Road-user participation in vehicle-data sharing systems

for the purpose of Dynamic Traffic Management

A.M. de Jong

A stated preference experiment to provide insight into the factors influencing road-user participation in vehicle-data sharing systems for the purpose of improved Dynamic Traffic Management

TU Delft

Road-user participation in vehicle-data sharing systems

for the purpose of Dynamic Traffic Management

by

A.M. de Jong

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on July 15, 2019 at 13:00 AM.

Student number:	4320506	
Project duration:	march 19, 2018 – July 15, 2019	
Thesis committee:	Prof. dr. ir. C. Chorus,	TU Delft, Chair
	Dr. E. J. E. Molin,	TU Delft, Supervisor
	H. Asghari,	TU Delft
	P. B. C. van Erp,	TU Delft

TU Delft

Preface

This thesis represents the final step for the completion of my Master Complex Systems Engineering and Management at the Delft University of Technology. With this thesis I aim to provide a meaningful scientific contribution regarding insight into choice behaviour of road-users when choosing to participate in a vehicle-data sharing system. My hope is that this thesis will help policy makers in the development of such vehicle-data sharing systems.

First, I would like to express my gratitude to my graduation committee for their support and patience during the course of this project. I would like to thank Associate Professor Eric Molin, my first supervisor, for his general support in both practical and theoretical matters, as well as the guidance offered during this period. I would like to thank Paul van Erp, my daily supervisor, for helping define the topic and direction of this thesis, and his help with more practical aspects. I would like to thank Assistant Professor Hadi Asghari for his insights into aspects related to privacy, and Professor Caspar Chorus for chairing my graduation committee.

Secondly I would like to thank my friends and family for their support over the course of my graduation. Special thanks to my parents for their patience, listening ears,and helping me during this period.

I hope you all enjoy reading the result,

Delft, Juli 2019,

Alexander Mark de Jong

Management Summary

Research problem

With the increasing amount of traffic on the Dutch road network problems relating to congestion and travel-time lost are increasing. The implementation of improved dynamic traffic management (DTM) measures may lead to an improvement in road-network performance. The input for DTM are traffic state estimations (TSE) relating to traffic flow, traffic densities, and traffic speed. Currently data is gathered using static detection methods such as loop detectors and camera's. Combining this stationary data with data gathered from moving observers is valuable for improving TSE. Collecting this data directly from road-users instead of third parties, may lead to a reduction in cost and an increase in data resolution.

The gathering of vehicle-based sensing data (VBSD) directly from road users is complex due to possible legal issues, privacy concerns and uncertainty regarding people's willingness to share their vehicle-data with the government. As such the objective of this research is *to further develop understanding of the effects of privacy factors and incentives on road-user participation in sharing their Vehicle-Based Sensing Data for the purposes of Dynamic Traffic Management.* In line with this objective the following research question is formulated:

RQ:*How do factors relating to privacy and incentives affect road-users participation in a vehicle-data sharing system for the purposes of Dynamic Traffic Management?*

Definition of privacy

In this study privacy has been defined as *an individual's' right to control the collection, access to and uses of information relating to places, bodies, and personal data.* Factors relating to privacy relevant to this specific study are nested in one of three privacy types, these being, privacy of location an space, privacy of location and action, and privacy of data and image. The definitions of these privacy types are; Privacy of location and space: *"The right to move about in public or semi-public space without being identified, tracked, or monitored."*, privacy of behaviour and action: *"The ability to behave in public, semi-public or one's private space without having actions monitored or controlled by others."*, and privacy of data and image *"Concerns about making sure that individuals' data is not automatically available to other individuals and organizations and that people can exercise a substantial degree of control over that data and its use."*

Process

Through literature review and structured interviews the factors related to privacy and incentives that may influence user participation are identified. These factors were operationalized for use in the stated preference experiment from which a logit model was estimated. Furthermore several socio-demographic factors were included in order to test their influences and improve the model fit. Furthermore it is tested whether previous exposure to alternatives offering monetary reward influences the participation levels when no compensation is offered.

For the stated preference experiment a survey was designed and a total of 98 valid responses were obtained. The sampling is a convenience sample and does include bias towards academically educated males in the age-group 18 - 35. The data was collected though online surveys and physical surveys administered at the Delft University of Technology, the Hague University, a local library, and public spaces. The physical surveys were obtained using a tablet.

Findings and Conclusions

Table 1 shows the willingness-to-accept values for the attributes relating to the different privacy constructs. These values represent the level of monetary compensation required for a person to accept the privacy harm associated with the attribute. The main factors influencing participation in a vehicle-data sharing system are related to privacy of data and image, specifically the sharing of data with third parties. Any choice situation in which data was shared with third parties resulted in a minority level of participation. The negative effect of sharing data with third parties is more pronounced with age. The sharing of data with emergency services and researchers was valued in a positive manner and increases levels of user participation. Age positively influences the willingness to share data for research purposes. This positive valuation would seem to imply that people are generally willing to participate in vehicle-data sharing systems where data is shared with emergency services and for the purpose of transport research, leading to improved emergency responses and opening possibilities for using fine-grained datasets for improving TSE and subsequently DTM measures.

Table 1: Willingness-to-Accept values relating to privacy attributes

Privacy Construct	Attribute	WtA (Euro's)
Location and space	Trip registration (User ID)	33.55
Behaviour and Action	Vehicles passed	1.95
	Accelerometer	12.67
Data and image	Sharing Emergency	-15.69
	Sharing Research	-19.26
	Sharing Third	124.14
	On-/off function	-13.06

With regards to incentives it was found that a lack of exposure to alternatives offering monetary reward had higher participation levels compared to when people were previously shown alternatives offering monetary compensation. This would suggest that participation levels in the initial deployment of a vehicle-data sharing system would benefit from not offering monetary compensation. The offering of monetary compensation was confirmed to increase levels of user participation. However the effect of monetary compensation is reduced for individuals who have high levels of privacy concern. The communication of social benefits was found to influence the level of participation, but the effectiveness is highly dependent on the age, sex, and education level of the recipient. With young, female, academically educated individuals being most sensitive to the communication.

The indicators relating to privacy concern and institutional trust were confirmed to measure their respective attitudes. With higher levels of privacy concern lowering levels of participation and higher levels of institutional trust increasing levels of participation. It would seen that privacy concern of individuals could be compensated for by increasing levels of institutional trust.

Other observations regarding socio-demographic factors indicate that higher levels of education reduce participation, high levels of yearly kilometers driven lower participation, and hatchback drivers have higher levels of participation.

In conclusion it can be stated road-users do indeed seem willing to participate in sharing their vehicle-data for the purposes of DTM, given the data is not shared with third parties. When the data collected from participants is sufficiently parsimonious it would seem that offering monetary compensation is not required. This in turn allows for the gathering of fine-grained data from moving observers at a low cost. In addition the general willingness to share data with emergency services and for transport research may in turn improve road conditions due to improved emergency responses and advancements in the field of transport research.

Contents

1 Introduction **1**
1.1 Knowledge gap: Effects of privacy factors and incentives on data-sharing 2
1.2 Research Objective 4
1.3 Research Questions 4
1.4 Methods. 5
1.4.1 Scoping and Classification of the SP experiment 5
1.4.2 Survey Design 5
1.4.3 Acceptable data-privacy conditions and compensation schemes 6

2 Literature review **7**
2.1 Method 7
2.2 Defining Privacy 8
2.2.1 Normative, descriptive, and reductionist perspectives 8
2.2.2 Control- and Use-Based definition of privacy rights. 9
2.2.3 Privacy Rights and Property Rights 9
2.2.4 Defining privacy in the context of Vehicle-based sensor data 9
2.3 Legal aspects of privacy 10
2.3.1 Principles 10
2.3.2 Consent 11
2.3.3 Information provided to data subject. 12
2.3.4 Rights of the data subject 12
2.4 Domain specific privacy and participation factors from a user perspective 13
2.4.1 Privacy concerns of users 14
2.4.2 Factors affecting user participation 16
2.5 Comparison with exploratory interview results 18
2.6 Summary and implications for experiment design 19
2.7 Conceptual model 20

3 Survey design and Analysis Methods **21**
3.1 Survey Design 21
3.1.1 Attributes and levels 22
3.1.2 Experimental design 24
3.1.3 Context Variables. 24
3.1.4 Socio-demographic Factors 26
3.2 Analysis 27
3.2.1 Logistic Regression 27
3.2.2 Willingness-to-Accept 28
3.2.3 Importance of attributes 28
3.2.4 T-test for order of alternatives 28
3.3 Attitudes relating to Trust and Privacy Concern. 29
3.3.1 Factor Analysis 29

4 Survey development and deployment process **31**
4.1 Target Audience 31
4.2 Survey development 31
4.2.1 Results of test questionnaire. 31
4.2.2 Survey pretest 32
4.3 Survey Deployment 32
4.3.1 Data collection procedure 32
4.3.2 Respondents and representativity 32

5 Results and Analysis **35**
5.1 Exploratory analysis . . . 35
5.1.1 Responses per choice situation . . . 35
5.1.2 Order effects . . . 36
5.1.3 Confirmatory factor analysis of Attitudes . . . 36
5.2 Logit model estimation . . . 37
5.2.1 Base Model . . . 37
5.2.2 Expanded Model . . . 39
5.2.3 Comparison of model fit between base model and expanded model . . . 42
5.2.4 Predicted behaviour . . . 42
5.2.5 Limitations . . . 44
5.3 Willingness-to-Accept . . . 44
5.4 Comparison of results to formulated hypothesis . . . 44

6 Discussion and conclusion **49**
6.1 Main findings . . . 49
6.2 Contributions . . . 54
6.2.1 Theoretical Contributions. . . . 54
6.2.2 Managerial Contributions. . . . 54
6.3 Limitations and further research . . . 55

Bibliography **57**

A Interviews **61**

B Slides of survey introduction film **67**

C Online survey **73**

D Test survey experimental design **85**

E Test survey SPSS output **87**

F Final survey experimental design **89**

G Base model SPSS output **91**

H Expanded model SPSS output **95**

I Factor analysis SPSS output **101**

J Scientific Article **107**

1 Introduction

The road network of the Netherlands is increasingly busy, with the amount of daily vehicles on the roads increasing with 16% between 2005 and 2016, and an increase of 3% over 2016 alone [van Infrastructuur en Waterstaat, 2017]. This observed growth on the already densely used road-network is leading to problems with congestion, with the travel-time lost increasing with 10% in 2016, compared to 2015 [van Infrastructuur en Waterstaat, 2017]. Congestion on the road network is increasing and so are the economic losses that resulting from the total amount of travel-time lost. Excessive transport times negatively impact the mobility of citizens and the attractiveness of the Netherlands for international freight transport. The economic damage for companies over 2015 as a result of congestion in cargo transport is estimated to be between 857M and 1,1 billion Euro, while in 2014 this was estimated to be between 655M to 852M Euro [EVO, 2016].

The implementation of improved Dynamic Traffic Management (DTM) measures may prevent congestion before it occurs, leading to an improvement in road-network performance. DTM may refer to measures regarding flow control, or dynamic routing. Flow control relates to measures in which the flow of traffic is managed through measures such as; variable speed limits which increase road safety and avoid congestion due to lower influx vehicles, and ramp metering which controls the amount of vehicles entering a road with the usage of traffic lights [Papageorgiou et al., 1991]. Dynamic routing refers to measures that influence the route a road-user takes based on the current road conditions, enabling routing around parts of the road-network which are congested or inaccessible. Routing information can be presented to road users using either text-based dynamic routing information panels (DRIPS), or graphical routing information panels (GRIPS). Combining both flow control and dynamic routing into a comprehensive system is integrated network management which is a traffic management approach that includes both traffic management and traffic information measures integrated and managed within a transport network [Ebner, 2017]. The input for DTM are estimations of traffic flow, traffic densities, and traffic speed [Seo et al., 2015].

Currently the data used for traffic state estimations (TSE) are mainly collected using stationary measuring instruments such as loop-detectors or video based technologies[Buch et al., 2011], however the placement of sensors across the road network is financially prohibitive due to the high costs associated with these stationary detectors and thus may not cover the entirety of the network. Another option is using Floating-Car Data (FCD), this is usually based on cellular and Satellite Navigation (SatNav) data [Leduc, 2008]. FCD is usually comprised of basic vehicle telemetry such as speed, direction and, most importantly, the position of the vehicle [Schäfer et al., 2002]. Vehicle-based sensing data (VBSD) refers to data generated by one vehicle as a sample to assess the overall traffic condition, this is an extension of the definition of floating car data (FCD). VBSD broadens this definition by including the data gathered by on-board sensors, e.g. the amount of cars passed left and right. Often-times datasets containing VBSD are purchased from private parties at a significant cost, these purchased datasets often have a high level of data-aggregation due to privacy aspects relating to the General Data Protection Regulation (GDPR) and fears over public outcry, such as when TomTom sold data to the Dutch government which included speeding data [Palmer, 2011].

The future development of the field of TSE has many exciting possibilities due to advances in methods and the increasing availability of mobile disaggregated traffic data generated by sensors integrated in the vehicles. Vlahogianni [Vlahogianni et al., 2014] states one of the challenges for the future to be "Using new technologies for collecting and fusing data". The use of traditional on-road sensors such as inductive loops is necessary, but is not always sufficient for the development of intelligent transport systems (ITS) [Leduc, 2008]. In order to better estimate the traffic state, relying on a combination of both stationary and moving observers is valuable [van Erp et al., 2019]. In particular the sampling resolution is important for estimation performance when large changes in traffic conditions occur [van Erp et al., 2018]. Collecting the data directly from road-users instead of third parties, may lead to a reduction in cost and an increase in data resolution. For the use of VBSD to provide meaningful datasets the penetration of these technologies is key. Herrera et al. [2010] suggested that a 2-3% market penetration of cell phones in the driver population is enough to provide accurate traffic measurements based on cellphone GPS data. The advent of autonomous vehicles and vehicles with more on-board sensors provides a large potential market penetration for VBSD, and access to new types of data. As such, the possibility of gathering VBSD directly from road users becomes interesting, albeit more complex due to possible legal issues, privacy concerns and uncertainty regarding people's willingness to share the data with the government.

1.1. Knowledge gap: Effects of privacy factors and incentives on data-sharing

The main gap that this research aims to address is the lack of understanding how factors relating to privacy concerns and potential compensation influence a road-users choice to participate in sharing their VBSD for the purposes of DTM. Privacy considerations of road-users will determine if the proposed gathering of VBSD can be successful regarding user participation. As such this subject will be a focal point in this study. In the following section a brief overview of the current regulation and privacy literature is presented.

EU Regulation

According to the EU General Data Protection Regulation (GDPR) [EU, 2016] citizens have the right to: information of the processing of their personal data, access to personal data, ask for correction of inaccurate or incomplete personal data, request that personal data be erased when it's no longer needed or if processing it is unlawful, object to the processing of personal data for marketing purposes, request the restriction of the processing of personal data in specific cases, receive their personal data and send it to another controller, request that decisions based on automatic processing concerning them or significantly affecting them are made by natural persons, not only by computers.

in particular several parts of the regulation are relevant in the context of this research proposal: art.1(32) states that consent must be given explicitly, (39) processing data should be lawful and fair, and (50) data can be for other purposes as long as this is within the intial scope of consent. Furthermore, art.1(65) states that "*...the further retention of the personal data should be lawful where it is necessary,..., for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller, ..., for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, or for the establishment, exercise or defence of legal claims.*"

Privacy considerations

As can be seen in the previous section the road-user needs to explicitly provide consent for the use of their data. There are two general types of transactions for providing consent, an information transaction or an composite transaction [Jentzsch, 2016]. With an information transaction, only information is exchanged, either incentivized with money or a social exchange, where the incentive is reciprocity. Examples are Google, Facebook and Twitter. In composite transactions the main exchange is that of a good or service. Within this transaction is an implicit information transaction that occurs in parallel. Examples of this are the online purchase of goods, and online banking and insurance transactions. When individuals decide whether to disclose personal data, they often do not consider the externalities this sharing produces [Jentzsch, 2016]. These externalities can positively or negatively affect the welfare of other individuals. in the context of VBSD gathering from road-users, the positive externalities represent the improvements in road network performance, while negative externalities could represent many different factors, for example the use of data to decide where to place speed cameras, as was the case by TomTom[1].

Using the information gathered for purposes such as the placement of speed cameras leads to increased identification and surveillance. This reduction in anonymity can lead to more pro-social behaviour [Bohnet and Frey, 1999] [Charness and Gneezy, 2008]. Similarly, surveillance introduces the feeling of being watched, leading to more pro-social behaviour [Nettle et al., 2013]. While more identification may lead to more pro-social behaviour it may lead to a reduction in acceptance [Jentzsch, 2016] and subsequently a reduced number of people willing to share their VBSD.

Due to this behaviour, the use of differential privacy seems promising. Differential privacy can be defined as works that take into account that: "*agents desire that not much of their information be revealed to any other agent via participation in (a) mechanism*" [Kearns et al., 2014]. A mechanism (or algorithm) is differentially private, if its output is insensitive to the change of a single input. In other words, an algorithm is said to be differentially private if by looking at the output, one cannot tell whether any individual's data was included in the original dataset or not. Differential privacy is a property of the algorithm and there is a trade off between privacy and the accuracy of statistics computed from the acquired data [Chen et al., 2016].

Eight strategies for privacy design have been derived [Hoepman, 2014], these are minimize, hide, separate, aggregate, inform, control, enforce, and demonstrate. Minimize states that the amount of personal data that is processed should be restricted to the minimal amount possible. Hide states that any personal data, and their interrelationships, should be hidden from plain view. Separate states that personal data should be processed in a distributed fashion, in separate compartments whenever possible. Aggregate states that personal data should be processed at the highest level of aggregation and with the least possible detail in which it is (still) useful. Inform states that data subjects should be adequately informed whenever personal data is processed. Control states that data subjects should be provided agency over the processing of their personal data. Enforce states that a privacy policy compatible with legal requirements should be in place and should be enforced. Finally, demonstrate states that it should be possible to demonstrate compliance with the privacy policy and any applicable legal requirements. These strategies may form a basis on which to base potential privacy configurations for the gathering and use of VBSD.

Incentives

Another factor affecting willingness to share VBSD is the inclusion of economic incentives, which introduce a monetary or other type of cash-value reward in exchange for sharing VBSD. These incentives activate self-interested behaviour and may undermine pre-existing intrinsic motivation [Grant, 2011]. This is a contradiction to the current paradigm in choice models that everyone has a price. Individuals are often characterized into one of three groups regarding privacy concerns; fundamentalist, pragmatist, or unconcerned [Kumaraguru and Cranor, 2005]. It could be argued that pragmatists are potentially more receptive to economic incentives than the fundamentalists and unconcerned. Ethical problems arise with incentivization. It can be argued that, for example, offering a high level of compensation to a poor person is coercive and has little to do with free choice. As such incentives may crowd-out other types of social exchange of personal data, other motivations of disclosure, and lead to pre-selection effects [Jentzsch, 2016].

[1]https://www.emerce.nl/nieuws/tomtom-verkoopt-gebruikersdata-door-politie

1.2. Research Objective

Based on the knowledge gap the research objective is formulated. The research objective functions as the input for the formulated research questions.

The objective of this research is *to further develop understanding of the effects of privacy factors and incentives on road-user participation in sharing their Vehicle-Based Sensing Data for the purposes of Dynamic Traffic Management.*

The results of this research contribute to the understanding of how policy makers can successfully implement a VBSD sharing system in order to improve data resolution and enable better TSE leading to more effective DTM measures.

1.3. Research Questions

In order to realize the stated research objective this section provides the research questions this study will answer. The following main research question (RQ) is formulated:

RQ:*How do factors relating to privacy and incentives affect road-users participation in a vehicle-data sharing system for the purposes of Dynamic Traffic Management?*

In order to answer the RQ, multiple sub-questions are formulated to provide answers to both the theoretical and practical aspects of this study. The main method in answering the research question is choice modelling using data gathered using stated preference (SP) experiments, in which respondents may choose to share their data or not. This method will make the trade-offs made by road-users explicit regarding data sharing, privacy, and compensation. In order to conduct a choice experiment which yields useful results the correct variables and classification levels need to be determined beforehand. As has been touched upon in previous sections, the main considerations regarding the feasibility of road-user data sharing are privacy concerns and compensation levels/schemes. From these considerations the following sub-questions are formulated to provide guidance in developing the choice experiment.

SQ1 *How do road-users perceive privacy regarding sharing their vehicle-based sensing data?*

SQ2 *What are the factors contributing to road-user participation in sharing vehicle-based sensing data?*

Through literature review and semi-structured interviews with and road-users the range of potential privacy configurations and compensation schemes can be narrowed down to a set that can be used for the construction of the SP experiment. After conducting the SP experiment the data is analyzed and information regarding the participation levels for different levels of privacy and compensation level/schemes. The sub-questions relevant to these issues are:

SQ3 *Which privacy configurations for sharing vehicle-based sensing data are acceptable for road-users?*

SQ4 *Which compensation schemes for sharing vehicle-based sensing data are acceptable for road-users?*

SQ5 *How do socio-demographic factors influence the acceptance of sharing road-vehicle data?*

The use of SP experiments is commonplace when determining willingness-to-pay, both outside the transport domain [Grabicki and Menges, 2017] [Schoot et al., 2017], as well as within [Yoon et al., 2017] [Lyu, 2017]. The method is strong in determining quantifiable relations between criteria from the choices respondents make, thus making it possible to quantify the value road-users place on their VBSD. The method also enables the exploration of yet unrealized alternatives, enabling decision makers to make informed decisions on potential future policy and services.

1.4. Methods

The following section outlines the flow of the proposed research as well as the different methods and data sources used to answer the research questions. Elaboration on the chosen qualitative and quantitative methods is ordered sequentially.

1.4.1. Scoping and Classification of the SP experiment

Before the SP experiment can be designed, the different classification levels to be included for privacy, compensation level, and compensation method are determined. This is done by answering SQ1 and SQ2.

The perception of privacy and factors contributing to participation in data-sharing systems are determined on the basis of literature review and exploratory interviews with both experts and road-users. Through the means of literature review a substantial amount of knowledge in the research area can be accrued and the state of knowledge can be determined [Wee and Banister, 2016]. Following the literature review several semi-structured interviews with experts and road-users are conducted the opportunity to answer questions arising from the literature review and compare statements between them. The use of semi-structured interviews as a means of data collection is well suited to the exploration of attitudes, values, beliefs and motives [Barriball, 1994]. By comparing statements from literature and interviews to each other the validity of chosen classification levels is safeguarded, this is referred to as triangulation [Jick, 1979].

1.4.2. Survey Design

After defining the classification levels to be included in the SP experiment the experimental setup is constructed. Chapter 3 will clarify on all methodological choices made to construct the survey. From the results of chapter 1 and 2 the different attribute levels are selected. Besides the attribute levels several socio-demographic factors are included. Attitude variables regarding a respondents importance of privacy and their Trust in the government are also included. The use of SP choice experiments enables the researcher to create hypothetical situation that do not exist, allowing for evaluation of potential future possibilities.

The use of these hypothetical choice sets allow the researcher to value attributes based on the choices made, in this case the trade-off between privacy harm and compensation. In this design the effect of cognitive limitations and fatigue of respondents needs to be taken into account, failure to do so may increase the measurement error [Johnson et al., 2013]. Measurement error may also come from respondent's heterogeneous interpretation of the alternatives or from inattention due to the hypothetical nature of the study. Thus clear descriptions that allow for unequivocal interpretation are required. The use of dominating alternatives in choice sets should be avoided to avoid biased results [Bliemer et al., 2017], through the use of binary choice sets resulting in either participation or non-participation the presence of dominant alternatives is circumvented.

In the process to develop the final survey several steps are taken to ensure the quality and validity. First a test version was be distributed to gather data in order to test the survey with regards to realism, length and clarity. Together with developing the test survey, expert validation of the survey has taken place, after which the final questionnaire was determined. Only surveys that have been filled in completely will be used in further analysis in order to avoid any potential bias stemming from imputing data. The process of gathering the data will be done using an online survey.

Important challenges to SP experiments exist, [Schläpfer, 2017] states these to be: cognitive limitations of the respondents, issues with incentives leading to strategic answers, dominant alternatives in choice sets [Bliemer et al., 2017], and hypothetical bias [Loomis, 2011] [Murphy et al., 2005]. Another potential issue is the fact that, for some respondents with a high value of privacy, the dominant strategy is to maximize payout and cushion privacy by disclosing manipulated information [Jentzsch, 2016].

1.4.3. Acceptable data-privacy conditions and compensation schemes

Analysis of the dataset by means of logistical regression provides the data required to answer SQ3, SQ4, and SQ5.

Answering which levels of data-privacy, methods of compensation, and the level of compensation are acceptable to the population is be conducted using logit models estimated using logistical regression, this method has been employed commonly in the past [Seo et al., 2017]. This method allows for determining the (Utility) coefficients of the different attributes, as well as the main effects and interaction effects of socio-economic factors.

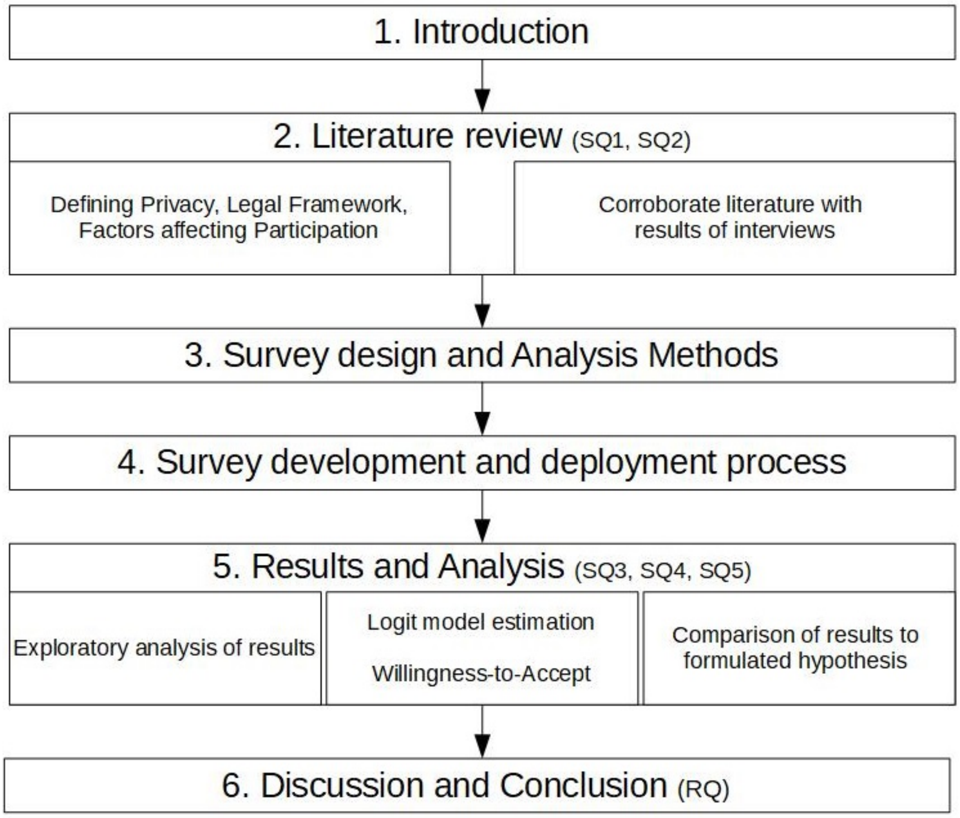


Figure 1.1: Thesis flowchart

2

Literature review

This thesis aims to answer the question: *How do factors relating to privacy and incentives affect road-users participation in a vehicle-data sharing system for the purposes of Dynamic Traffic Management?* The context for which is the voluntary collection of vehicle-based sensing data from road-users for the purpose of DTM. This chapter aims to answer the following research questions in order to provide a picture of the different aspects relevant to the design of the stated-choice experiment.

SQ1 *How do road-users perceive privacy regarding sharing their vehicle-based sensing data?*

SQ2 *What are the factors contributing to road-user participation in sharing vehicle-based sensing data?*

In order to provide founded answers to these questions the chapter is divided into several parts. First a literature review is conducted on the nature of privacy in order to arrive at a definition that fits with the current research. As many different views on this definition exist it is important to arrive at a definition that is unequivocal and encompasses all aspects relevant to the research. Secondly the legal framework within which any data collection will take place is explored. Knowledge of the legal framework is essential in understanding what the design space of any vehicle-based sensing data collection system is. Understanding this legal framework will aid in creating an experiment which simulates a real life situation as close as possible. Thirdly an overview of domain specific factors from a users perspective is provided regarding privacy issues and factors affecting participation (including incentives) in a vehicle-based sensing data collection system. Next, the results from the literature review are compared to the results from six exploratory interviews. Following the results of the interviews the results are summarized and implications for the experimental design are presented. Finally the conceptual model is presented.

2.1. Method

The literature used in this chapter was found using both web of knowledge and Google scholar. Using a combination of different keywords including (but not limited to): privacy, floating car data, connected vehicles, vehicle-based sensing data. Furthermore, both forward and backwards snowballing was employed as the amount of relevant literature from the searches alone was not sufficient. As the literature on privacy in the context of vehicle-based sensing data sharing is limited literature regarding connected vehicles and web-based services has been included with regards to privacy.

The six interviews are conducted using a fixed set of questions, the results of which can be found in Appendix A. The structured interviews are used to provide more insight into the related issues. The interview questions are based on the results of the literature review and concern the sharing of data with different parties, the use of (monetary) incentives, and motivations for sharing. The interviews were recorded and transcribed. The results of these interviews are then compared to the results of the literature review.

2.2. Defining Privacy

Definitions of privacy can be nested in descriptive or normative terms [Moore, 2008], where it can be viewed as a condition or moral claim on others to refrain from certain activities. Others view privacy as a derivative notion that rests upon rights of property and liberty. The following brief summary of different perspectives indicates the breadth of the different definitions that have been proposed.

Privacy has been defined in a myriad of ways over the last few centuries [Morton, 1998]. With it being defined as ”the right to be left alone” by Warren and Brandeis[Brandeis and Warren, 1890]. The legal scholar William Prosser separating privacy into four distinct torts. “Intrusion: Intruding (physically or otherwise) upon the solitude of another in a highly offensive manner. Private facts: Publicizing highly offensive private information about someone which is not of legitimate concern to the public. False light: Publicizing a highly offensive and false impression of another. Appropriation: Using another’s name or likeness for some advantage without the other’s consent.” [Prosser, 1960] Another perspective is the description of privacy in terms of information control, as described by Alan Westin and others [Westin, 1968]. William Parent advocated that ”Privacy is the condition of not having undocumented personal knowledge about one possessed by others.”[Parent, 1983] Yet another perspective is taken by Julie Inness is that privacy is defined as ”the state of possessing control over a realm of intimate decisions, which include decisions about intimate access, intimate information, and intimate actions.” A more recent definition is by Judith Wagner DeCew, stating the ”realm of the private to be whatever types of information and activities are not, according to a reasonable person in normal circumstances, the legitimate concern of others.” [DeCew, 1997]

2.2.1. Normative, descriptive, and reductionist perspectives

In defining privacy, two relevant distinctions have been widely discussed. The first is the distinction between descriptive and normative concepts of privacy. The second being a reductionist and non-reductionist perspective on privacy.

The distinction between normative and descriptive pertains to whether a definition of privacy: describes a state or condition where privacy is obtained (descriptive), or implies moral obligations or claims (normative).

The distinction between reductionist and non-reductionist views on privacy pertains to whether privacy is derived from other rights (reductionist) or whether it is a distinct right itself (non-reductionist). Reductionists argue that privacy is derived from other rights such as life, liberty, and property rights. From this perspective there is no overarching concept, but rather an amalgamation of several different core concepts. For example, Frederick Davis argued from a reductionist perspective that ”If truly fundamental interests are accorded the protection they deserve, no need to champion a right to privacy arises. Invasion of privacy is, in reality, a complex of more fundamental wrongs. Similarly, the individual’s interest in privacy itself, however real, is derivative and a state better vouchsafed by protecting more immediate rights.” [Davis, 1959] In contrast, the non-reductionist views privacy as related to, yet separate from other rights or moral concepts.

Moore [Moore, 2008] states that the distinction between normative and non-normative is important and crucial for conceptual coherence and that it is proper to define privacy along normative and descriptive lines. Assuming a normative definition of privacy, without consideration for the justification of rights involved it is not clear if privacy is reducible to other rights, or if other rights can be reduced to privacy rights [Parent, 1983]. Given the nature of privacy, it is not surprising there are close ties between it and notions such as liberty and self-ownership rights.

2.2.2. Control- and Use-Based definition of privacy rights

In the context of a control and use-based definition of privacy rights Moore [Moore, 2008] defines a privacy right as an access control right over oneself and to information about oneself. Privacy rights include an control or use feature, meaning that privacy rights allow exclusive use and control over personal information and specific bodies or locations. The term ”control” can be given a normative or descriptive representation. A descriptive representation of control would probably relate to the power to physically manipulate an object or intangible good. A normative representation of control would relate to moral claims that should hold independent of the condition. The use of descriptive control-based definitions of privacy has been attacked by William Parent, who argues:

”*All of these definitions should be jettisoned. To see why, consider the example of a person who voluntarily divulges all sorts of intimate, personal, and undocumented information about herself to a friend. She is doubtless exercising control...But we would not and should not say that in doing so she is preserving or protecting her privacy. On the contrary, she is voluntarily relinquishing much of her privacy. People can and do choose to give up privacy for many reasons. An adequate conception of privacy must allow for this fact. Control definitions do not.*” [Parent, 1983]

In this case, by yielding control to others the condition of privacy is diminished. However, yielding control over access doesn’t mean that control over use is automatically provided. This means that descriptive views on access and use are not invalidated by Parent’s arguments. Parent further states that those who defend a control definition of privacy may be worried about a right to privacy rather than the condition of privacy. The definition of privacy given by Parent is: ”*Privacy is the condition of not having undocumented personal knowledge about one possessed by others.*”[Parent, 1983]

Parent’s view of privacy ignores the physical or locational aspects of privacy. An example of this would be the following situation: A person with severe amnesia wanders into your room when you are sleeping and pets you on the head. This is completely separate from having (un)documented information and many would argue that that this is a violation of privacy. Given that no information is involved it would fall outside the definition given by Parent. Furthermore the definition given does not include a use dimension. This use dimension pertains to the fact that while other parties may have the means to invade your privacy, wholly outside of your control, it does not mean that your privacy has been breached until the means are employed for this purpose. An example of this could be neighbors owning a sensitive listening device capable of picking up conversations in your house. In this situation, a control based definition of privacy no longer holds, yet until your neighbors actually use the device to listen to your conversations no privacy rights have been violated. Conditional privacy is present when others do not have access, while a right to privacy enables control over access and use.

2.2.3. Privacy Rights and Property Rights

Property rights and privacy rights both control access, thus privacy rights may be a special form of property rights [Moore, 2008]. Thompson, as a reductionist, agrees that privacy rights are a special form of property rights, stating that: ”... the right to privacy is itself a cluster of rights, and it is not a distinct cluster of rights but itself intersects with...the cluster of rights which owning property consists in.” [Thomson, 1975] Property rights come in several forms, such as intellectual property, characterized as non-physical property, for which the rights-holder has control over physical manifestations. In this context, privacy can be viewed as a right to control access to locations and ideas, independent of context. Coupled with this, property rights can be construed as being relevant in determining the boundaries within which privacy rights are applicable. Understanding privacy as having accessibility and control over use dimensions, it is logical that there is overlap with property rights.

2.2.4. Defining privacy in the context of Vehicle-based sensor data

Different perspectives have been shown on what the definition of privacy is. For the purposes of this research it is important to have a definition of privacy explicitly stated so as to avoid the myriad of different interpretations that are associated with it.

Three definitions that are relevant to this research are privacy defined as; (1) the individual's ability to control the collection and use of personal information [Westin, 1968] [Hann et al., 2001]; (2) a right to control access to and uses of-places, bodies, and personal information [Moore, 2008]; and (3) the desire individuals have in sustaining a 'personal space' free from interference by other people and organizations [Derikx et al., 2016][Clarke, 1999].

Together, these three definitions encompass the range perspectives required to properly define privacy in the context of users sharing vehicle-based sensing data. Following definitions (1) and (2), any situation in which a person freely discloses information satisfies the conditions of a right to privacy, if the information is only used for the scope for which consent was given. Definition (3) concerns the fact people want to minimize the amount of data they share. In the context of this research this means that (1) and (2) provide sufficient definition of privacy assuming a person acts out of free will in choosing to (not) share their data. As such the definition of privacy that is used in this research is:

Definition of privacy:*Privacy is defined as an individual's' right to control the collection, access to and uses of information relating to places, bodies, and personal data.*

2.3. Legal aspects of privacy

Besides gaining consent from people to gather and use their data there are many legal requirements to keep in mind. As of 25 may 2018 the new EU regulation on the protection of natural persons with regard to the processing of personal data and on the free movement of such data has come into effect (regulation (EU) 2016/679), also known as the General Data Protection Regulation (GDPR) [EU, 2016]. This regulation replaces the old Directive 95/46/EC. This new regulation offers better protection to a natural persons privacy. The GDPR is important to understanding the design space for systems using vehicle-based sensing data, as the GDPR sets the boundaries of what is legal. Because this legal context shapes the design space of such a given system, and subsequently also any experiments designed to simulate participation in such a system, the GDPR is a prominent part of this chapter.

Within the GDPR, the person providing their data is defined as the data-subject, and the party using and processing their data is defined as the Controller. The regulation is not applicable to activities that fall outside the scope of Union Law, this includes national security and "processing of personal data by Member States when carrying out activities in relation to the common foreign and security policy of the Union"[art. 16]. Similarly, the GDPR does not apply to activities by the competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offenses or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security. These elements are covered by the Directive (EU) 2016/680, which also came into effect on may 25th 2018. Furthermore, the regulation does not apply to activities that involve no professional or commercial activity, such as the use of social media. However, the processing of this kind of information does fall under the GDPR.

2.3.1. Principles

Article 5 of the GDPR provides principles relating to personal data. The article states that personal data must be: processed lawfully, fairly and in a transparent manner(Lawfulness, transparency and fairness); collected for a specific purpose and used in a manner not incompatible with such purpose (Purpose Limitation); adequate, relevant and limited to what is necessary for the processing (Data Minimization); accurate and kept up to date (Accuracy); stored for no longer than necessary and the security of the data (Storage Limitation); processed in a manner that ensures appropriate security of personal data, including protection against unauthorized or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organizational measures (Integrity and Confidentiality). The practical application of these principles may not be immediately clear, as such each of the principles is given a practical explanation.

The first principle described, "lawfulness, fairness and transparency", consists of three elements; lawfulness, fairness, and transparency. In order for data gathering and processing to be lawful, the conditions listed in Article 52(1) of the EU Charter and the requirements for justified interference under Article 8(2) of the European Charter of Human Rights (ECHR) need to be accounted for when determining if data processing meets the lawfulness criteria. The processing of data is lawful when it is in accordance with the law; pursues a legitimate purpose and is necessary in a democratic society in order to achieve a legitimate purpose. Article 6 of the GDPR defines the lawfulness of processing data when the following conditions obtain;

(a) the data subject has given consent to the processing of his or her data for one or more specific purposes; (b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract; (c) processing is necessary for compliance with a legal obligation to which the controller is subject; (d) processing is necessary in order to protect the vital interests of the data subject or of another natural person; (e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller; (f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child. [EU, 2016]

The principle of fairness can be said to obtain when there is openness and honesty about the processor's identity; people are informed how the processor intends to use any personal data collected from them; subject's data is handled only in ways that they would reasonably expect; subject's data is not used in ways that unjustifiably have a negative effect on them. The principle of transparency pertains to the controller communicating to the data subject the way in which the data is used and/or processed. This principle obtains when the controller communicates these aspects truthfully and wholly.

The second principle described is that of purpose limitation. This principle obtains when the purpose for which the data is to be processed is clearly defined before processing. Another principle described is that of data minimization, this principle obtains when only the quantity of data required for for a particular processing activity should be gathered. The principle of accuracy "stipulates that personal data should be valid with respect to what it is intended to describe, and relevant and complete with respect to the purposes for which it is intended to be processed." [Bygrave, 2014] The principle of storage limitation is to ensure that the identification of data subjects is only possible for the duration of the processing of the data. Thus the principle of storage limitation obtains when data subjects can only be identified during the processing of the data. The principle of Integrity and Confidentiality pertains to the controller being charged with the responsibility of complying and demonstrating compliance with the principles above. This is a new provision in the GDPR with the aim of making the Controller responsible for the protection and enforcement of the aforementioned principles and other provisions of the GDPR.

2.3.2. Consent

Article 7 of the GDPR concerns the giving of consent of the data subject to the processor. Art. 7(1) states that the controller shall be able to demonstrate that the data subject has consented to processing of his or her personal data. The elements of valid consent are [Salami, 2017]; the data subject must have been under no pressure when consenting; the data subject must have been duly informed about the object and consequences of consenting; and the scope of consent must be reasonably concrete. All these requirements must be met for consent to be considered valid. Art. 7(2) further states that if the request for consent also concerns other matter, it needs to be presented in a manner that is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language. Any part of such a declaration which constitutes an infringement of the GDPR shall not be binding. Furthermore, Art. 7(3) goes on to state that consent can be withdrawn at any time by the data subject. When consent is withdrawn, it does not affect the lawfulness of data processing prior to withdrawing consent. Withdrawing consent needs to be as easy as giving consent. Article 8 goes on to provide conditions for obtaining a Child's consent, however, as the research in question will limit itself to road users (with presumable a driving license) it is not necessary to obtain consent from minors, and thus falls outside the scope of this research.

2.3.3. Information provided to data subject

The GDPR provides regulation as to what information is to be provided to the data subject where personal data are collected. The information to be shared with the data subject when personal data are collected is provided in Art. 13 of the GDPR. The information to be shared with the data subject is; the identity and contact details of the controller and, where applicable, the controller's representative; the contact details of the data protection officer, where applicable; the purposes of the processing for which the personal data are intended as well as the legal basis for the processing; the legitimate interests pursued by the controller or by a third party; the recipients or categories of recipients of the personal data, if any; and whether the controller intends to transfer personal data to a third country or international organization and the legal basis for such transfer. Other information to be provided include the period of storage of the personal data; the existence of the data subject's right to rectification or erasure of personal data or restriction of processing as well as the right to data portability; the existence of the right to withdraw consent at any time without affecting the lawfulness of processing based on consent before its withdrawal; the right to lodge a complaint with a supervisory authority; whether the provision of personal data is a statutory or contractual requirement and the consequences of failure to provide such data; and the existence of automated decision-making, including profiling. The Controller is obliged to provide information about his intention to process personal data for purposes other than it was obtained.

2.3.4. Rights of the data subject

The GDPR provides certain rights to data subjects to exercise control over their data, these have been briefly touched upon in the previous paragraph. In this section the rights of the data subject are expounded upon to provide insight into these rights.

Right of access by the data subject

First is the right of access by the data subject [Art.15]. This states that the data subject shall have the right to obtain confirmation from the controller whether or not personal data concerning him/her is being processed. And if so, access to the following information: the purposes of the processing; the categories of personal data concerned; the recipients or categories of recipients to whom the personal data have been or will be disclosed, in particular recipients in third countries or international organizations; the envisioned period for which the data will be stored, or, if not possible, the criteria used to determine that period; the existence of the right to request from the controller rectification or erasure of personal data or restriction of processing of personal data concerning the data subject or to object to such processing; the right to lodge a complaint with a supervisory authority; the existence of automated decision making, including profiling, and meaningful information about the logic involved, as well as the significance and the envisioned consequences of such processing for the data subject.

Right to rectification

The right to rectification is described in Art.16 of the GDPR. This concerns the right of the data subject to obtain rectification of inaccurate personal data. Furthermore, the subject has the right to have incomplete personal data completed.

Right to erasure

The right to erasure, also named 'the right to be forgotten', is described in Art. 17 of the GDPR. It states that a data subject has the right to obtain from the controller the erasure of personal data concerning him/her without undue delay, and the controller is obligated to honor this request when one of the following grounds apply: the personal data is no longer necessary in relation to the purposes for which data was collected or otherwise processed; the data subject withdraws consent on which the processing is based and there is no other legal ground for processing; objection by the data subject; unlawful processing; the personal data needs to be erased for compliance with Union or Member State law.

In the case the controller has made the data public, and is required to erase the personal data, the controller is obliged to take reasonable steps to inform controllers processing the data that the data subject requests the erasure of links to, and copies of their data by said controllers.

Exceptions to the right to erasure are present. The aforementioned grounds for erasure are not applicable when data processing is necessary for the purposes: for exercising the right of freedom of expression and information; for compliance with legal obligation requiring processing by Union or MS law or for the performance of a task carried out in the public interest or in exercise of official authority vested in controller; for reasons of public health; for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes [in accordance with Art. 89(1)], when the right to erasure would likely render impossible or seriously impair the achievement of the objectives of that processing; or for the establishment, exercise or defense of legal claims.

Right to restriction of processing

The right to restriction of processing refers to the the right the subject has to restrict processing of data of the Controller when one of the following applies: the accuracy of personal data is contested by the data subject, for a period enabling the Controller to verify the accuracy of the personal data; the processing is unlawful and the data subjects opposes the erasure of the personal data and requests the restriction of their use instead; the controller no longer needs the personal data for the purposes of the processing, but they are required by the data subject for the establishment, exercise or defence of legal claims; the data subject has objected to processing to processing pursuant to article 21(1)[the right to object] pending the verification whether legitimate grounds of the controller override those of the data subject.

Right to data portability

The right to portability (Art. 20) refers to the data subject's right to receive the personal data concerning them, which they have provided to a Controller. This will have to provided to the data-subject in a structured, commonly used and machine-readable format and have the right to transmit the data to another controller without hindrance of the original controller. Where it is technically feasible the data-subject may have the data directly transmitted from one controller to another.

Right to object

The final right of the data subject is the right to object (Art. 21). This states that data subjects has the right to object, on grounds relating to their particular situation, at any time to processing of personal data concerning them which is either: kept in a form that permits identification for longer than is necessary for the purposes for which the personal data are processed [Art. 6(1)]; processed in a manner that does not ensure appropriate security of the personal data [Art. 6(1)]. The Controller may only use the data if it demonstrates compelling legitimate grounds for the processing which override the interests, rights and freedoms of the data subject or for the establishment, exercise or defence of legal claims. Furthermore, the data subject always has the right to object to processing of personal data for marketing purposes, including profiling to the extent that it is related to direct marketing.

2.4. Domain specific privacy and participation factors from a user perspective

In this section the domain specific factors regarding participation in a VBSD data-sharing system are viewed from a user perspective. The user has the choice to participate in sharing their VBSD with a governmental party for the purposes of DTM. Examples of the types of data gathered could be location, time, speed, and the number of vehicles passed on either side. First an overview is provided of the privacy concerns users have according to literature, followed by a review of the other factors affecting user participation, such as control, transparency, trust, social factors, and the use of incentives. From the review of the literature hypothesis are formed which are compared to the results of the stated preference experiment in chapter 4 and 5.

2.4.1. Privacy concerns of users

When speaking in broad terms about privacy (not specifically connected vehicles) it has been found three types of people exist regarding their general attitude towards privacy and their willingness to share data[Westin, 1991]. These groups are the privacy fundamentalists (17%), pragmatists (56%), and marginally concerned (27%) [Ackerman et al., 1999]. The privacy fundamentalists are very concerned about their privacy and are not willing to provide their data, even when privacy protection measures are in place. The pragmatists are less concerned about data use than fundamentalists, but often had specific concerns and tactics for addressing them, with each person employing different heuristics to evaluate potential privacy risks. Furthermore pragmatists can often be swayed to share their data in the presence of privacy protection measures. The marginally concerned are generally willing to share their data although they express a mild general concern for their privacy. An overview if and how these groups are represented in the context of connected vehicles is not clear. According to an international survey 69.3% of people were concerned about privacy in connected cars [Schoettle and Sivak, 2014].

The privacy interests that people experience with regards to connected vehicles are myriad. Privacy interests can be affected by various activities, i.e. (1) information collection, (2) information processing, (3), information dissemination, and (4) invasion [Solove, 2006].

Information collection

The act of collecting data from a person equates to a privacy risk as this data can be accessed, processed and disseminated. Furthermore, the data is vulnerable to breaches of security in which the data can be accessed by third parties. The amount and nature of information collected negatively affects the willingness of users to participate in a data sharing system. This is shown in earlier research concerning the acceptability of Event Data Recorders (EDR) in vehicles for the purpose of speed control. An EDR can be compared to a ”black box” in airplanes, which collects data, this includes data such as speed and location, and can be used to identify users. It has been stated that the lack of privacy is the main element of the acceptability of the EDR [Eyssartier, 2015] and that addressing these privacy issues can increase acceptability levels to some extent [Eriksson and Bjørnskau, 2012]. The reason that EDR were considered to lack privacy is because of the amount of information collected on the device compared to other alternatives in the study such as Intelligent Speed Adaptation (ISA) and section control (measuring average speed over a section of road) for the achievement of the same goal of controlling road-users speed.

Users seem to be critical towards parties active on the private market [Walter and Abendroth, 2018]. This is due to the privacy risk associated with each kind of disclosed data. Disclosure of confidential data, such as for personal information or identifiable location data, equates to a high privacy risk, and thus users tend to be delicate in disclosing this data. Similarly, when the privacy risk for disclosing data is low users tend to be tolerant in their data disclosure [Endo et al., 2016]. The type and frequency of disclosed data by users are changed for the privacy risk level of sharing that data. Furthermore, a system or measure will be acceptable to users if the type and frequency of data collected is perceived as efficient and fair. At the opposite, a system or measure will be deemed unacceptable if the measures are perceived as a lack of privacy [Eyssartier, 2015]. This highlights the fact that parsimony in the types of data collected and with whom it is shared is very important to the acceptance of a service requiring user data. For when more data is required than necessary, a user may feel a lack of privacy and thus decide not to participate. From this information the following hypothesis are construed:

H1 *Users are more critical towards parties active on the private market.*

H2 *Higher levels of data collection lead to reduced levels of user participation.*

Information processing

Besides the type and frequency of data gathered from users, the manner in which it is processed and used is also important for users' perception of privacy. In a study on the adoption of an electronic toll system for vehicles it was found that consumer response and implied preferences demonstrated a clear resistance to consumer tracking, potential profiling and problems relating to government tracking [Riley, 2008]. Consumer tracking refers to the lack of privacy of location, in which the Controller can track the data-subject geographically, and profiling refers to the creation of user profiles based on observed behaviour. In particular the aspect of government tracking is due to the potential uses of the gathered information, which in certain cases has been used in a court of law in matters unrelated to the goal of the data collection, such as divorce cases. Many users in this study were unwilling to exchange personal information without the guarantee of their privacy.

Consumers seem to value privacy of behaviour and action (profiling) more than privacy of location and space (geographical tracking) [Kehr et al., 2015]. This is further demonstrated by [Derikx et al., 2016], who concluded (in the context of usage-based car insurance) that if the processing of user data would have included the calculation of risk profiles in subsequently linking insurance cost to driving behaviour, the dis-utility found for sharing this data may have been higher. From this information the following hypothesis is construed:

H3 *Users value privacy of behaviour and action more than privacy of location and space.*

Information dissemination

The dissemination of information refers to the act of the Controller sharing information with other parties. Examples of this could be companies sharing their user information with other companies or governmental parties such as the police.

It is clear that there is a negative utility associated with sharing user information with additional parties, as this increases the privacy risks associated with the collected information. Simply put, the more parties granted access to the collected data and information, the lower user acceptance becomes [Walter and Abendroth, 2018]. An explanation of this could be because consumers trust the party they share their data with initially more than unnamed third parties. According to Kehr et. al. the usage of personal data for personalized offerings from the data collector (regarding car insurance) is positively evaluated, while third party advertisements have a clear negative utility [Kehr et al., 2015]. An explanation of this may be the institutional trust, which is higher for the party that users directly share their data with as opposed to unnamed third parties. The issue of trust is expanded upon further in this chapter. From this information the following hypothesis is construed:

H4 *Users value the sharing of data with additional parties negatively.*

Invasion

There are two types of invasion; intrusion, and decisional interference [Solove, 2006]. Intrusion involves the invasion into one's life, and disturbes daily activities , alters routines and destroys solitude. Protection from intrusion involves protecting from unwanted social invasions, affording "the right to be left alone". Intrusion is related to disclosure, as disclosure in often made possible by intrusive information gathering. Examples of this are for example: spam, telemarketing, taking unwanted pictures, the feeling of being observed. In the context of vehicle data, intrusion could take the form of unease due to users feeling they are constantly being observed, which affects their behaviour. It is plausible that measures which infringe upon the users "right to be left alone" will have a negative effect on participation.

The risk of decisional interference is related to the risk of data leakage. Where information from one system is used in a very different context. An example of this would be the use of electronic toll data in a divorce case, such as described in [Riley, 2008]. The difference of decisional interference compared to information dissemination or data leakage is that decisional interference specifically refers to the situation in which this dissemination of information leads to differences in decisions affecting the data-subject, Information literacy and experience of data leakage raise user's awareness of privacy risks, and subsequently affect the participation of users in a service [Endo et al., 2016].

Privacy categories

While the four types privacy concerns set out by Clarke are valid types of privacy harm, the potential privacy issues relating to the issues of technological advancements are not sufficiently covered in these four categories. An expanded categorization of Clarke's privacy categories is defined by Finn into a total of seven categories [Finn et al., 2013]. These include the following categories, (1) privacy of the person, (2) privacy of behavior and action, (3) privacy of personal communication, (4) privacy of data and image, (5) privacy of thoughts and feelings, (6) privacy of location and space, and (7) privacy of association. Of these seven definitions three in particular are useful for the categorization of privacy in the context of this study and are listed as follows. Privacy of behaviour and action, defined as *"The ability to behave in public, semi-public or one's private space without having actions monitored or controlled by others."* Which relates to the ability to behave in public, semi-public, or private space without having actions monitored or controlled by others. Privacy of location and action refers to the ability of a person to keep their location private and is defined as: *"The right to move about in public or semi-public space without being identified, tracked, or monitored."* [Finn et al., 2013]. Privacy of data and image is defined as *"Concerns about making sure that individuals' data is not automatically available to other individuals and organizations and that people can exercise a substantial degree of control over that data and its use."* [Finn et al., 2013]. These privacy categories are used to categorize the different privacy factors affecting participation in vehicle-data sharing systems under the definition of privacy that has been determined.

2.4.2. Factors affecting user participation

Beside the perception of privacy risks, other factors are influential in the decision to participate in sharing data. These include factors such as control over data, transparency, trust, social factors, and incentives. The selection of incentives are both monetary (e.g. discount, cash compensation) or non-monetary (e.g. use of a service, decreased travel time). The influence of all these different factors is discussed in the following paragraphs.

Control

Users are sensitive to the ability to control access to their data, and are easy to disclose their personal data when they have knowledge and ownership of their data[Sheehan and Hoy, 2000]. This is manifest in having the ability to control access to their data by means of an on/off functionality, a means of controlling frequency of data transfer, and it is important for users to have proper service and data disclosure consent [Endo et al., 2016]. Failure to provide the necessary measures for users to control access to their data will likely lower user participation. From this information the following hypothesis is construed:

H5 *Users are more willing to disclose data when they have more control over the collection of their data.*

Transparency

Transparency is one of the most important aspects for users when considering sharing their data. Transparency is a factor that can compensate for possible privacy concerns by providing the user with an overview why the collected data is necessary, how it will be used and who will have access to it [Walter and Abendroth, 2018]. Transparency should not be employed simply for its own sake, respect for transparency should be seen as the manner in which the claim that a service creates value is justified, rather than it creating harm, wrong or injustice [Elia, 2009]. This is highlighted by results from interviews in which participants felt uncomfortable in the cases of intensive data consumption, but were willing to disclose this information by means of transparent usage communication [Walter and Abendroth, 2018]. Proper service and disclosure consent is important for enhancing the user's privacy setting acceptance [Endo et al., 2016]. Without protections such as service and disclosure consent transparency is unlikely to produce lasting trust and participation [Elia, 2009].

Trust

Trust has been identified as a strong incentive for people to disclose their data. Transparency is a manner in which trust is built between users and companies [Kang and Hustvedt, 2014] in combination with attention to data security and privacy risk mitigation. Enhancing the perception of security and privacy is one of the most important elements in building trust between user and system [Riley, 2008] and companies can garner high acclaim for mitigating privacy risk for users [Endo et al., 2016]. A company merely being transparent without mitigating privacy risks, and subsequently experiencing data leaks does not foster trust, as users will be hesitant to trust a company that does not provide a secure environment for their data. Privacy policy by itself does not contribute to the perception of corporate reliability or user participation, as most users do not read privacy policies. Similarly privacy seals do not fulfill their desired function as users do not understand its meaning [Endo et al., 2016]. Thus a clear manner in which to communicate the privacy policies and protections towards users is essential. The manner in which this perception can be influenced is by being consistent in providing a secure environment for users' data in which their rights are protected reliably, and this is communicated in a consistent and transparent manner.

Social factors

Users create a distinction in what is acceptable for private and professionally used vehicles, in which they are more reluctant to use connected car technologies in their private car compared to cars for professional use [Walter and Abendroth, 2018].

The social acceptability of this system is about the attitudes potential users have without the use of it. With regards to these attitudes it seems that education level has an impact on how potential users perceive the sharing of data. Highly educated people seem to value privacy concerns higher, as shown in research regarding internet technologies [Sheehan and Hoy, 2000] and thus may be less willing to disclose data. While potential users without higher education seem to require higher monetary compensation in order to share their data [Derikx et al., 2016]. It is however not known what other demographic factors will influence participation in the system.

Furthermore, people seem to look to others as to whether to participate or not, as something will be more acceptable if other people use it and they agree with it [Eyssartier, 2015]. This means that if the perception is cultivated that a lot of people participate and agree with the system it will lead to higher participation rates. From this information the following hypothesis are construed:

H6 *Higher educated people are less willing to disclose data.*

H7 *Lower educated people require higher levels of monetary compensation in order to participate.*

Incentives

Incentives refer to stimuli that move a road-user to participate in sharing their vehicle-based sensing data, both monetary and non-monetary. In broad terms it refers to the expected benefits people receive that compensate for the anticipated costs.

Regarding monetary compensation it is unclear if this will affect user participation in the system as different authors have found different results. The results of [Walter and Abendroth, 2018] did not find any positive effect on data disclosure when offering monetary compensation. The generally accepted belief is that monetary compensation do increase user participation [Endo et al., 2016] [Riley, 2008]. This is further shown in a study in which a monetary compensation was found to significantly increase user participation in usage-based car insurance [Derikx et al., 2016]. In particular Derikx et. al. found that lower education levels correlated to a substantially higher monetary compensation for them to participate.

According to a study into the perception of connected vehicles [Schoettle and Sivak, 2014], the majority of those surveyed stated that safety was the most important aspect of connected vehicles (versus mobility or environment). In a different study regarding ICT measures that limit speed, such as section control, intelligent speed adaption, and event data recorders, it was found that acceptance of these measures was increased if road safety was viewed as an important collective good, and speed as an important risk factor. This could also be applied to the current case for sharing vehicle-based sensing data, in which the safety and mobility can be seen as the collective good and delays (either due to congestion or road accidents) can be seen as the risk factor.

Social benefit has the same effect as monetary compensation in increasing disclosure rates. Herein the intrinsic value of a service and the service experience also promotes data disclosure [Endo et al., 2016]. From this information the following hypothesis are construed:

H8 *Monetary compensation increases user participation.*

H9 *Awareness of social benefits increases user participation.*

2.5. Comparison with exploratory interview results

In order to further explore the opinions of people on the matter of privacy regarding vehicle data sharing a series of six exploratory interviews were performed. The results of these interviews are not representative for the whole population due to the limited sample size. However, these interviews can be used to provide further insight into the opinions people hold and as a method of corroborating the findings of the literature to. The transcripts of these interviews can be found in Appendix A.

The interviewees have opinions across the range of attitudes towards privacy, with one being marginally concerned, two being privacy fundamentalists, and three being privacy pragmatists. All respondents are of the opinion that sharing vehicle data is a good idea when they benefit from it in the form of reduced travel time. Disagreements arise about which parties may access the data and for what purpose. One of the questions touched upon using the data for law enforcement and national safety purposes. While respondents agree that the data may be used for specific cases with a clear purpose, most disagree with police monitoring the data and ticketing road users based on this data. Transparency is specifically stated by one respondent to be very important in this context.

Regarding the sharing of the collected data with other parties such as universities, research agencies, and other third parties the following observations were made. Four out of six reported that they would not be willing to share this data with any commercial party, one would only share it to relevant commercial parties and one would be willing to share it with commercial parties if the information was aggregated. There is a strong distrust towards commercial parties accessing the data, and they feel the government should not share their data with commercial parties. One respondent even stating that the government should be independent and this is not compatible with commercial goals in this context. For the purposes of research the respondents are more willing to share their data, as long as it is anonymous.

The next questions pertain to whether monetary incentives would be required for them to participate in a vehicle data sharing system. Surprisingly all six respondents were of the opinion that sharing their data for improved traffic conditions would be reason enough to participate. With travel time and safety being listed as important factors. However, if there would be significant risks to them due to police monitoring many would choose not to participate. Regarding the nature of any monetary compensation for participating, there was no consensus on what would be a good manner in which to do this. Three out of six noted that a discount on the road tax would be a good method, however this would not be applicable to those who drive a company car, as noted by another respondent. Other ideas included direct monetary compensation based on their data contribution (what types of data and how many kilometers on the road), and public transport budget (although this is not applicable to everyone due to company public transport cards).

Across the respondents it is clear that they value transparency, safety and control of their data. Several reasons to stop participating were noted to be; a breach of security such as data being hacked, privacy risks of commercial parties using the data, and continual monitoring by the police.

The results of the the interviews are in line with the literature. Trust, transparency, control are important factors in the decision. Regarding the use of monetary compensation there is no clear consensus, other than that is would not be the primary reason for them to participate. This lack of consensus is also present in the literature and will be researched further in this thesis to provide clarification.

2.6. Summary and implications for experiment design

For the purpose of this research privacy has been defined as the right to control the collection, access to and uses of places, bodies, and personal information. Any system that gathers data from data-subjects, in this case road-users, needs to comply by the legal framework set out by the GDPR and be processed lawfully, fairly and in a transparent manner. The data-subject must provide explicit consent for the collection, processing and dissemination of their data. Important rights for data-subjects are requesting removal of their (unprocessed) data and requesting access to the data gathered on them. The legal framework outlined by the GDPR provides the design space in which to design realistic choice scenarios, as any system for the purpose of gathering data from road-users will be obligated to comply with the GDPR.

SQ1 *How do road-users perceive privacy regarding sharing their vehicle-based sensing data?*
It seems that people are sensitive to the perceived privacy risk level of sharing their data, with parties active on the private market perceived as higher risk. Higher evaluation of privacy risk is associated with higher education levels and higher levels of information literacy. The perception of data gathering to be fair and efficient is important to user participation. At the opposite, a system or measure will be deemed unacceptable if the measures are perceived as a lack of privacy or a high risk of privacy. This perceived risk can be lowered through transparent communication, strong user protection agreements and a pro-active stance towards data security, these measures foster trust between the user and the controller. When more data is required than necessary, a user may feel a lack of privacy and thus decide not to participate. This highlights the fact that parsimony in the types of data collected and with whom it is shared is very important to the acceptance of a service requiring user data. Parsimony of data collection is in line with the GDPR principle of data minimization. Furthermore, the value placed upon privacy of behaviour and action seems to be higher than the value placed upon privacy of location and space. As such, people are quite negative towards the creation of personal user behaviour profiles. Furthermore, users may be wary of a data collection system due to the nature of constant observation, however there is no concrete data to provide an indication as to if and how large this will impact participation.

SQ2 *What are the factors contributing to road-user participation in sharing vehicle-based sensing data?*

There are multiple factors contributing to road-user participation in sharing their VBSD. Education seems to play an important role, with high levels of education related to higher value of privacy concerns, and lower levels of education related to a higher level of monetary compensation required. Incentives can be both monetary and non-monetary. It seems that offering monetary compensation to participants, e.g. discounts, can increase the rate of participation, with lower levels of education associated with a higher level of monetary compensation for participation. There is however no consensus on the effectiveness of monetary compensation across the literature, and as such this needs to be tested in the experiment. Non-monetary compensation associated with participation are the social benefits provided in the form of increased mobility and safety on the road. These social benefits could have the same effect as monetary compensation. Cultivating the perception that delays and accidents are risks and participation in the system lowers these risk through increased safety and mobility may increase the effectiveness of these non-monetary compensation.

The creation of realistic choice scenarios in accordance with the framework set out by the GDPR, sensitivity to the dimensions of privacy perception,the inclusion of monetary incentives, and the inclusion of non-monetary incentives as the communication of social benefits will enable the creation of realistic choice scenarios.

2.7. Conceptual model

From the knowledge gleaned from both the literature review and the exploratory interviews, a conceptual model is formed to describe the stated choice experiment. The proposed model includes socio-demographic factors and latent variables in the form of attitudes.

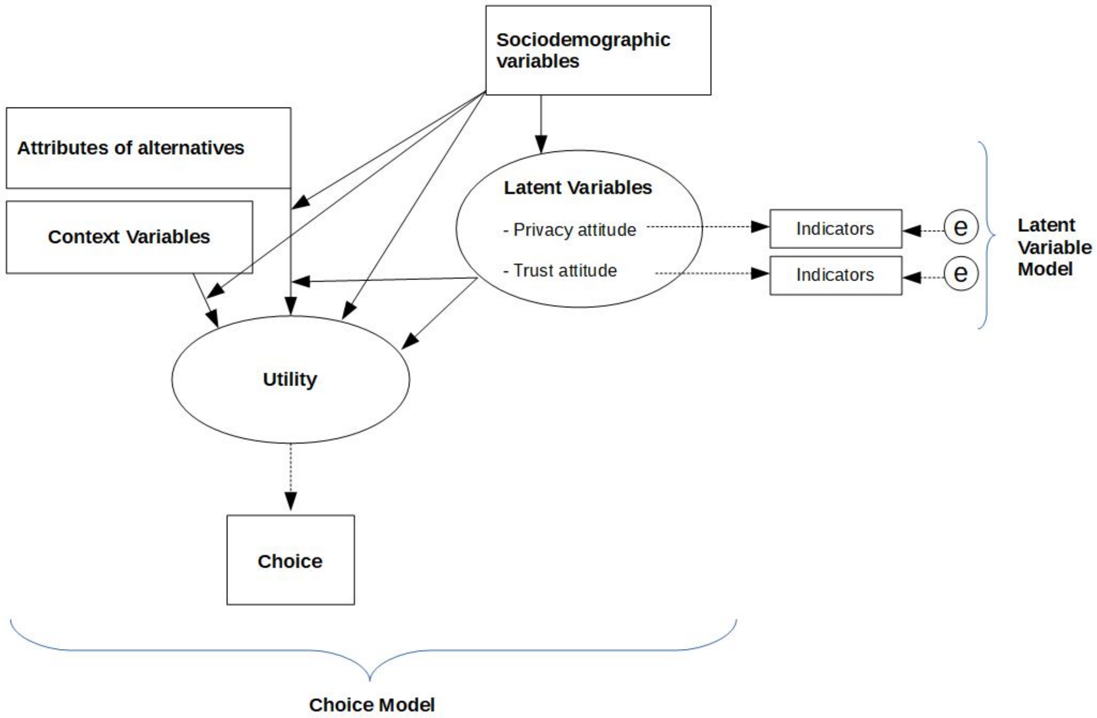


Figure 2.1: Conceptual model

First a selection is made of socio-demographic factors to include in the model, from the literature review it seems that education level, age and gender may provide information. A higher level of education seems to correlate with increased privacy risk perception [Sheehan and Hoy, 2000]. A lower level of education may lead to a higher level of monetary compensation required for participation [Derikx et al., 2016]. The assumptions regarding education level will be tested. Furthermore age and gender are included, to see if significant relationships are present in the data.

The latent variables to be included in the model are attitudes related to privacy concern and institutional trust. These are the privacy index, which measures a respondents attitude concerning the importance of privacy, and the (dis)trust index, which measures the respondents (dis)trust in the government [Kumaraguru and Cranor, 2005]. Each index consists of several statements to which respondents must answer whether they agree of or disagree on a Likert scale. These indexes are adapted and translated to dutch for use in the local environment. The details of these indexes are discussed in chapter 3. The inclusion of these two indexes provides insight into the attitudes of respondents, which is assumed to influence the evaluation of proposed alternatives and the decision whether to participate.

The attributes to be included are divided into two categories, the first concerns the data and its usage, the second concerns the position of the road-user. The first category contains the following attributes: the type of information collected, the manner in which information is processed, and the dissemination of information. The second category contains the following attributes: the users' measure of control over the data and the use of incentives both monetary and non-monetary.

3
Survey design and Analysis Methods

The aim of this study is to gain insight in how factors relating to privacy and incentives affect road-user participation in a vehicle-data sharing system. To this end a survey is developed employing the use of a stated preference experiment to gain insight into the choice behaviour of respondents. The conceptual model visible in figure 2.1 highlights the hypothesized links between the different elements affecting choice behaviour. Direct effects are present from the attributes of alternatives, context variables, and attitudes relating to privacy and trust. Furthermore it is hypothesized that socio-demographic factors and attitudes may influence the valuation of the attributes present in choice situations and may also affect the effect of context variables.

This chapter will provide insight into how the specific elements of this conceptual model are defined in the context of the survey. First the attributes and context variables of the stated preference are defined. After which the methodology and analyses used in this experiment are presented. Following this the construction of the attitude variables and the use of factor analysis is explained.

3.1. Survey Design

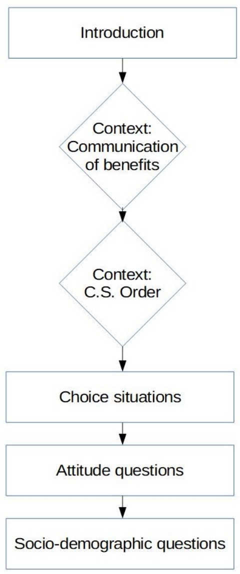


Figure 3.1: Survey Flowchart

The survey has four distinct configurations, one of which is presented to the respondent, this is dependent on the context variables relating to the communication of benefits and the order of the choice situations. First a respondent is randomly assigned to one of two contexts regarding communication towards expected benefits. Following this, the respondent is randomly assigned to one of two contexts for A/B testing the effect of order on the acceptability of non-monetary alternatives. After scenario assignment the respondents are presented with a total of twelve different choice situations, after which the attitude statements from table 3.7 are presented. Finally the respondents are asked the questions regarding socio-demographic factors, as shown in table 3.5.

There are a great many different methods for measuring respondents' choice behaviour. The most common of which are choice based stated-preference. Each choice situation presents two possible outcomes, participation or non-participation. In choice-based stated preference models participants are required to choose one of the presented alternatives in a given number of choice situations. Through the option to include the current situation (non-participation) in the choice situations it allows for modelling choice behaviour based on the attractiveness of the competitive alternative (participation) [Karniouchina et al., 2009]. Its main assumption is that respondents act in a utilitarian fashion with the aim of maximizing the utility derived from their choice. The utility of each alternative with is based upon a respondents valuation of each alternative's characteristics, also called attributes. These attributes each have different attribute levels, to which respondents assign a certain (part-worth) utility to each level of the attribute.

In the subsequent sections the selection of attributes and their respective levels is presented. Followed by an explanation of the chosen experimental design, this design encompasses the attribute levels present in the choice situations present in the survey. After this this the additional context variables are included, these refer to the communication of social benefits and the order in which choice situation are presented to the respondent. After which the socio-demographic factors included in the survey are presented and discussed.

3.1.1. Attributes and levels

Table 3.1: Operationalized attributes

	Factor	**Attribute**	**Levels**
Privacy Type			
Location and space	Information collection	Trip registration	Per road / Personal account
Behaviour and action	Information collection	Number of passed vehicles	No / Yes
	Information collection	Accelerometer data	No / Yes
Data and image	Information Dissemination	Sharing with emergency services	No / Yes
	Information Dissemination	Sharing with researchers	No / Yes
	Information Dissemination	Sharing with third parties	No / Yes
	Control	On-/functionality	No / Yes
Compensation type			
Monetary	Monetary compensation	Euro's per month	0 / 20 / 40 / 60

Privacy of location and space

Privacy of location refers to the ability of a person to keep their location private and is defined as: *"The right to move about in public or semi-public space without being identified, tracked, or monitored."* [Finn et al., 2013].

In the context of this experiment this attribute is operationalized as the level in which the gathered data can be attributed to the individual. Two levels have been identified. The first level is the least privacy sensitive, here a new id is created for the participants data per road, meaning that any given trip will have multiple unique id's attached to it, thus making it difficult to identify any individual from the data over their entire trip. The second level is the most privacy sensitive, here all trips are tied to a users' personal account, this means that all trips an individual makes can be traced back to that individual. These levels are chosen so as to measure the effect of reduced privacy of location and space, while being parsimonious is the amount of levels presented to the respondent. It is expected that the registration of trips on a personal account will be valued negatively compared to registration per road.

H10 *Trip registration is valued more negatively when tied to a personal account.*

Privacy of behaviour and action

Privacy of behaviour and action is defined as: *"The ability to behave in public, semi-public or one's private space without having actions monitored or controlled by others."* [Finn et al., 2013].

The level of privacy of behaviour and action is operationalized as the observation of road behaviour. Driving behaviour could be measured using a in-car motion sensor that measures the G-forces, and thus registers abrupt steering movements, acceleration and deceleration. Including this would enable the competent road authorities to gain insight into the behaviours of road users and may alert them to emergency situations (in the case of very rapid deceleration). The inclusion of such a device in the car would harm the users privacy of behaviour and action. This attribute and its operationalization was used previously in research regarding usage-base insurance in the Netherlands [Derikx et al., 2016].

An addition to this is including measuring the number of vehicles a car passes while driving (both ways of traffic). Measuring the amount of vehicles passed provides data which can be used to effectively estimate the traffic flow which is valuable for improving TSE [van Erp et al., 2019]. This data could be collected through the use of cameras or other sensors on the outside of the vehicle. It is expected that the inclusion of more sensors and measurements will be valued negatively by user due to increased privacy harm.

H11 *The addition of more sensors gathering data is valued negatively.*

Privacy of data and image

Privacy of data and image is defined as *"Concerns about making sure that individuals' data is not automatically available to other individuals and organizations and that people can exercise a substantial degree of control over that data and its use."* [Finn et al., 2013].

Privacy of data and image is operationalized as reuse of data generated by users participating in the data sharing system. The parties with which the data may be shared are emergency services, research agencies, and commercial parties. The selection for these particular parties to potentially share data with covers the breadth of potential parties in a real-world situation. What data is shared with these parties depends on how data is stored and which sensors are gathering the data, as presented in the previous sections.

First is the sharing the data with emergency services like the police, firefighters, and health services for the purposes of reacting to emergency situations. The data will not be shared with the police for the purpose of ticketing drivers and will only be used to improve responses to potential accidents and emergency situations. As the sharing of data with emergency services may improve safety on the road, a positive valuation is expected.

H12 *Sharing data with emergency services is valued positively.*

Second is the sharing of data with researchers, which refers to sharing data with universities and knowledge agencies for the purposes of research within the transport domain. No direct benefits for road-users exist due to the sharing of this data, and as such it is expected to be valued negatively due to potential privacy harm.

H13 *Sharing data with researchers is valued negatively.*

The third attribute concerns the sharing of data with third parties, such as commercial organizations. While the sharing of potentially identifiable data with unnamed third parties is not compatible with the principles of the GDPR as explained, it is included for experimental purposes. The sharing of data with unnamed third parties represents a situation with the highest uncertainty and privacy risk for users and can be said to represent the upper bound with regards to privacy risk. It is expected that respondents will value the sharing of data with third parties negatively.

H14 *Sharing data with third parties is valued negatively.*

The users measure of control over their data is included in the experiment as this may be an important factor for people in deciding whether to participate in a given system or not. People may not want to divulge their trip data all the time. The level of control is operationalized the ability to turn the data collection system on or off. In order to avoid people interpreting the option to turn the data collection off while enjoying a potential monetary compensation, the additional condition of 80% trips registered must hold for eligibility for monetary compensation. The experiment will include on-/off control over data collection as an attribute. It is expected that the inclusion of an on-/off functionality will be valued positively (H5).

Compensation

The inclusion of monetary incentives allows for insights whether respondents are willing to make trade-offs regarding privacy and monetary compensation by participating in a vehicle-data gathering system where they would otherwise not be willing to participate. A study on whether people are willing to engage in usage-based car insurance in the Netherlands [Derikx et al., 2016] used monetary compensation levels of 0, 10, 20 euros per month respectably. These levels were determined based on half the average insurance cost per month as this was deemed a reasonable level of compensation. Due to the similarity in the use-case the levels of monetary compensation have been determined to be up to half the monthly road tax cost. An average vehicle will have a monthly tax cost of around 50 Euros for gasoline vehicles and 100 Euros for Diesel vehicles. Additionally the inclusion of data sharing with different parties may require a higher compensation. As such the levels of monetary compensation have been defined as 0,20,40, and 60 euro's per month. It is expected that offering monetary compensation will increase participation (H8).

3.1.2. Experimental design

The experiment has been designed using an orthogonal design as can be seen in table 3.2, with a total of 12 choice situations. The experimental setup in which respondents have a binary choice whether to participate in a given alternative or not lends itself to an orthogonal design.The choice was made not to use a full factorial design as this would lead to a very large amount in choice situations, thus requiring a larger set of respondents. As such the use of an orthogonal design of the smallest possible size is deemed correct for the purpose of this experiment. The experimental design was constructed using Ngene.

Table 3.2: Experimental design

Variable	Attribute	Levels / Choice situations:	1	2	3	4	5	6	7	8	9	10	11	12
Registration	a.p_reg	0= per road 1= personal account	0	1	0	1	0	0	0	1	1	1	0	1
Passed vehicles	a.p_sense	0 = not present 1= present	1	0	0	1	0	1	0	0	1	1	1	0
Accelerometer	a.p_beha	0 = not present 1= present	1	1	0	1	0	0	1	1	0	0	1	0
Sharing Emergency services	a.s_emer	0 = no sharing 1 = sharing	0	1	1	1	0	1	1	0	1	0	0	0
Sharing Research	a.s_rese	0 = no sharing 1 = sharing	0	1	1	1	1	0	0	0	0	1	1	0
Sharing Third parties	a.s_third	0 = no sharing 1 = sharing	1	0	1	1	0	0	1	0	0	1	0	1
On- /off functionality	a.c_onoff	0 = no 1 = yes	1	0	1	1	1	0	0	1	1	0	0	0
Monthly compensation	a.c_comp	Euro's per month	0	0	0	60	60	40	60	40	20	40	20	20

3.1.3. Context Variables

Within the experiment two different types of context variables are present. The first is the communication of social benefits in the survey introduction, the second is the order in which choice situations are presented to the respondent. The second the order in which respondents are presented with choice situations without monetary compensation.

Communication of benefits

At the start of the survey the respondent is randomly assigned to one of two groups and may be presented with additional information regarding the expected social benefits of participation, both in the explanatory text and the accompanying introduction movie. In these scenarios special emphasis will be placed on the benefits of the data-sharing initiative. The benefits are stated to be improved traffic management leading to less congestion, less frequent delays on-route and increased safety on the road. The inclusion of this context variable makes it possible to gain insight into the effectiveness of communicating these benefits. The attribute relating to this context variable is included in the estimated utility function.

Additional text for communication of benefits:
Met de verzamelde voertuigdata zullen de metingen en schattingen van het verkeer worden verbeterd. Dit maakt het mogelijk om verkeersstromen beter te managen, met als gevolg minder vaak en minder ernstig last van files en opstoppingen. Daarnaast wordt verwacht dat het verzamelen van voertuigdata tot een veiligere wegsituatie zal leiden doordat nooddiensten sneller en effectiever op kunnen treden wanneer zich onveilige situaties voor doen.

Table 3.3: Communication of benefits

Context	**Highlights**
(0) Base	System purpose
(1) System purpose + Benefits	Improved traffic management, delay frequency & severity, road safety

Order of alternatives

The order in which respondents are faced with choice situation may influence their evaluation, specifically regarding choice situations without monetary compensation. As people may be less inclined to share their data for free after being presented with choice situations where monetary compensation was present. The inclusion of A/B testing will test whether being presented with choice situations that include monetary compensation significantly influences the valuation of choice situations without monetary compensation. To this end the survey randomly assigns one of two possible groups to respondent. Group A will present the choice situations without monetary compensation first in random order followed by the remaining choice situations in random order, while group B will present all choice situations in a randomized order. As this effect is only applicable to the choice situations without monetary compensation it cannot be included in the utility function. As such effect will be tested using a t-test over all observations which contain no monetary compensation, which are choice situation 1, 2, and 3. It is expected that group A, who are presented with the non-monetary choice situations first will choose to participate more often than group B.

H15 *People are less likely to disclose data for free after being presented with other options that offer monetary compensation.*

Table 3.4: Order of choice situations

Group	**Order**
(A)	First CS 1,2,3 in random order, then CS 4-12 in random order
(B)	All CS in random order

Table 3.5: Socio-demographic factors

Attribute	Levels	Coding
Year of Birth	[Number]	Single variable
Gender	[Man / Vrouw / Anders]	Dummy Coding
Education Level	[Basisonderwijs / VMBO / HAVO / VWO / HBO / WO]	Dummy Coding
Car Ownership	[Privé auto (eigenaar) / Privé auto (gebruiker) / Leaseauto / Huurauto / Bedrijfsauto / Geen beschikking over een auto]	Dummy Coding
Car Type	[Sedan / Hatchback / Stationwagen / Terreinauto / SUV / MPV / Coupé / Cabriolet / Sportwagen / Bestelbus]	Dummy Coding
Km driven per year	[<5.000km / 5.000 <15.000 / 15.000 <30.000 / 30.000 <45.000 / 45.000 <60.000 / >60.000]	Single variable

3.1.4. Socio-demographic Factors

Socio-demographic factors are included in the experiment to provide insight whether these factors structurally affect respondents choices. The included socio-demographic factors are visible in table 3.5. Each of these factors is discussed in the following section.

Age

The first factor to include is age, the level that people perceive privacy risk seems to differ with age. With the privacy risk perception being higher in older generations than younger generations, and old generations tending to be delicate in disclosing their data [Endo et al., 2016].

H16 *Age negatively affects the willingness to disclose data.*

Gender

The inclusion of gender is a factor that is often used in experiments. In this case the literature seems to indicate that gender should not make a significant difference in the observed choices [Ackerman et al., 1999]. It is however included in order to test this hypothesis.

H17 *Gender does not cause a significant difference in observed choices.*

Education level

The inclusion of education level in the socio-demographic factors is grounded in literature. It seems that respondents with lower education levels may require higher monetary compensation in order to participate in a given data collection system (H8). Higher education levels have been linked to a higher disutility towards factors negatively affecting their privacy [Derikx et al., 2016] (H7).

Car ownership

The inclusion of car ownership as a socio-demographic factor has partly been discussed at the beginning of this chapter. The inclusion of this factor may provide insight whether the type of ownership of a vehicle affects the perception of privacy and valuation placed on the different experiment attributes. It seems that people are more willing to share vehicle data when this vehicle is a company vehicle as opposed to a privately owned vehicle [Walter and Abendroth, 2018]. However it is not known whether there are significant differences in choices between vehicle owners and people who borrow vehicles to use.

Car type

At the moment it is unknown if there is any effect that the type of vehicle people use influences their choices regarding data-sharing. This factor is included in order to shed light on this and explore if there is any significant difference in answers between different types of vehicle owners.

Driven kilometers per year
The inclusion of the kilometers per year driven by a respondent provides insight into whether users who travel more value privacy issues differently and whether their reaction to monetary compensation differs from people who travel less.

3.2. Analysis

3.2.1. Logistic Regression

“Stated-choice models offer an approach to investigate, estimate and predict the behavior of potential and actual participants in a controlled experimental framework to proposed or uncertain changes in attributes of goods or services in an existing or hypothetical situation” [Louviere et al., 2000]

In stated-choice experiments respondents are asked to choose one alternative from a given set of alternatives. Its main assumption is that respondents act in a utilitarian fashion with the aim of maximizing the utility derived from their choice. The utility of each alternative with is based upon a respondents valuation of each alternative’s characteristics, also called attributes. These attributes each have different attribute levels, to which respondents assign a certain (part-worth) utility to each level of the attribute. On the basis of these part-worth utilities it is possible to estimate utility functions that describe the choice behaviour through the use of logistic regression. These utility functions describe the total utility as a linear addition of the different utility components in the context of the experiment. In this study the outcome is binary and as such binary logistic regression is employed.

Assumptions of using logit models need to be discussed before it can be applied to the case of sharing road-vehicle data. The first assumption is that people distinguish alternatives based on its characteristics, defined as ’attributes’. These attributes each have their own, predefined values, defined as ’attribute levels’. It is further assumed that each participant derives a certain (part worth) utility value related to each attribute level. The total sum of part worth utilities of the attribute levels present in a given alternative is the total utility of that alternative. The goal of the respondent is to maximize their derived utility, by choosing the alternative that suits them best. A core assumption to logit models is the assumption of Independence of Irrelevant Alternatives (IIA), the IIA property arises from the assumption of independent random errors and equal variances for the choice alternatives, implying that the odds of choosing a certain alternative over another alternative must be constant regardless of whatever other alternative is present [Ben-Akiva et al., 1985].

The following formulas are used for the logit model:

(1) $V_{ijk} = \beta_k X_{ijk}$

(2) $V_{ij} = \sum_k V_{ijk}$

(3) $U_{ij} = V_{ij} + \varepsilon_{ij}$

(4) $P(Y|X) = \frac{e^{V_{ij}}}{1+e^{V_{ij}}}$

Formula (1) calculates the ‘part worth utility’ V_{ijk}, with β as the coefficient (weight), X as attribute value, I = individual, j = profile and k = attribute. The following formula (2) calculates the structural utility V_{ij} of a specific alternative as the sum of the part worth utilities of a alternative V_{ij}. Formula (3) calculates the total utility U_{ij} of a specific profile consisting of the structural utility, and a random utility component to catch everything that could not be explained by the model. SPSS 21 is the software used to perform the logistic regression.

Formula (4) calculates the odds of an individual picking a certain alternative, here $P(Y|X)$ is the odds that alternative Y is chosen, given X. V_{ij} is the structural utility that respondent i derives of alternative j, and e is the base constant of the natural logarithms (+-2,72).

3.2.2. Willingness-to-Accept

While the stated choice model is initially utility driven, the inclusion of one or more monetary attributes expressed in monetary terms makes it possible to convert these utility values to monetary values [Ben-Akiva et al., 1985]. The Willingess-to-Accept (WTA) represents the minimum amount of money required for a person to trade away a good or put up with a negative factor [Hanemann, 1991]. In this case privacy is the good to be traded away. As shown in formula (5), this is done by dividing the β of a given attribute by the β of the compensation attribute, thus translating the values from utility to euro's and taking the negative of this value. In this manner it is possible to estimate monetary values placed upon different attribute levels.

(5) $WtA = -\frac{\beta_k}{\beta_c}$

3.2.3. Importance of attributes

The importance of each attribute is defined as "the relative contribution of an attribute to the utility" [Hensher et al., 2005]. The importance of each attribute is calculated by dividing the range of each attribute by the sum of all attribute ranges [Orme, 2010].

3.2.4. T-test for order of alternatives

Testing whether the odds of participation in choice situations without monetary reward are affected by the order in which the choice situations are presented is done by means of a t-test. in the survey respondents are placed into one of two groups, Group A is presented with the choice situations without monetary reward first in a randomized order, while Group B is presented with a fully random order of choice situations. This t-test is conducted on a subset of the dataset containing the choice situations 1 through 3, which are the choice situations without monetary reward. The t-test is performed to determine whether significant differences are present between the groups together with a Levene's test to determine whether the variances between the groups differ significantly.

Table 3.6: Westin's Indexes

Privacy Index (4 point Likert Scale)
- 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 (5 point likert scale)
- 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 harm us

Table 3.7: Modified privacy indexes

Privacy Concern (5 point Likert Scale)
- Bij de keuze om gebruik te maken van een dienst speelt privacy een grote rol in die overweging.
- Ik ben op de hoogte welke gegevens van mij worden verzameld en met wie deze gedeeld worden.
- Wanneer ik een handige app of dienst wil gebruiken weegt het voordeel zwaarder dan de privacy aspecten.

Institutional Trust (5 point likert scale)
- De overheid kan vertrouwd worden om de belangen van de burgers te behartigen.
- Huidige wetgeving en de manier waarop organisaties met privacy omgaan biedt een voldoende niveau van privacy bescherming voor de gebruiker.
- De overheid kan vertrouwd worden om privacy gevoelige gegevens veilig te houden.

3.3. Attitudes relating to Trust and Privacy Concern

The experiment measures two latent variables that concern the attitude of the respondent towards privacy issues and trust in the government respectively. These are included as it is hypothesized that:

H18 *Higher levels of privacy concern decrease levels of data disclosure.*

H19 *Higher levels of institutional trust increase levels of data disclosure.*

The Distrust index and the Privacy index are two indexes created by the long-term privacy researcher Alan Westin, who for over forty years was responsible for the majority of privacy polling in the United States. These indexes have been used to measure privacy attitudes of people and can be found in table 3.6. These indexes however do not measure exactly what needs to be measured in the current experiment, namely; peoples attitude on privacy concern, and the trust people place in the government. As such two new metrics which take Westin's work as an inspiration have been constructed. These can be seen in table 3.7. In the survey, respondents are asked indicator questions, three regarding the attitude of privacy concern, and three regarding trust in the government. A 5-point Likert scale is used for rating, ranging from ”strongly disagree” to ”strongly agree”. The latent attitude variable is calculated by adding the scores of the different indicators. One exception to this is the third indicator for privacy (ap3), as its score will be inverted before calculation due to the inverse direction of the statement.

3.3.1. Factor Analysis

Factor analysis is a technique that is used to reduce a large number of variables into fewer numbers of factors [Schulze et al., 2015]. This technique extracts maximum common variance from all variables and puts them into a common score. In this study factor analysis is used as a tool to evaluate whether the statements relating to the attitudes actually measure these attitudes. Factor analysis using the principal axis factoring method is specifically used for measuring latent variables for which a causal relationship is hypothesized. These being privacy concern and institutional trust, for which it is hypothesized that the indicator statements as shown in table 3.7 are linked to their respective factors, as shown in figure 3.2. The score on an indicator can be described with the formula:

$x_i = L_i F + e_i$

Here x_i is the score x on an indicator i. L_i is the factor loading of i on its related factor, representing . F is the position of a respondent on the latent factor. e_i represents the error term which encompasses all unexplained variance.

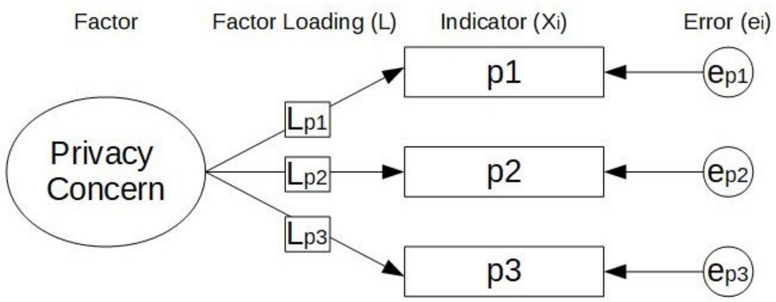


Figure 3.2: Factormodel

The appropriateness of using factor analysis will be evaluated using the Bartlett Test of sphericity, testing whether a correlation matrix is significantly different to an identity matrix, and the Kayser-Meyer-Olkin (KMO) measure of sampling adequacy. The KMO is an index for comparing the magnitude of the observed correlation coefficients. A KMO score of less than .5 is unnaceptable, and indicates that a factor analysis is not appropriate. Furthermore the communalities of the indicators should be $> .25$ for a sufficient portion of each indicators variance that can be explained by the extracted factors. The amount of factors extracted is determined by the number of eigenvalues >1. An eigenvalue >1 indicates that the factor explains more variance than a single observed variable. From the analysis the factor loadings are determined, which represent the correlation coefficient from the factor to the indicator. Following this a orthogonal rotation is applied in the form of varimax rotation. This is done to maximize the factor loading on one factor while minimizing it on other factors and aids in the interpretation of the results.

4

Survey development and deployment process

Within this chapter the process of the development of the survey and its subsequent deployment are discussed. First the target audience is defined. Secondly the process of the survey development is discussed. After which the survey deployment is presented in terms of respondent selection and the representativity of the acquired set of respondents.

4.1. Target Audience

The survey concerns the collection of VBSD on the Dutch road network. As such the target audience are Dutch road-users. As such the selection of respondents excludes respondents who are underage and those that never use a car as a driver. While this does reduce the potential pool of respondents it is necessary in order to obtain answers from respondents that have experience using road vehicles. By excluding this group of non-users the goal is to reduce measurement error. However, this alone does not make a distinction between respondents who own a vehicle or respondents who borrow a vehicle, as such this is included in the socio-demographic factors.

4.2. Survey development

In this section the process by which the final survey was determined is discussed. The steps taken were the dissemination of a test questionnaire, and the survey pretest for the final questionnaire.

4.2.1. Results of test questionnaire

Before the finalized version of the questionnaire was developed (which is described in this chapter) a test-questionnaire was used to receive feedback. This test questionnaire had a larger experimental design with 24 choice situations, three levels to the scenarios regarding the highlighting of benefits (a distinction between personal and societal benefits), and included an additional attribute relating to the control users had over the sharing of their data.

The test questionnaire was gained a total of twelve respondents, two of which were observed while taking the questionnaire. The respondents reported that the relative large amount of choice situations in the test questionnaire (24) was longer than would be reasonable for respondents to complete. As such the design was reduced to a total of 12 choice situations. The introduction of the different scenarios is easy for respondents to gloss over without reading, as such a short video explaining the experiment and scenario is added. The video also quickly introduces the format of the experiment as this was somewhat unclear as per the feedback of respondents.

Using SPSS a preliminary analysis has been performed. While most attributes are not significant with such a small dataset it does provide some insight whether variables have the correct sign (positive or negative) and which variables are likely to be most important. It was observed that the experiment does not result in respondents answering no to every choice situation, the results show that 53 percent of observed choices resulted in participation. The results with the trust statements differ and were expected to be positive, as increased trust should indicate increased participation. Regarding the attributes of presented alternatives in the choice situations it is clear not all signs can be determined due to the limited dataset. However several attributes are very clear. Firstly, the level of trip registration, which concerns the level at which data is identifiable is negative (taking level 0, segment id as the baseline instead of level 2, user id) and significant. This is in line with expectations. The gathering of additional data such as vehicles passed and motion sensor cannot clearly be determined to be positive or negative. The sharing of data with emergency services is valued mildly positive, whereas sharing with researchers is undetermined. Sharing with third parties is valued strongly negative and is significant even with the limited dataset available. Sharing with third parties resulted in a 73% rejection of the choice situation. The control attributes are valued positively, as is expected. Furthermore the compensation is valued positively. These signs are as expected.

4.2.2. Survey pretest

After gathering feedback from the test questionnaire, the feedback was discussed with three experts at the Delft University of Technology [1]. After several adjustments and feedback gained from several test participants the survey is finalized. In order to reduce the complexity for respondents and the duration of the experiment the levels related to trip registration were reduced from three to two, and the additional attribute relating to additional control over the dissemination of the data is removed. Due to these changes the amount of choice situations was reduced from 24 to 12. This change allows for a more understandable survey which is faster to complete. The duration of the survey is approximately 10 minutes.

4.3. Survey Deployment

The final survey was deployed over a period of nine days to gather respondents. In this section the procedure for collecting responses is presented. Followed by a discussion regarding the representativity of the collected sample.

4.3.1. Data collection procedure

Respondents were reached by providing the survey online and sharing the link with various people, and the collection of surveys using a tablet at several locations. These locations included the Delft University of Technology, the Hague University, a public library, and public spaces. The method for recruiting respondents was mainly through the use of the researchers social network, and direct contact, as such the sample is a convenience sample that does not take into account representativity in its respondent selection.

4.3.2. Respondents and representativity

A total of 124 completed responses were obtained, of which 98 were retained after filtering out respondents who do not drive a road vehicle. This corresponds with a margin of error (MOE) of 9,1% at a 95% confidence interval. The MOE is the amount of random sampling error in the survey's result. Smaller samples have a larger margin of error [Myers et al., 2006][2]. Most studies consider a maximum MOE of 5% as acceptable, highlighting the disclaimer to interpret the results with care [Simon, 2006].

Due to the sample being a convenience sample that does not account for representativity it

[1] Professor, Associate Professor, and Assistant professor at TPM, all members of the thesis committee

[2] Formula for MOE, with: p = probability of population participating (0,69), n = sample size (98) and N = population size (11.176.150)

$$MOE = 1.96\sqrt{\frac{p(1-p))}{n}} * \sqrt{\frac{N-n}{N-1}}$$

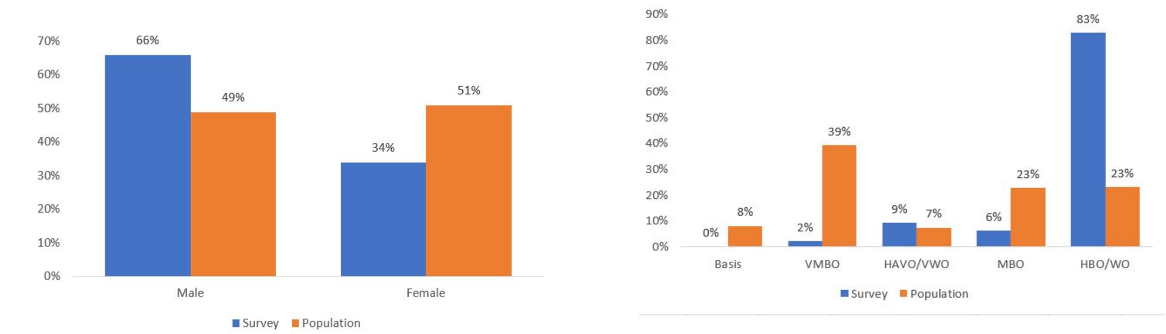


Figure 4.1: Gender and Education

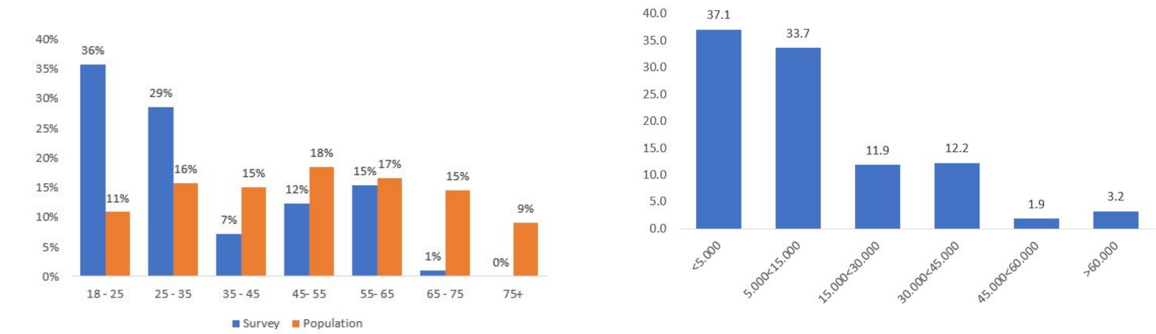


Figure 4.2: Age and km/year

can be said that the sample is not representative. The target audience is a subset of the Dutch population, however statistics showing the socio-demographic factors of road-users were not obtainable, as such the Dutch population is used for comparison. Specifically comparisons are made concerning, age, gender, education level, and average mileage per year.

The distribution of gender in the respondents is 66% male and 34% female, compared to the Dutch population with 49% Male and 51% Female [CBS, 2018]. Looking at education levels as shown in figure 4.1 it can be seen that 83% of respondents are highly educated at academic (HBO/WO) level, in contrast to only 23% of the population. The age distribution of the respondents can be seen in figure 4.2. While a diverse age group of respondents has been observed there is an over-representation of the age groups 18 - 25 and 25 - 35. Concerning the amount of kilometers driven on a yearly basis, figure 4.2 shows the percentage of responses corresponding to their categories. The average Dutch amount of kilometers traveled on a yearly basis is 13.000 km/year [CBS, 2018], this is close to the average of the observed responses, which is 13.700 km/year.

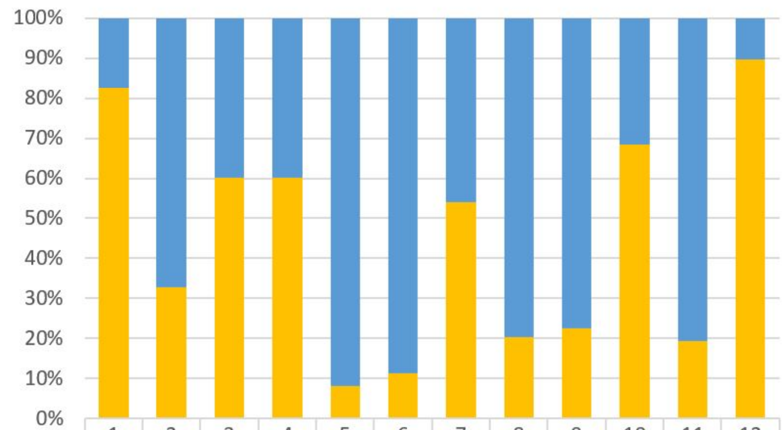
From these comparisons it is clear that the sample contains biases regarding age, gender, and education. Further bias may be present due to a measure of self-selection, by which respondents who are willing to participate in a data sharing system may choose to fill out the survey more often than people who are not willing to participate in a data sharing system. Because these biases can influence the results these biases need to be taken into account when interpreting the results of the analyses in the next chapter.

5

Results and Analysis

5.1. Exploratory analysis

in this section several exploratory analyses over the collected sample are performed. First the responses per choice situation are presented and discussed, presenting a overview how the survey was answered by the respondents. Following this the effects of presenting respondents with different orders of choice situations is analyzed using a t-test and the results are discussed. Thirdly a factor analysis is performed on the attitude statements to determine whether these statements do indeed measure their respective attitudes.



	1	2	3	4	5	6	7	8	9	10	11	12
Percent Yes	17%	67%	40%	40%	92%	89%	46%	80%	78%	32%	81%	10%
Percent No	83%	33%	60%	60%	8%	11%	54%	20%	22%	68%	19%	90%

Figure 5.1: percentage odds of participation per choice situation

5.1.1. Responses per choice situation

The results of the survey are visualized in figure 5.1 and table 5.1 contains an overview the different attribute levels present within each of the choice situations. Some observations can be made on the basis of this overview of this data. Most importantly it can be noted that the choice situations where data is shared with third parties always result in a minority share of the respondents choosing to participate, most notable is choice situation 12 with a 90% rejection rate. The opposite is true when data is not shared with third parties, with a majority share choosing to participate in sharing their vehicle-data. This holds true for all choice situations even those where compensation is not present. The highest levels of participation are observed in choice set 5 and 6, with 92% and 89% levels of participation

Table 5.1: Answers per choice situation

Choice sit.	Count Yes	Count No	Percent Yes	Percent No	Regis-tration	Passed vehicles	Acceler-ometer	Emergency	Research	Third Parties	On-/off	Compe-nsation
1	17	81	17%	83%	0	1	1	0	0	1	1	0
2	66	32	67%	33%	1	0	1	1	1	0	0	0
3	39	59	40%	60%	0	0	0	1	1	1	1	0
4	39	59	40%	60%	1	1	1	1	1	1	1	60
5	90	8	92%	8%	0	0	0	0	1	0	1	60
6	87	11	89%	11%	0	1	0	1	0	0	0	40
7	45	53	46%	54%	0	0	1	1	0	1	0	60
8	78	20	80%	20%	1	0	1	0	0	0	1	40
9	76	22	78%	22%	1	1	0	1	0	0	1	20
10	31	67	32%	68%	1	1	0	0	1	1	0	40
11	79	19	81%	19%	0	1	1	0	1	0	0	20
12	10	88	10%	90%	1	0	0	0	0	1	0	20

respectively. These choice situations do not share data with third parties, register vehicle movements per road and offer financial compensation.

In relation to research question 3, *"Which privacy configurations for sharing vehicle-based sensing data are acceptable for road-users?"*. The results suggest that when data is not shared with third parties a majority of respondents is willing to share their data. Regarding research question 4, *"Which compensation schemes for sharing vehicle-based sensing data are acceptable for road-users?"*. It seems that all levels of compensation can be deemed acceptable to respondents, as long as the data is not shared with third parties. The offering of compensation does however positively influence the percentage of respondents willing to share their data.

5.1.2. Order effects

In the survey, the respondents were randomly assigned one of two different groups regarding the order in which choice situations were presented. With group A being presented the choice situations without compensation first in a random order, and group B presented with a fully random order.

Table 5.2: Group statistics regarding order effects

Group	N	Mean	Std. Deviation	Std. Error Mean
A	180	.47	.500	.037
B	114	.33	.473	.044

Table 5.3: t-test for equality of means (equal variances not assumed)

t	df	sig. (2 tailed)	Mean difference	Std. Error Difference	95% C.I. of difference Lower	Upper
2.301	250.292	.022	.133	.058	.019	.247

The Levene's test has p=0.000, which implies the variance between the two groups differs significantly. The means of the respondents' is shown in table 5.2, with group A choosing to participate in 47% of the cases and group B in 33% of the cases. The t value as seen in table 5.3 is t=2,301, indicating the size of the difference in participation relative to the variation in the sample data. The difference between the scenarios is significant ($p<0,05$). This means that when presented with a potential vehicle-data sharing system, respondents are significantly more likely to agree to sharing their data for free when they have not previously been exposed to alternatives that offer monetary compensation.

5.1.3. Confirmatory factor analysis of Attitudes

Confirmatory factor analysis is performed on the attitude indicator statements relating to privacy concern and institutional trust in order to test whether these indicators are indeed related to their respective attitude factors. The Barlett test resulted in a significance level of 0.00, implying the correlation matrix differs significantly for the identity matrix and factor analysis is suitable. The results of the KMO measure of sampling adequacy is .672, which is deemed acceptable. Furthermore all communalities are > 0.25 and thus a sufficient portion of each indicators variance can be explained by the extracted factors. Two factors with an eigenvalue of > 1 were extracted accounting for $63,5\%$ of the variance.

The two extracted factors do indeed only contain the expected indicators. With factor 1 representing institutional trust, and factor 2 representing privacy concern, as shown in table 5.4. The rotated factor loadings of at1 (.796), at2 (.705), and at3 (.688) have large factor loadings on factor 1, as such this factor describes Institutional Trust. The rotated factor loadings of ap1 (.593), ap2 (.574), and ap3 (-.621) load on factor 2, Privacy Concern. These factor loadings are considered sufficient (> .5) but are not considered high (> .7), indicating that privacy concern is measured by its related indicator but with smaller correlation coefficients as compared to factor 1.

Thus it can be stated that the indicators do indeed measure their respective attitudes. With institutional trust more strongly correlated to its indicators than the indicators correlating to privacy concern.

Table 5.4: indicator scores and results of factor analyses (Factor loadings < 0.2 have been excluded)

			C.I. of 95%			Rotated Factor Loading	
Indicators	**Mean**	**Std. Deviation**	**Lower bound**	**Upper Bound**	**Communalities**	**1 (Institutional Trust)**	**2 (Privacy Concern)**
ap1	4.04	1.169	3.81	4.27	0.351		.593
ap2	3.42	1.059	3.21	3.63	0.337		.574
ap3	3.15	1.207	2.91	3.39	0.407		-.621
at1	3.40	1.038	3.19	3.60	0.640	.796	
at2	2.98	1.134	2.76	3.20	0.504	.705	
at3	3.05	1.101	2.83	3.27	0.483	.688	

5.2. Logit model estimation

Through the use of binary logistic regression the logit models are estimated. First the base model is estimated which includes only the attributes varied in the SP experiment and the importance of the different attributes are discussed. Following this the results of the expanded model are presented and discussed, which include the socio-demographic factors, and attitudes. Subsequently the two models are compared to each other regarding model fit. Finally the expanded model is applied to predict the odds of participation in different configurations to provide insight into the expected choice behaviour of road-users.

5.2.1. Base Model

The base model which only includes the attributes present in the choice situations is obtained and the results can be seen in table 5.5. From the estimated β values an overview of the part worth utilities is constructed, as seen in table 5.6. These part worth utilities show the β values for each level present in each attribute. The importance of each attribute as shown in 5.6 is defined as ”the relative contribution of an attribute to the utility” [Hensher et al., 2005] and is included in table 5.6. Each attribute is discussed and compared to relevant hypothesis. Regarding model fit the Chi^2 value is considered significant at df=9, indicating a significant improvement over the null model. The Mc Faddan R^2 value is 0.238, indicating an excellent model fit [Domencich and McFadden, 1975].

Constant

The constant represents the utility in a choice situation where all attributes are zero. The positive value of $\beta = 1.014$ implies that the odds of participating in the reference alternative are greater than non-participation, with the odds of participation calculated at 73%. This indicates that respondents are generally willing to disclose data for the purpose of DTM.

Trip registration

Trip registration contains two levels as can be seen in table 5.6. The registration of trips is a highly significant attribute, $p = .000$, with a coefficient $\beta = -0.664$. The relative importance of

Table 5.5: Base Model

	B	S.E.	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Communication of benefits	-.059	.141	.675	.943	.715	1.242
Trip registration*	-.664	.148	.000	.515	.385	.688
Vehicles passed	-.039	.148	.794	.962	.720	1.285
Accelerometer	-.251	.147	.089	.778	.583	1.039
Sharing Emergency*	.311	.147	.035	1.364	1.023	1.821
Sharing Research*	.381	.148	.010	1.464	1.096	1.956
Sharing Third*	-2.458	.152	.000	.086	.064	.115
On-/off function	.259	.144	.073	1.295	.977	1.718
Monthly compensation*	.020	.003	.000	1.020	1.013	1.027
Constant*	1.014	.227	.000	2.756		

trip registration is 11.84% and ranks third in importance. This negative evaluation is due to the increased privacy harm of trip registration on a personal account, as it is easier to trace movements back to the individual.

Measuring Vehicles Passed
The measuring of vehicles passed refers to the presence of sensors gathering data on the amount of vehicles passed in both direction. The attribute is not significant ($p = .962$), with a coefficient of $\beta = -0.39$. The relative importance of vehicles passed is only 0.30% and it ranks last out of the attributes. While the sign of the β is negative as was expected, the non-significance and low importance of this attribute means that the inclusion of these sensors does not odds of participation in a meaningful manner.

Accelerometer
The inclusion of gathering accelerometer data of a vehicle has a coefficient of $\beta =$ p-value of $p = .089$ and is considered non-significant. The relative importance of this attribute is 4.47% and is ranked in 7th place. The expected sign for the accelerometer was negative and this is observed in the coefficient.

Data sharing with emergency services
The attribute relating to the sharing of data with emergency services is considered significant with $p = .035$ and $\beta = 0.311$. The relative importance of data sharing with emergency services is 4.47% and it is ranked at 5th place.

Data sharing for research
The attribute relating to the sharing of data for the purposes of research, such as universities and research agencies, is considered significant with $p = .010$ and the coefficient is estimated at $\beta = 0.381$. The relative importance of this attribute 6.79% and is ranked 4th in importance.

Data sharing with third parties
The attribute relating to the sharing of data with third parties is considered highly significant with $p = .000$ and has a coefficient of $\beta = -2.458$. The relative importance of this attribute is 43.81% and is ranked as the most important attribute. Sharing data with unnamed third parties is valued strongly negative by respondents as the sharing of data with unnamed third parties on the private market represents the highest level of privacy risk and uncertainty.

On-/off functionality
The attribute relating to the existence of on-/off functionality has a coefficient of $\beta = 0.259$. This attribute is considered marginally significant when only taking into account the p-value of $p = .073$. The relative importance is 4.62% and it is ranked at 6th place. The expected sign for this attribute was positive, and this true for the estimated coefficient. The inclusion of an on-/off function positively influence the odds of participation.

Table 5.6: Part Worth Utilities

Attribute level	Code	Estimate value		PWU	Range	Importance	Rank
Registration*		*ap_reg*			0.664	11.84%	3
Per road	0		-0.664	0.000			
Personal Account	1			-0.664			
Passed vehicles		*ap_sense*			0.039	0.70%	9
No	0		-0.039	0.000			
Yes	1			-0.039			
Accelerometer		*ap_beha*			0.251	4.47%	7
No	0		-0.251	0.000			
Yes	1			-0.251			
Sharing Emergency services*		*as_emer*			0.311	5.54%	5
No	0		0.311	0.000			
Yes	1			0.311			
Sharing Research*		*as_rese*			0.381	6.79%	4
No	0		0.381	0.000			
Yes	1			0.381			
Sharing third parties*		*as_third*			2.458	43.81%	1
No	0		-2.458	0.000			
Yes	1			-2.458			
On-/Off functionality		*ac_onoff*			0.259	4.62%	6
No	0		0.259	0.000			
Yes	1			0.259			
Monthly Compensation*		*ac_comp*			1.188	21.18%	2
€0.-	0		0.020	0.000			
€20.-	20			0.396			
€40.-	40			0.792			
€60.-	60			1.188			
Scenarios			-0.059		0.059	1.05%	8
No mention of benefits	0			0			
Mention of benefits	1			-0.059			
		Constant		1.014			

Financial compensation

Financial compensation is a highly significant attribute with $p = .000$ and has a coefficient of $\beta = 0.020$ per euro of monthly compensation. The relative importance of this attribute is 21.18% and is ranked as the 2nd most important attribute. Offering monetary compensation positively influences the odds of participation.

Communication of social benefits

The context variable concerning the communication of social benefits is considered non-significant with $p = .943$ and has a coefficient of $\beta = -.059$. The relative importance is $1,05\%$ and is ranked at 8th place. While this was expected that communicating social benefits would directly influence participation odds it seems that no direct effect is present.

5.2.2. Expanded Model

An expanded logit model is estimated to account for socio-demographic factors, attitudes, and interaction effects between attributes and the variables relating to socio-demographic factors and attitudes. First the the selection of socio-demographic variables is discussed relating to the exclusion of certain categories based on the observed frequency in the sample. Next the selection of the included interaction terms is discussed. Subsequently the results of the estimated model are discussed.

In this expanded model not all categories relating to car ownership and car type were included due to low observed frequencies in the sample. The frequencies of car ownership are presented in table 5.7, all categories with less than 5% of respondents have been excluded. The frequencies of car type are presented in table 5.8, similarly all categories with less than 5% of respondents have been excluded. The exclusion of these categories with low levels of respondents reduces the amount of variables in the model improving the ease of interpretation and model fit.

Table 5.7: Car Ownership type frequency

Car Ownership	Frequency	Percent
Company Car	5	5.1
Rental	2	2.0
Lease	8	8.2
Private car (owner)	34	34.7
Private car (user)	49	50.0
Total	98	100.0

Table 5.8: Car type frequency

Car Type	Frequency	Percent
Van	4	4.1
Cabriolet	2	2.0
Coupé	4	4.1
Hatchback	30	30.6
MPV	6	6.1
Sedan	27	27.6
Sportscar	1	1.0
Stationwagen	15	15.3
SUV	6	6.1
All-terrain vehicle	3	3.1
Total	98	100.0

The included interaction terms have been determined by first estimating all interaction terms between socio-demographic factors and attributes, and between attitudes and attributes. From this the significant interactions were kept while the non-significant interactions were removed. As such all interaction terms shown in table 5.9 are significant.

The estimated coefficients of the expanded model are presented in table 5.9. Regarding model fit the Chi^2 value is considered significant at df=30, indicating a significant improvement over the null model. The Mc Faddan R^2 value is 0.332, indicating an excellent model fit [Domencich and McFadden, 1975].

Attribute main effects compared to base model

Compared to the base model several observations can be made regarding the attributes present in the choice situations. Interaction effects that are observed are discussed together with their respective socio-demographic factors or attitudes in the subsequent sections.

Firstly the communication of social benefits is still positive an non-significant, but significant interaction terms are present with age, sex, and education level. The presence of an accelerometer is significant in the expanded model where it was non-significant in the base model. On the other hand the sharing of data for research is non-significant in the expanded model where it was significant in the base model. A significant interaction effect between sharing data for research and age was observed. Furthermore the main effect of sharing data with third parties was less pronounced in the expanded model, this being compensated with a significant interaction between sharing with third parties and age. Finally the main effect of compensation is higher in the expanded model, with this being compensated with a significant negative interaction with privacy concern.

Socio-demographic factors

Regarding the inclusion of socio-demographic factors several observations can be made. Firstly age does not appear to play a significant role from their direct effect. It does however have significant interaction effects. As such it can be stated that age by itself does not

Table 5.9: Expanded Model

		B	S.E.	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Context variable	Communication Benefits	0.525	0.471	0.265	1.690	0.672	4.251
Attributes	Trip registration *	-0.793	0.161	**0.000**	0.453	0.330	0.620
	Vehicles passed	-0.051	0.159	0.750	0.950	0.696	1.299
	Accelerometer *	-0.321	0.159	**0.043**	0.726	0.532	0.990
	Sharing Emergency *	0.345	0.160	**0.031**	1.412	1.033	1.931
	Sharing Research	-0.347	0.379	0.361	0.707	0.336	1.487
	Sharing Third *	-2.019	0.394	**0.000**	0.133	0.061	0.288
	On-/off function *	0.326	0.156	**0.037**	1.385	1.020	1.881
	Monthly compensation *	0.054	0.014	**0.000**	1.056	1.027	1.085
Attitudes	Privacy Concern *	-0.096	0.049	**0.051**	0.908	0.825	1.001
	Institutional Trust *	0.140	0.030	**0.000**	1.150	1.085	1.220
Socio-dem	Age	0.005	0.012	0.685	1.005	0.982	1.028
	Sex	-0.193	0.228	0.397	0.824	0.527	1.289
	Ownership_Owner	-0.718	0.600	0.232	0.488	0.150	1.582
	Ownership_User	-0.426	0.608	0.483	0.653	0.198	2.149
	Ownership_Lease	-0.100	0.690	0.885	0.905	0.234	3.500
	Ownership_Company	-1.057	0.688	0.124	0.347	0.090	1.337
	car_sedan	0.453	0.272	0.096	1.573	0.922	2.682
	car_hatchback *	0.550	0.265	**0.038**	1.733	1.031	2.914
	car_stationwagen	0.295	0.307	0.337	1.343	0.736	2.451
	car_suv	-0.018	0.384	0.963	0.983	0.463	2.086
	car_mpv	-0.476	0.408	0.244	0.622	0.279	1.384
	Yearly driven km *	-0.181	0.084	**0.031**	0.835	0.708	0.984
	High Education *	-0.778	0.211	**0.000**	0.459	0.304	0.695
Interactions	Age * Communication of benefits *	-0.033	0.012	**0.005**	0.967	0.945	0.990
	Sex * Communication of benefits *	0.888	0.388	**0.022**	2.429	1.135	5.201
	High Education * Communication of benefits *	0.841	0.321	**0.009**	2.318	1.236	4.348
	Age * Sharing Research *	0.023	0.010	**0.026**	1.023	1.003	1.045
	Age * Sharing Third *	-0.026	0.011	**0.018**	0.974	0.954	0.995
	Privacy Concern * Monthly Compensation *	-0.003	0.001	**0.021**	0.997	0.994	1.000
Constant	Constant	1.550	0.931	0.096	4.712		

significantly affect the odds of participation directly, but mainly through interactions with communication of benefits, sharing data for research purposes, and sharing data with third parties.

Sex does not play a significant role from its main effect, but does affect the odds of participation through interaction with the communication of benefits. This result implies that females are more sensitive to this communication of social benefits than males are.

The results show that the form of ownership and car type are not considered significant, with the exception of the car type hatchback, which has a significant positive effect on the odds of participation.

The amount of km driven on a yearly basis negatively affects the odds of participation, implying that people who drive more are less inclined to share their data, potentially due to a higher perceived privacy risk as a result of an increase in data that can be shared.

Regarding education it is observed that high levels of education[1] reduce the odds of participation. Implying that higher educated people may be more concerned about privacy issues when deciding to share their data. The dis-utility of high education levels can be compensated through the communication of social benefits, as the effect of interaction between high education and communication of social benefits is higher than the dis-utility of higher education.

Attitudes

The attitudes relating to Privacy Concern and Institutional Trust are significant and are in the expected directions. Higher levels of privacy concern lower the odds of participation ($\beta = -.096$), while higher levels of Institutional Trust raise the odds of participation ($\beta = .140$).

As the scales for both attitudes are equal and the coefficient for Institutional Trust is higher than for Privacy Concern it can be said that privacy concerns can be compensated

[1] Academic levels of education (HBO/WO)

for by high levels of institutional trust. This implies that people with high levels of privacy concern will be willing to disclose data when they sufficiently trust the other party.

A negative interaction effect between Privacy Concern and monetary compensation is present in the expanded model. This interaction effect shows that when Privacy Concern is high in an individual the positive effect of offering monetary compensation is reduced. This implies that the more a person is concerned with their privacy, the less susceptible they are to the offering of monetary rewards.

5.2.3. Comparison of model fit between base model and expanded model

Ascertaining whether the expanded model improves upon the base model is done through the comparison of various goodness-of-fit measures. These include the Mc Fadden R^2, the Bayesian information criterion (BIC), and the Akaike information criterion (AIC).

The scores on the various goodness-of-fit measures are shown in table 5.10. When comparing the two models the Mc Fadden R^2 is higher in the expanded model, implying an improvement in model fit. This is further substantiated by comparing the BIC and AIC of both models, with the expanded model scoring lower on these measures. This lower scoring on the BIC and AIC similarly implies an improvement in model fit. Thus the expanded model does indeed improve model fit compared to the base model.

Table 5.10: Goodness of fit measures for base and expanded model

Goodness of fit measure	Base Model	Expanded Model
Chi^2	384*	535*
Mc Fadden R^2	0.238	0.332
BIC	1251.597	1138.788
AIC	1251.694	1139.069

5.2.4. Predicted behaviour

The expanded model has been used to predict the odds of participation for a diverse group of individuals in the choice situations that were present in the experiment. These predictions can be found in table 5.12.

The user profiles are determined using two values for the socio-demographic factors, attitudes, and the inclusion of the communication of benefits. Where possible a low and high value for each factors has been used. Regarding age it was arbitrarily determined that 18 years old would constitute a low age and 50 would constitute a high age in this context. For the attitudes relating to Privacy Concern and Institutional Trust a low level was determined to be a score of 6 and high a score of 12. These scores correspond to an average scoring on their respective indicators of 2 and 4. In the case of car ownership and car type it is not possible to determine a high and low score, as such the two most common categories are used in the prediction. For ownership this includes an individual being a car user or a car owner, and for car type this includes the hatchback and sedan.

The results presented in 5.12 can be used to provide an insight into the expected odds of participation for different user profiles in different choice situations and may prove useful for policy makers in designing VBSD sharing systems.

Table 5.11: Willingness-to-Accept values relating to attributes

Privacy Construct	Attribute	WtA (Euro's)
Location and space	Trip registration (User ID)	33.55
Behaviour and Action	Vehicles passed	1.95
	Accelerometer	12.67
Data and image	Sharing Emergency	-15.69
	Sharing Research	-19.26
	Sharing Third	124.14
	On-/off function	-13.06

Table 5.12: Predicted odds of participation in each choice situation based on user profiles (Expanded model)

							Observed choices in sample		**Choice sit.**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
									Count No	81	32	59	59	8	11	53	20	22	67	19	88
									Count yes	17	66	39	39	90	87	45	78	76	31	79	10
									Percent No	83%	33%	60%	60%	8%	11%	54%	20%	22%	68%	19%	90%
									Percent Yes	17%	67%	40%	40%	92%	89%	46%	80%	78%	32%	81%	10%
							Configuration of choice situations		**Trip registration**	0	1	0	1	0	0	0	1	1	1	0	1
									Vehicles passed	1	0	0	1	0	1	0	0	1	1	1	0
									Accelerometer	1	1	0	1	0	0	1	1	0	0	1	0
									Sharing Emergency	0	1	1	1	0	1	1	0	1	0	0	0
									Sharing Research	0	1	1	1	1	0	0	0	0	1	1	0
									Sharing Third	1	0	1	1	0	0	1	0	0	1	0	1
									On-/off function	1	0	1	1	1	0	0	1	1	0	0	0
User profiles									**Monthly compensation**	0	0	0	60	60	40	60	40	20	40	20	20
Communication of benefit	**sex [0=male]**	**age**	**eduhigh [1=hbo/wo]**	**privacy**	**trust**	**car owner [0=user, 1=owner]**	**type [0=hatchback, 1=sedan]**	**km per yr [0= <5000, 5=>60.000]**	**Predicted odds of participation**	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***11***	***12***
0	0	18	0	6	6	0	0	0	***1***	35%	77%	54%	76%	99%	97%	83%	93%	92%	52%	91%	34%
1	0	50	1	12	6	1	0	0	***2***	4%	53%	15%	14%	91%	75%	11%	50%	57%	7%	70%	3%
0	0	50	1	6	12	0	1	0	***3***	25%	91%	61%	81%	100%	98%	76%	95%	94%	59%	97%	25%
1	0	18	0	12	12	1	1	0	***4***	34%	76%	53%	51%	97%	95%	62%	86%	89%	33%	87%	26%
0	0	50	0	12	12	0	0	5	***5***	18%	87%	51%	50%	98%	95%	42%	86%	89%	32%	93%	13%
1	0	18	1	6	12	1	0	5	***6***	35%	77%	55%	77%	99%	97%	84%	93%	92%	52%	91%	35%
0	0	18	1	12	6	0	1	5	***7***	9%	39%	18%	17%	85%	78%	24%	54%	61%	9%	57%	6%
1	0	50	0	6	6	1	1	5	***8***	4%	55%	16%	34%	97%	87%	28%	69%	67%	15%	78%	4%
0	1	50	1	12	12	1	1	5	***9***	7%	68%	25%	24%	95%	85%	19%	66%	71%	13%	82%	5%
1	1	18	0	12	12	0	1	5	***10***	48%	85%	68%	66%	98%	97%	75%	92%	93%	48%	92%	39%
0	1	18	0	6	6	1	0	5	***11***	16%	55%	30%	54%	97%	93%	65%	82%	81%	28%	79%	16%
1	1	50	1	6	6	0	0	5	***12***	11%	78%	36%	61%	99%	95%	53%	87%	86%	34%	92%	11%
0	1	18	1	6	6	1	1	0	***13***	14%	51%	27%	50%	96%	92%	61%	80%	79%	25%	76%	14%
1	1	50	0	6	6	0	1	0	***14***	16%	85%	48%	71%	99%	97%	65%	91%	91%	45%	95%	16%
0	1	50	0	12	12	1	0	0	***15***	18%	87%	50%	49%	98%	94%	41%	85%	88%	31%	93%	13%
1	1	18	1	12	12	0	0	0	***16***	60%	90%	77%	76%	99%	98%	83%	95%	96%	60%	95%	51%

5.2.5. Limitations

The main limitation of the logit model is its bias due to an over-representation of males, highly educated people, and the age group 18 - 35 years old. Due to this bias inherent in the sample the interaction effects may not be as strong as suggested in the estimated logit model. Furthermore the possibility of self-selection effects, by which people who are more inclined to participate in a data-sharing system are more likely to participate in the survey, may cause the odds of participation to be overestimated. Similarly, the respondents were asked whether they would be willing to participate in a given data-sharing system, not whether they would actually participate, this formulation of the question may also cause overestimation of the odds of participation. These limitations need to be taken into account when interpreting the results of the model.

5.3. Willingness-to-Accept

The WtA is used to quantify the amount of monetary compensation required for an individual to trade away their privacy regarding the attributes present in the choice situations. Positive values imply that compensation is required for acceptance. The results of the base model are used as input for determining the values shown in table 5.11.

These results highlight that the level of trip registration requires higher levels of compensation than the inclusion of data from additional sensors. Demonstrating the fact that privacy of location and space is valued at a higher price than privacy of behaviour and action.

An extremely high WtA value is observed for sharing data with third parties. It is important to note that the WtA value for sharing with third parties is likely to be the highest possible value for sharing with third parties, as the specific parties and purpose of the data sharing is unknown, representing a situation of great uncertainty, low trust, and thus a high level of privacy risk. When the data would potentially be shared in real-life situations it would be on an opt-in basis for a specific purpose and in line with the principles set out by the GDPR. Thus the WtA estimated in table 5.11 can be assumed to be the maximum value associated with sharing data with third parties.

5.4. Comparison of results to formulated hypothesis

With the results obtained from the analysis presented in this chapter it is possible to evaluate all the hypothesis that are formulated in chapter 2 and 3, an overview of all hypothesis and results is shown in table 5.13. These hypothesis are used to help answer the research questions of this study in chapter 6.

Table 5.13: Overview of hypothesis and result (+ = confirmed, - =rejected, ∼ =debatable)

Hypothesis	Description	Result
H1	Users are more critical towards parties active on the private market.	+
H2	Higher levels of data collection lead to reduced levels of user participation.	+
H3	Users value privacy of behaviour and action more than privacy of location and space.	-
H4	Users value the sharing of data with additional parties negatively.	∼
H5	Users are more willing to disclose data when they have more control over the collection of data.	+
H6	Higher educated people are less willing to disclose data.	+
H7	Lower educated people require higher levels of monetary compensation in order to participate.	-
H8	Monetary compensation increases user participation.	+
H9	Awareness of social benefits increases user participation.	+
H10	Trip registration is valued more negatively when tied to a personal account.	+
H11	The addition of more sensors gathering data is valued negatively.	+
H12	Sharing data with emergency services is valued positively.	+
H13	Sharing data with researchers is valued negatively.	-
H14	Sharing data with third parties is valued negatively.	+
H15	People are less likely to disclose data for free after being presented with other options that offer monetary compensation.	+
H16	Age negatively affects the willingness to disclose data.	∼
H17	Gender does not cause a significant difference in observed choices.	∼
H18	Higher levels of privacy concern decrease levels of data disclosure.	+
H19	Higher levels of institutional trust increase levels of data disclosure.	+

H1 *Users are more critical towards parties active on the private market*

Confirmed: The results show that sharing data with third parties is valued strongly negative, while sharing data with governmental parties is actually valued in a positive manner. The negative valuation of third parties in increasingly more negative with age. As such it can be stated that users are indeed more critical towards parties active on the private market.

H2 *Higher levels of data collection lead to reduced levels of user participation*

Confirmed: Looking at the results of the expanded model it can be seen that the inclusion of more data types, such as accelerometer data, leads to lower odds of participation. Furthermore increased levels of trip registration, registration of complete trips tied to a user account, also reduces the odds of participation. Thus it can be confirmed that higher levels of data collection will lead to reduced levels of user participation.

H3 *Users value privacy of behaviour and action more than privacy of location and space.*

Rejected: This hypothesis is rejected using the WtA values obtained in 5.3. Privacy of behaviour and action is operationalized in the form of sensors measuring driving behaviour, being an accelerometer and sensing the amount of vehicles passed in both directions. Privacy of location and space is measured by the level of trip registration. Comparing these results it is clear that the WtA value for privacy of location and space (€33,55) is higher than the sum of WtA values relating to privacy of behaviour and action(€14,62), thus H3 is rejected.

H4 *Users value the sharing of data with additional parties negatively.*

Debatable: The sharing of data with (unnamed) third parties negatively affects the odds of participation. However, sharing data with emergency services or for the purpose of research positively affects the odds of participation. While it may hold true in the general sense that more parties accessing a users' data will be valued negatively this is not always the case, and as such the hypothesis cannot be confirmed nor rejected.

H5 *Users are more willing to disclose data when they have more control over the collection of data.*

Confirmed: A measure of users' control over data collection was operationalized as the inclusion of an on-/off function in the data collection system, by which the user may choose not to collect and share their data for at most 20% of trips. This on-/off function had a significant positive effect on the odds of haring their data.

H6 *Higher educated people are less willing to disclose data.*

Confirmed: Using the expanded logit model it can be observed that academic levels of education reduce the odds of disclosing data. Thus H6 is confirmed.

H7 *Lower educated people require higher levels of monetary compensation in order to participate.*

Rejected: For the expanded logit model all interactions between attributes and socio-demographic factors were analyzed. The interaction effect between education level and monetary compensation was estimated and was found to be non-significant, thus H7 is rejected.

H8 *Monetary compensation increases user participation.*

Confirmed: From the results of the logit models it is clear that the offering of monetary compensation increases the odds of participation. Thus H8 is confirmed.

H9 *Awareness of social benefits increases user participation.*

Confirmed: The communication of social benefits was included as a context variable in the SP experiment. The results presented in the expanded logit model confirm that awareness of the social benefits increases the odds of participation.

H10 *Trip registration is valued more negatively when tied to a personal account.*

Confirmed: From the results of the logistic regression is is clear that trip registration tied to a personal account negatively affects the odds of participation. Registration of trips tied to a personal account is valued more negatively than when trips are registered anonymously per road, thus H10 is confirmed.

H11 *The addition of more sensors gathering data is valued negatively*

Confirmed: The expanded model has estimated that the inclusion of more sensors is valued negatively, in particular the inclusion of an accelerometer was noted to be significant. Thus it can be stated that the addition of more sensors gathering data is valued negatively and H11 is confirmed.

H12 *Sharing data with emergency services is valued positively.*

Confirmed: From the estimated logit models it can be observed that sharing data with emergency services is valued positively and increases the odds of participation as such H12 is confirmed.

H13 *Sharing data with researchers is valued negatively*

Rejected: Sharing data with researchers positively affects the odds of participation, as observed in both the base and expanded model. Due to this positive evaluation H13 is rejected.

H14 *Sharing data with third parties is valued negatively*

Confirmed: Sharing data with third parties is observed to be the single most determining factor affecting the odds of participation. The presence of sharing data with third parties is valued negatively, and as such H14 is confirmed.

H15 *People are less likely to disclose data for free after being presented with other options that offer monetary compensation*

Confirmed: In the experiment the order in which respondents were presented with choice situations was varied, with one group first being presented the choice situations without monetary reward. Using a t-test (see 5.1.2) it was determined that respondents were significantly less prone to participate in the data sharing system after they were already presented with other choice situations in which monetary compensation was offered, as such H15 is confirmed.

H16 *Age negatively affects the willingness to disclose data*

Debatable: In the expanded model it is observed that age does not have a significant main effect on the odds of participation. However significant interaction effects exist. Age positively influence the willingness to share data for research purposes, and negatively influences the willingness to share data with third parties. As such H16 can neither be confirmed nor rejected.

H17 *Gender does not cause a significant difference in observed choices*

Debatable: While the main effect of a respondents gender is non-significant, a significant interaction effect between gender and the communication of social benefits is observed. As such, while gender does not directly influence the observed choices, it does affect the observed choices when the social benefits have been communicated to the respondent.

H18 *Higher levels of privacy concern decrease levels of data disclosure.*

Confirmed: As is shown in table 5.9 higher levels of privacy concern reduce the odds of participating in the data sharing system. Thus H18 is confirmed.

H19 *Higher levels of institutional trust increase levels of data disclosure.*

As is shown in table 5.9 higher levels of Institutional Trust increase the odds of participating in the data sharing system. Thus H19 is confirmed.

6 Discussion and conclusion

In this chapter of the report the outcomes, relevance, and limitations of the conducted study are discussed. Section 6.1 presents the main findings of the research. Followed by the implications of this study for theory and practice in section 6.2. Elaboration on the limitations of this study are presented in section 6.3, followed by areas for future research.

6.1. Main findings

The objective of this research is *to further develop understanding of the effects of privacy factors and incentives on road-user participation in sharing their Vehicle-Based Sensing Data for the purposes of Dynamic Traffic Management.* The main research question that supports this objective is:

RQ: *How do factors relating to privacy and incentives affect road-users participation in a vehicle-data sharing system for the purposes of Dynamic Traffic Management?*

Answering the research question is done by answering the formulated sub-questions related to it, after which the findings are summarized and the main research question is answered. The tools employed in this study for gathering the required data include literature review, structured interviews, and a stated preference experiment.

SQ1 *How do road-users perceive privacy regarding sharing their vehicle-based sensing data?*

Based on a review of the literature a definition of privacy is constructed for this study: *Privacy is defined as an individual's' right to control the collection, access to and uses of information relating to places, bodies, and personal data.*

Due to the fact that participation in a data-sharing system requires the registration of all trips it can be stated that control is traded away in this situation, as such it is possible to view privacy as a trade-able good.

From the review of the literature and several main categories affecting the perception of privacy have been identified and based on the results of the analysis the effects of these categories has been verified. The main categories affecting the perception of privacy are identified to be the collection of information, processing of information, dissemination of information, and invasion.

Higher levels of information collection are valued negatively by users. This is confirmed by the negative coefficients observed in the estimated logit model relating to the inclusion of accelerometer data and data relating to the amount of vehicles passed, with higher levels of data collection reducing the odds of participation. Implying that people are more hesitant to disclose data when they perceive it to be more than necessary, highlighting the need for parsimonious data collection.

Processing information is also important to users' perception of privacy, and processing of information related to consumer tracking meets resistance from road-users. This is highlighted by results of the estimated logit model which noted a strong negative effect on the odds of participation when the collected data was linked to a personal account. Implying many users would be unwilling to disclose data without the guarantee of their anonymity.

The dissemination of data is generally believed to be valued negatively according to literature, as a higher number of parties with access to the data implies increased privacy risks. The results of the choice models show that the dissemination of data with third parties active on the private market is the single most determining factor negatively affecting the odds of data disclosure. Contrary to the literature the sharing of data with emergency services and for research purposes was valued positively by respondents, this was observed in the interviews and the estimated logit models. These results imply that users are more critical towards parties active on the private market.

The last category influencing the perception of privacy is that of invasion, which related to intrusions upon the ”wish to be left alone” and the risk associated with data leakage, both of which are perceived negatively by users. In both literature and the interviews the aspect of government tracking is seen as problematic when this information can be used against them in matters unrelated to the collection purpose. These results highlight the a users’ need for clear boundaries of data usage in line with the principle of purpose limitation as set out in the GDPR. Implying a clear communication of the scope for which the data can be used is essential for users’ to be willing to disclose their data.

SQ2 *What are the factors contributing to road-user participation in sharing vehicle-based sensing data?*

From the results of the literature review, interviews, and the estimated logit model a myriad of factors contributing to user participation are identified. The attributes varied in the choice situations of the stated preference experiment are discussed, followed by an overview of the effects of the measured attitudes.

Table 6.1: Ranking of attributes varied the stated preference experiment

Attribute	Rank	Importance	Effect
Sharing data with (unnamed) third parties	1	43.81%	-
Monthly monetary compensation	2	21.18%	+
Level of trip registration	3	11.84%	-
Sharing data for research purposes	4	6.79%	+
Sharing data with emergency services	5	5.54%	+
On-/off functionality	6	4.62%	+
Collection of accelerometer data	7	4.47%	-
Communication of social benefits	8	1.05%	+
Collection of # of vehicles passed data	9	0.70%	-

From the estimated logit model including the attributes varied in the stated preference experiment a ranking of these attributes and their relative importance is determined, for which the results are presented in table 6.1.

The sharing of data with unnamed third parties is the single most important factor influencing user participation. The sharing of data with unnamed third parties for unknown purposes is not realistic as it would violate certain principles of the GDPR relating to consent, purpose limitation and transparency. As such the sharing of data with unnamed third parties represents a situation where privacy risk is highest and should be interpreted as a maximum.

The second most determining attribute contributing to participation is the offering of monthly monetary compensation. In the literature the effect of compensation for trading away privacy is not unequivocal, with some studies finding no positive effect on disclosure rated. In the stated preference experiment respondents were offered a monthly monetary compensation of up to €60.-. The results of this study do show that offering monetary rewards can positively increase disclosure rates to a large degree. This implies that offering monetary compensation is indeed a correct tool for increasing participation.

The third attribute affecting participation is level of trip registration. Trip registration was operationalized in two levels, separate registration for each road (0), or registration linked to a user account (1). The results of the estimated logit model confirm the literature, which states that increased levels of privacy risk associated with being personally identifiable negatively influence participation.

Contrary to the assumptions of the literature the sharing of data for research purposes and with emergency services has a positive effect on participation. From the literature is was expected that higher levels of information dissemination due to increased levels of privacy risk. This contradiction to the literature would indicate a violation of the utilitarian perspective employed in this study which states that users aim to maximize their utility. As such it is possible a different decision rule may be employed by people when deciding to share data for research and with emergency services, possibly due to a moral element in the decision making.

The presence of on-/off functionality by which a user may turn off the collection and sharing of their VBSD for 20% of trips was noted to positively influence participation. This result is in line with the literature and highlights the fact that increased levels of control from a user perspective increase the odds of participation.

Increased levels of data collection were expected to negatively influence participation on the basis of the literature and this effect is confirmed by the results of the logit model. Interestingly the collection of accelerometer data from a users' vehicle is ranked higher than the collection of data by observing other vehicles passed. This could imply that users value their own privacy over the privacy of others. The small negative effect of the measuring of the amount of vehicles is an important observation, as this implies that collecting this data will only negatively influence user participation to a very limited degree.

According to the literature the communication of benefits was expected to positively influence participation. However the direct effect of this attribute is ranked lowest of all attributes in the estimated logit model including only the attributes varied in the state preference experiment. The effect of the communication of benefits is however strongly dependent on interactions with socio-demographic factors such as age, sex, and education. This large heterogeneity is accounted for in the expanded logit model and its interactions are discussed in more detail in the answering of SQ5.

From the literature it was assumed that higher levels of privacy concern would negatively influence participation and higher levels of institutional trust would increase participation. Through the use of indicator statements in the survey related to these attitudes their respective effects were estimated and found to have significant effect in the hypothesized directions. Significant interaction between privacy concern and the valuation of monetary compensation was observed, by which higher levels of privacy concern would reduce the value placed upon monetary compensation. The significant effect of Institutional trust would imply that fostering trust between the data collecting party and potential users through ensuring data security and transparent communication is valuable in increasing user participation.

SQ3 *Which privacy configurations for sharing vehicle-based sensing data are acceptable for road-users?*

In order to answer this question, a privacy configuration is deemed acceptable when a majority of users would be willing to participate in a given system. Not accounting for heterogeneity in socio-demographic factors the following conclusions can be drawn on the basis of the base logit model including only the attributes of the choice situations.

The short answer to this question is that any configuration of attributes will result in a majority of users estimated to be willing to participate, given that data is not shared with unnamed third parties.

The situation with the lowest utility without sharing data with third parties is when; benefits are not communicated, trips are registered linked to a user account, additional data is gathered regarding amount of vehicles passed and accelerometer, and no on-/off functionality is present. In this configuration the odds of participation are estimated to be 51%. If in this configuration data is also shared with third parties the odds of participation fall to a mere 8%.

In the situation where the expected utility is maximized the odds of participation are a lot higher. In this configuration trips are registered per road, no additional data is gathered, data is shared with emergency services and researchers, and an on-/off functionality is present. in this situation it is predicted that 87% of people would be willing to participate.

For the attributes a willingness-to-accept has also been calculated, as seen in table 5.11. These values represent the amount of monetary compensation required to accept the negative effects of these attributes. Registration of trips on a user account, compared to anonymous trip registration per road is valued at a WtA of €33,55. The inclusion of extra data collection; accelerometer €12,67, vehicles passed €1,95. Most notable is the WtA for sharing data with unnamed third parties, valued at a WtA of €124,14.

These results highlight the strong privacy concerns users have with the sharing of data with third parties. The effect on acceptability of sharing data with with emergency services could provide opportunities in the design of VBSD sharing systems for the improvement of responses to emergency situations. Similarly the positive effect of sharing data with researchers could provide many opportunities for using the data gathered to improve DTM measures.

SQ4 *Which compensation schemes for sharing vehicle-based sensing data are acceptable for road-users?*

Direct monetary compensation increases odds that a road-user will deem participating in a vehicle-data system as acceptable. This is confirmed by the results of the estimated logit model. However offering alternative compensation methods may also increase participation levels.

The first non-monetary scheme of compensation is simply not offering financial compensation at all. When road-users have not yet been exposed to vehicle-data sharing systems offering monetary the level of participation in systems without compensation is significantly higher than when road-users have been previously exposed to alternatives that do offer monetary reward. In the survey group A was presented first presented with choice situation without monetary compensation in a random order, while group B was presented with a fully randomized order of choice situations. Through means of a t-test it was determined that the observed choices differed significantly between these groups for choice situations without monetary compensation, with the mean odds of participation for group A being 47% and group B being 33%. These results would imply that people have higher odds of participating for free when they have not exposed to alternative systems that offer monetary compensation.

The perception of non-monetary compensation in the form of perceived social benefits resulting from participation can be improved through the communication of the benefits. These social benefits may include improvements relating to congestion, faster travel times, and increased effectiveness of emergency services. A large heterogeneity exists in the effectiveness of this communication and is accounted for by interaction effects in the expanded model. The groups most affected by the communication of these social benefits are young, highly educated, and female. These results imply that while the communication of benefits is a valuable tool to increase participation, it is highly dependent on the socio-demographic factors of the recipient.

SQ5 *How do socio-demographic factors influence the acceptance of sharing road-vehicle data?*

The effect of socio-demographic variables has been estimated in the expanded logit model, for which the results are presented in table 5.9. This includes the main effects and significant interaction effects between attributes and socio-demographic factors. Here the main effects and interaction effects of each socio-demographic factor are discussed. The socio-demographic factors are age, sex, education, yearly driven km, car ownership, and car type.

Age was not found to have a significant direct effect on choice behaviour. However, age does express itself through significant interactions with the communication of benefits, sharing data with researchers, and sharing data with third parties. These interactions result in the following effects; communication of benefits is valued more negatively with age, sharing data with researchers is valued more positively with age, and sharing data with third parties is valued more negatively with age. These results imply that the communication of benefits is more effective in younger people than older people, people are more willing to share data with researchers as they get older, and people are less willing to share data with third parties as they get older.

Respondents' sex was not found to have a significant direct effect on choice behaviour. Sex does interact significantly with the communication of benefits, with females placing a greater value on the communication of social benefits compared to males. This would imply that communication of social benefits would be more effective in increasing participation levels for females than it is for males.

Education was found to significantly affect choice behaviour directly. Significant interaction effects were also found between education and the communication of social benefits. Academic levels of education have a negative effect on the odds of participation. Academic levels of education also increase the positive effect of the communication of social benefits. These results imply that while academically educated people are less willing to disclose their data that this can be more than compensated for by providing them with clear communication as to the social benefits of participating in a given vehicle-data sharing system.

Socio-demographic factors relating to the car are discussed here together and include; yearly driven km, car ownership, and car type. The amount of yearly driven km was found to have a direct negative effect on participation, potentially due to increased perception of privacy risk due to an increased amount of data provided. Regarding car ownership type no significant effects were found in the estimated logit model. With regards to car type only hatchback drivers were found to have a significant positive direct effect on participation. These results imply that people who drive more are less willing to disclose data, while hatchback drivers are more willing to disclose data.

Summarizing

Privacy factors influencing user participation in a vehicle-data sharing system are related to the collection of data, processing of data, dissemination of data, and risks associated with privacy invasion. The collection of additional types of data has a negative effect on participation, with accelerometer data valued more negatively than the observation of the amount of vehicles passed. The risk of personal identification in the processing of data has a negative influence on participation which is notably higher than the combined effect of including extra sensors. However the most notable observations are in the dissemination of data. The dissemination of data with unnamed third parties is the most important factor for participation, with a very large negative effect and a WtA value of €124,14 per month. Sharing data with researchers and emergency services has a significant positive effect on participation levels. Age was found to interact with the dissemination of data, with older people being less willing to share data with third parties, and more willing to share data with researchers. Regarding the risk of privacy invasion it was found that high levels of privacy concern reduce participation levels, while high levels of Institutional Trust increase participation levels to a greater degree. Demonstrating that privacy concerns can be compensated through the building of trust between the data collector and participant.

The use of monetary incentives has been found to increase participation levels significantly. The positive effect of monetary compensation is reduced in individuals who have high levels of privacy concern. Alternate methods of compensation may also result in sufficient levels of participation. In particular people are significantly more likely to disclose data for free when they have not yet been presented with alternatives that offer monetary compensation. Furthermore the communication of social benefits may also provide inventive for people to participate in a VBSD sharing system, although the effect of the communication of social benefits is highly dependent on the socio-demographic factors of the recipient of the communication, people most sensitive to this communication are young, academically educated, and female.

6.2. Contributions

This section elaborates on the theoretical and managerial contribution of this thesis. The theoretical implications involve the scientific contributions of this study. The managerial implications are related to the contributions for direct practical purposes.

6.2.1. Theoretical Contributions

Utilitarian perspective on privacy concerns in the context of vehicle-data sharing system

While previous literature demonstrates that utilitarian trade-offs between perceived benefits and privacy related harms can occur, no such study has been performed relating in the context of sharing vehicle based sensing data with a governmental party for the purposes of short-term traffic prediction. While it has been stated in literature that the decentralized collection of data through road-vehicles has potential benefits, the feasibility of such a system in utilitarian terms has never been researched. With this study it is possible estimate whether a hypothesized vehicle-data collection system would be accepted and adopted by road-users. Thus this study aims to bridge the gap between hypothesized benefits and the practical implementation of such a vehicle-data collection system in a utilitarian manner.

Development of attitude measure for Privacy concern and Institutional Trust

The use of attitude metrics relating to privacy and trust has been extensively employed in different fields of privacy research. The adaptation of these attitude measures to the specific domain of vehicle-data sharing systems in the Netherlands represents a valuable step for measuring attitudes relating to Privacy concern and Institutional Trust in future studies relating to privacy concerns in the transport domain. The inclusion of these attitude measures may provide the basis for a standardized Dutch index for Privacy concern in general and Institutional Trust in the specific context of data collection by governmental parties.

Inclusion of both monetary and non-monetary compensation schemes

Generally in utilitarian studies in the transport domain benefits are only defined in monetary terms. The inclusion of non-monetary benefits in this study such as the communication of benefits and its significant interactions with socio-demographic factors. As well as the inclusion of effects relating to the order in which alternatives are presented provide an alternative method of providing compensation to road-user for participating. In particular, the effects of the order in which alternatives are presented highlight the fact that the assumed Independence of irrelevant alternatives that is present in logit models does not hold true in the context of participation in vehicle-data sharing systems. These alternate methods of providing social benefit as a method of compensation can be used in future research relating to vehicle-data sharing systems to provide alternatives to direct monetary compensation.

6.2.2. Managerial Contributions

The expanded logit model estimated in this study can be used to predict the odds of participation of a given person on the basis of a range of socio-demographic factors. This tool can be used by policy makers to aid in the design of feasible VBSD collection systems.

Road-users are generally willing to share their vehicle-data, provided data is not shared with third parties. Especially when they are not personally identifiable. This study also highlights the positive effect on participation from sharing their data with emergency services and for research purposes. Communication of social benefits has been found to increase participation from road-users who are young, highly educated, and female. With the sharing of data for research purposes increasing the participation of older road-users in particular.

Recommendations for implementing a vehicle-data sharing system are to start without offering monetary benefit, as people have been found to be more willing to disclose data for free when they have not yet been presented with alternatives offering monetary compensation. Clear communication of the purpose and the social benefits can increase participation levels, with the greatest effect of this communication being observed in recipients who are young, academically educated, and female. Being parsimonious in data collection will result in the least amount of privacy harm and the risk of the system being perceived as unfair and inefficient. A sufficiently parsimonious data-sharing system would allow for the collection of data without monetary compensation, which is plausible due to the general willingness to share vehicle-data. As a result it becomes possible to acquire fine-grained data from moving observers at a low cost. However if participation levels are not sufficient without monetary reward, the use of monetary reward can be implemented.

6.3. Limitations and further research

The first important limitation of this study is the bias present in the the survey sample. There is over representation present for highly educated people and people in the 18-35 year age interval. An imbalance in the male to female ratio is also present to a lesser degree. Due to this bias the estimated interaction effects may not be as strong as predicted. Additionally it is not known to what degree self-selection effects are present, whereby respondents who are generally more willing to participate may have been over-represented in the sample.

In addition, the limited sample size may have lead the inability to estimate the effect of socio-demographic variables at a significant level. A larger sample size may have improved significance levels for these factors, specifically on the factors relating to car ownership and car type.

A further limitation of this study is that all effects estimated are assumed to be linear in nature. Which may not be the case. This relates in particular to the estimation of monetary benefits, and the estimated interaction effects. Future studies are recommended to include testing for non-linearity on both main and interaction effects to account for non-linear effects in the valuation of these attributes.

The fourth limitation of this study is the unspecified nature of sharing with third parties. In the current study it is intentionally left unspecified to estimate the maximum level of privacy harm related to sharing data with an unknown number of unknown parties for unknown purposes. The inclusion of specific third parties and/or applications would improve the predictive power of the estimated model. Potentially users may even value specific third parties in a positive manner based on their level of trust and clarity of the purpose the data is shared for.

Attitudes relating to Privacy concern and Trust in the government have been included in the study and have been analyzed using factor analysis. The effect of socio-demographic variables on these attitudes have not been considered. In order to gain insight into the relationship between these attitudes and socio-demographic factors it is recommended that future studies apply structural equation modelling.

Finally, the reason behind the positive valuation of sharing data with emergency services and researchers is not known. This could be indicative of a moral component in the decision making process, which was not included in this study. As such it is recommended that future studies include an exploration of the moral component of decision making. Using the framework described by Chorus [2015] may provide a good starting point for defining and measuring the moral component of decision making in this context.

Bibliography

Ackerman, M. S., Cranor, L. F., and Reagle, J. Privacy in e-commerce: examining user scenarios and privacy preferences. In *Proceedings of the 1st ACM conference on Electronic commerce*, pages 1–8. ACM, 1999.

Barriball, K. L. Collecting data using a semi-structured interview: a discussion paper. *Journal of Advanced Nursing*, 19:328–335, 1994.

Ben-Akiva, M. E., Lerman, S. R., and Lerman, S. R. *Discrete choice analysis: theory and application to travel demand*, volume 9. MIT press, 1985.

Bliemer, M. C., Rose, J. M., and Chorus, C. G. Detecting dominance in stated choice data and accounting for dominance-based scale differences in logit models. *Transportation Research Part B: Methodological*, 102:83–104, 2017.

Bohnet, I. and Frey, B. S. Social distance and other-regarding behavior in dictator games: Comment. *American Economic Review*, 89(1):335–339, 1999.

Brandeis, L. and Warren, S. The right to privacy. *Harvard law review*, 4(5):193–220, 1890.

Buch, N., Velastin, S. A., and Orwell, J. A review of computer vision techniques for the analysis of urban traffic. *IEEE Transactions on Intelligent Transportation Systems*, 12(3): 920–939, 2011.

Bygrave, L. A. *Data privacy law: an international perspective*, volume 63. Oxford University Press Oxford, 2014.

CBS. Bevolking; kerncijfers. 2018.

Charness, G. and Gneezy, U. What's in a name? anonymity and social distance in dictator and ultimatum games. *Journal of Economic Behavior & Organization*, 68(1):29–35, 2008.

Chen, Y., Chong, S., Kash, I. A., Moran, T., and Vadhan, S. Truthful mechanisms for agents that value privacy. *ACM Transactions on Economics and Computation (TEAC)*, 4(3):13, 2016.

Chorus, C. G. Models of moral decision making: Literature review and research agenda for discrete choice analysis. *Journal of choice modelling*, 16:69–85, 2015.

Clarke, R. Introduction to dataveillance and information privacy, and definitions of terms. *Roger Clarke's Dataveillance and Information Privacy Pages*, 1999.

Davis, F. What do we mean by right to privacy. *SDL Rev.*, 4:1, 1959.

DeCew, J. W. *In pursuit of privacy: Law, ethics, and the rise of technology*. Cornell University Press, 1997.

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

Domencich, T. A. and McFadden, D. Urban travel demand-a behavioral analysis. Technical report, 1975.

Ebner, C. Reducing congestion with integrated network management (inm). *Conference for European Directors of Roads*, 2017.

Elia, J. Transparency rights, technology, and trust. *Ethics and Information Technology*, 11 (2):145–153, 2009.

Endo, T., Nawa, K., Kato, N., and Murakami, Y. Study on privacy setting acceptance of drivers for data utilization on connected cars. In *Privacy, Security and Trust (PST), 2016 14th Annual Conference on*, pages 82–87. IEEE, 2016.

Eriksson, L. and Bjørnskau, T. Acceptability of traffic safety measures with personal privacy implications. *Transportation Research Part F: Traffic Psychology and Behaviour*, 15(3):333–347, 2012.

EU. Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46. *Official Journal of the European Union (OJ)*, 59:1–88, 2016.

EVO, Fenedex, T. Economische wegwijzer 2016 - Alarmerende sijging fileschade. 2016.

Eyssartier, C. Acceptability of driving an equipped vehicle with drive recorder: the impact of the context. *IET intelligent transport systems*, 9(7):710–715, 2015.

Finn, R. L., Wright, D., and Friedewald, M. Seven types of privacy. In *European data protection: coming of age*, pages 3–32. Springer, 2013.

Grabicki, F. and Menges, R. Status quo bias and consumers' willingness to pay for green electricity: A discrete choice experiment with real economic incentives. In *Entscheidungsunterstützung in Theorie und Praxis*, pages 169–190. Springer, 2017.

Grant, R. W. *Strings attached: Untangling the ethics of incentives*. Princeton University Press, 2011.

Hanemann, W. M. Willingness to pay and willingness to accept: how much can they differ? *The American Economic Review*, 81(3):635–647, 1991.

Hann, I.-H., Hui, K.-L., Lee, T. S., and Png, I. P. The value of online information privacy: An empirical investigation. *New York Times*, 1, 2001.

Hensher, D. A., Rose, J. M., and Greene, W. H. *Applied choice analysis: a primer*. Cambridge University Press, 2005.

Herrera, J. C., Work, D. B., Herring, R., Ban, X. J., Jacobson, Q., and Bayen, A. M. Evaluation of traffic data obtained via gps-enabled mobile phones: The mobile century field experiment. *Transportation Research Part C: Emerging Technologies*, 18(4):568–583, 2010.

Hoepman, J.-H. Privacy design strategies. In *IFIP International Information Security Conference*, pages 446–459. Springer, 2014.

Jentzsch, N. State-of-the-art of the economics of cyber-security and privacy. *Browser Download This Paper*, 2016.

Jick, T. D. Mixing qualitative and quantitative methods: Triangulation in action. *Administrative science quarterly*, 24(4):602–611, 1979.

Johnson, F. R., Lancsar, E., Marshall, D., Kilambi, V., Mühlbacher, A., Regier, D. A., Bresnahan, B. W., Kanninen, B., and Bridges, J. F. Constructing experimental designs for discrete-choice experiments: report of the ispor conjoint analysis experimental design good research practices task force. *Value in Health*, 16(1):3–13, 2013.

Kang, J. and Hustvedt, G. Building trust between consumers and corporations: The role of consumer perceptions of transparency and social responsibility. *Journal of Business Ethics*, 125(2):253–265, 2014.

Karniouchina, E. V., Moore, W. L., van der Rhee, B., and Verma, R. Issues in the use of ratings-based versus choice-based conjoint analysis in operations management research. *European Journal of Operational Research*, 197(1):340–348, 2009.

Kearns, M., Pai, M., Roth, A., and Ullman, J. Mechanism design in large games: Incentives and privacy. In *Proceedings of the 5th conference on Innovations in theoretical computer science*, pages 403–410. ACM, 2014.

Kehr, F., Kowatsch, T., Wentzel, D., and Fleisch, E. Blissfully ignorant: the effects of general privacy concerns, general institutional trust, and affect in the privacy calculus. *Information Systems Journal*, 25(6):607–635, 2015.

Kumaraguru, P. and Cranor, L. F. Privacy indexes: a survey of westin's studies. 2005.

Leduc, G. Road traffic data: Collection methods and applications. *Working Papers on Energy, Transport and Climate Change*, 1(55), 2008.

Loomis, J. What's to know about hypothetical bias in stated preference valuation studies? *Journal of Economic Surveys*, 25(2):363–370, 2011.

Louviere, J. J., Hensher, D. A., and Swait, J. D. *Stated choice methods: analysis and applications*. Cambridge university press, 2000.

Lyu, S. O. Which accessible travel products are people with disabilities willing to pay more? a choice experiment. *Tourism Management*, 59:404–412, 2017.

Moore, A. Defining privacy. *Journal of Social Philosophy*, 39(3):411–428, 2008.

Morton, B. In pursuit of privacy: Law, ethics, and the rise of technology., 1998.

Murphy, J. J., Allen, P. G., Stevens, T. H., and Weatherhead, D. A meta-analysis of hypothetical bias in stated preference valuation. *Environmental and Resource Economics*, 30(3): 313–325, 2005.

Myers, L., Sirois, M., Kotz, S., Read, C., Balakrishnan, N., and Vidakovic, B. Encyclopedia of statistical sciences, 2006.

Nettle, D., Harper, Z., Kidson, A., Stone, R., Penton-Voak, I. S., and Bateson, M. The watching eyes effect in the dictator game: It's not how much you give, it's being seen to give something. *Evolution and Human Behavior*, 34(1):35–40, 2013.

Orme, B. Interpreting the results of conjoint analysis. *Getting Started with Conjoint Analyis: Strategies for Product Design and Pricing Research*, 2:77–88, 2010.

Palmer, M. Subscribe to the ft to read: Financial times tomtom sorry for selling driver data to police, Apr 2011. URL `https://www.ft.com/content/3f80e432-7199-11e0-9b7a-00144feabdc0`.

Papageorgiou, M., Hadj-Salem, H., Blosseville, J.-M., et al. Alinea: A local feedback control law for on-ramp metering. *Transportation Research Record*, 1320(1):58–67, 1991.

Parent. Privacy, morality, and the law. *Philosophy & Public Affairs*, 12:269–288, 1983.

Prosser, W. L. Privacy. *California Law Review*, 48:383, 1960.

Riley, P. F. The tolls of privacy: An underestimated roadblock for electronic toll collection usage. *Computer Law & Security Review*, 24(6):521–528, 2008.

Salami, E. A. An analysis of the general data protection regulation (eu) 2016/679. 2017.

Schäfer, R.-P., Thiessenhusen, K.-U., Brockfeld, E., and Wagner, P. A traffic information system by means of real-time floating-car data. 2002.

Schläpfer, F. Stated preferences for public services: A classification and survey of approaches. *Journal of Economic Surveys*, 31(1):258–280, 2017.

Schoettle, B. and Sivak, M. A survey of public opinion about connected vehicles in the us, the uk, and australia. In *Connected Vehicles and Expo (ICCVE), 2014 International Conference on*, pages 687–692. IEEE, 2014.

Schoot, T., Pavlova, M., Atanasova, E., and Groot, W. Preferences of bulgarian consumers for quality, access and price attributes of healthcare services—result of a discrete choice experiment. *The International journal of health planning and management*, 32(1), 2017.

Schulze, R., Hilger, A.-L., and Engelberg, P. M. Factor analysis and latent variable models in personality psychology. 2015.

Seo, T., Kusakabe, T., and Asakura, Y. Estimation of flow and density using probe vehicles with spacing measurement equipment. *Transportation Research Part C: Emerging Technologies*, 53:134–150, 2015.

Seo, T., Bayen, A. M., Kusakabe, T., and Asakura, Y. Traffic state estimation on highway: A comprehensive survey. *Annual Reviews in Control*, 43:128–151, 2017. ISSN 13675788. doi: 10.1016/j.arcontrol.2017.03.005. URL `http://dx.doi.org/10.1016/j.arcontrol.2017.03.005`.

Sheehan, K. B. and Hoy, M. G. Dimensions of privacy concern among online consumers. *Journal of public policy & marketing*, 19(1):62–73, 2000.

Simon, M. K. *Dissertation and scholarly research: recipes for success: a practical guide to start and complete your dissertatin, thesis, or formal research project.* Kendall/Hunt Publishing Company, 2006.

Solove. A taxonomy of privacy. *University of Pennsylvania law review*, pages 477–564, 2006.

Thomson, J. J. The right to privacy. *Philosophy & Public Affairs*, pages 295–314, 1975.

van Erp, P. B., Knoop, V. L., and Hoogendoorn, S. P. Macroscopic traffic state estimation using relative flows from stationary and moving observers. *Transportation Research Part B: Methodological*, 114:281–299, 2018.

van Erp, P. B., Knoop, V. L., and Hoogendoorn, S. P. On the value of relative flow data paul. *Transportation Research Part C: Emerging Technologies*, 2019.

van Infrastructuur en Waterstaat, M. Mobiliteitsbeeld 2016-mobiliteitsbeeld en kerncijfers mobiliteit 2017-kennisinstituut voor mobiliteitsbeleid. 2017.

Vlahogianni, E. I., Karlaftis, M. G., and Golias, J. C. Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies*, 43: 3–19, 2014. ISSN 0968090X. doi: 10.1016/j.trc.2014.01.005. URL `http://dx.doi.org/10.1016/j.trc.2014.01.005`.

Walter, J. and Abendroth, B. Losing a private sphere? a glance on the user perspective on privacy in connected cars. In *Advanced Microsystems for Automotive Applications 2017*, pages 237–247. Springer, 2018.

Wee, B. V. and Banister, D. How to write a literature review paper? *Transport Reviews*, 36 (2):278–288, 2016.

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

Westin, A. F. Harris-equifax consumer privacy survey 1991. *Atlanta, GA: Equifax Inc*, 1991.

Yoon, T., Cherry, C. R., and Jones, L. R. One-way and round-trip carsharing: A stated preference experiment in beijing. *Transportation Research Part D: Transport and Environment*, 53:102–114, 2017.

A
Interviews

Interview 1

Leeftijd: 53, Geslacht: V, Opleidingsniveau: MBO

Hoe vaak gebruikt u de auto?

4 dagen in de week, gemiddeld.

Voor welke doeleinden gebruikt u de auto?

Meeste voor werk. Niet veel last van files e.d., maar als ik er in sta vind ik het heel irritant. Ik ontwijk ze nu met behulp van een navigatie systeem.

De data die gedeeld zou kunnen worden zijn bijvoorbeeld, de locatie, snelheid en het aantal gepasseerde auto's. Zou u bereid zijn deze data te delen? Waarom wel/niet?

Waarom wel, ik heb niets te verbergen, ze mogen weten waar ik ben. Als je erg op privacy gesteld bent doe je dat niet, maar ben er zelf niet moeilijk in, zeker als je weet dat het voor deze doeleinden wordt gebruikt. In de dagelijkse zin niet veel met privacy bezig. Met het delen van de data, geen probleem. Je staat er niet zo bij stil wat er met je gegevens gedaan kan worden, daar ben ik misschien wat naief in. Als iedereen mij kan volgen is dat niet fijn, maar als het alleen bij de relevante instantie komt is het geen probleem. Daarnaast kunnen mensen misschien toch bij de data, eventueel moet ik dan zeggen, niet herkenbaar. Herkenbaarheid is alleen ok als je zeker weet dat het bij die instantie blijft.

Indien het mogelijk zou zijn om u te herkennen in de data, mogelijk ten behoeve van de nationale veiligheid of de politie, zou u daarmee akkoord gaan? Waarom?

Ja, dat zou ik ook geen probleem vinden, want als ze dat gebruiken voor het oplossen van misdaden is dat een goed doel. Continu monitoren is geen probleem, want flitspalen staan er ook al. Laat de hardrijders lekker betalen. Ik heb ook bijna nooit verkeersboetes, voor mij staat veiligheid op de weg voorop.

Wat zou u er van vinden als de data (wel of niet herkenbaar?) met andere partijen gedeeld zou worden, zoals onderzoeksinstituten, universiteiten en andere relevante partijen?

Het hangt er vanaf wat voor onderzoeken het zijn, maar commercieel zou ik het niet delen. Herkenbaarheid maakt uit, commerciel mogen ze de gegevens alleen onherkenbaar.

Een partij als tomtom zou bijv. wel anonieme data mogen. Het belangrijkste is dat de data binnen het domein van verkeer wordt gehouden.

Is het verbeteren van de verkeers voorspellingen en zodoende de verkeersstroom reden genoeg voor u om uw data te delen? Waarom wel/niet?

Ja, het verbeteren van het verkeer is nodig in nederland, dus daar zou ik aan meewerken. Er zou geen compensatie nodig zijn, als je er geen moeite voor hoeft te doen.

Als men gecompenseerd zou moeten worden, op welke manier zou u dat wenselijk vinden?

Een goede compensatie zou bijvoorbeeld zijn een korting op de wegenbelasting, de hoogte van de compensatie is een percentage van wat je betaalt en dat het aantal kilometers er in meegenomen wordt. Daarnaast zou het mogelijk moeten zijn zelf te kiezen welke data je deelt, en afhankelijk hiervan, een bepaald bedrag. Het aan/uit kunnen zetten van de data die je deelt heeft een positief effect op het aantal mensen die mee zouden doen. Het geven van een compensatie zou mensen meer bereidheid geven de data te delen. Geen data die ik absoluut niet zou delen.

Belangrijkste redenen om het te doen is dat het voor iedereen gunstig is, als ze de vertragingen kunnen minderen is dat het waard. Zo ook met het oplossen van misdaden.

Belangrijskte redenen om het niet te doen is als je iets te verbergen hebt en ook de privacy ook natuurlijk, als andere partijen zouden kunnen inbreken in het systeem. Het aantonen van de veiligheid van een syteem is onmogelijk. Een opt-in opt-out systeem zou in ieder geval wel nodig zijn. Maar je stopt er mee of je gaat door, niet voor een paar uur, veel te gevoelig voor strategisch gedrag. Eens in de zes maanden kunnen switchen.

Interview 2

Leeftijd: 25, Geslacht: M, Opleidingsniveau: WO

Hoe vaak gebruikt u de auto?

4 dagen in de week, totaal zo'n 10 keer.

Voor welke doeleinden gebruikt u de auto?

Woon-/werk verkeer. Valt wel mee hoe vaak last van vertragingen en files.

De data die gedeeld zou kunnen worden zijn bijvoorbeeld, de locatie, snelheid en het aantal gepasseerde auto's. Zou u bereid zijn deze data te delen? Waarom wel/niet?

Ja, als er geen sancties aan hangen. Positie en ingehaalde auto's prima. Snelheid constant mete, prima als er geen sancties/boetes aan hangen. Omdat ik toch wel een zwaardere voet heb. Het belangrijkste is dat er alleen voordelen aan het delen van de data hangen (betere voorspellingen)

Indien het mogelijk zou zijn om u te herkennen in de data, mogelijk ten behoeve van de nationale veiligheid of de politie, zou u daarmee akkoord gaan? Waarom?

Ik zou er mee akkoord zijn als er geen nadelige gevolgen zijn voor mij, plus als het niet standaard ingezien kan worden met herkenning. Dus alleen met een geldige reden bij een lopend onderzoek met criminele zaken. (moord/vermissingen/diefstal) Dan ook niet als bijzaak boetes uitdelen.

Wat zou u er van vinden als de data (wel of niet herkenbaar?) met andere partijen gedeeld zou worden, zoals onderzoeksinstituten, universiteiten en andere relevante partijen?

Ze mogen mijn data hebben maar dan wel anoniem. Omdat je anders geen gevoel meer hebt wie wel/niet data heeft. Ik vind het niet nodig dat zij weten dat ik mijn boodschappen bij die precieze winkel doe. Ik zie de voordelen voor derde partijen in de vorm van big data, maar de persoon hoeven ze niet te weten. Ook voor verzekeringen niet persoonlijk herkenbaar, maar patronen herkennen dat zou mogen. Maar niet van een verzekering afsluiten en dat ze in kunnen zien dat ik te hard rij. Per datum ben ik een andere identiteit.

Is het verbeteren van de verkeers voorspellingen en zodoende de verkeersstroom reden genoeg voor u om uw data te delen? Waarom wel/niet?

Ja, maar ik zie nog niet hoe ze de stromen gaan verbeteren. Het is altijd mooi meegenomen als je er geld voor zou krijgen, maar het komt van de belastingcenten, dus dan gaat de belasting omhoog. Belangrijkste is dat ze verzekeren dat je er geen nadelige gevolgen aan hangen. Mochten ze gaan monitoren dan moet er wel wat tegenover staan. Niet bereid boetes op basis van de data te krijgen, ondanks een compensatie mechanisme.

Als men gecompenseerd zou moeten worden, op welke manier zou u dat wenselijk vinden?

Een soort reductie op de wegenbelasting, want je doet iets voor de wegen, voor het verkeer. Als je er voor gecompenseerd worden en er zijn mogelijke nadelige gevolgen, dan moet het hoger zijn dan een enkele keer een boete krijgen.

Het (hard rijden) gebeurt sneller dan je denkt. Bijv. een boete in het kwartaal "gratis" kunnen veroorloven, zoals een beetje te hard rijden. Want als je veel rijdt is het nuttig om te weten wie er rijden, maar dan ga je een keer een boete krijgen. Een compensatie in het aantal Km is misschien nuttiger, hoe meer Km je rijdt hoe nuttiger je data is.

Een dergelijke regeling zou zeker het gedrag beinvloeden, zoals een continue trajectcontrole. Mijn naam en gegevens zou ik nooit delen behalve in een criminele zaak.

Interview 3

Leeftijd: 61, Geslacht: M, Opleidingsniveau: WO

Hoe vaak gebruikt u de auto?

1 keer per week

Voor welke doeleinden gebruikt u de auto?

Boodschappen, familiebezoek en naar de boot te gaan. Zondag middag op de a4 wel eens last van de files.

De data die gedeeld zou kunnen worden zijn bijvoorbeeld, de locatie, snelheid en het aantal gepasseerde auto's. Zou u bereid zijn deze data te delen? Waarom wel/niet?

In principe wel, als ik weet waar het verder gebruikt zal worden. Ik heb al een slimme tomtom. Als het herkenbaar is, ga ik aarzelen, ik symphatiseer met het idee voor verkeers doorstroming. Op moment dat mijn naam er aan hangt, dan wordt het een ander verhaal, eerste reactie is: dan geef ik eht niet zomaar vrij. Dan moeten daar garanties aan hangen dat het niet tegen mij gebruikt kan worden. Ik weet niet hoe ze dat kunnen aantonen, dat is aan de partij die de data wilt. Als het anoniem en niet traceerbaar is, dan heeft het mijn enorme sympathie.

Indien het mogelijk zou zijn om u te herkennen in de data, mogelijk ten behoeve van de nationale veiligheid of de politie, zou u daarmee akkoord gaan? Waarom?

Daar zou ik wel mee akkoord kunnen gaan, als het aannemelijk is dat het alleen daarvoor is. Enerzijds, voor boetes zijn er ook camera's die alleen voor observatie en niet boetes gebruikt worden. Als er reden voor is zou ik het wel goed vinden, net als foulleren, ze mogen het niet zomaar zonder reden. Niet boetes uitdelen op basis van continue observatie. Wel als aangegeven trajectmetingen eventueel. Wel voor het oplossen van misdrijven. Transparantie heel belangrijk.

Wat zou u er van vinden als de data (wel of niet herkenbaar?) met andere partijen gedeeld zou worden, zoals onderzoeksinstituten, universiteiten en andere relevante partijen?

Commercieel aarzel ik mee, het idee van de ANWB van een sensor plaatsen voor korting, sympathiek, maar ik heb gehoord dat dat de andere kant op werkt. Als ik niet herkenbaar ben minder bezwaar, niet heel goed over nagedacht, maar niet voor zuiver commerciele partijen, wel relevante partijen die wat bijdragen. Uni's en onderzoekspartijen, geen probleem als het anoniem is, niet-anoniem vind ik het niet leuk omdat er te makkelijk persoonlijke informatie uit te halen is.

Is het verbeteren van de verkeers voorspellingen en zodoende de verkeersstroom reden genoeg voor u om uw data te delen? Waarom wel/niet?

Wederom, als het anoniem is heb ik daar geen bezwaar tegen, het mag ook individueel herkenbaar zijn maar mijn naam moet er niet aan gekoppeld zijn. Van mij zou de financiele prikkel niet nodig zijn, voor mij is het interessant genoeg en denk ik dat het een richting is waar we sowieso heen moeten. Bij boetes door continue monitoring zou ik niet meedoen, ook al zou hier een financiele beloning tegenover staan.

Als men gecompenseerd zou moeten worden, op welke manier zou u dat wenselijk vinden?

Het feit dat ik terugkoppeling krijg over de data zou ik al leuk vinden. Als daar een financiele tegemoetkoming tegenoverstaat is dat leuk, maar niet een voorwaarde.

Het zelf kunnen bepalen wat je deelt, als ik uit kan zetten dat de politie mij continu kan volgen is dat een goed idee. Als ik het daarmee kan tunen dat wat we hebben besproken doorgestuurd akn worden dan is dat goed.

Als commerciele partijen continu mee zouden kunnen kijken dan is dat een dealbreaker, net zoals als het niet aangetoond kan worden dat het voldoende afgeschermd kan worden. De belangrijkste punten zijn transparantie, controle en veiligheid.

Interview 4

Leeftijd: 22, Geslacht: V, Opleidingsniveau: WO

Hoe vaak gebruikt u de auto?

Verschilt heel erg, maar in de toekomst zo'n 2-3 keer per week.

Voor welke doeleinden gebruikt u de auto?

Om naar mn stage te gaan en naar een rugby wedstrijd te gaan. Op wedstrijd dagen geen last van file, ik verwacht wel onderweg naar mn stage last te hebben van files en vertragingen.

De data die gedeeld zou kunnen worden zijn bijvoorbeeld, de locatie, snelheid en het aantal gepasseerde auto's. Zou u bereid zijn deze data te delen? Waarom wel/niet?

Locatie is best privacy gevoelig, op dit moment niet heel erg, maar het kan een probleem zijn. Als het echt gekoppeld is aan je auto zou ik het misschien niet delen. (herkenbaarheid)

Indien het mogelijk zou zijn om u te herkennen in de data, mogelijk ten behoeve van de nationale veiligheid of de politie, zou u daarmee akkoord gaan? Waarom?

Nee, denk het niet, omdat je ook ergens onzichtbaar en anoniem moet kunnen zijn. Bij misdrijven en ongelukken wel bereid toegang te geven tot die data, maar niet voor boetes e.d.

Wat zou u er van vinden als de data (wel of niet herkenbaar?) met andere partijen gedeeld zou worden, zoals onderzoeksinstituten, universiteiten en andere relevante partijen?

Commerciele partijen sowieso niet. Ook niet als het niet herkenbaar is, dat voelt niet veilig met al die schandalen zoals bijv. facebook, zeker met alle hackers die er tegenwoordig zijn. Het is heel belangrijk om zeker te weten dat je data veilig is. Voor mij overtuigen is open communicatie van wat ze er mee doen en het inzien van je eigen data belangrijk. Ook het kiezen wat wel en niet te delen is belangrijk, feedback terug krijgen zoals bijv de resultaten van een onderzoek, zodat je weet waar je het voor doet.

Voor derde partijen ook onderscheid tussen wel/niet gerelateerde bedrijven, ANWB is best betrouwbaar en ze helpen je op de weg. Een tomtom bijvoorbeeld niet. Een compensatie voor het delen met derde partijen is wel op zn plaatst. Een bedrijf wat billboards plaatst langs de weg zou ik bijv. niet mee delen.

Voor onderzoek, niet herkenbaar,

Is het verbeteren van de verkeers voorspellingen en zodoende de verkeersstroom reden genoeg voor u om uw data te delen? Waarom wel/niet?

Ja opzich wel, want ik denk wel dat het heel veel zou oplossen qua files. Een financiele prikkel zou fijn zijn, maar als je 20m minder in de file staat is dat ook een winst.

Als men gecompenseerd zou moeten worden, op welke manier zou u dat wenselijk vinden?

Eventueel geld om in het OV te reizen, omdat je dan het OV stimuleert en het OV goedkoper maakt om in te reizen. Continue politie monitoring bij financiele compensatie, nog steeds niet akkoord, want je wilt niet dat de politie over je schouder meekijkt. Als het oorlog is zou ik ook geen data delen. Maar er moet vooral gekeken worden hoe veilig het is en dat het niet gehackt kan worden. Als de overheid merkt dat er veel aanvallen zijn, dan even het systeem uitzetten en geen gegevens ontvangen. Als het een keer misgaat zou ik er helemaal mee stoppen.

Niet heel erg bewust van privacy, beetje laks. Toch vertrouw je partijen als facebook een beetje, zelfs als die bedrijven de dinfo hebben denk ik, ja, maar wat kan je er eigenlijk mee.

Interview 5

Leeftijd: 64, Geslacht: M, Opleidingsniveau: HBO

Hoe vaak gebruikt u de auto?

1 of 2 keer per week

Voor welke doeleinden gebruikt u de auto?

Meestal voor winkelen en soms om familie te bezoeken, soms ook werkverkeer Niet vaak last van vertragingen, wel vaker last bij werkverkeer

De data die gedeeld zou kunnen worden zijn bijvoorbeeld, de locatie, snelheid en het aantal gepasseerde auto's. Zou u bereid zijn deze data te delen? Waarom wel/niet?

Ja, volgens mij doe ik dat al, ik gebruik TomTom navigatie, die deelt al data met de centrale server. Als je er voordeel van hebt zoals kortere vertragingen, dan heeft dat zeker zin om dat te delen. Als tomtom geen verbinding heeft vind ik dat erg frustrerend.

Indien het mogelijk zou zijn om u te herkennen in de data, mogelijk ten behoeve van de nationale veiligheid of de politie, zou u daarmee akkoord gaan? Waarom?

Ik heb in principe niets te verbergen, dus ik kan daarmee akkoord gaan. Dat is anders als het voor private partijen is. Niet voor continue monitoring of beboeten van mensen. Bestuurder moet kunnen beslissen om snelheid te verhogen, zoals bijvoorbeeld inhalen of uit een moeilijke situatie komen als je ingesloten raakt. Dus het zou niet goed zijn daar gelijk een prent voor te geven. Wel om te traceren waar je bent als er noodsituaties zijn. Oplossen van misdrijven is ook een goede reden. Evt contact leggen met mensen, als het waardevolle informatie kan hebben. Onafhankelijke instantie zoals OM zou moeten beslissen of data zomaar in te zien is.

Wat zou u er van vinden als de data (wel of niet herkenbaar?) met andere partijen gedeeld zou worden, zoals onderzoeksinstituten, universiteiten en andere relevante partijen?

Onderzoek is prima, als het anoniem is. Voor mij geen reden daar persoonlijke gegevens aan te koppelen. Derde partijen, commerciele partijen, absoluut niet. Die gebruiken dat voor reclame e.d. Dat vind ik niet juist. Overheid is en moet onafhankelijk zijn, derde partijen hebben andere commerciele belangen en proberen reclame te verpreiden.

Is het verbeteren van de verkeers voorspellingen en zodoende de verkeersstroom reden genoeg voor u om uw data te delen? Waarom wel/niet?

Opzich wel ja, dat is het belangrijkste argument, dat je goede voorspelling kan doen. De data die je moet delen zal ook de bestemming moeten bevatten. Als iedereen van a naar b moet, dan kan je beter spreiden, bijv over a4 en a13. Maatschappelijk belang is reden genoeg om de data te delen.

Als men gecompenseerd zou moeten worden, op welke manier zou u dat wenselijk vinden?

Lagere belasting, wegenbelasting. Maak je beter gebruik van de infrastructuur, dus dat kan betekenen dat je daar minder geld in hoeft te stoppen.

Reactie op OV tegoed: Niet zoveel zoden aan de dijk in mijn geval, daar gebruik ik de auto te weinig voor. Meestal gebruik ik de fiets. Heb ook al een OV business card vanuit werk.

Andere partijen (derde partijen) tege vergoeding: Nee, teveel gesteld op de privacy. Overheid is maatschappelijk belang. Commercieel belang wil ik zo min mogelijk data mee delen.

Snellere reistijden en minder oponthoud is de voornaamste reden om mee te doen (mobiliteit). Als er nood is, dat de politie je kan vinden (veiligheid).

Inbreuk op privacy is de voornaamste reden om het niet te doen.

Interview 6

Leeftijd: 25, Geslacht:V, Opleidingsniveau: HBO

Hoe vaak gebruikt u de auto?

Dagelijks, 7 keer per week

Voor welke doeleinden gebruikt u de auto?

Vooral voor mijn werk, boodschappen en uitjes. Vaak last van vertragingen, ik reis voor mijn werk het hele land door. Ik be drie dagen in de week bij klanten, dus ik reis veel, en dan sta ik in de spits standaard vast.

De data die gedeeld zou kunnen worden zijn bijvoorbeeld, de locatie, snelheid en het aantal gepasseerde auto's. Zou u bereid zijn deze data te delen? Waarom wel/niet?

Ik zou het anoniem delen op moment dat het me helpt om niet meer in de file te staan. Mijn exacte locatie op het exacte moment weet ik niet, als het echt anoniem is dan wel. Een app die dat doet gebruik ik dan wel (ways), daarmee kan ik zelf ook zien waar files ontstaan.

Ik zou niet willen dat ze een gebruiker van me maken, ik wil prima info delen maar niet een vaste gebruiker zijn. Per rit een ID zou mij geruster stellen.

Indien het mogelijk zou zijn om u te herkennen in de data, mogelijk ten behoeve van de nationale veiligheid of de politie, zou u daarmee akkoord gaan? Waarom?

Nee, daar zou ik niet mee akkoord gaan, politie moet dat alleen doen als er een verdachte is. Het monitoren voor boetes absoluut niet. Als het een duidelijk doel heeft om een misdaad op te lossen, maar niet op voorhand de data vrijgeven. Het is een inbreuk op mn privacy omdat ik niet vind dat iemand die info over mij hoort te hebben omdat ze het a) niet nodig hebben en b) ze het altijd tegen me kunnen gebruiken. Het is gewoon mijn privacy.

Wat zou u er van vinden als de data (wel of niet herkenbaar?) met andere partijen gedeeld zou worden, zoals onderzoeksinstituten, universiteiten en andere relevante partijen?

Nee sowieso niet mee akkoord gaan als het herkenbaar is.

Als het anoniem is, dan ook niet voor de commerciele partijen. Ik vind het niet nodig dat zij een database opbouwen met deze gegevens over mij. Bijv. wel als het eigenbelang voor mij groter is dat het commerciele belang zoals bij ways, zoals waar files, flitspalen controles, dan is het voor mij heel nuttig, en daarmee help je ook anderen.

Voor onderzoeksdoeleinden tot op zekere hoogte, zolang het anoniem is.

Is het verbeteren van de verkeers voorspellingen en zodoende de verkeersstroom reden genoeg voor u om uw data te delen? Waarom wel/niet?

Ja, puur eigenbelang, want ik sta heel veel uur per week in de file.

Als men gecompenseerd zou moeten worden, op welke manier zou u dat wenselijk vinden?

Dat zou ik niet zo kunnen bedenken. Wegenbelasting betaal ik niet, leaseauto. OV krediet, tot op zekere hoogte. Privacy is belangrijker dan een compensatie, ik wil niet men weet hoe ik beweeg. Anoniem zou ik het evt wel voor een korting doen. Tegen compensatie wel bereid zijn te delen met commerciele partijen? Weet ik niet. Het ligt er aan hoe ze mij kunnen overtuigen, sowieso wil ik inzicht in welke informatie hebben van mij en wat ze er mee doen. Als het anoniem is en ze hebben een goede reden dan zou ik het overwegen, daarnaast moet ik het makkelijk kunnen vinden en aan/uit te zetten. Ik wil niet het idee hebben dat ik gebruikt word en dat alles standaard aan staat.

Absoluut niet delen voor commerciele doeleinden waar ik zelf niet beter van wordt. Als ik er een misdrijf mee kan oplossen dan zou ik het delen, dat dient een hoger doel. Bij een ongeluk, naderhand zou kunnen. In geval zelf bij het ongeluk betrokken, ook akkoord, (als bijv. airbag open gaat) want daar ben ik zelf bij gebaat.

Gesteld op privacy, transparantie is heel belangrijk, eigenbelang is heel belangrijk.

B

Slides of survey introduction film

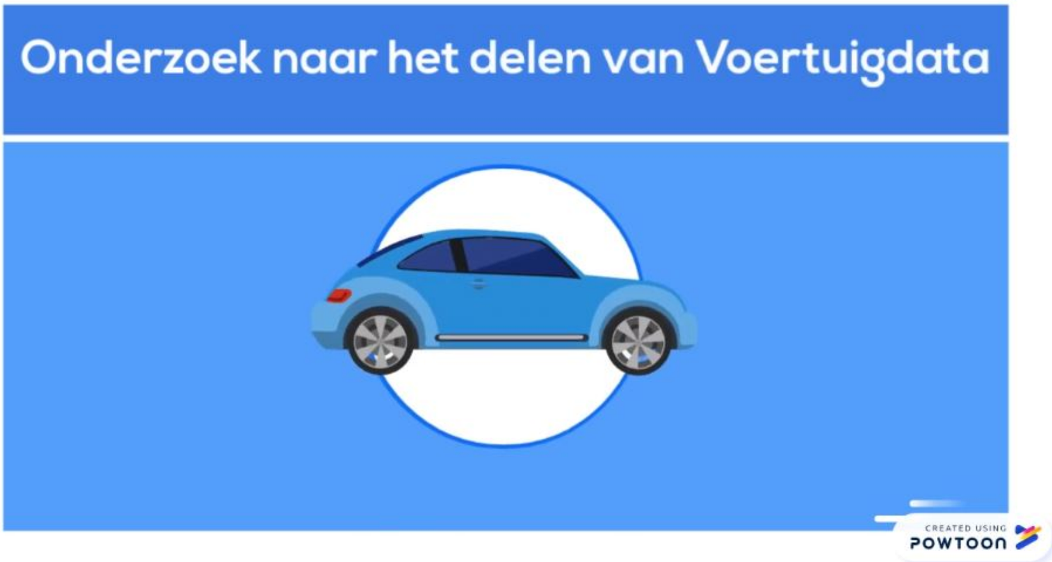


Figure B.1: Slide 1

Figure B.2: Slide 2, Communication of social benefits, only shown to part of the respondents as context variable

Figure B.3: Slide 3

Figure B.4: Slide 4

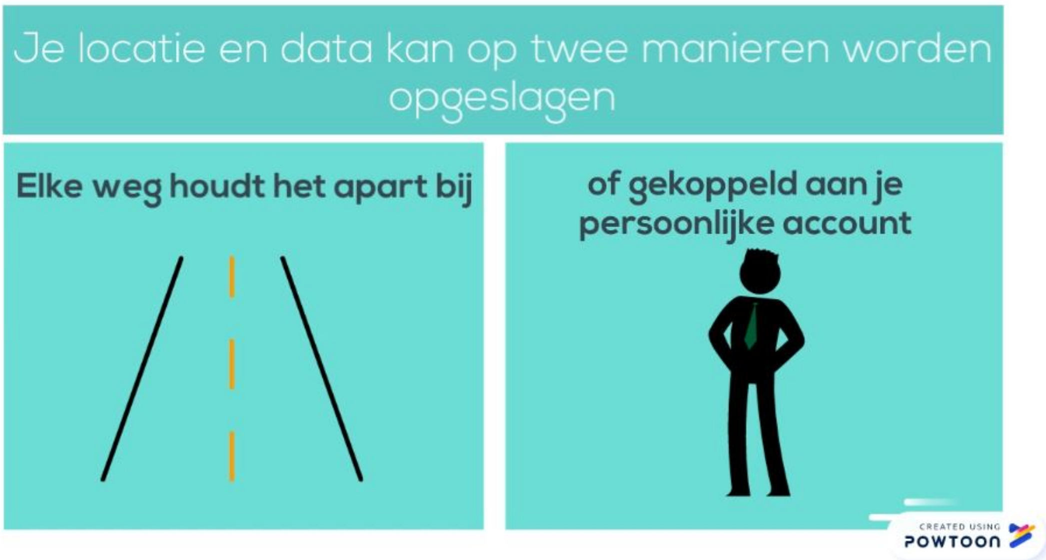


Figure B.5: Slide 5

Figure B.6: Slide 6

Figure B.7: Slide 7

Figure B.8: Slide 8

Figure B.9: Slide 9

C
Online survey

Figure C.1: Example of choice situation with all icons shown

Onderzoek delen van voertuigdata

(untitled)

Deze enquête is onderdeel van mijn master thesis aan de TU Delft. Deze enquête verkent de voorkeuren die automobilisten hebben bij de keuze om wel of niet voertuigdata te delen met de overheid. Momenteel wordt data over het verkeer vooral verzameld door lussen in de weg. Dit geeft slechts beperkt inzicht in de hoeveelheid verkeer. Een beter inzicht wordt verkregen als auto's informatie zouden registreren over het verkeer. In deze enquête dient u aan te nemen dat ook uw auto is uitgerust met sensors en dat uw auto dus ook verkeersdata kan registreren. Uw toestemming is nodig om deze data te kunnen gebruiken. De uitkomsten van deze enquete zullen gebruikt worden om adviezen te formuleren over mogelijke implementatie van een data-deel systeem.

U wordt verzocht de vragen aandachtig te lezen en naar waarheid in te vullen.

De enquete duurt ongeveer 10 - 15 minuten en de informatie verkregen via deze enquête zal alleen gebruikt worden voor wetenschappelijke doeleinden en is strikt anoniem.

Hartelijk bedankt voor uw deelname,

Alex de Jong

(untitled)

Bekijk hier het filmpje voor uitleg

Mocht u in het filmpje iets gemist hebben, dan kunt u hier de verschillende elementen nakijken.

In deze enquete zie je steeds een mogelijk data-deel systeem voor het verzamelen van voertuigdata. Ga er van uit dat uw auto de nodige sensoren heeft. De voorgestelde systemen hebben een verzameling aan elementen: welke data wordt verzameld, met wie deze data wordt gedeeld, welke controle u heeft over uw data en mogelijk een vergoeding. Bij elk voorgesteld systeem heeft u steeds de optie om wel of niet mee te doen.

Verzamelde data

Locatie data: De locatie wordt altijd bijgehouden. De locatie en andere data kan op twee manieren worden opgeslagen: ***per weg of gekoppeld aan uw persoonlijke account***.

Aantal gepasseerde voertuigen: Dit betekent dat sensoren bijhouden hoe veel voertuigen u passeert in beide richtingen.

Versnellingsmeter: De versnellingsmeter meet de mate waarin een auto optrekt of afremt. Dus niet de snelheid van de auto.

Data delen

De verzamelde data kan mogelijk met drie partijen gedeeld worden: **nooddiensten, onderzoeksinstellingen, en derde partijen**.

Nooddiensten: Zoals de ambulance, brandweer en politie. (Niet voor boetes)

Onderzoek: Partijen zoals Universiteiten en Kennisinstituten.

Derde partijen: Alle andere partijen, zoals bedrijven en andere organisaties.

Extra opties

Aan-/Uit- Functionaliteit: Maakt het mogelijk om de dataverzameling tijdelijk uit te zetten. Voor deelname moet in minimaal 80% van de ritten data worden verzameld.

Compensatie: Mogelijk wordt er in ruil voor het delen van uw data een maandelijkse compensatie geboden. Dit zal gebeuren in de vorm van een maandelijks bedrag in euro's.

Bekijk hier het filmpje voor uitleg

Mocht u in het filmpje iets gemist hebben, dan kunt u hier de verschillende elementen nakijken.

In deze enquete zie je steeds een mogelijk data-deel systeem voor het verzamelen van voertuigdata. Ga er van uit dat uw auto de nodige sensoren heeft. De voorgestelde systemen hebben een verzameling aan elementen: welke data wordt verzameld, met wie deze data wordt gedeeld, welke controle u heeft over uw data en mogelijk een vergoeding. Bij elk voorgesteld systeem heeft u steeds de optie om wel of niet mee te doen.

Met de verzamelde voertuigdata zullen de metingen en schattingen van het verkeer worden verbeterd. Dit maakt het mogelijk om verkeersstromen beter te managen, met als gevolg minder vaak en minder ernstig last van files en opstoppingen. Daarnaast wordt verwacht dat het verzamelen van voertuigdata tot een veiligere wegsituatie zal leiden doordat nooddiensten sneller en effectiever op kunnen treden wanneer zich onveilige situaties voor doen.

Verzamelde data

Locatie data: De locatie wordt altijd bijgehouden. De locatie en andere data kan op twee manieren worden opgeslagen: ***per weg of gekoppeld aan uw persoonlijke account***.

Aantal gepasseerde voertuigen: Dit betekent dat sensoren bijhouden hoe veel voertuigen u passeert in beide richtingen.

Versnellingsmeter: De versnellingsmeter meet de mate waarin een auto optrekt of afremt. Dus niet de snelheid van de auto.

Data delen

De verzamelde data kan mogelijk met drie partijen gedeeld worden: **nooddiensten, onderzoeksinstellingen, en derde partijen**.

Nooddiensten: Zoals de ambulance, brandweer en politie. (Niet voor boetes)

Onderzoek: Partijen zoals Universiteiten en Kennisinstituten.

Derde partijen: Alle andere partijen, zoals bedrijven en andere organisaties.

Extra opties

Aan-/Uit- Functionaliteit: Maakt het mogelijk om de dataverzameling tijdelijk uit te zetten. Voor deelname moet in minimaal 80% van de ritten data worden verzameld.

Compensatie: Mogelijk wordt er in ruil voor het delen van uw data een maandelijkse compensatie geboden. Dit zal gebeuren in de vorm van een maandelijks bedrag in euro’s.

---

c1

1. Zou u onder deze omstandigheden bereid zijn om uw voertuigdata te delen? *

- Ja
- Nee

c2

2. Zou u onder deze omstandigheden bereid zijn om uw voertuigdata te delen? *

- Ja
- Nee

c3

3. Zou u onder deze omstandigheden bereid zijn om uw voertuigdata te delen? *

- Ja
- Nee

c4

4. Zou u onder deze omstandigheden bereid zijn om uw voertuigdata te delen? *

- Ja
- Nee

c5

5. Zou u onder deze omstandigheden bereid zijn om uw voertuigdata te delen? *

- Ja
- Nee

c6

6. Zou u onder deze omstandigheden bereid zijn om uw voertuigdata te delen? *

- Ja
- Nee

c7

7. Zou u onder deze omstandigheden bereid zijn om uw voertuigdata te delen? *

- Ja
- Nee

c8

8. Zou u onder deze omstandigheden bereid zijn om uw voertuigdata te delen?

- Ja
- Nee

c9

9. Zou u onder deze omstandigheden bereid zijn om uw voertuigdata te delen? *

- Ja
- Nee

c10

10. Zou u onder deze omstandigheden bereid zijn om uw voertuigdata te delen? *

- Ja
- Nee

c11

11. Zou u onder deze omstandigheden bereid zijn om uw voertuigdata te delen? *

- Ja
- Nee

c12

12. Zou u onder deze omstandigheden bereid zijn om uw voertuigdata te delen? *

- Ja
- Nee

cc1

13. Zou u onder deze omstandigheden bereid zijn om uw voertuigdata te delen? *

- Ja
- Nee

cc2

14. Zou u onder deze omstandigheden bereid zijn om uw voertuigdata te delen? *

- Ja
- Nee

cc3

15. Zou u onder deze omstandigheden bereid zijn om uw voertuigdata te delen? *

- Ja
- Nee

16. **Bij de keuze om gebruik te maken van een dienst speelt privacy een grote rol in die overweging.** *

	1	2	3	4	5	
Helemaal mee oneens	○	○	○	○	○	**Helemaal mee eens**

17. **Ik zorg altijd dat ik weet welke gegevens van mij worden verzameld en met wie deze gedeeld worden.** *

	1	2	3	4	5	
Helemaal mee oneens	○	○	○	○	○	**Helemaal mee eens**

18. **Wanneer ik een handige app of dienst wil gebruiken weegt het voordeel zwaarder dan de privacy aspecten.**
*

	1	2	3	4	5	
Helemaal mee oneens	○	○	○	○	○	**Helemaal mee eens**

(untitled)

19. **De overheid kan vertrouwd worden om de belangen van de burgers te behartigen.**
*

	1	2	3	4	5	
Helemaal mee oneens	○	○	○	○	○	**Helemaal mee eens**

20. **Huidige wetgeving biedt een voldoende niveau van privacy bescherming voor de gebruiker.**
*

	1	2	3	4	5	
Helemaal mee oneens	○	○	○	○	○	**Helemaal mee eens**

21. **De overheid kan vertrouwd worden om privacy gevoelige gegevens veilig te houden.** *

	1	2	3	4	5	
Helemaal mee oneens	○	○	○	○	○	**Helemaal mee eens**

Persoonskenmerken

Persoonlijke kenmerken

22. Wat is uw geboortejaar? *

23. Wat is uw geslacht? *

- Man
- Vrouw
- Anders / zeg ik liever niet

24. Wat is uw hoogst genoten opleiding? *

- Basisonderwijs
- Middelbaar Onderwijs, VMBO
- Middelbaar Onderwijs, HAVO
- Middelbaar Onderwijs, VWO
- MBO
- HBO
- WO

25. Wat geldt voor u. Ik heb een... *

- Privé auto (eigenaar)
- Privé auto (gebruiker)
- Leaseauto
- Huurauto
- Bedrijfsauto
- Ik heb geen beschikking over een auto

26. Hoe veel kilometer legt u per jaar af? *

- Minder dan 5.000 km
- 5.000 tot 15.000 km
- 15.000 tot 30.000 km
- 30.000 tot 45.000 km
- 45.000 tot 60.000 km
- Meer dan 60.000 km

27. Over welk type auto beschikt u? (indien u meerdere auto's bezit, kies degene met de meeste kilometers)
*

- Sedan
- Hatchback
- Stationwagen
- Terreinauto
- SUV
- MPV
- Coupé
- Cabriolet
- Sportwagen
- Bestelbus

(untitled)

Bedankt voor het invullen van de enquete, uw input wordt zeer gewaardeerd!

Mocht u naar aanleiding van deze enquete feedback willen geven kunt u hieronder feedback achter laten.

28. Heeft u op- of aanmerkingen over de enquete?

Thank You!

Bedankt voor uw input!

Mocht u deze enquête willen delen met anderen dan kunt u de onderstaande link gebruiken om deze gemakkelijk te delen.
Het delen van deze enquête wordt zeer op prijs gesteld.

https://www.surveygizmo.eu/s3/90121460/Onderzoek-delen-van-voertuigdata

D

Test survey experimental design

Table D.1: Test survey experimental design

Choice situation	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
a.p_reg	0	0	2	1	1	2	0	0	1	1	1	1	2	2	2	0	0	0	0	1	1	2	2	2
a.p_sense	0	1	1	0	1	0	0	1	1	0	0	1	0	1	0	0	1	0	1	1	0	1	0	1
a.p_beha	0	1	0	1	0	1	1	1	0	0	0	1	1	1	0	1	0	0	0	1	1	1	0	0
a.s_emer	0	1	1	0	1	0	1	0	0	0	1	0	1	0	1	1	0	0	1	1	1	0	0	1
a.s_rese	0	0	1	1	1	0	1	1	1	0	0	0	0	1	1	1	1	0	0	0	1	0	1	0
a.s_third	0	0	0	1	1	1	1	1	0	0	1	1	0	1	1	0	0	1	1	0	0	0	0	1
a.c_share	0	1	0	1	1	0	0	1	1	1	0	0	1	0	1	1	0	1	0	0	0	1	0	1
a.c_onoff	0	1	0	0	1	1	0	1	0	1	0	1	1	0	1	0	1	0	1	0	1	0	1	0
a.c_comp	0	0	0	0	0	0	40	40	20	40	60	60	20	20	60	60	60	20	20	40	20	60	40	40

E

Test survey SPSS output

Table E.1: Model Summary test questionnaire

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	268.847a	0.303	.406

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Table E.2: Classification table test questionnaire

Classification Table (a)					
Observed			Predicted		
			choice		Percentage Correct
			1	2	
Step 1	choice	1	113	29	79.6
		2	30	92	75.4
	Overall Percentage				77.7
a. The cut value is .500					

Table E.3: Estimated coefficients test questionnaire

Variables in the Equation									
		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1a	**ap_reg**	.559	.197	8.067	1	.005	1.749	1.189	2.571
	ap_sense	-.162	.303	.286	1	.593	.851	.470	1.540
	ap_beha	.044	.310	.020	1	.887	1.045	.569	1.919
	as_emer	-.346	.311	1.237	1	.266	.707	.384	1.302
	as_rese	-.187	.307	.372	1	.542	.829	.454	1.514
	as_third	2.607	.324	64.860	1	.000	13.553	7.187	25.558
	ac_share	-.476	.310	2.364	1	.124	.621	.339	1.140
	ac_onoff	-.483	.307	2.482	1	.115	.617	.338	1.125
	ac_comp	-.011	.007	2.484	1	.115	.989	.975	1.003
	Constant	-.932	.471	3.911	1	.048	.394		
a. Variable(s) entered on step 1: ap_reg, ap_sense, ap_beha, as_emer, as_rese, as_third, ac_share, ac_onoff, ac_comp.									

F

Final survey experimental design

Table F.1: Experimental design final questionnaire

Choice sit.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Count Yes	20	83	48	48	114	110	52	99	98	34	99	12
Count No	104	41	76	76	10	14	72	25	26	90	25	112
Percent Yes	16%	67%	39%	39%	92%	89%	42%	80%	79%	27%	80%	10%
Percent No	84%	33%	61%	61%	8%	11%	58%	20%	21%	73%	20%	90%
a.p_reg	0	1	0	1	0	0	0	1	1	1	0	1
a.p_sense	1	0	0	1	0	1	0	0	1	1	1	0
a.p_beha	1	1	0	1	0	0	1	1	0	0	1	0
a.s_emer	0	1	1	1	0	1	1	0	1	0	0	0
a.s_rese	0	1	1	1	1	0	0	0	0	1	1	0
a.s_third	1	0	1	1	0	0	1	0	0	1	0	1
a.c_onoff	1	0	1	1	1	0	0	1	1	0	0	0
a.c_comp	0	0	0	60	60	40	60	40	20	40	20	20

G
Base model SPSS output

Logistic Regression

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	1176	100.0
	Missing Cases	0	.0
	Total	1176	100.0
Unselected Cases		0	.0
Total		1176	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
0	0
1	1

Block 0: Beginning Block

Classification Table[a,b]

			Predicted		
			Responses		Percentage Correct
	Observed		0	1	
Step 0	Responses	0	0	519	.0
		1	0	657	100.0
	Overall Percentage				55.9

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	.236	.059	16.119	1	.000	1.266

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Scenario	.124	1	.725
		ap_reg	11.205	1	.001
		ap_sense	.003	1	.953
		ap_beha	.279	1	.597
		as_emer	7.619	1	.006
		as_rese	3.314	1	.069
		as_third	300.136	1	.000
		ac_onoff	1.521	1	.217
		ac_comp	24.121	1	.000
	Overall Statistics		348.323	9	.000

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	384.357	9	.000
	Block	384.357	9	.000
	Model	384.357	9	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1229.694[a]	.279	.373

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	4.888	8	.770

Contingency Table for Hosmer and Lemeshow Test

		Responses = 0		Responses = 1		
		Observed	Expected	Observed	Expected	Total
Step 1	1	88	83.352	10	14.648	98
	2	81	80.144	17	17.856	98
	3	67	71.648	31	26.352	98
	4	59	61.460	39	36.540	98
	5	59	55.957	39	42.043	98
	6	53	54.438	45	43.562	98
	7	32	31.144	66	66.856	98
	8	20	24.065	78	73.935	98
	9	22	21.833	76	76.167	98
	10	38	34.957	256	259.043	294

Classification Table[a]

			Predicted		
			Responses		Percentage Correct
Observed			0	1	
Step 1	Responses	0	407	112	78.4
		1	181	476	72.5
	Overall Percentage				75.1

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Scenario	-.059	.141	.176	1	.675	.943	.715	1.242
	ap_reg	-.664	.148	20.161	1	.000	.515	.385	.688
	ap_sense	-.039	.148	.068	1	.794	.962	.720	1.285
	ap_beha	-.251	.147	2.897	1	.089	.778	.583	1.039
	as_emer	.311	.147	4.459	1	.035	1.364	1.023	1.821
	as_rese	.381	.148	6.659	1	.010	1.464	1.096	1.956
	as_third	-2.458	.152	261.983	1	.000	.086	.064	.115
	ac_onoff	.259	.144	3.225	1	.073	1.295	.977	1.718
	ac_comp	.020	.003	35.075	1	.000	1.020	1.013	1.027
	Constant	1.014	.227	19.910	1	.000	2.756		

a. Variable(s) entered on step 1: Scenario, ap_reg, ap_sense, ap_beha, as_emer, as_rese, as_third, ac_onoff, ac_comp.

H

Expanded model SPSS output

Logistic Regression

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	1176	100.0
	Missing Cases	0	.0
	Total	1176	100.0
Unselected Cases		0	.0
Total		1176	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
0	0
1	1

Block 0: Beginning Block

Classification Table[a,b]

			Predicted		
			Responses		Percentage Correct
Observed			0	1	
Step 0	Responses	0	0	519	.0
		1	0	657	100.0
	Overall Percentage				55.9

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	.236	.059	16.119	1	.000	1.266

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Scenario	.124	1	.725
		ap_reg	11.205	1	.001
		ap_sense	.003	1	.953
		ap_beha	.279	1	.597
		as_emer	7.619	1	.006
		as_rese	3.314	1	.069
		as_third	300.136	1	.000
		ac_onoff	1.521	1	.217
		ac_comp	24.121	1	.000
		aptot	32.414	1	.000
		attot	28.042	1	.000
		Age	11.062	1	.001
		Sex	.118	1	.731
		e_own	4.898	1	.027
		e_user	8.970	1	.003
		e_leas	.607	1	.436
		e_comp	3.028	1	.082
		car_sedan	.070	1	.792
		car_hatchback	1.584	1	.208
		car_stationwagen	3.481	1	.062
		car_suv	.297	1	.586
		car_mpv	17.802	1	.000
		Kmyr	3.063	1	.080
		EduHigh	.716	1	.397
		Age by Scenario	6.210	1	.013
		Scenario by Sex	.665	1	.415
		EduHigh by Scenario	.694	1	.405
		Age by as_rese	.687	1	.407
		Age by as_third	273.130	1	.000
		ac_comp by aptot	5.585	1	.018
	Overall Statistics		445.695	30	.000

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	535.813	30	.000
	Block	535.813	30	.000
	Model	535.813	30	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1078.238[a]	.366	.490

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	13.153	8	.107

Contingency Table for Hosmer and Lemeshow Test

		Responses = 0		Responses = 1		
		Observed	Expected	Observed	Expected	Total
Step 1	1	110	110.559	8	7.441	118
	2	95	98.536	23	19.464	118
	3	95	85.866	23	32.134	118
	4	64	72.510	54	45.490	118
	5	56	57.174	62	60.826	118
	6	47	39.059	71	78.941	118
	7	27	26.003	91	91.997	118
	8	17	16.166	101	101.834	118
	9	4	9.600	114	108.400	118
	10	4	3.527	110	110.473	114

Classification Table[a]

			Predicted		
			Responses		Percentage Correct
	Observed		0	1	
Step 1	Responses	0	390	129	75.1
		1	128	529	80.5
	Overall Percentage				78.1

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1[a]	Scenario	.485	.470	1.065	1	.302	1.624	.647	4.077
	ap_reg	-.787	.160	24.093	1	.000	.455	.333	.623
	ap_sense	-.051	.159	.105	1	.746	.950	.696	1.297
	ap_beha	-.318	.158	4.041	1	.044	.727	.533	.992
	as_emer	.359	.160	5.043	1	.025	1.432	1.047	1.958
	as_rese	-.345	.379	.829	1	.363	.708	.337	1.488
	as_third	-2.009	.393	26.073	1	.000	.134	.062	.290
	ac_onoff	.337	.156	4.660	1	.031	1.400	1.031	1.901
	ac_comp	.054	.014	14.602	1	.000	1.055	1.027	1.085
	aptot	-.092	.049	3.480	1	.062	.912	.829	1.005
	attot	.143	.030	22.844	1	.000	1.154	1.088	1.224
	Age	.004	.012	.143	1	.705	1.004	.982	1.028
	Sex	-.158	.228	.479	1	.489	.854	.547	1.335
	e_own	-.809	.600	1.819	1	.177	.445	.138	1.443
	e_user	-.446	.608	.538	1	.463	.640	.194	2.108
	e_leas	-.380	.680	.313	1	.576	.684	.180	2.591
	e_comp	-1.173	.687	2.912	1	.088	.310	.080	1.190
	car_sedan	.398	.270	2.168	1	.141	1.489	.877	2.530
	car_hatchback	.511	.264	3.745	1	.053	1.666	.993	2.795
	car_stationwagen	.267	.306	.761	1	.383	1.306	.717	2.381
	car_suv	-.002	.383	.000	1	.996	.998	.471	2.117
	car_mpv	-.500	.407	1.509	1	.219	.607	.273	1.347
	Kmyr	-.095	.078	1.498	1	.221	.909	.780	1.059
	EduHigh	-.744	.210	12.533	1	.000	.475	.315	.717
	Age by Scenario	-.033	.012	7.732	1	.005	.968	.946	.990
	Scenario by Sex	.890	.388	5.261	1	.022	2.434	1.138	5.207
	EduHigh by Scenario	.857	.321	7.142	1	.008	2.355	1.257	4.415
	Age by as_rese	.023	.010	4.888	1	.027	1.023	1.003	1.044
	Age by as_third	-.026	.011	5.646	1	.017	.975	.954	.995
	ac_comp by aptot	-.003	.001	5.179	1	.023	.997	.994	1.000
	Constant	1.532	.936	2.677	1	.102	4.626		

a. Variable(s) entered on step 1: Scenario, ap_reg, ap_sense, ap_beha, as_emer, as_rese, as_third, ac_onoff, ac_comp, aptot, attot, Age, Sex, e_own, e_user, e_leas, e_comp, car_sedan, car_hatchback, car_stationwagen, car_suv, car_mpv, Kmyr, EduHigh, Age * Scenario , Scenario * Sex , EduHigh * Scenario , Age * as_rese , Age * as_third , ac_comp * aptot .

Factor analysis SPSS output

Factor Analysis

Correlation Matrix[a]

		ap1	ap2	ap3	at1	at2	at3
Correlation	ap1	1.000	.341	-.366	-.030	-.084	-.033
	ap2	.341	1.000	-.370	-.170	-.035	-.132
	ap3	-.366	-.370	1.000	.131	.196	.163
	at1	-.030	-.170	.131	1.000	.571	.554
	at2	-.084	-.035	.196	.571	1.000	.491
	at3	-.033	-.132	.163	.554	.491	1.000
Sig. (1-tailed)	ap1		.000	.000	.150	.002	.127
	ap2	.000		.000	.000	.113	.000
	ap3	.000	.000		.000	.000	.000
	at1	.150	.000	.000		.000	.000
	at2	.002	.113	.000	.000		.000
	at3	.127	.000	.000	.000	.000	

a. Determinant = .280

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.672
Bartlett's Test of Sphericity	Approx. Chi-Square	1490.543
	df	15
	Sig.	.000

Communalities

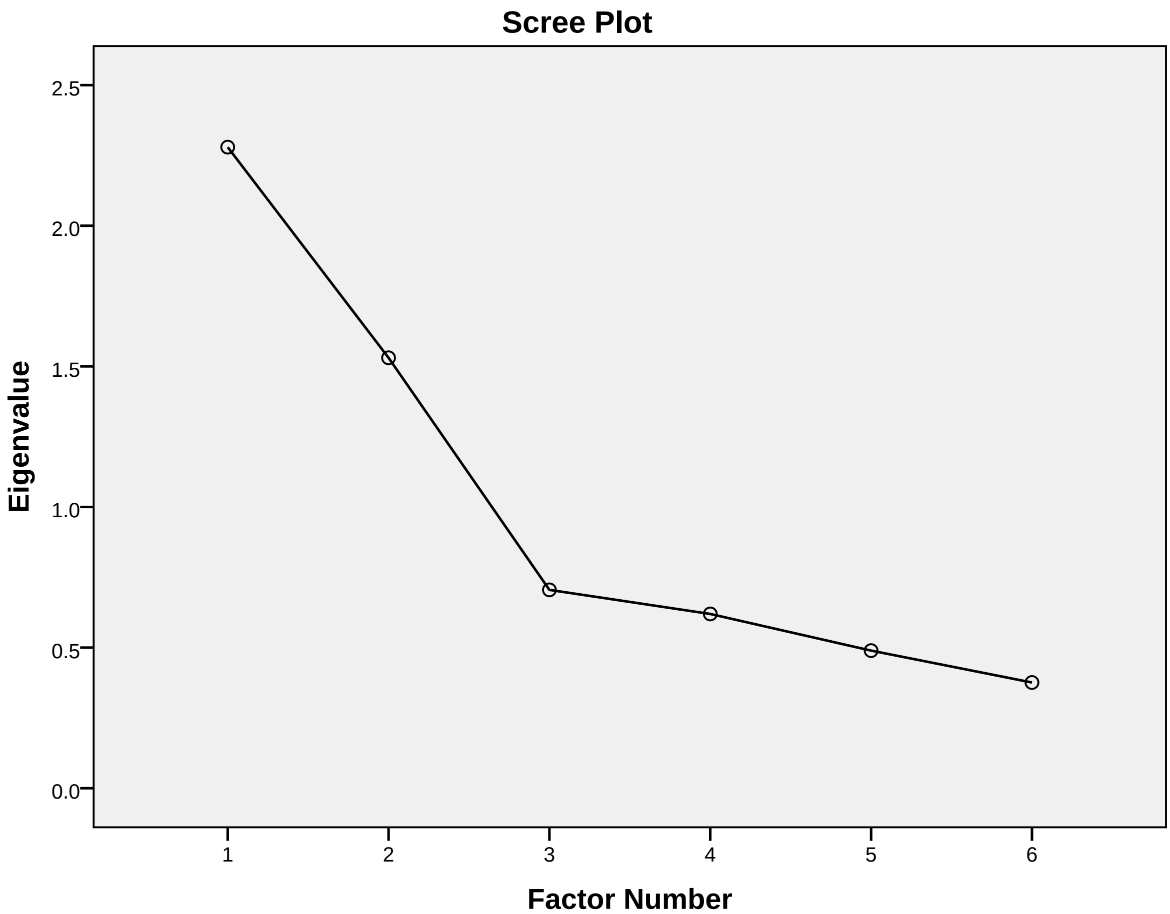
	Initial	Extraction
ap1	.190	.351
ap2	.222	.337
ap3	.232	.407
at1	.441	.640
at2	.397	.504
at3	.359	.483

Extraction Method: Principal Axis Factoring.

Total Variance Explained

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.279	37.992	37.992	1.789	29.818	29.818	1.633	27.222	27.222
2	1.530	25.507	63.499	.933	15.555	45.373	1.089	18.151	45.373
3	.705	11.754	75.253						
4	.620	10.328	85.581						
5	.489	8.154	93.735						
6	.376	6.265	100.000						

Extraction Method: Principal Axis Factoring.



Factor Matrix[a]

	Factor	
	1	2
ap1	-.249	.538
ap2	-.325	.481
ap3	.398	-.499
at1	.754	.267
at2	.675	.221
at3	.663	.209

Extraction Method: Principal Axis Factoring.

a. 2 factors extracted. 11 iterations required.

Rotated Factor Matrix[a]

	Factor	
	1	2
ap1	.004	.593
ap2	-.089	.574
ap3	.147	-.621
at1	.796	-.080
at2	.705	-.088
at3	.688	-.094

Extraction Method: Principal Axis Factoring.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

Factor Transformation Matrix

Factor	1	2
1	.904	-.427
2	.427	.904

Extraction Method: Principal Axis Factoring.
Rotation Method: Varimax with Kaiser Normalization.

Factor Score Coefficient Matrix

	Factor	
	1	2
ap1	.047	.345
ap2	.027	.320
ap3	.008	-.382
at1	.474	.028
at2	.299	.011
at3	.283	.001

Extraction Method: Principal Axis Factoring.
Rotation Method: Varimax with Kaiser Normalization.

Factor Score Covariance Matrix

Factor	1	2
1	.781	-.052
2	-.052	.622

Extraction Method: Principal Axis Factoring.
Rotation Method: Varimax with Kaiser Normalization.

J

Scientific Article

Road-user participation in vehicle-data sharing systems for the purpose of dynamic traffic management, a stated preference study

A.M. de Jong (4320506)
Complex System Engineering & Management - Delft University of Technology

July 1, 2019

Abstract

This study aims to provide insight into how factors relating to privacy and incentives influence people's willingness to participate in sharing their vehicle based sensing data with governmental parties for the purposes of improved dynamic traffic management in the Netherlands. Through the use of a stated preference experiment data is gathered in order to estimate a discrete choice model using binary logistic regression. Respondents are most likely willing to share their data when trip registration is not personally identifiable and this data is not shared with third parties. Sharing of data with emergency services and for research purposes actually increases the odds of participation. Furthermore, potential users who have not been exposed alternatives which offer monetary reward are more likely to participate for free. Clear communication of the purpose and the social benefits of participation is important for obtaining sufficient levels of participation without offering monetary reward. Being parsimonious in data collection will result in the least amount of privacy harm and avoid the perception of a system as unfair and inefficient.

Keywords: Stated Preference Experiment, Discrete choice modelling, Dynamic traffic management, vehicle data sharing

1 Introduction

The road network of the Netherlands is increasingly busy, with the amount of daily vehicles on the roads increasing with 16% between 2005 and 2016, and an increase of 3% over 2016 alone [van Infrastructuur en Waterstaat, 2017]. This observed growth on the already densely used road-network is leading to problems with congestion, with the travel-time lost increasing with 10% in 2016, compared to 2015 [van Infrastructuur en Waterstaat, 2017]. Congestion on the road network is increasing and so are the economic losses that resulting from the total amount of travel-time lost. Excessive transport times negatively impact the mobility of citizens and the attractiveness of the Netherlands for international freight transport. The economic damage for companies over 2015 as a result of congestion in cargo transport is estimated to be between 857M and 1,1 billion Euro, while in 2014 this was estimated to be between 655M to 852M Euro [EVO, 2016].

The implementation of improved Dynamic Traffic Management (DTM) measures may prevent congestion before it occurs, leading to an improvement in road-network performance. The input for DTM are estimations of traffic flow, traffic densities, and traffic speed [Seo et al., 2015].

Currently the data used for traffic state estimations (TSE) are mainly collected using stationary measuring instruments such as loop-detectors or video based technologies[Buch et al., 2011], however the placement of sensors across the road network is financially prohibitive due to the high costs associated with these stationary detectors and thus may not cover the entirety of the network. Another option is using Floating-Car Data (FCD), this is usually based on cellular and Satellite Navigation (SatNav) data [Leduc, 2008]. FCD is usually comprised of basic vehicle telemetry such as speed, direction and, most importantly, the position of the vehicle [Schäfer et al., 2002]. Vehicle-based sensing data (VBSD) refers to data generated by one vehicle as a sample to assess the overall traffic condition, this is an extension of the definition of floating car data (FCD). VBSD broadens this definition by including the data gathered by on-board sensors, e.g. the amount of cars passed left and right. Often-times datasets containing VBSD are purchased from private parties at a significant cost, these purchased datasets often have a high level of data-aggregation due to privacy aspects relating to the General Data Protection Regulation (GDPR) and fears over public outcry, such as when TomTom sold data to the Dutch government which included speeding data [Palmer, 2011].

The future development of the field of TSE has many exciting possibilities due to advances in methods and the increasing availability of mobile disaggregated traffic data generated by sensors integrated in the vehicles. Vlahogianni [Vlahogianni et al., 2014] states one of the challenges for the future to be ”Using new technologies for collecting and fusing data”. The use of traditional on-road sensors such as inductive loops is necessary, but is not always sufficient for the development of intelligent transport systems (ITS) [Leduc, 2008]. In order to better estimate the traffic state, relying on a combination of both stationary and moving observers is valuable [van Erp et al., 2019]. In particular the sampling resolution is important for estimation performance when large changes in traffic conditions occur [van Erp et al., 2018]. Collecting the data directly from road-users instead of third parties, may lead to a reduction in cost and an increase in data resolution. For the use of VBSD to provide meaningful datasets the penetration of these technologies is key. Herrera et al. [2010] suggested that a 2-3% market penetration of cell phones in the driver population is enough to provide accurate traffic measurements based on cellphone GPS data. The advent of autonomous vehicles and vehicles with more on-board sensors provides a large potential market penetration for VBSD, and access to new types of data. As such, the possibility of gathering VBSD directly from road users becomes interesting, albeit more complex due to possible legal issues, privacy concerns and uncertainty regarding people's willingness to share the data with the government.

The main gap that this research aims to address is the lack of understanding how factors relating to privacy concerns and potential compensation influence a road-users choice to participate in sharing their VBSD for the purposes of DTM. Privacy considerations of road-users will determine if the proposed gathering of VBSD can be successful regarding user participation. As such this subject will be a focal point in this study.

In order to address the knowledge gap in how factors relating to privacy and incentives influence people's willingness to participate in sharing their VBSD with the government, the following research question is defined:

How do factors relating to privacy and incentives affect road-users participation in a vehicle-data sharing system for the purposes of Dynamic Traffic Management?

Section 2 provides the background of the study. Section 3 provides the method, followed by the results in section 4. Section 5 discusses the findings and concludes the article.

2 Theoretical background

Privacy as a trade-able interest

Privacy has been defined in a myriad of ways over the last few centuries [Morton, 1998]. In conjunction with the utilitarian perspective of this study, control- and Use-based privacy definitions are particularly well suited. Three definitions that are relevant to this research are privacy defined as; (1) the individual's ability to control the collection and use of personal information [Westin, 1968] [Hann et al., 2001]; (2) a right to control access to and uses of-places, bodies, and personal information [Moore, 2008]; and (3) the desire individuals have in sustaining a 'personal space' free from interference by other people and organizations [Derikx et al., 2016][Clarke, 1999].

Together, these three definitions encompass the range perspectives required to properly define privacy in the context of users sharing vehicle-based sensing data. Following definitions (1) and (2), any situation in which a person freely discloses information satisfies the conditions of a right to privacy, if the information is only used for the scope for which consent was given. Definition (3) concerns the fact people want to minimize the amount of data they share. In the context of this research this means that (1) and (2) provide sufficient definition of privacy assuming a person acts out of free will in choosing to (not) share their data. As such the definition of privacy that is used in this research is: *Privacy is defined as an individual's' right to control the collection, access to and uses of information relating to places, bodies, and personal data.*

Adopting a utilitarian view to the control and use based privacy definition noted above, users' can choose to allow access to and uses of information relating to places, bodies, and personal data if they so choose. The utilitarian perspective assumes that people are be willing to trade away levels of privacy if the positive utility resulting from this is higher than the negative utility due reduced levels of privacy.

Factors influencing the perceived level of privacy

The privacy interests that people experience with regards to connected vehicles are myriad. Privacy interests can be affected by various activities, i.e. (1) information collection, (2) information processing, (3), information dissemination, and (4) invasion [Solove, 2006]. With increased levels for these four activities corresponding to an increased level of privacy harm.

The categorization of these factors influencing privacy perception is chosen on the basis of the seven types of privacy, as laid out by Finn [Finn et al., 2013]. These seven types of privacy are: 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.

Within the domain of sharing VBSD with governmental parties, three of these categories are particularly affected, these are the privacy of location and space, privacy of behaviour and action, privacy of data and image. Privacy of location and space is affected by the tracking of movements in the form of trip registration. Privacy of behaviour and action is affected by the presence of sensors measuring driving behaviour. Privacy of data and image is affected by the sharing of information with other parties, such as emergency services or private third parties.

Besides the activities directly causing privacy harm, other factors also contribute to the perception of privacy. These factors are; the level of control over the collection of data [Sheehan and Hoy, 2000], transparency how data will be used and who will have access to it [Walter and Abendroth, 2018], and trust placed in the data-collecting party [Kang and Hustvedt, 2014].

Reduced levels of privacy perception can be compensated through the use of both monetary and non-monetary methods of compensation. The generally accepted belief is that monetary compensation do increase user participation [Endo et al., 2016] [Riley, 2008]. Offering monetary compensation for privacy harm has been demonstrated to be feasible in earlier studies regarding Dutch road users [Derikx et al., 2016]. Non-monetary compensation takes the form of social benefit. Social benefit has the same effect as monetary compensation in increasing disclosure rates. Herein the intrinsic value of a service and the service experience also promotes data disclosure [Endo et al., 2016].

3 Methods

A stated preference experiment is conducted to evaluate effects of activities affecting privacy interests, compensation, and personal attitudes relating to Privacy Concern and Trust in the government. Stated choice experiments are a statistical approach often used to determine consumer preferences based on trade-offs made between different alternatives and is often used in market research [Louviere et al., 2000]. The method is utilitarian as it assumes that people will value the alternatives based by making implicit trade-offs. From these trade-offs it is possible to obtain the preferences of the respondent. Based on the choices observes by a respondent it is possible to estimate the perceived utilities for each attribute present in the choice situation.

Using binary logistic regression the coefficients of the choice model are estimated. The Main effects of all variables and significant interaction effects between alternative attributes and socio-demographic factors and attitudes are included in the model estimation.

Confirmatory factor analysis is used to analyze whether the attitude indicators do indeed measure their respective attitudes relating to Institutional trust and Privacy Concern.

Operationalization

The operationalized attributes that are varied in the survey can be found in table 1. This section will provide an overview of the different operationalized attributes.

Privacy of location and space is operationalized in to the attribute Trip registration. The registration of trips, which includes the storage location data and data gathered by other potential sensors, has two levels. The level for which the user is not identifiable is that of registration per road, this means that a vehicle has multiple associated identifiers per trip corresponding to the roads travelled. The second level is the trip registration on a personal account, here all trips are stored on one personal account.

Privacy of behaviour and action is operationalized with factors relating to extra data collection. The number of vehicles passed refers to the use of on-board sensors to track the amount of vehicles passed in both directions. Accelerometer data refers to the presence of an on-board motion sensor.

Privacy of Data and image is operationalized with the following factors. Corresponding to the dissemination of data are the sharing of data with, emergency services, researchers, and unnamed third parties. The level of control is operationalized as the presence of on-/off functionality, which may not be turned off in more than 20% of the trips if the user wishes to receive monetary compensation.

Compensation is operationalized in two factors, monetary compensation and social benefits. Monetary compensation takes the form of a monthly compensation for participation. Social benefits is operationalized as the communication of social benefits. It should be noted that communication of social benefits is done at the start of the survey, by including extra information relating to the benefits of participation, both in the text as well as the introductory film. The social benefits are stated to be improved traffic management leading to less congestion, less frequent delays on-route and increased safety on the road.

The latent variables regarding an individual's attitude towards Privacy Concern and Institutional Trust are measured by the indicators shown in table 2. Each of the two attitudes is measured by three statements that are rated on a 5-point Likert scale. The summation of the respective indicator scores represent to score for its latent variable, herein the score for ap3 is inverted.

In addition several socio-demographic variables have been included in the survey, these include age, gender, yearly driving amount, car ownership type, and car type.

Table 1: Operationalized attributes

	Factor	Attribute	Levels
Privacy Type			
Location and space	Information collection	Trip registration	Per road / Personal account
Behaviour and action	Information collection	Number of passed vehicles	No / Yes
	Information collection	Accelerometer data	No / Yes
Data and image	Information Dissemination	Sharing with emergency services	No / Yes
	Information Dissemination	Sharing with researchers	No / Yes
	Information Dissemination	Sharing with third parties	No / Yes
	Control	On-/functionality	No / Yes
Compensation type			
Monetary	Monetary compensation	Euro's per month	0 / 20 / 40 / 60
Non-monetary	Social benefits	Communication of social benefits	No/Yes

Table 2: Attitude Indicators

Privacy Concern (5 point Likert Scale)
- Bij de keuze om gebruik te maken van een dienst speelt privacy een grote rol in die overweging.
- Ik ben op de hoogte welke gegevens van mij worden verzameld en met wie deze gedeeld worden.
- Wanneer ik een handige app of dienst wil gebruiken weegt het voordeel zwaarder dan de privacy aspecten.

Trust towards government (5 point likert scale)
- De overheid kan vertrouwd worden om de belangen van de burgers te behartigen.
- Huidige wetgeving en de manier waarop organisaties met privacy omgaan biedt een voldoende niveau van privacy bescherming voor de gebruiker.
- De overheid kan vertrouwd worden om privacy gevoelige gegevens veilig te houden.

Categories for car ownership type and car type with fewer than 5% of the sample are excluded. Categorical variables such as car ownership and car type are included as dummy variables. The socio-demographic factors are included in the logistic regression, including any significant interaction effects with the attributes shown in table 1. The socio-demographic variables as shown in table 3 are all binary, with the exception of yearly km driven, which is an ordinal variable [1] , and age which is a scalar.

Sample

The survey concerns the collection of VBSD on the Dutch road network. As such the target audience are Dutch road-users. As such the selection of respondents excludes respondents who are underage and those that never use a car as a driver.

Respondents were reached by providing the survey online and sharing the link with various people, and the collection of surveys using a tablet at several locations. These locations included the Delft University of Technology, the Hague University, a public library, and public spaces. The method for recruiting respondents was mainly through the use of the researchers social network, and direct contact, as such the sample is a convenience sample that does not take into account representativity in its respondent selection.

[1]($<$ 5.000 km/yr, 5.000 $<$ 15.000, 15.000 $<$ 30.000, 30.000 $<$ 45.000, 45.000 $<$ 60.000, $>$ 60.000)

A total of 124 completed responses were obtained, of which 98 were retained after filtering out respondents who do not drive a road vehicle. This corresponds with a margin of error (MOE) of 9,1% at a 95% confidence interval. The MOE is the amount of random sampling error in the survey's result. Smaller samples have a larger margin of error [Myers et al., 2006][2]. Most studies consider a maximum MOE of 5% as acceptable, highlighting the disclaimer to interpret the results with care [Simon, 2006].

Due to the sample being a convenience sample that does not account for representativity it can be said that the sample is not representative. The target audience is a subset of the Dutch population, however statistics showing the socio-demographic factors of road-users were not obtainable, as such the Dutch population is used for comparison. Specifically comparisons are made concerning, age, gender, education level, and average mileage per year.

The distribution of gender in the respondents is 66% male and 34% female, compared to the Dutch population with 49% Male and 51% Female [CBS, 2018]. Regarding education levels 83% of respondents are highly educated at academic (HBO/WO) level, in contrast to only 23% of

[2]Formula for MOE, with: p = probability of population participating (0,69), n = sample size (98) and N = population size (11.176.150)

$MOE = 1.96\sqrt{\frac{p(1-p))}{n}} * \sqrt{\frac{N-n}{N-1}}$

Table 3: Estimated Discrete Choice model

						95% C.I.for EXP(B)	
		B	**S.E.**	**Sig.**	**Exp(B)**	**Lower**	**Upper**
Context variable	Communication Benefits	0.525	0.471	0.265	1.690	0.672	4.251
Attributes	Trip registration *	-0.793	0.161	**0.000**	0.453	0.330	0.620
	Vehicles passed	-0.051	0.159	0.750	0.950	0.696	1.299
	Accelerometer *	-0.321	0.159	**0.043**	0.726	0.532	0.990
	Sharing Emergency *	0.345	0.160	**0.031**	1.412	1.033	1.931
	Sharing Research	-0.347	0.379	0.361	0.707	0.336	1.487
	Sharing Third *	-2.019	0.394	**0.000**	0.133	0.061	0.288
	On-/off function *	0.326	0.156	**0.037**	1.385	1.020	1.881
	Monthly compensation *	0.054	0.014	**0.000**	1.056	1.027	1.085
Attitudes	Privacy Concern *	-0.096	0.049	**0.051**	0.908	0.825	1.001
	Institutional Trust *	0.140	0.030	**0.000**	1.150	1.085	1.220
Socio-dem	Age	0.005	0.012	0.685	1.005	0.982	1.028
	Sex	-0.193	0.228	0.397	0.824	0.527	1.289
	Ownership_Owner	-0.718	0.600	0.232	0.488	0.150	1.582
	Ownership_User	-0.426	0.608	0.483	0.653	0.198	2.149
	Ownership_Lease	-0.100	0.690	0.885	0.905	0.234	3.500
	Ownership_Company	-1.057	0.688	0.124	0.347	0.090	1.337
	car_sedan	0.453	0.272	0.096	1.573	0.922	2.682
	car_hatchback *	0.550	0.265	**0.038**	1.733	1.031	2.914
	car_stationwagen	0.295	0.307	0.337	1.343	0.736	2.451
	car_suv	-0.018	0.384	0.963	0.983	0.463	2.086
	car_mpv	-0.476	0.408	0.244	0.622	0.279	1.384
	Yearly driven km *	-0.181	0.084	**0.031**	0.835	0.708	0.984
	High Education *	-0.778	0.211	**0.000**	0.459	0.304	0.695
Interactions	Age * Communication of benefits *	-0.033	0.012	**0.005**	0.967	0.945	0.990
	Sex * Communication of benefits *	0.888	0.388	**0.022**	2.429	1.135	5.201
	High Education * Communication of benefits *	0.841	0.321	**0.009**	2.318	1.236	4.348
	Age * Sharing Research *	0.023	0.010	**0.026**	1.023	1.003	1.045
	Age * Sharing Third *	-0.026	0.011	**0.018**	0.974	0.954	0.995
	Privacy Concern * Monthly Compensation *	-0.003	0.001	**0.021**	0.997	0.994	1.000
Constant	Constant	1.550	0.931	0.096	4.712		

the population. While a diverse age group of respondents has been observed there is an over-representation of the age groups 18 - 25 and 25 - 35. Concerning the amount of kilometers driven on a yearly basis. The average Dutch amount of kilometers traveled on a yearly basis is 13.000 km/year [CBS, 2018], this is close to the average of the observed responses, which is 13.700 km/year.

From these comparisons it is clear that the sample contains biases regarding age, gender, and education. Further bias may be present due to a measure of self-selection, by which respondents who are willing to participate in a data sharing system may choose to fill out the survey more often than people who are not willing to participate in a data sharing system. Because these biases can influence the results these biases need to be taken into account when interpreting the results of the analyses.

4 Results

Order effects

The results of the t-test for investigating the order effects are visible in table 4 and table 5. The Levene's test has p=0.000, which implies the variance between the two groups differs significantly. The means of the respondents' is shown in table 4, with group A choosing to participate in 47% of the cases and group B in 33% of the cases. The t value as seen in table 5 is t=2,301, indicating the size of the difference in participation relative to the variation in the sample data. The difference between the scenarios is significant (p<0,05). This means that when presented with a potential vehicle-data sharing system, respondents are significantly more likely to agree to sharing their data for free when they have not previously been exposed to alternatives that offer monetary compensation.

Confirmatory factor analysis

The Barlett test resulted in a significance level of 0.00, implying the correlation matrix differs significantly for the identity matrix and factor analysis is suitable. The results of the KMO measure of sampling adequacy is .672, which is deemed acceptable. Furthermore all communalities are > 0.25 and thus a sufficient portion of each indicators variance can be explained by the extracted factors. Two factors with an eigenvalue of > 1 were extracted accounting for $63,5\%$ of the variance.

The two extracted factors do indeed only contain the expected indicators. With factor 1 representing institutional trust, and factor 2 representing privacy concern, as shown in table 6. The rotated factor loadings of at1 (.796), at2 (.705), and at3 (.688) have large factor loadings on factor 1, as such this factor describes Institutional Trust. The rotated factor loadings of ap1 (.593), ap2 (.574), and ap3 (-.621) load on factor 2, Privacy Concern. These factor loadings are considered sufficient ($>$ .5) but are not considered high ($>$ .7), indicating that privacy concern is measured by its related indicator but with smaller correlation coefficients as compared to factor 1.

Thus it can be stated that the indicators do indeed measure their respective attitudes. With institutional trust more strongly correlated to its indicators than the indicators correlating to privacy concern.

Discrete Choice model

A discrete choice model is estimated to analyze the choice behaviour of the respondents [Ben-Akiva et al., 1985]. Binary logistic regression is performed using SPSS on a dataset containing all predefined choice sets and all respondents choices from the survey. Table 3 provides the an overview of all estimated coefficients.

The estimated coefficients of the expanded model are presented in table 3. Regarding model fit the Chi^2 value is considered significant at df=30, indicating a significant improvement over the null model. The Mc Faddan R^2 value is 0.332, indicating an excellent model fit [Domencich and McFadden, 1975].

The coefficients of the main effects relating to trip registration, accelerometer data, sharing with emergency services, sharing with third parties, on-/off functionality and monthly compensation are all significant. The attitudes relating to privacy concern and Institutional Trust are significant. For the socio-demographic factors, hatchback ownership, education level, and yearly driven km are significant. Significant interaction terms have been found regarding the communication of social benefit and Age, Sex, and Education level. Age also has significant interactions with sharing data with researchers and with third parties. Furthermore a significant interaction effect was found between privacy concern and monthly compensation.

The most important factor influencing the decision to participate is the sharing of data with third parties. With an extremely high negative utility associated with it. Interaction with age implies that the older people are, the more negative the valuation of sharing with third parties is.

Sharing data with emergency services and for research purposes is valued positively. Although the dissemination of data is associated with negative utility in general, it seems that respondents ascribe a positive value to this. The positive β of sharing with researches increases with age.

The effects of the communication of social benefits is dependent on age, sex, and education level. With age decreasing the valuation of this communication. Females tend to value the communication of benefits more than males. Higher education levels are also associated with

Table 4: Group statistics regarding order effects

Group	N	Mean	Std. Deviation	Std. Error Mean
A	180	.47	.500	.037
B	114	.33	.473	.044

Table 5: t-test for equality of means (equal variances not assumed)

					95% C.I. of difference	
t	**df**	**sig. (2 tailed)**	**Mean difference**	**Std. Error Difference**	**Lower**	**Upper**
2.301	250.292	.022	.133	.058	.019	.247

a higher valuation of the communication of benefits.

None of the forms of car ownership are significant. Some observations can still be made. Most notable is the order in which the coefficients are ranked, with company cars being the least willing to participate and lease cars being the most willing to participate.

For the variables regarding car type it is observed that hatchback ownership has a significant effect on the odds of participation. However none of the other car types are considered significant.

The amount of kilometers driven on a yearly basis is measured in ordinal categories, and is considered significant. With the odds of participation dropping when respondents travel more on a yearly basis.

Education level is significant and it implies that higher educated individuals are less likely to participate in a data-sharing system. Education level also has a significant interaction effect with the communication of benefits. With academically educated people being more sensitive to the communication of benefits.

Higher levels of privacy concern result negatively affect the odds of participation, and is considered significant. Higher levels of privacy concern also interact with the valuation of monthly compensation. This interaction implies that high levels of privacy concern lead to a lower evaluation of the utility of monetary compensation. In contrast to privacy concern, high levels of institutional trust positively influence the odds of participation in a data-sharing system.

5 Conclusion and discussion

The most important factor in relation to participation is the sharing of data with unnamed third parties. Followed in order by monetary compensation, trip registration, sharing with researchers, sharing with emergency services, and the presence of on-/off functionality. Sharing data with unnamed third parties lowers the perception of transparency and data safety, leading to high perceived privacy risks which in turn reduces the likelihood of road-user participation. Through transparency in communication, parsimonious data collection, and data safety, trust can be fostered in the relationship between the participant and data-controller. Combined with non-monetary compensations schemes it is possible for road-users to view the vehicle-data system as acceptable without the need for monetary compensation. In situations where the data collection is not considered parsimonious, trust is low, and social benefits are not communicated clearly. It will be necessary to offer monetary rewards to compensate for the perceived privacy harm in order to increase participation to acceptable levels.

Recommendations for implementing a vehicle-data sharing system are to start without offering monetary benefit. Potential users who have not been shown alternatives which offer monetary reward the odds of participation are expected to be higher than predicted by the discrete choice model, this increased level of participation is lost once alternatives which offer monetary reward have been presented. Clear communication of the purpose and the social benefits of participation is important for obtaining sufficient levels of participation without offering monetary reward. Being parsimonious in data collection will result in the least amount of privacy harm and avoid the perception of a system as unfair and inefficient. If participation levels are not sufficient without monetary reward, the use of monetary reward can be implemented.

All effects estimated are assumed to be linear in nature. Which may not be the case. This relates in particular to the estimation of monetary benefits, and the estimated interaction effects. Future studies are recommended to include testing for non-linearity on both main and interaction effects to account for non-linear effects in the valuation of these attributes.

The unspecified nature of sharing with third parties is considered a limitation. In the current study it is intentionally left unspecified to estimate the maximum level of privacy harm related to sharing data with an unknown number of unknown parties for unknown purposes. The inclusion of specific third parties and/or applications would improve the predictive power of the model. Potentially users may even value specific third parties in a positive manner based on their level of trust and clarity of the purpose the data is shared for.

The use of attitude metrics relating to privacy and

Table 6: indicator scores and results of factor analyses (Factor loadings < 0.2 have been excluded)

			C.I. of 95%			**Rotated Factor Loading**	
Indicators	**Mean**	**Std. Deviation**	**Lower bound**	**Upper Bound**	**Communalities**	**1 (Institutional Trust)**	**2 (Privacy Concern)**
ap1	4.04	1.169	3.81	4.27	0.351		.593
ap2	3.42	1.059	3.21	3.63	0.337		.574
ap3	3.15	1.207	2.91	3.39	0.407		-.621
at1	3.40	1.038	3.19	3.60	0.640	.796	
at2	2.98	1.134	2.76	3.20	0.504	.705	
at3	3.05	1.101	2.83	3.27	0.483	.688	

trust has been extensively employed in different fields of privacy research. The adaptation of these attitude measures to the specific domain of vehicle-data sharing systems in the Netherlands represents a valuable step for including Privacy concern and Trust in the government attitudes in future studies relating to privacy concerns in the transport domain. The inclusion of these attitude measures may provide the basis for a standardized Dutch index for Privacy concern in general and Trust in government in the specific context of data collection by governmental parties. The effect of socio-demographic variables on these have not been considered. In order to gain insight into the relationship between these attitudes and socio-demographic factors it is recommended that future studies apply structural equation modelling.

References

Ben-Akiva, M. E., Lerman, S. R., and Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand*, volume 9. MIT press.

Buch, N., Velastin, S. A., and Orwell, J. (2011). A review of computer vision techniques for the analysis of urban traffic. *IEEE Transactions on Intelligent Transportation Systems*, 12(3):920–939.

CBS (2018). Bevolking; kerncijfers.

Clarke, R. (1999). Introduction to dataveillance and information privacy, and definitions of terms. *Roger Clarke's Dataveillance and Information Privacy Pages*.

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

Domencich, T. A. and McFadden, D. (1975). Urban travel demand-a behavioral analysis. Technical report.

Endo, T., Nawa, K., Kato, N., and Murakami, Y. (2016). Study on privacy setting acceptance of drivers for data utilization on connected cars. In *Privacy, Security and Trust (PST), 2016 14th Annual Conference on*, pages 82–87. IEEE.

EVO, Fenedex, T. (2016). Economische wegwijzer 2016 - Alarmerende sijging fileschade.

Finn, R. L., Wright, D., and Friedewald, M. (2013). Seven types of privacy. In *European data protection: coming of age*, pages 3–32. Springer.

Hann, I.-H., Hui, K.-L., Lee, T. S., and Png, I. P. (2001). The value of online information privacy: An empirical investigation. *New York Times*, 1.

Herrera, J. C., Work, D. B., Herring, R., Ban, X. J., Jacobson, Q., and Bayen, A. M. (2010). Evaluation of traffic data obtained via gps-enabled mobile phones: The mobile century field experiment. *Transportation Research Part C: Emerging Technologies*, 18(4):568–583.

Kang, J. and Hustvedt, G. (2014). Building trust between consumers and corporations: The role of consumer perceptions of transparency and social responsibility. *Journal of Business Ethics*, 125(2):253–265.

Leduc, G. (2008). Road traffic data: Collection methods and applications. *Working Papers on Energy, Transport and Climate Change*, 1(55).

Louviere, J. J., Hensher, D. A., and Swait, J. D. (2000). *Stated choice methods: analysis and applications*. Cambridge university press.

Moore, A. (2008). Defining privacy. *Journal of Social Philosophy*, 39(3):411–428.

Morton, B. (1998). In pursuit of privacy: Law, ethics, and the rise of technology.

Myers, L., Sirois, M., Kotz, S., Read, C., Balakrishnan, N., and Vidakovic, B. (2006). Encyclopedia of statistical sciences.

Palmer, M. (2011). Subscribe to the ft to read: Financial times tomtom sorry for selling driver data to police.

Riley, P. F. (2008). The tolls of privacy: An underestimated roadblock for electronic toll collection usage. *Computer Law & Security Review*, 24(6):521–528.

Schäfer, R.-P., Thiessenhusen, K.-U., Brockfeld, E., and Wagner, P. (2002). A traffic information system by means of real-time floating-car data.

Seo, T., Kusakabe, T., and Asakura, Y. (2015). Estimation of flow and density using probe vehicles with spacing measurement equipment. *Transportation Research Part C: Emerging Technologies*, 53:134–150.

Sheehan, K. B. and Hoy, M. G. (2000). Dimensions of privacy concern among online consumers. *Journal of public policy & marketing*, 19(1):62–73.

Simon, M. K. (2006). *Dissertation and scholarly research: recipes for success: a practical guide to start and complete your dissertatin, thesis, or formal research project*. Kendall/Hunt Publishing Company.

Solove (2006). A taxonomy of privacy. *University of Pennsylvania law review*, pages 477–564.

van Erp, P. B., Knoop, V. L., and Hoogendoorn, S. P. (2018). Macroscopic traffic state estimation using relative flows from stationary and moving observers. *Transportation Research Part B: Methodological*, 114:281–299.

van Erp, P. B., Knoop, V. L., and Hoogendoorn, S. P. (2019). On the value of relative flow data paul. *Transportation Research Part C: Emerging Technologies*.

van Infrastructuur en Waterstaat, M. (2017). Mobiliteitsbeeld 2016-mobiliteitsbeeld en kerncijfers mobiliteit 2017-kennisinstituut voor mobiliteitsbeleid.

Vlahogianni, E. I., Karlaftis, M. G., and Golias, J. C. (2014). Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies*, 43:3–19.

Walter, J. and Abendroth, B. (2018). Losing a private sphere? a glance on the user perspective on privacy in connected cars. In *Advanced Microsystems for Automotive Applications 2017*, pages 237–247. Springer.

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