're happy when that happens. But maybe you're not so happy when your tax returns land in someone else's hands, but you always have to ask the question: what can they do with it? So people get scared because they think, "all those people have my banking information". But yet they throw out all those letters they get with their name and address on it in the garbage. Who knows if their banking information might also be on that? So people are concerned, but I think that any time when you're trying to convince people to do things, you have to show them the benefit. The benefits have to outweigh the risks.

That's also what I am investigating, presenting people with the benefits and risks and giving them a choice whether or not to share their data. Being in such a group talking about connected cars, do you think participatory sensing could be of relevance to connected cars?

Of course. They already do collect data, lots of data. In the United States and I think also here in Europe. In your car, you can connect to the Onboard Diagnostics system, so there is a little computer that can tell what's wrong with the car. It can tell you other information too about the car. It can tell your speed, and all sorts of stuff like that. There are insurance companies, that give you something to plug into the car and it that way they can see how fast you're driving, and they can see what type of driving profile you have. They can determine if you're a crazy driver or not, and you're high risk. But then again, the people who are going to do that, are the ones that are not high risk people and want to save money. That's connected car, allowing somebody to see that.

Recently, somebody was at the BMW dealership and said his convertible roof was broken and wanted it to be fixed for free, because it is a defective product. But they did not want to fix it for free, since he was driving too fast when he opened it. They showed it to him on the on-board computer, the diagnosis tells that he was driving too fast when opening it. So this is where the car is already collecting data.

I was recently in an Audi, and it had to ask me a lot of questions on the screen, which was really annoying. Having to click a lot can be bothering. It can make some functions of a product a little bit less attractive, especially in the car.

Like I said, everything has to have a benefit. With the cars, with this swarm intelligence, if the cars are talking to each other, and they say when there is an accident, if that can save them three hours' time because the navigation immediately gives a new route to drive, a lot of people would like to do that.

People like safety. You get asked for feedback all the time. Same thing with safety information. if a car gets in an accident, maybe when you buy a car they ask you if they can use your information when you get in an accident, for future safety research.

I know someone from VW who used to go to accidents immediately. They had an agreement with the authorities so they could see the accident and do research about how they can make it better. That's not fun, and can be a very hard job to do. It can be even easier if a car just sends this information.
For automotive driving, they don't use a lot of cameras anymore, they use a lot of sensors. So that's probably better, because cameras are always a big thing.

You have a car that already knows it is you or you have a preference and the seat rises, it remembers things. Those things people like. Because what are they going to do if they steal that information? What are they going to do with it? That's information that nobody really cares about if people know.

I.3. Interview C, Transport operator

2 December 2021

Ik ben nu sinds 2,5 jaar business analist bij GVB. Ik zit zelf in het commercie team. Dus ik houd me heel erg veel bezig met data met betrekking tot OV chipkaart. Dus echt ritten, reizen, hoe men door het OV beweegt. Ik houd me een beetje bezig met het hele dataspectrum, dus van hoe willen we data verzamelen tot richting advanced analytics gaan van wat kun je daar nou mee doen. Dus we zijn ook met zo'n drukte-indicator bezig om in de app te kunnen laten zien hoe druk het is. En zo steeds een beetje meer niet alleen data gebruiken voor de standaard BI rapportages die iedereen kent, maar ook om gewoon echt data te gebruiken om direct richting reiziger ofwel mensen van GVB die in contact staan met de reiziger te brengen, bestuurders, service medewerkers, dat soort zaken, om die al wat meer gebaseerd op data adviezen te geven. Dus dat is waar ik me mee bezig houd. We zijn inmiddels ook een digitaal programma gestart, het gaat nu ook iets groter worden bij GVB allemaal om ook data in te gaan zetten. Dat hangt af van digital natuurlijk hoe we die data in de digitale omgeving kunnen krijgen om vervolgens dat te brengen op de plek waar het moet zijn. Dus dat is wel een leuke periode nu hier om daar in te zitten. Dus ik ben ook niet degene die daar besluiten over maakt, over het inzetten per se van zulk soort apparatuur, maar wel heel veel advies daarover geeft en ook meekijkt van wat moeten we wel doen, wat moeten we niet doen, en af en toe ook dingen uitzet. We hadden bijvoorbeeld, jij had het over sensoren, maar we hadden infrarood telsystemen op onze voertuigen zitten, en die hebben we sinds begin dit jaar uitgezet, omdat het te weinig extra informatie gaf.

En ook om privacy redenen bijvoorbeeld?

Nee, volledig niet. Want dit waren gewoon infrarood dingen, maar het probleem was bij ons vooral dat we het op slechts 13% van onze voertuigen zitten, en dat werd heel erg random ingezet, maar hij kon bijvoorbeeld ook niet zo goed omgaan met omleidingen. In Amsterdam hebben we nogal eens omleidingen. Dus de bruikbaarheid was uiteindelijk misschien nog 5% van alle ritten die je dan overhield. En daar kon je eigenlijk niet zo heel veel meer mee dan wat we met de OV chipkaart zelf ook al kunnen. Dus dan heb je echt iets nodig dat je zwartrijders in beeld kan brengen, waar je echt goede conclusies uit kunt trekken. Misschien wel waar jij mee bezig bent ook. Als genoeg mensen er gebruik van maken, want daar zit bij dat stuk ook wel ons issue natuurlijk.

Wat voor data halen jullie op dit moment bijvoorbeeld uit de OV chipkaart?

Eigenlijk precies alle reisbewegingen, dus inchecken, uitchecken, waar dat gebeurd is, op welk tijdstip dat gebeurd is, op welk poortje dat precies gebeurd is, met welke kaart. De kilometers die gereisd zijn, de kilometerprijzen die er aan vast hangen, al dat soort zaken. Het nadeel daarvan is natuurlijk dat je alleen die data binnenhaalt, van de mensen die ingecheckt en uitgecheckt hebben. Dus je bent wel afhankelijk van het gedrag van mensen ook. Met sensoren heb je dat natuurlijk minder.

Waar gebruiken jullie die data voor?

Heel veel. Het voornaamste is misschien wel om onze vervoersanalyses te maken, dus ook om onze vervoerplannen te maken, en dienstregelingen te maken, ook hoe we door de stad heenrijden, welke voertuigen waar ingezet worden. Al dat soort zaken. Dus dat is echt heel erg in de operatie. Maar ook waar je marketingcampagnes in gaat zetten, en je reiziger ook gewoon kennen. Dat je gaat snappen welke mensen op welke manier door het OV heenreizen. En daar je producten ook weer op probeert af te stemmen. Een drukte-indicator bijvoorbeeld maken.

Ik las op jullie website bijvoorbeeld dat jullie op voertuigen aangeven hoeveel mensen erin zitten, of hoe de capaciteit is. Kun je daar iets meer over vertellen?

Volgens mij doen we het nog niet op dit moment, maar komt het wel. Wat we willen is inderdaad dat we de bezetting weergeven in het voertuig, van hoe druk het is, hoeveel mensen erin zitten. Als je bijvoorbeeld de metro hebt, dan heb je natuurlijk heel vaak dat mensen zich opproppen in één bak, terwijl het aan de andere kant helemaal leeg is. Dus ook met name om te gaan stimuleren dat mensen zich verspreiden door de voertuigen heen. Dat ze op het scherm gaan tonen: hier is het druk, daar niet, hier zo veel mensen. Dus met name daarvoor.

En hoe zouden jullie dat meten?

Nu voornamelijk met het gewicht in de voertuigen. Dus elk voertuig geeft ons het gewicht door op alle plekken. Daaruit kun je afleiden hoe druk het is en waar het drukker is. Dus dat is wat we daar nu nog voor hebben. Want zeker in de metro is dat natuurlijk lastiger omdat je incheckt voordat je het voertuig in gaat op het station. Terwijl in bussen en trams, zeker in trams, we hebben nu ook een aantal trams rijden die gekoppeld rijden. Maar dat weet je natuurlijk gewoon dat is dan op dat voertuig ingecheckt. Dus dan hebben we eigenlijk niks anders nodig.

Al weet je dan natuurlijk niet de specifieke plek waar mensen zitten.

Nee. Bij de nieuwe trams kunnen we er volgens mij wel ook gewichten uithalen, maar dat durf ik niet helemaal zeker te zeggen. Zo ver zijn we nog niet. Maar dat zou op basis daarvan kunnen. Of de zaken waar jij ook mee bezig bent natuurlijk, dat zou daar heel erg nuttig voor kunnen zijn.

Wat hebben jullie verder voor projecten op het gebied van data verzameling en smart mobility? Zijn jullie daar mee bezig?

Nog niet extreem, heel eerlijk gezegd. Want wij zitten hier wel over na te denken, van wat kun je daarin, wat mag je daar ook in, wat wil je daarin. Er zitten natuurlijk best wel veel juridische dingen ook aan. En wat wil je ook eventueel wel of niet aan je reiziger vragen. Dus dat is nog best wel lastig. Ik denk dat we nu vooral inderdaad aan het kijken zijn, hoe kunnen we de dingen die we al hebben gebruiken om er wel slimmer in te zitten, maar dat gaat echt met name om reisinformatie en spreiding van mensen op dit moment.

Voor de reisinformatie op de website of in de app, gebruiken jullie daar op dit moment ook bijvoorbeeld sensoren voor, of verzamelde data?

Durf ik niet precies te zeggen, hoe de reisinformatie kant in elkaar zit. Sowieso alles wat we uit OV chipkaart krijgen is nog niet real-time. Dus dat is allemaal sowieso met een dag vertraging, dus dat doen we alleen op basis van voorspellingen. We hebben bijvoorbeeld wel als een bus vol is, dan geeft een bestuurder wel door dat het vol is. En daar wordt dan ook wel aan de achterkant door de verkeersleiding wat mee gedaan. En als er vertraging is, bijvoorbeeld, dat komt wel rechtstreeks uit de voertuigen. Dus daar is wel gewoon voertuigdata die rechtstreeks in verbinding staat, dat is waarschijnlijk wat je ziet ook op de website en in de app, als er echt vertraging is. Dan is het echt voertuigdata, op basis van geo waar die aangeeft waar hij is.

Werken jullie ook samen met andere vervoersbedrijven bijvoorbeeld, om data uit te wisselen of op andere manieren?

Qua uitwisselen eigenlijk bijna niet. Daar blijf je natuurlijk ook een beetje met de marktwerking en dat soort zaken zitten. Dus data uitwisselen dat doen we eigenlijk bijna niet. We proberen wel wat samen te werken, bijvoorbeeld met andere stadsvervoerders, omdat die tegen dezelfde problematiek of uitdagingen aanlopen. Dus wat we dan meer doen is dat we wel als we een project doen, dat we kijken qua aanpak en qua wat voor variabelen gebruik je, wat is voor jullie nuttig, dat we die kennis uitwisselen. Maar niet data zelf. Wat we wel uitwisselen, is voor de drukte-indicator bijvoorbeeld. De indicator die we hebben is een 1'tje, 2'tje, of 3'tje, voor vol, matig, druk. Dat gaat naar een openbaar loket toe, en daar kan iedereen gebruik van maken. En niet allen vervoerders, maar ook andere bedrijven die dat zouden willen. En wij werken wel steeds meer samen met de gemeente Amsterdam. Het is wel mooi om onze informatie aan hen te geven. Ze gaan een pilot doen in de Kalverstraat ook met sensoren en dergelijke om te kijken voor druktespreiding en dat gaan zij dan ook weer aan ons leveren, dat wij daar ook weer iets mee kunnen doen eventueel, richting onze reizigers.

Misschien is het nu nog niet heel erg van toepassing, maar stel je gaat meer data verzamelen, denken jullie dan ook al na over de ethische kant daarvan, en hoe het zit op het gebied van privacy?

Ja, zeker. En bijvoorbeeld over het verzamelen van data via mobiele telefoons van reizigers, daar hebben we wel eens over nagedacht, en ook wel eens over geïnformeerd gekregen. Maar dan bleek dat we daarvan uiteindelijk van slechts 40% van de mensen gebruik konden maken. Daarvan hebben wij toen gezegd, dat is voor ons de informatie eigenlijk van te weinig toegevoegde waarde nu, om te kunnen verantwoorden dat we dit gaan vragen van onze reizigers. Maar zeker nemen we dat meteen mee in onze vraagstukken, van wat we willen. Wij willen heel erg gaan werken vanuit toepassingsgebied, en dan kijken wat hebben we nodig, en dan zullen we denk ik meer tegen dat soort vraagstukken aan gaan lopen. En daar dan ook over gaan bepalen wat we er mee gaan doen. Maar daar hebben we nog niet een heel plan voor liggen of iets dergelijks.

Je zei net dat als je die mobiele telefoons zou gebruiken dat het een te laag percentage zou zijn. Komt dat dan omdat te weinig mensen daar gebruik van zouden maken?

Ja. Omdat dat percentage zo laag lag van mensen die dat zouden willen, ook alles aan hebben staan wat aan moet staan. Al dat soort zaken. Voor ons was het vooral belangrijk om naast de OV chipkaart iets te gaan hebben om inzicht te krijgen in bijvoorbeeld zwartrijden. Dan is dat eigenlijk net een te laag percentage, om dat dan goed in beeld te krijgen. Ik denk dat we daar dan een iets hoger gebruik voor nodig zouden hebben.

Misschien ook lastig, als mensen zwartrijden zullen ze ook niet zo'n app willen gebruiken.

Daar zullen ze niet zo snel aan meewerken inderdaad. Er zullen er misschien een aantal zijn die dat dan per ongeluk wel doen, afhankelijk van hoe je het gaat vragen, maar dan is weer de vraag, wil je dat dan op die manier doen. Hoe transparant wil je zijn? En als je dat heel transparant doet, dan krijg je die zwartrijders waarschijnlijk nog steeds niet in beeld. Dus dat is dan wel een soort van privacy, maar ook een ethisch issue. Dus dat zijn best wel lastige vragen daarin. Wat voor toepassingen zie jij hiervoor?

Ja, goede vraag. Ik denk dat crowdsensing wel een goede manier kan zijn om inzicht te krijgen in hoeveel mensen op een bepaalde plek zijn bijvoorbeeld. Voor zwartrijden bijvoorbeeld denk ik dat dat lastig is, omdat mensen wel gemotiveerd moeten worden om er aan deel te nemen. Wat ik tot nu toe al heb gezien in mijn onderzoek is dat mensen vooral niet zo snel video's en geluidsopnames zullen delen. Voor locatie of bewegingsdata is de bereidheid wel groter.

Is dat op GPS? Want daar zat namelijk nog een tweede issue voor ons. Bij de NS is dit heel goed bruikbaar, want je bent op een bepaalde plek en die is vaak heel ver weg van een ander station. Terwijl bij ons echt alles door elkaar heen loopt. En dat er dan ook nog mensen vlak langs een tram lopen of fietsen, dat dat net niet uit elkaar te houden was. Dus daar zit qua GPS voor stadsvervoer ook nog best wel een issue. Ik denk dat dat voor streekvervoer en treinen iets minder het geval is. Naast een trein loopt niemand, of fietst niemand.

Extra lastig om het dan goed in beeld te krijgen inderdaad. Dan zou je eigenlijk aan de gebruiker moeten vragen, wat zulk soort toepassingen heb je ook, dat de app aan de gebruikers feedback vraagt. En dat vraagt natuurlijk wel iets meer van de gebruiker, dus daar zou je dan iets anders tegenover moeten zetten. In mijn onderzoek kijk ik bijvoorbeeld wat het effect is van een kleine financiële vergoeding per maand, als je ervoor kiest om data te delen.

Oh dat is wel interessant inderdaad. Want als er genoeg gebruik van gemaakt wordt, denk ik wel dat het heel nuttig kan zijn als extra data source. Maar dan moet je er inderdaad wel een soort van meer op kunnen vertrouwen dat je daar voldoende informatie uithaalt. Ik ben wel benieuwd of mensen daar meer toe bereid zouden zijn, met bijvoorbeeld een financiële vergoeding.

Wat denk jij dat ervoor nodig zou zijn om zoiets te implementeren? Stel jullie zouden dit willen implementeren, wat zou je daar op technisch en organisatorisch gebied voor nodig hebben?

Ik denk eerlijk gezegd dat op technisch gebied dingen vrij snel kunnen. Omdat er heel veel van dit soort apparatuur al in voertuigen zit om hiermee te spreken en dergelijke. En GPS natuurlijk sowieso. Dus ik verwacht dat die kant best wel eenvoudig zou moeten zijn. En is dit dus qua technische implementatie een redelijk goedkope oplossing. Je moet het natuurlijk binnen kunnen halen, maar daar hebben we tegenwoordig inmiddels ook een integratieplatform voor staan die die data makkelijk binnenhaalt en die we kunnen bewerken naar onze data lakes toe, om het te kunnen gebruiken. Dus dat is wat je net als bij elke andere data source ook moet doen. We moeten het juridisch goed inrichten. Daar zit denk ik één van de belangrijkste stukken, dat je dat heel goed afdicht, en daar ook goede communicatie over doet naar je reizigers en de mensen die dit willen. Dus op die plek zit echt wel werk. En vervolgens misschien inderdaad een stukje werving van mensen om dat te gaan doen. En dat je erover moet gaan nadenken, van wil je mensen daar iets voor geven, of hoe wil je ermee omgaan? Dus die ook. En dan natuurlijk het laatste stukje van ons hoe gebruiken we het, wat kunnen we ermee, wat zegt dat dan, als we dit zo volgen, dat moet je gewoon leren kennen.

Ik vind dat je sowieso als publieke aanbieder verplicht bent om de risico's heel helder naar iedereen te brengen als je dit soort dingen gaat doen. Dus dat is wel één van de belangrijkste dingen. En ik denk sowieso op het gebied van smart mobility, en alles wat we tegenwoordig meer smart doen, dat communicatie misschien wel één van de belangrijkste dingen is, en transparantie. Naast natuurlijk de juridische vastlegging, maar die spreekt vaak voor zich. Maar de communicatie is denk ik net zo belangrijk.

Jullie zijn bijvoorbeeld al bezig met de druktemeter. Wat voor projecten hebben jullie nog meer voor ogen in de toekomst op het gebied van smart mobility?

Uiteindelijk wil je ook naar zelfsturende voertuigen toe en dergelijke. Op onze metro's is daar al wel een soort van voorbereiding, maar dat systeem werkt nog altijd niet op dit moment. Maar dat systeem is geïmplementeerd, en als dat straks allemaal wel goed werkt dan zou je daar in principe al "bestuurderloos" op moeten kunnen rijden. Die kant wil je stapje voor stapje meer gaan ondernemen. Dat is in de openbaar vervoerswereld misschien op dit moment wel het meest gewenste of vooruitstrevende op het gebied van smart mobility.

Als GVB gaan we ook met MaaS bezig. Eerst zouden we als consortium daar ook een geheel platform voor bouwen, maar daar doen we als GVB niet meer aan mee. Maar we willen wel zeker dat aangaan. En uiteindelijk via onze services, via onze app, daar dingen over laten inzien, over welke mogelijkheden er nog meer zijn. Eventueel al snel met deelfietsen ook iets gaan doen, omdat dat natuurlijk ook heel erg aansluit op stadsvervoer. Dus daar zijn we nu nog heel erg mee bezig, om erover na te denken van hoe gaan we hierop aansluiten. Dus daar gaan we zeker wat mee doen, maar daar hebben we nu nog geen concrete plannen voor. Daar zijn allemaal gesprekken over met ook weer MaaS aanbieders, maar de vraag is hoe wil je daar dan in samenwerken, en op welke manier, en hoe gaan OV daar ook een goede plek in geven. We moeten als vervoerder zorgen dat we een goede plek in het MaaS-landschap krijgen. Ik denk dat dat een grote uitdaging is. Maar hoe en wat, daar kan ik gewoon echt nog niks over zeggen, want zover zijn we nog niet helaas. Het gebied van drukte en spreiding is echt wel een belangrijk thema, zeker met corona is dat natuurlijk veel belangrijker geworden en dat zal een belangrijk thema blijven. Dus daar zijn we op zoek naar hoe kunnen we nog meer, daar willen we echt real-time, daar willen we echt bovenop gaan zitten. Ook heel veel volle voertuigen, dat wordt veroorzaakt doordat er een ander voertuig is uitgevallen, of verlaat is of iets dergelijks, of juist te vroeg weg is gereden. Dat zijn dingen die in voorspelmodellen soms lastiger te vatten zijn. Als je hele ingewikkelde modellen gaat bouwen kun je daar misschien nog iets mee doen, maar daar zijn we nog niet qua kennis en modelbouw en dergelijke. Dus dan wil je eigenlijk gewoon real-time erop zitten, en die informatie doorgeven. Dus wat we binnen hebben, gewichten van voertuigen, maar ook real-time data doorgeven, dat is één van de belangrijkste dingen. De kans bestaat dat we dan in de toekomst op zoek gaan naar andere manieren daarvoor.

En zo proberen we ook wel persoonlijker te worden. Dus richting de reiziger persoonlijkere informatieverstrekking. Misschien ook richting mensen die een beperking hebben, die het ingewikkeld vinden, hoe kun je daar ook juist weer slimmer mee omgaan? En ik denk dat dat deels kan door dingen persoonlijk te maken. Drukte is voor iemand met een beperking misschien wel weer iets anders dan voor een student, die het niet uitmaakt dat het druk is. Op die manieren zijn we er naar op zoek om het voor de reiziger beter te maken.

En slimme voertuigen, ook om servicemedewerkers en bestuurders in staat te stellen om meer te informeren. We willen sneller kunnen reageren op verstoringen. Dat is misschien nog wel een interessante. We hebben inmiddels elektrische bussen die heel veel informatie hebben. Zijn we ook nog aan het ontsluiten, dus ook daar is het nog een soort van beginfase. Maar dat je echt al beter en sneller op verstoringen kan reageren zodat de reiziger daar zo min mogelijk last van heeft. Dus dat je al snel een andere bus kunt laten komen, of ziet dat er iets gaat gebeuren, of dat een elektrische bus niet voldoende elektriciteit meer heeft om te rijden en stil komt te staan. Dat je daar al op in kunt spelen van tevoren, zowel met het oplossen als informatie naar de reiziger. Zoals ik al zei zijn we net begonnen met het digitale programma waarin we een soort road map willen gaan maken van wat zijn nou de thema's en de onderwerpen die we als eerste willen gaan aanpakken op basis van data en nieuwe intelligentie.

I.4. Codes with group codes

Code	Code Group 1	Code Group 2	Code Group 3	Code Group 4	Code Group 5
Benefits of sharing	Challenges of smart mobility				
Citizen involvement	Challenges of smart mobility				
Cooperation by providers			Data collection practices		
Cooperation with other parties			Data collection practices		
Cooperation with transport operators			Data collection practices		
Crowd management		Current smart mobility projects			
Crowd management issues	Challenges of smart mobility				
Crowd sensing	Challenges of smart mobility				
Data collection			Data collection practices		
Data collection for research			Data collection practices		
Data collection principles			Data collection practices		
Data collection techniques			Data collection practices		
Data exchange			Data collection practices		
Data use	Challenges of smart mobility				
Ethical considerations				Ethical aspects of implementing crowd sensing	
Ethical innovation				Ethical aspects of implementing crowd sensing	
Heavy transport		Current smart mobility projects			
Implementation smart mobility projects					Technical aspects of implementing crowd sensing
Incentivizing travellers	Challenges of smart mobility				
Infra-red systems		Current smart mobility projects			
Intrinsic motivation	Challenges of smart mobility				
Legal aspects of implementing crowd sensing				Ethical aspects of implementing crowd sensing	
Onboard Diagnostics		Current smart mobility projects			
Parking		Current smart mobility projects			
Personalised services	Challenges of smart mobility				
Platform development					Technical aspects of implementing crowd sensing
Privacy				Ethical aspects of implementing crowd sensing	
Public transport		Current smart mobility projects			
Responding to incidents	Challenges of smart mobility				

Risks	Challenges of smart mobility				
Smart energy management	Challenges of smart mobility				
Smart mobility management	Challenges of smart mobility				
Smart mobility projects		Current smart mobility projects			
Smartphone applications		Current smart mobility projects			
Technical aspects of implementing crowd sensing					Technical aspects of implementing crowd sensing
Traffic accident information	Challenges of smart mobility				
Transparency in data collection				Ethical aspects of implementing crowd sensing	
Travel information	Challenges of smart mobility				
Trust	Challenges of smart mobility				

Appendix J. Probabilities of choice sets

In Table 41, the probabilities of the “yes” alternative being chosen are calculated for each choice set. For calculating these probabilities, the estimates from model 1 are used. The right column displays the difference in percentage points (pp) between the acceptance rate predicted by the model and the acceptance rate in the sample.

Table 41. Probabilities of choice sets (base model)

Choice set	Probability (yes)	Percentage (sample)	Deviation
1	47%	63%	16 pp
2	26%	24%	2 pp
3	23%	34%	11 pp
4	17%	21%	4 pp
5	89%	89%	0 pp
6	39%	47%	8 pp
7	9%	5%	4 pp
8	38%	45%	7 pp
9	65%	68%	3 pp
10	85%	63%	22 pp
11	52%	45%	7 pp
12	10%	5%	5 pp
13	24%	25%	1 pp
14	22%	13%	10 pp
15	74%	73%	2 pp
16	40%	43%	3 pp
17	39%	38%	2 pp
18	53%	68%	15 pp
19	16%	10%	6 pp
20	28%	20%	8 pp
21	36%	30%	6 pp
22	33%	45%	12 pp
23	13%	15%	2 pp
24	28%	28%	1 pp
25	69%	68%	1 pp
26	13%	15%	2 pp
27	25%	30%	5 pp
28	20%	53%	33 pp
29	41%	38%	3 pp
30	31%	21%	10 pp
31	43%	40%	3 pp
32	41%	40%	1 pp
33	61%	47%	14 pp
34	39%	32%	7 pp
35	49%	47%	2 pp
36	21%	23%	2 pp
		Mean absolute s.d.:	**7 pp**

Appendix K. Acceptance rates for use cases

This appendix elaborates on the acceptance rates for choice scenarios derived from the use cases that were based on the conducted interviews. Table 42 presents the acceptance rates for the scenarios defined for the crowd management use case.

Table 42. Acceptance rates for crowd management in a city scenarios

Crowd management in a city (Interview A)					
Factor	**Scenario 1.1 (base)**	**Scenario 1.2**	**Scenario 1.3**	**Scenario 1.4**	**Scenario 1.5**
Monetary reward	€60 per month	None	€60 per month	€60 per month	€60 per month
Effort	Low	Low	Low	Low	Low
Risk of re-identification	10%	10%	30%	20%	10%
Types of data	Location data, motion data	Location data, motion data	Location data, motion data	Location data, motion data	Location data, motion data
Data use	Governmental institution aiming to improve mobility	Governmental institution aiming to improve mobility	Governmental institution aiming to improve mobility	Governmental institution aiming to improve mobility	Academic institution investigating transport modes
Acceptance	**80%**	**60%**	**62%**	**72%**	**84%**

In Table 43, the acceptance rates for scenarios relating to safety research are displayed.

Table 43. Acceptance rates for safety research using car accident information scenarios

Safety research using car accident information					
Factor	**Scenario 2.1 (base)**	**Scenario 2.2**	**Scenario 2.3**	**Scenario 2.4**	**Scenario 2.5**
Monetary reward	€60 per month	€60 per month	€60 per month	€60 per month	€60 per month
Effort	Low	Low	Low	Moderate	High
Risk of re-identification	10%	30%	10%	10%	10%
Types of data	Location data, motion data, contextual data, multimedia data	Location data, motion data, contextual data, multimedia data	Location data, motion data, contextual data	Location data, motion data, contextual data	Location data, motion data, contextual data
Data use	Corporate institution aiming to improve products or services	Corporate institution aiming to improve products or services	Corporate institution aiming to improve products or services	Corporate institution aiming to improve products or services	Corporate institution aiming to improve products or services
Acceptance	**49%**	**29%**	**58%**	**41%**	**26%**

Finally, Table 44 presents the acceptance rates for choice scenarios derived from the public transport use case.

Table 44. Acceptance rates for real-time travel information in public transport scenarios

Real-time travel information in public transport					
Factor	**Scenario 3.1 (base)**	**Scenario 3.2**	**Scenario 3.3**	**Scenario 3.4**	**Scenario 3.5**
Monetary reward	None	€20 per month	€60 per month	€60 per month	€60 per month
Effort	Moderate	Moderate	Moderate	Low	Low
Risk of re-identification	10%	10%	10%	10%	10%
Types of data	Location data, motion data	Location data, motion data	Location data, motion data	Location data, motion data, contextual data	Location data, motion data
Data use	Governmental institution aiming to improve mobility	Governmental institution aiming to improve mobility	Governmental institution aiming to improve mobility	Governmental institution aiming to improve mobility	Governmental institution aiming to improve mobility
Acceptance	**43%**	**51%**	**66%**	**73%**	**80%**