Engaging the crowd in sensing for smart mobility: A discrete choice experiment

Ria Johanna van den Boogert, and Aaron Yi Ding*, Member, IEEE

* Corresponding Author: Aaron Yi Ding

Abstract With rising numbers of people living in cities leading to increasing congestion and pollution, mobile crowdsensing applications form a potential solution to make transport systems smarter and more efficient. However, sharing data comes with the risk of private information being disclosed. Therefore, a clear incentive is necessary to motivate smart device users to share data about their activities and their environment. Taking a choice modelling approach, this study aims to identify factors related to incentives and privacy that explain choice behavior of users in crowdsensing applications. We find that the effort required by users is a main factor influencing the willingness to share data. 47% of respondents (n=125) indicated to be highly concerned about their privacy. However, the risk of re-identification was found to be the least important factor to respondents, a finding which could be explained by the Privacy Paradox. Our findings imply that a trade-off has to be made by developers of crowdsensing applications between the richness of information on one hand, and the privacy risks and participation rate of users on the other hand. We propose three practical principles for designing effective and value-sensitive crowdsensing applications for smart mobility, which are 1) Tailor-made applications, 2) Transparency by design, and 3) Ensuring attractiveness of applications. Furthermore, our study provides a basis for further research on user preferences in smart mobility applications, which will become increasingly important in the light of current challenges in the field of mobility.

Index Terms — **Smart mobility services, Crowdsensing, Choice Modelling, Willingness to share data, Privacy Calculus**

Note: Experiment input for this paper is based on research conducted as part of an industry-based project between September 2021 until January 2022. The “raw material” of the study as MSc thesis can be found on repository.tudelft.nl.

I. Introduction

By 2050, it is expected that 70% of the world’s population will live in cities and surrounding regions [1]. The growth we see in cities all around the world has direct effects on climate change, rising emission and pollution levels, as well as on infrastructure and transport requirements [2]. Smart mobility is one of the critical features contributing to smart and sustainable development [3]. Current challenges relating to congestion, accidents, and scarceness of space lead to increased delays, energy expenditure, and pollution [4]. This raises the need for better planning of traffic and infrastructure. Emerging solutions in the field of smart transportation systems, smart charging, and Mobility as a Service (MaaS) ask for new ways to acquire large amounts of data. These data can be used for analyzing and predicting mobility flows and make public and private transport more efficient, safe, and sustainable.

Crowdsensing is a novel paradigm in the field of Internet of Things, enabling both public and professional users to gather, analyze, and share data about the urban environment using built-in sensors and applications in smart mobile devices [5]. Considering that over 94% of the population has access to a mobile network in 138 countries [6], there is a huge potential in obtaining real-time data from smartphones and other smart devices. Shit [7] argues for the relevance of crowdsensing methods for the realization of intelligent transport systems. Data collected from applications running on travelers’ smart devices can contribute to better predictions of traffic flows and of the traveler situation, leading to valuable information for transport operators, authorities, and travelers.

Yet, according to Ribeiro et al. [4] the impact of users and their readiness to get involved in these new opportunities is a crucial and little addressed element regarding the digitization of transport. A key challenge in unlocking the potential of crowdsensing applications for smart mobility is the identification of robust incentives that enhance participation of individuals [8]-[9]. However, sharing data can lead to risks related to privacy. Sensing measurements might be tagged with location information or may enable the

This paper was submitted for review on [date].

R.J. van den Boogert was with the Faculty of Technology, Policy and Management, Delft University of Technology, Jaffalaan 5 2628 BX Delft, The Netherlands.

A.Y. Ding is with the Department of Engineering Systems and Services, Faculty of Technology, Policy and Management, Delft University of Technology, Jaffalaan 5 2628 BX Delft, The Netherlands (e-mail: aaron.ding@tudelft.nl).

identification of personal routines and habits [10]-[11]. For users, a clear benefit should therefore exist in order to encourage them to share their personal data about their activities and environment. Incentive mechanisms can motivate users to participate, but also require a quantification of privacy [12]. General studies on the motivations of volunteers to engage in crowdsensing tasks and the effectiveness of incentives across different contexts are still lacking [13]. Furthermore, the specific privacy concerns of users, which can be linked to their different characteristics, have to be further researched [14].

This paper explores what factors related to privacy and incentives affect the willingness of smart device users to contribute to crowdsensing systems for smart mobility. Through the lens of a discrete choice experiment, we aim to identify the trade-off that these individuals make between potential costs and benefits of participating in sensing applications. The primary contribution of this paper is the empirical insight it provides into trade-offs concerning data sharing in smart mobility services. An additional, secondary contribution concerns the method that is applied. Choice modelling is a method not often applied to the topic of crowdsensing. Thus, this study expands the research field of crowdsensing for smart mobility by providing new insights on behavior and user preferences regarding crowdsensing systems. Moreover, these insights are applied to practical use cases in the field of smart mobility. From these scientific contributions, a clear relevance follows for policy and society, by contributing to the societal debate regarding data sharing and privacy. Insights in user perceptions can help organizations collecting data to make ethical choices, taking into account users and their preferences. This can lead to the development of effective smart mobility services while protecting values like trust and privacy.

II. Background

A. *Challenges in Smart Mobility*

Three semi-structured interviews were conducted with various parties in the field of smart mobility in order to get insight in the current challenges relating to digitization in the mobility sector. Each one of these interviews leads to one use case illustrating a challenge with regard to smart mobility, as displayed in Table 1. The main modes of transportation that the use cases are focused on are described in the third column.

Table 1. Conducted interviews

Interview	**Use case derived from interview**	**Mode(s) of transportation**
Interview A with a municipality	Crowd management in a city	Walking, cycling, public transport
Interview B with a transport operator in a city	Real-time travel information in public transport	Metros, trams
Interview C with a research group on connected cars	Traffic information for car drivers	Vehicles

The first interview concerns a municipality, aiming for digitization of crowd management. Although cameras have been used to count the number of people at a specific moment in time, there is a wish for additional data to get a more detailed view on crowdedness. Real-time data on crowdedness could be compared with historical data to prepare, monitor, and control traffic flows, for example by providing alternative routes to travelers. A use case related to the first one is a public transport operator wishing to predict crowdedness in vehicles and communicate capacity information to travelers. To this end, data from public transport cards, weights in vehicles, and infrared sensors have been used. However, these solutions do not suffice in providing detailed information. With data collected real-time, capacity requirements at specific locations can be determined, travel behavior can be understood, and products can be improved. The challenge of data collection not only concerns public, but also private mobility. In the third interview, sharing data with respect to traffic was mentioned. For example, traffic information can be optimized by signaling accidents and proposing alternative routes to car drivers nearing an accident. This can lead to increased safety and decreased traffic delays.

B. *Crowdsensing for Smart Mobility*

Crowdsensing applications can contribute to the identified challenges in the field of smart mobility [15]. With real-time data collected from travelers' smart devices, traffic flows as well as travel behavior and travel demand can be predicted [7]. Also, this allows transport operators to deliver more personalized services to travelers [16]. By gaining more insight in the current traffic system, transportation authorities can improve transport policies. For example, they can optimize congestion charges, taxation, and subsidies, contributing to smoother and more sustainable transportation [15].

However, major challenges in crowdsensing include trust and privacy issues, as well as the provision of appropriate incentives [9]. Several studies highlight the importance of incorporating privacy-preserving mechanisms into the design of crowdsensing incentives [8]-[17]-[18]. Specifically, a major concern with respect to privacy is maintaining user-level control over sensitive sensor data [19]. Here, privacy related to crowdsensing is defined as "the guarantee that participants maintain control over the release of their sensitive information. this includes the protection of information that can be inferred from both the sensor readings as well as from the interaction of the users with the participatory sensing system”. Data captured by crowdsensing systems can reveal the identity of an individual based on location data and other data attributes and thus violate privacy [9]. A key challenge in crowdsensing is the identification of robust incentives that ensure participation of individuals, while taking into account privacy risks [8].

Previous studies have addressed incentive mechanisms for *vehicular crowdsensing* (VCS) applications [20]-[21]-[22], in which vehicles act as probes that collect information about the environment. However, in this research we choose to focus specifically on *mobile crowdsensing* applications that use sensors from mobile devices such as smartphones, smart watches or smart bracelets. These applications have a huge potential coverage since the application can be used by anyone owning a smartphone or other smart device, and solutions can be applied to different modes of both public and private transport. However, research on the trade-offs that users make when sharing data using the sensors on their smartphone is still limited [13]. As described in Section II-A, three different types of potential applications of mobile crowdsensing are identified. In Section V, The results of our study are specifically applied to these use cases in the field of smart mobility.

C. The Benefit-Cost Trade-off

A theory relevant in this context of privacy risks and behavior of consumers is Privacy Calculus Theory. This model was first proposed by Laufer & Wolfe [24]. This theory states that individuals are more likely to disclose personal information if the benefits exceed the costs of data sharing [25]. Before making a decision whether or not to share data, consumers weigh the risks and benefits to assess the outcomes, and react accordingly [26]. Privacy risks are related to the expected loss of personal information to external parties or loss of control over personal information. Benefits of information disclosure can be provided in the form of financial rewards, personalization, and social adjustment benefits [27]. In previous studies, Privacy Calculus Theory has mainly been applied in the context of individuals’ self-disclosure on social networks or on websites. However, Privacy Calculus Theory as the theoretical basis in the context of IoT applications has been limited [28]. A conceptual model that applies the Privacy Calculus Theory to the context of crowdsensing systems is presented in Figure 1.

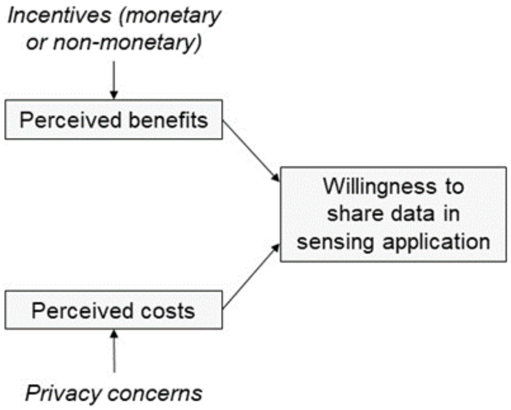

Incentives (monetary or non-monetary)
Perceived benefits
Willingness to share data in sensing application
Perceived costs
Privacy concerns

Figure 1. Privacy Calculus

In this benefit-cost trade-off, perceived benefits are either monetary or non-monetary incentives, which motivate individuals to engage in crowdsensing systems. Perceived costs are concerns about the disclosure of location or other personal data. Individuals are assumed to weigh these incentives and privacy concerns and make a decision on whether or not to participate in the crowdsensing system according to this trade-off.

D. Basics of Choice Modelling

This trade-off is made explicit in our research by applying a *choice modelling* approach. Choice modelling has been widely used in the field of travel behavior for identifying preferences for travel options that are not revealed in the market [29], for example to explore the choice travelers make between different travel routes [30]. Other examples of the application of choice modelling in the field of transportation are studies determining travelers’ willingness to pay for advanced public transport information services [31], the willingness to pay for safety improvements in passenger air travel [32], and the willingness to adopt Mobility as a Serivce (MaaS) in metropolitan areas [33].

A discrete choice model describes the choices of decision-makers between different alternatives [34]. The *Multinomial Logit (MNL)* model, which is the most well-known discrete model, is derived by assuming that a decision-maker faces a choice among a certain amount of alternatives. When choosing an alternative, the decision-maker obtains a certain level of *utility* (or satisfaction). The utility of an alternative is composed of a systematic part (V_i), which can be measured by the researcher, and an unobserved part (ε_i), which is an error term representing unobserved factors, heterogeneity in tastes, or randomness in choices.

The utility of an alternative i is defined as:

$$U_i = V_i + \varepsilon_i \tag{1}$$

where:
i = alternative, e.g., scenario 1, scenario 2
U_i = utility of alternative i
ε_i = unobserved utility of alternative i (error term).

In this equation, the systematic part is defined as:

$$V_i = \sum_m \beta_m x_{im} \quad (2)$$

where:
m = attribute, e.g., monetary reward, type of data
β_m = attribute weight for an attribute m in alternative i (to be estimated in the model)
x_{im} = attribute value of attribute m for alternative i, e.g., €20, €40.

According to *Random Utility Maximization* (RUM) theory, which is underlying the MNL model, the decision-maker will choose the alternative providing the greatest utility [34]-[35].

These choices can only be predicted up to a probability because of the error term, *i.e.,* a higher systematic utility means there is a higher probability of the alternative being chosen. This probability is determined as:

$$p_i = \frac{e^{V_i}}{1+e^{V_i}} \quad (3)$$

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

In the MNL model, the assumption is that no correlation exists between the choices made by an individual. However, in reality, repeated choices made by the same individual are correlated, caused by variation in preferences and tastes across individuals, as well as (partial) stability in preferences and tastes within the individual, across time. Since this implies the assumption that the dataset contains more information than it does in reality, we correct for this error by estimating *Mixed Logit* (ML) models for panel data in addition, which are able to capture utility-correlation between consecutive choices of respondents [34]. In order to obtain the ML model, an additional error component is added to the model, capturing (part of) the correlation between choices made by the same individual. This error component is drawn from as a zero mean continuous, normal distribution, in such a way that only the variance (σ) has to be estimated.

Our research aims to explain choice behavior in terms of underlying factors. Since there is a lack of understanding in the preferences of potential contributors to crowdsensing systems, a choice modelling approach is suitable to address the knowledge gap. Approaches suggested in literature to design effective incentive mechanisms for crowdsensing applications have mostly relied on purely game-theoretic approaches. Users in crowdsensing systems might, however, exhibit different behavior. Modelling approaches can therefore complement previous studies by exploring the design space of user behavior [9]. The underlying factors that are used to explain choice behavior in crowdsensing applications for smart mobility are discussed in the next paragraph.

E. Potential factors influencing the willingness to share data

Individuals have a varying degree of perceived privacy concerns and attach a different value to their privacy [27]. This potentially influences the described benefit-cost trade-off. Kumaraguru & Cranor [28], reviewing works on privacy indexes by Westin between 1978 and 2004, mention three categories referring to different groups regarding privacy concerns: 1) Privacy Fundamentalist, 2) Privacy Pragmatist, and 3) Privacy Unconcerned. In order to derive this Privacy Index, the statements as displayed in Table 2 are used [36]. Respondents agreeing with the first statement and disagreeing with the second and third statements, are considered Privacy Fundamentalists. Privacy Unconcerned respondents are respondents who disagree with the first statement and agree with the second and third statements. All other respondents are categorized as Privacy Pragmatists. Since these attitudes towards privacy could influence the willingness of potential users to share data in a crowdsensing application, these are taken into account in the experiment.

Table 2. Privacy Index

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

A literature review was conducted in order to identify factors influencing the benefit-cost trade-off made by individuals when deciding whether or not to share data in crowdsensing applications for smart mobility. Since including too many factors in a choice experiment can lead to increased choice difficulty [37], five out of fourteen factors were selected to be included in the choice experiment, based on their prevalence in literature and the possibility to be influenced by policy or design.

Monetary reward

Bhatnagar & Kumra [38] found a significant positive impact of extrinsic, monetary rewards on the willingness to share IoT product data (n = 337). This is confirmed by research by Turland & Slade [39], concluding that participation rates are significantly higher when providing even a small monetary benefit. Furthermore, recent work shows that rewarding drivers with a small incentivization budget when taking minor detours towards roads with a higher sensing demand, can lead to significant improvements in spatio-temporal coverage, especially for minor roads [22].

Required effort

Effort is seen as the time spent by performing sensing tasks. The expectation is that a higher effort has a negative effect on the willingness to participate. Salim & Haque [11] distinguish three levels regarding user engagement in crowdsensing systems. The lowest level (aware and consent) means that participants are aware of their participation and provide consent for data being collected, but their interaction with the system remains minimal or passive. Engaged users are more actively involved in the system and interact with the system by adding their feedback and experiences. At the collaborative level, which is the highest level of participation, users actively contribute more data and aim for a better coverage in data collection activities [11]. For example, participants can be asked to derive from their initial planned trajectory in order to achieve a better-balanced sensing coverage [20]-[22]. It is expected that a higher engagement level leads to a lower willingness to share data. Besides privacy concerns, this is therefore an additional "perceived cost" that is expected to influence the Privacy Calculus.

Risk of re-identification

The protection of users' identity is a core aspect for privacy. In a study by Schomakers *et al.* [40], anonymization is discovered as the most important factor that influences users' decision to share data (n = 126). Collected mobility data are potentially sensitive, since they could be used to reconstruct information about individual participants, such as commute patterns, routines, or private locations. For example, collection of GIS (Geographical Information System) coordinates simplifies the process of identifying the exact location of drivers, but also increases risks regarding privacy and security [41]. Thus, the risk of re-identification potentially influences the privacy risk as perceived by the user.

Types of data

Several sensors embedded in smart devices can be used for data collection. These sensors include sensors for localization (GPS, Wi-Fi, Bluetooth), physical motion sensors (accelerometer, gyroscope), environmental or contextual sensors (temperature and humidity sensors, barometer), and multimedia sensors (camera, microphone) [11]-[42]. According to Christin *et al.* [43], time and location data, sound samples, pictures and videos, acceleration, and environmental data can be potential threats to privacy when being shared with unauthorized parties. Especially when linked with other information provided by individuals, sharing microphone and camera data can be a threat to privacy [38]. The types of data being collected may therefore influence the perceived privacy risk. The willingness of people to share data in a crowdsensing system may depend on the kind of data being collected [40]-[45].

Data use

Another factor potentially influencing the willingness of individuals to share data, is the party with whom the data are shared [15]-[46]-[47]. Users are found to be more reluctant to share data with corporate institutions compared to academic institutions [43]. This finding is confirmed by research by Aitken *et al.* [48], showing that participants have greater support for data usage by the public sector compared to usage by the private sector. However, Turland & Slade [39] find an opposite effect in a study on crowdsensing for farm management. Here, users are more concerned about sharing data with the government compared to sharing with private organizations, echoing concerns about government surveillance. Related to the data collecting party, the purpose for which the data are used can play a role for individuals when considering participation in sensing tasks [8]-[40]-[48].

Factor	**Levels**
Monetary reward	€0/month
	€20/month
	€40/month
	€60/month
Effort	Low
	Moderate
	High
Risk of re-identification	10%
	20%
	30%
Types of data	Time and location data
	Time and location data, Motion data
	Time and location data, Motion data, Contextual data
	Time and location data, Motion data, Contextual data, Multimedia data
Data use	Governmental institution aiming to improve mobility
	Academic institution aiming to investigate transport modes
	Corporate institution aiming to improve products or services
	Societal organization aiming to address local issues related to mobility

III. Methodology

A. Stated Choice Experiment

Data collection is required to model factors influencing choice behavior in crowdsensing applications. A choice can be made between collecting *revealed preference (RP)* data, mirroring the actual choices people made in real-life situations, or *stated preference* data, by presenting respondents with hypothetical choice scenarios [34]. Since crowdsensing is an emerging field for which no historical data are available yet, stated preference data were collected instead of revealed preference data. These data were collected by conducting a *stated choice experiment,* in which respondents have to choose whether they would share data or not in a presented situation.

B. Experiment Design

The hypothetical choice situations, also referred to as *choice sets,* were constructed by operationalizing the five factors identified in section II into *attributes.* Each attribute is varied in different *attribute levels.* These levels are based on the literature review as described in section II. The numerical levels require some additional explanation. Several previous studies use monetary benefits that range between €5 and €75 per month [49]. In order to maintain equal distance between the attribute levels, we use a comparable range from €0 and €60 to vary the *monetary reward.* The risk of re-identification is derived from the *k*-anonymity factor as used in previous research [40]. To increase the comprehensiveness of the attribute for respondents without knowledge on anonymization techniques, we reframed this attribute as the *risk of re-identification.* Since a large gap was observed in previous research between attribute levels because of unrealistic values [40], we use three levels with a slight difference in risk of re-identification. This risk of re-identification is roughly based a *k*-anonymity of 1 out of 10 (10%), 1 out of 5 (20%), and 1 out of 3 (30%). Since "complete anonymization" does not exist in reality, such a level was not used.

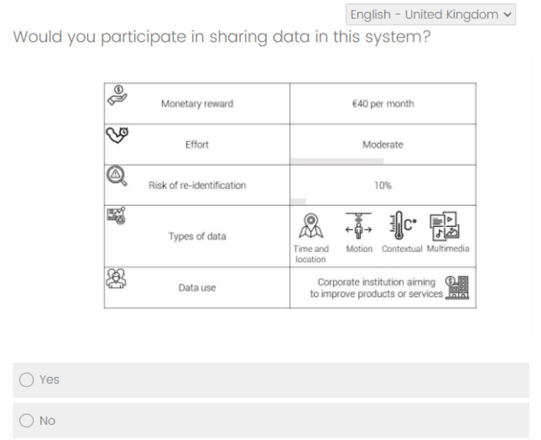
All attributes and attribute levels varied in the experiment are presented in Table 3.

Table 3. Attributes and attribute levels

Based on the attributes and attribute levels, the software *NGene* was used to construct choice sets for the choice experiment. The software finds the minimum number of choice scenarios, making sure that attribute levels have zero correlation, all pairs of attribute levels occur equally often across all pairs of alternatives, and each level occurs an equal amount of times for each alternative (also known as an *orthogonal design*) [50]. Taking into account these conditions, *NGene* found 36 choice sets, which is a number of scenarios that can still be perceived as too exhausting for a single respondent. Therefore, *blocking* was applied, by dividing the choice sets into three smaller blocks of choice sets. This means that each respondent was presented with only 12 scenarios. Respondents were randomly assigned to block 1, block 2 or block 3.

C. Survey Design

The stated choice experiment was conducted through an online survey. The survey included an introduction to the topic and the experiment, after which 12 hypothetical choice situations were presented. An example of a choice scenario as presented to respondents is displayed in Figure 2.



English - United Kingdom

Would you participate in sharing data in this system?

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

○ Yes

○ No

Figure 2. Example of a choice situation in the choice experiment

At the end of the survey, the statements related to the Privacy Index, as well as questions related to personal characteristics were presented, including age, gender, education, income level, digital behavior, and altruism.

IV. RESULTS ANALYSIS

A. Descriptive analytics

The target population comprised all smart device owners being 18 years or older. A total number of 125 appropriate records were collected. Of these 125 respondents, 38 participated in block 1 (30,4%), 40 in block 2 (32,0%), and 47 in block 3 (37,6%). In total, the "yes" option was chosen in 39% of the cases, and the "no" option was chosen in 61% of the cases. In order to see how respondents perceive privacy, they were categorized using Westin's Privacy Index, as described in Section II-E. We found that, in our sample, 6% of respondents belong to the Privacy Unconcerned category, 47% to the Privacy Pragmatists, and 47% to the Privacy Fundamentalists (Figure 3). When compared to percentages from previous surveys reported by Woodruff *et al.* [51], it appears that the Privacy Unconcerned percentage is almost similar (5-10%), the Privacy Pragmatist percentage is slightly higher (40-58%), and the Privacy Fundamentalist percentage is slightly lower (34-49%) in previous research.

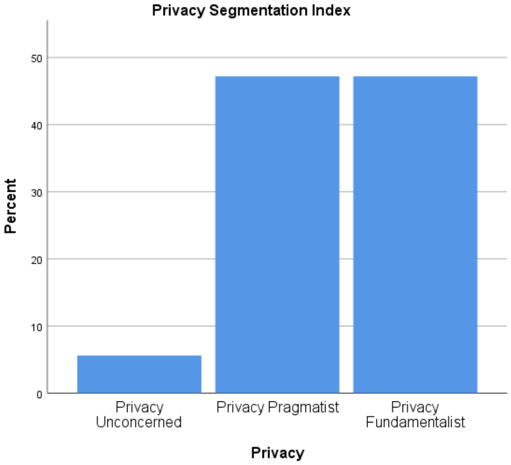


Figure 3. Categories according to Privacy Index

Using the obtained data, several models were estimated. This estimation process is discussed in Section IV-B. Based on the performance of the estimated models, the model that fits the data best is selected for further analysis in Section IV-C. Section IV-D elaborates on the final model that is selected. In Section IV-E*,* conclusions are drawn on the relative importance of factors on the willingness to share data, after which these results are discussed more elaborately in Section IV-F.

B. *Model estimations*

First, an MNL model was estimated including only the *main effects* of the factors, which is the simplest model and considered the base model. The systemic utility of this model is as follows:

$$V_{yes} = \beta_{yes} + \beta_{mon} \cdot mon + \beta_{eff} \cdot eff + \beta_{rid} \cdot rid + \beta_{tod} \cdot tod + \beta_{use_{aca}} \cdot (use == 1) + \beta_{use_{cor}} \cdot (use == 2) + \beta_{use_{soc}} \cdot (use == 3)$$

$$V_{no} = 0 \qquad (4)$$

where:

V_{yes} = the systematic utility of sharing data
V_{no} = the systematic utility of sharing no data
β_{yes} = the base utility (constant) of choosing the “yes” option
β_{mon} = the marginal utility of the factor monetary reward
β_{eff} = the marginal utility of the factor effort
β_{rid} = the marginal utility of the factor risk of re-identification
β_{tod} = the marginal utility of the factor type of data
$\beta_{use_{aca}}$ = the marginal utility of the factor data use by an academic institution
$\beta_{use_{cor}}$ = the marginal utility of the factor data use by a corporate institution
$\beta_{use_{soc}}$ = the marginal utility of the factor data use by a societal organisation.

Since the factor *data use* has categorical levels, this factor is *dummy coded.* Table 4 defines how this factor is coded. The parameters $use_{aca},$ $use_{cor},$ and use_{soc} are estimated by comparing them to the reference category, which is data use by a governmental institution.

Table 4. Dummy coding of data use factor

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

Besides the base model, a variety of other models were estimated, e.g., by including the effects of privacy perceptions or personal characteristics on the willingness to share data, also known as *interaction effects.* Furthermore, several Mixed Logit (ML) models were estimated, in order to capture heterogeneity in choices. We used 500 *Halton* draws to estimate the ML models, since these draws provide better coverage compared to using random draws [34]. Out of the estimated models, only the 14 most interesting models were included in the research, based on their performance and new insights they provide. All models were estimated using the Apollo package, which is a statistical tool in *R.* For every model, the estimation outcomes include the parameter estimates, measures for goodness of fit, and standard errors associated with the parameter estimates.

C. *Model performance*

An essential step in discrete choice modelling is the selection process of models. Several ways exist that allow comparing different models. However, the model that is most useful for a given dataset depends highly on the purpose and context of the research [50].

The first strategy to compare models statistically is based on the estimated *Log-Likelihood* values. These values provide information on how well a model explains the data. For each model, the *Likelihood Ratio Statistic (LRS)* was calculated, which is obtained by (5):

$$LRS = -2 \cdot (LL_A - LL_B) \qquad (5)$$

where:

LL_A = Log-Likelihood of null model
LL_B = Log-Likelihood of estimated model.

The *LRS* is used to evaluate whether the model performs better than “throwing a dice”. Specifically, the statistic indicates how well the model performs compared to a model in which all parameters are set to zero (the *null model*) [34].

If the *LRS* is higher than the threshold according to the χ^2 table, the conclusion can be drawn that the estimated model is better than the null model at a given significance level. A higher *LRS* indicates a better model fit. Since the *null-Log-Likelihood (LL(0))* is similar for all models in this case, a higher *LL(final*) indicates a better performance.

A second statistic for scoring and comparing models is the *Bayesian Information Criterion (BIC)*. The *BIC* value is based on the Log-Likelihood of the models and gives a penalty to more complex models that include a higher amount of parameters. A lower BIC value is considered better than a higher BIC value [34].

Lastly, *McFadden's Rho-squared* is a widely used measure for the goodness-of-fit of discrete choice models, measuring the uncertainty that is explained by the model, which is defined as (6)

$$McFadden's\ pseudo\text{-}R^2 = 1 - \frac{\ln(LL)}{\ln(LL_0)} \qquad (6)$$

The obtained value is always in the range of [0,1] and a higher value represents a better model fit. Note that this measure should be used in a relative sense and there is no rule of thumb for what is a "good fit" [50].

All 14 models were evaluated based on their model fit. It appeared that the MNL models with interaction effects did not explain the data better than the base model. The ML model with an additional error term performed better than the MNL model and was thus further inspected. The ML model with an error term for all parameters had the best model fit. The performance metrics for these three models are presented in Table 5.

Table 5. Performance of estimated model

The base model, which is an MNL model only including the five parameters, explains 15.53% of the initial uncertainty. The Likelihood Ratio Test resulted in a value of 322.96. This is higher than the χ^2 value with 8 degrees of freedom, which is equal to 15.507. Thus, we conclude that the estimated model fits the data better than the model of throwing a dice. The Mixed Logit model with an additional error term added to each parameter appeared to perform best on the data, compared to the other estimated models. In this model, unobserved heterogeneity is captured and corrected for. The model has a Rho-square of 0.3499, meaning that it explains 35% of the initial uncertainty. Also, the BIC for this model is relatively low, and the final Log-Likelihood is relatively high, meaning that the Mixed Logit model with all parameters random performs better when compared to the other models. Thus, this model is selected as the final model for further analysis.

D. Parameter estimates

The parameter estimates of the final model, including the standard errors of these parameters, are presented in Table 6. The second column presents the parameters estimated in the model. The third column (estimate) shows the estimated weight of the factors, which is defined as the utils gained or lost by 1 unit increase of the attribute. The fourth column displays the standard errors associated with the parameter estimates, illustrating the variation of the estimate across the sample. This standard error is a measure of uncertainty about the true β. The fifth and sixth column present the *t*-Test results. The *t*-ratios, which are based on the parameter divided by its standard error, are used for determining if the attributes have an effect on choices in the population. Based on these *t*-ratios, the *p*-values are computed. Factors with an indicated *p*-value being higher than 0.05 are considered statistically insignificant, meaning that no effect can be observed in the population. More specifically, *p* can be seen as the probability that the null-hypothesis that the true β (in the population) = 0, is true. If this probability is smaller than 0.05, this null hypothesis is rejected.

The parameter for data use by societal organizations appears to be insignificant (p>0.05). This means we have too little evidence to reject the null-hypothesis (i.e., that β = 0 in the population). However, since our research has a design goal rather than a social science question and this estimate is still the best guess for the parameter, we choose to keep the parameter included in the model. All sigma's (the error terms) are significant. Thus, it can be observed that more heterogeneity is captured in this model, which has extra error terms for all parameters. The monetary reward parameter has a positive sign, representing a positive effect on the willingness to share data, while the *effort, risk of re-identification* and *type of data* parameters have a negative sign, which is as expected.

Table 6. Estimates ML model with all parameters random

Description	LL(final)	LL(0)	BIC	Rho-square
MNL with main effects	-878.24	-1039.72	1814.98	0.1553
ML model with error term	-735.61	-1039.72	1537.03	0.2925
ML model with all parameters random	-675.93	-1039.72	1468.87	0.3499

Attribute	Parameter	Estimate	s.e.	Rob.t.rat.(0)	p(1-sided)
Monetary reward	$\boldsymbol{\beta_{mon}}$	0.9956	0.1898	4.682	0.0000
Effort	$\boldsymbol{\beta_{eff}}$	-1.9692	0.2885	-6.225	0.0000

Risk of re-identification	$\boldsymbol{\beta_{rid}}$	-1.1258	0.2402	-4.845	0.0000
Types of data	$\boldsymbol{\beta_{tod}}$	-0.9455	0.1583	-5.382	0.0000
Data use by academic institution	$\boldsymbol{\beta_{useaca}}$	0.9819	0.3910	2.353	0.0093
Data use by corporate institution	$\boldsymbol{\beta_{usecor}}$	-2.2735	0.5754	-3.184	0.0000
Data use by societal organisation	$\boldsymbol{\beta_{usesoc}}$	-0.2708	0.3287	-0.713	0.2379
Base utility (constant)	$\boldsymbol{\beta_{yes}}$	2.1037	0.4666	4.491	0.0000
Error term base utility	$\boldsymbol{\sigma_{yes}}$	2.8825	0.4840	5.475	0.0000
Error term monetary reward	$\boldsymbol{\sigma_{mon}}$	1.4645	0.2493	4.949	0.0000
Error term effort	$\boldsymbol{\sigma_{eff}}$	1.2758	0.2346	5.346	0.0000
Error term risk of re-identification	$\boldsymbol{\sigma_{rid}}$	1.6647	0.3079	-4.832	0.0000
Error term types of data	$\boldsymbol{\sigma_{tod}}$	0.6276	0.1465	4.611	0.0000
Error term data use by academic institution	$\boldsymbol{\sigma_{useaca}}$	1.6863	0.4099	3.925	0.0000
Error term data use by corporate institution	$\boldsymbol{\sigma_{usecor}}$	3.7651	0.7929	3.807	0.0000
Error term data use by societal organisation	$\boldsymbol{\sigma_{usesoc}}$	1.4526	0.4781	2.523	0.0058

E. *Model interpretation*

The parameter estimates cannot be interpreted directly. Therefore, by using the utility ranges of the attributes, the *relative importance* of each factor in the decision to share data in a crowdsensing application is calculated. The relative importance can be obtained by calculating the utility contribution of an attribute as a percentage of the sum of utility contributions. The utility contribution is calculated by multiplying the estimate with the maximum value of the attribute. According to the utility contributions, effort is the most important factor that affects the benefit-cost trade-off by individuals (26%), as presented in Figure 4. The other factors playing a role in the decision, in order of importance, are types of data (21%), data use (18%), monetary reward (18%), and risk of re-identification (17%).

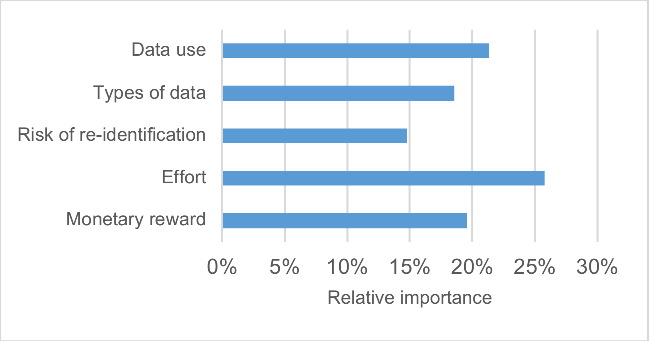


Figure 4. Relative importance of attributes (ML model)

F. *Discussion of results*

The factors that were included in the model are discussed in the following paragraphs, in order of importance.

Effort

Heiskala et al. [15] notes that users may feel overburdened when applications ask them to report observations. No previous study was found that included both required effort and privacy-related attributes in a choice experiment. From the results, however, it can be concluded that effort does play a highly important role in consumers' decisions regarding data sharing, and is regarded as more important than the types of data being collected, the data use, or the risk of re-identification.

Data use

When making a decision whether or not to participate, users take into account the party collecting the data and for what purpose. According to the results, the potential parties collecting data, ranked from most accepted to least accepted are 1) An academic institution aiming to investigate transport modes, 2) A governmental institution aiming to improve mobility, 3) A societal organization aiming to address local issues regarding mobility, and 4) A corporate institution aiming to improve products or services. Data collection by academic institutions is most accepted, while data collection by corporate institutions for improving products and services is least accepted. These findings are in line with previous studies on data sharing, in which was found that people claim higher rewards when sharing data with corporate institutions when compared to academic institutions [40]-[43].

Monetary reward

The finding that people are more likely to share data when receiving a higher financial reward in return is in line with research by Derikx *et al.* [52] and Schomakers *et al.* [40]. These studies also found that individuals are more willing to share data when a financial compensation is offered in return.

Train [34] describes a function to calculate the Value of Time. This is defined as the extra cost that a person would be willing to incur in order to save time. This Value of Time is calculated by using the estimated coefficients of cost and various time components. Inspired by this definition, we can specify a function that calculates the *Value of Privacy (VoP)*. This function is defined in (7).

$$VoP = \frac{\frac{\delta V}{\delta RID}}{\frac{\delta V}{\delta MON}} = \frac{\beta_{rid}}{\beta_{mon}} \qquad (7)$$

Here, β_{rid} presents the estimate for the risk of re-identification attribute, and β_{mon} presents the estimate for the monetary reward attribute. According to the base model, the Value of Privacy is equal to 1.41 €/pp, meaning that people want to receive an amount of €14.10 per month if the risk of re-identification is increased by 10 percentage points.

Types of data

In order to determine if an unequal utility contribution from each attribute level can be observed for the types of data factor, the factor was tested for non-linearity effects. In Figure 5, the results are displayed. The utility contribution displayed on the y-axis is a measure to compare the impact that attribute levels have on the choice made by individuals. Since types of data has a negative impact on individuals' choices, a higher utility contribution indicates that the attribute level has a higher negative impact on the utility (satisfaction) of people. From Figure 5, a large increase in disutility can be observed when moving from the third to the fourth level, which means multimedia data is shared in addition. This means that people are more sensitive to the types of data attribute when more types of data are being shared.

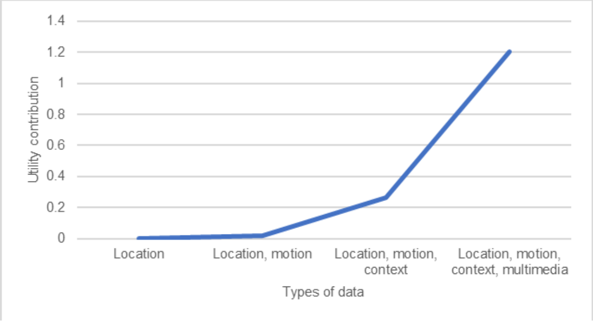


Figure 5. Utility contribution of *types of data* attribute levels

Our finding that there is a large gap in acceptance between collection of location, motion and context data and the collection of multimedia data in addition to these data, confirms previous research on sharing data from smart home applications [47]. The collection of motion data in addition to location data is highly accepted by users, while the collection of multimedia causes a large decrease in acceptance.

Risk of re-identification

The risk of re-identification appeared to be least important to participants in our study. From the statements categorizing individuals in Westin's Privacy Index, it appears that only a small number of people were categorized as Privacy Unconcerned, while a percentage of 47% of respondents were in the "Privacy Fundamentalist" group. This means that people in our study indicate to be concerned about their privacy and about how their data is handled. However, this is not clearly reflected in the choices they make, since the risk of re-identification factor has a relatively low importance.

In order to see if people being highly concerned about their privacy assign a higher importance to the risk of re-identification factor, the preferences of the Privacy Fundamentalists were further examined by estimating an MNL model for only this group of people. When comparing the estimated parameters to the estimates for the whole sample, it appears that the Privacy Fundamentalist group assigns a slightly higher importance to privacy-related factors (risk of re-identification, types of data, data use). However, these differences are insignificant in the population.

Since 47% of respondents indicate being highly concerned about their privacy, the fact that the risk of re-identification is regarded as the least important factor seems unexpected. This phenomenon can be explained by a concept known as the "Privacy Paradox", describing that on one hand, people express their concerns about the handling of their personal data, while at the same time, they often choose to share their data voluntarily and rarely make an effort to actively protect their data [53].

V. Applications for Smart Mobility

From the conducted interviews described in Section II, it appears that parties are already collecting data. However, these data is often not real-time. By involving users, mobile crowdsensing could be a valuable data source in addition to existing sources of data. Yet, the challenges of how to take into account users' preferences, and how to achieve a sufficient coverage, remain. In order to get insight in what factors could be influenced by parties in the field of smart mobility (municipality, car manufacturers, and transport operators) to design crowdsensing applications with the desired participation rate, the use cases as described in Section II-A were used for the purpose of defining choice situations, with attributes as varied in our choice experiment. Using the results from the estimated base model, the probabilities of acceptance for these combinations of attribute levels not included in the choice experiment, can be calculated. Table 7 shows an example of three different scenarios for implementing crowdsensing as a solution to digitize crowd management in a city.

Table 7. Participation rates for crowd management in a city

Factor	**Scenario 1**	**Scenario 2**	**Scenario 3**
Monetary reward	€60 per month	€60 per month	€60 per month
Effort	Low	Moderate	Low
Risk of re-identificati on	10%	10%	10%
Types of data	Location data, motion data	Location data, motion data	Location data, motion data, contextual data
Data use	Governmen tal institution aiming to improve mobility	Governmen tal institution aiming to improve mobility	Governmen tal institution aiming to improve mobility
Participati on rate	**80%**	**66%**	**73%**

By analyzing the expected participation rates, two main trade-offs for designing crowdsensing applications for smart mobility were identified. One finding was that a trade-off has to be made between the reliability of collected data and the participation rate. Requiring more inputs from users can enhance the reliability of the data, since feedback is provided about, e.g., the precise location of an individual. However, this also requires more effort from users, leading to a lower participation rate. This finding aligns with and adds to the Technology Acceptance Model originated by Davis [54], stating that perceived ease of use influences the likelihood of an application being adopted by a user.

Another trade-off needs to be made between the richness of information and the privacy of users. Collecting more data, like contextual data and multimedia, can provide more information about the context and travelling situation of a specific user [55]. Especially for applications in transport modes like trams or buses, which are often driving next to other vehicles in a crowded city, more accurate data are needed to determine the exact location and current mode of transport of a user. Since research by Masoud *et al*. [42] argued that the accuracy of location information can be improved by collecting contextual data, this could be an effective way to gain accurate insights. Also, multimedia data can be collected to characterize places more easily by using location-tagged images and videos and can further enrich the obtained information. However this also increases the risk of a data leading to a specific individual. Consequently, the collection of additional data results in a lower participation rate. Depending on the specific requirements for the smart mobility application, as well as the specific mode of transport for which the application is being used, these trade-offs can be evaluated differently.

VI. Discussion

A. Scientific and societal implications

Since little research has been conducted before on user behavior and incentives in mobile crowdsensing systems, our study adds to this field of research. Future research can build upon this study by investigating the user-friendliness of crowdsensing applications and further investigating the effort attribute, which appeared to be an important factor for potential users. Also, our study provides an insight in the monetary value users assign to their privacy when presented with different choice scenarios. Further research should be conducted on the risks of crowdsensing applications and user perceptions of these risks.

Besides scientific implications, this study also has societal implications. Our research goes beyond the technological aspects of crowdsensing applications and provides insights in values and trade-offs that come with developing crowdsensing applications for smart mobility. Taking into account the preferences of users as discovered in our research allows for a value-sensitive design of crowdsensing applications. Based on our study, practical guidelines can be derived for designing crowdsensing applications. We propose three principles:

- Tailor-made applications, by giving users control over the data they want and do not want to share, thus meeting differing preferences among users.
- Transparency by design, by being transparent to users about potential risks and involving their concerns in the design process from the beginning.
- Ensuring attractiveness of applications, by minimizing the user-burden and giving social incentives such as awards.

B. Limitations

When interpretating the results of this study, several limitations should be noted. Three limitations are mentioned.

The first limitation concerns the setup of the experiment. In order to limit the length of the survey, the amount of attributes was reduced to only 5 factors. However, 14 factors in total were identified in the literature research. Because of this simplification, other factors potentially influencing the willingness to share data are unaddressed in this study. Although the choice of factors was made carefully, considering a higher amount of factors or attribute levels may lead to a richer explanation of the choice behavior. Also, when setting up the experiment, the ranges of the monetary reward and risk of re-identification attribute levels were determined based on a careful review of previous experiments. However, choosing a higher or lower upper limit for these levels may have led to different results on the relative importance of these factors. Future research could investigate the amount of money that participants would want to receive in turn for sharing their data, to determine the most realistic range for this attribute. Previous research found that people even want to *pay* money when receiving relevant personalised promotions from their insurance company [52]. A future experiment could investigate if people would want to pay for participating in a sensing application, if being provided with a useful service.

Secondly, we should mention that the risk of re-identification attribute gives a limited indication of the degree of privacy protection. The risk of re-identification was based on the principle of k-anonymization. In our study, the risk of re-identification appeared to be the least important attribute that influences the decision of individuals to share data, a finding which is not in line with research by Schomakers et al. [40]. This could be due to two reasons. First, in the experiment set-up, we varied the attribute between 10% and 30%, which is a limited range. Although "full" anonymization is not really possible (which would be the 0% level), in reality the range could be larger, like mentioned in Section VI-A. Furthermore, the question can be asked whether participants fully understood the meaning of the risk of re-identification and what effect it can have on their privacy. In the survey, we provided an explanation of the attribute. However, the risk, described in percentages, could still have been a rather intangible attribute to respondents.

Lastly, the choice for applying choice modelling as the main method in this research requires some reflection. A discrete choice experiment makes the trade-offs explicit that play a role in the motivation of individuals with respect to data sharing, and allows for a prediction of future choices of individuals. The discrete choice experiment indirectly recovers the values behind people's choices, which provides insights on ethical aspects to incorporate in application design. The choice modelling approach is unique in assigning a numerical value to the weighing of these underlying motivating factors. However, the ethical aspects regarding privacy and trust should not be overlooked. Even if consumers indicate they want to sacrifice a part of their privacy in turn for some benefit, the question should always be asked if certain data should indeed be collected [56].

VII. Conclusion

This paper explored factors influencing the willingness of smart device owners to share data in mobile crowdsensing applications for smart mobility. Implementing crowdsensing applications is a potential solution contributing to smarter and more efficient mobility systems. When choosing whether or not to share data in a crowdsensing application, we found that the required effort of participation is regarded the most important factor by respondents. The risk of re-identification was found to be the least important factor in our sample. However, this does not mean that privacy is not important to smart device users. A relatively high amount of respondents (47%) indicated being concerned about their privacy. The phenomenon known as the Privacy Paradox could have played a role, stating that people tend to indicate being highly concerned about their privacy, while this is not reflected in their actual choices.

This study provides a new understanding of user preferences in crowdsensing systems for smart mobility. While previous research investigated the benefit-cost trade-off looking at privacy- and money-related factors, we show that the required effort is an additional factor which is highly important to users. Also, new insights were gained in the preferences of people that are highly concerned about their privacy and the handling of their personal data, by using the indexes as defined by Westin. Besides understanding the user side of crowdsensing applications, we were able to get insight in the challenges experienced by parties in the field of smart mobility when implementing such applications. By conducting interviews with relevant parties, we gained an understanding of current developments in the smart mobility sector. Through combining both quantitative and qualitative approaches, our study explores the expected participation of potential users in different scenarios, which were derived from realistic use cases. The insights provided by this study form a basis for further research on perceived benefits and privacy perceptions of users, contributing to the design value-sensitive and effective smart mobility services, which are becoming more critical for ensuring the efficiency, safety, and sustainability of transportation systems.

Acknowledgment

The authors thank AKKA Netherlands, in particular dr. R. Kherrazi, for providing their support during the research process in the period from September 2021 to January 2022. Furthermore, we thank prof.dr.ir. C.G. Chorus and T. Szép, MSc. (TU Delft) for providing methodological advice during the research process.

References

[1] J. Jin, J. Gubbi, S. Marusic, and M. Palaniswami. "An information framework for creating a smart city through internet of things," *IEEE Internet of Things Journal*, vol. 1, no. 2, pp. 112–121, 2014.

[2] T. Paalosmaa and M. Shafie-Khah. "Feasibility of innovative smart mobility solutions: a case study for Vaasa," *World Electric Vehicle Journal*, vol. 12, pp. 188, 2021.

[3] S. Paiva, M.A. Ahad, G. Tripathi, N. Feroz, and G. Casalino. "Enabling technologies for urban smart mobility: Recent trends, opportunities and challenges," *Sensors*, vol. 21, no. 6, pp. 1–45, 2021.

[4] P. Ribeiro, G. Dias, and P. Pereira. "Transport systems and mobility for smart cities," *Appl. Syst. Innov.*, vol. 4, no. 3, 2021.

[5] N.B. Truong, G.M. Lee, T.W. Um, and M. MacKay. "Trust evaluation mechanism for user recruitment in mobile crowd-sensing in the Internet of Things," *IEEE Trans. on Info. Forensics and Security*, vol. 14, no. 10, pp. 2705–2719, 2019.

[6] International Telecommunication Union. (2016). Mobile network coverage by country. Available: https://www.theglobaleconomy.com/rankings/Mobile_network_coverage/

[7] R. C. Shit, "Crowd intelligence for sustainable futuristic intelligent transportation system: A review," *IET Intell. Transp. Syst.,* vol. 14, no. 6, pp. 480–494, 2020.

[8] R. I. Ogie, "Adopting incentive mechanisms for large-scale participation in mobile crowdsensing: from literature review to a conceptual framework," *Hum.-centr. Comp. and Inform. Sciences*, vol. 6, no. 1. Springer Berlin Heidelberg, 2016.

[9] M. Riahi, R. Rahman, & K. Aberer, "Privacy, trust and incentives in participatory sensing". *Underst. compl. syst.*, pp. 93–114, 2017.

[10] R. Du, S. Member, P. Santi, M. Xiao, S. Member, A.V. Vasilakos, and C. Fischione. "The sensable city : a survey on the deployment and management for smart city monitoring," *IEE Comm. Surveys & Tutorials.,* vol. 21, no. 2, pp. 1533–1560, 2019.

[11] F. Salim and U. Haque. "Urban computing in the wild: a survey on large scale participation and citizen engagement with ubiquitous computing, cyber physical systems, and Internet of Things," *Internat. Journal of Hum. Comp. Studies*, vol. 81, pp. 31–48, 2015.

[12] S. Bennati, I. Dusparic, R. Shinde, and C.M. Jonker. "Volunteers in the smart city: comparison of contribution strategies on human-centered measures," *Sensors (Switzerland),* vol. 18, no. 11, 2018.

[13] F. Restuccia, S.K. Das, and J. Payton. "Incentive mechanisms for participatory sensing: survey and research challenges," *ACM Transact. on Sensor Networks,* vol. 12, no. 2, pp. 1–40, 2016.

[14] X. Kong, X. Liu, B. Jedari, M. Li, L. Wan, and F. Xia. "Mobile crowdsourcing in smart cities: technologies, applications, and future challenges," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8095–8113, 2019.

[15] M. Heiskala, J. P. Jokinen, and M. Tinnilä, "Crowdsensing-based transportation services - An analysis from business model and sustainability viewpoints," *Res. Transp. Bus. Manag.*, vol. 18, pp. 38–48, 2016.

[16] Z. Xiao, H. Lim, and L. Ponnambalam, "Participatory Sensing for Smart Cities : A Case Study on Transport Trip Quality Measurement," *IEEE Trans. on Industr. Inform.,* vol. 13, no. 2, 2018.

[17] V. Kotovirta, T. Toivanen, R. Tergujeff, and M. Huttunen, "Participatory Sensing in Environmental Monitoring - Experiences," *Proc. - 6th Int. Conf. Innov. Mob. Internet Serv. Ubiquitous Comput. IMIS 2012*, pp. 155–162, 2012.

[18] H. Ma, D. Zhao, and P. Yuan, "Opportunities in mobile crowd sensing," *Infocommunications J.*, vol. 7, no. 2, pp. 32–38, 2015.

[19] D. Christin, A. Reinhardt, S. S. Kanhere, and M. Hollick, "A survey on privacy in mobile participatory sensing applications," *J. Syst. Softw.*, vol. 84, no. 11, pp. 1928–1946, 2011.

[20] A. Chakeri, X. Wang, Q. Goss, M.I. Akbas, and L.G. Jaimes, "A platform-based incentive mechanism for autonomous vehicle crowdsensing, *IEEE Open Journal of Intell. Transp. Syst.,* vol. 2, pp.13-23, 2021.

[21] J.A. Khan and Y. Ghamri-Doudane, "ROVERS: Incentive-based recruitment of connected vehicles for urban big data collection," *IEEE Trans. On Veh. Technology,* vol. 86, no. 6, 2019.

[22] S. Di Martino and L.L.L. Starace, "Towards uniform urban map coverage in vehicular crowd-sensing: a decentralizated incentivization solution", *IEEE Open Journal of Intell. Transp. Syst.,* vol. 3, pp. 695-708, 2022.

[23] R. Laufer and M. Wolfe, "Privacy as a Concept and a Social Issue: A Multidimensional Developmental Theory," *J. Soc. Issues*, vol. 33, pp. 22–42, 1977.

[24] Y. Wang, X. Jia, Q. Jin, and J. Ma, "QuaCentive: a quality-aware incentive mechanism in mobile crowdsourced sensing (MCS)," *J. Supercomput.*, vol. 72, no. 8, pp. 2924–2941, 2016.

[25] T. Dinev and P. Hart, "An Extended Privacy Calculus Model for E-Commerce Transactions," *Inf. Syst. Res.*, vol. 17, no. 1, pp. 61–80, 2006.

[26] H. J. Smith, T. Dinev, and H. Xu, "Information privacy research: An interdisciplinary review," *MIS Q. Manag. Inf. Syst.*, vol. 35, no. 4, pp. 989–1015, 2011.

[27] E. Princi and N. C. Krämer, "Acceptance of smart electronic monitoring atwork as a result of a privacy calculus decision," *Informatics*, vol. 6, no. 3, 2019.

[28] P. Kumaraguru and L. F. Cranor, "Privacy Indexes: A Survey of Westin's Studies," Carnegie Mellon University, 2005.

[29] D. Hensher, "Stated preference analysis of travel choices: the state of practice," *Inst. of Transp. Studies,* vol. 21, pp. 107-133, 1994.

[30] C.G. Chorus and M. Bierlaire, "An empirical comparison of travel choice models that capture preferences for comprise alternatives," *Transportation,* vol. 40, pp. 549-562, 2013.

[31] E. Molin and C.G. Chorus, "The need for advanced public transport information services when making transfers," *EJTIR,* vol. 9, no. 4, pp. 397-410, 2009.

[32] E. Molin, J. Blangé, O. Cats and C.G. Chorus, "Willingness to pay for safety improvements in passenger air travel," *J. of Air Trans. Management,* vol. 62, pp. 165-175, 2017.

[33] I. Lopez-Carreiro, A. Monzon, M.E. Lopez-Lambas, "Comparison of the willingness to adopt MaaS in Madrid (Spain) and Randstad (The Netherlands) metropolitan areas," *Transp. Res. Part A: Policy and Practice,* vol. 152, pp. 275-294, 2021.

[34] K. Train, *Discrete choice methods with simulation*. Cambridge University Press, 2003.

[35] D. McFadden, "Economic choices," *Am. Econ. Rev.*, vol. 91, no. 3, pp. 351–378, 2001.

[36] A. F. Westin, "Privacy On & Off the Internet: What Consumers Want," 2001.

[37] A. Pearce *et al.*, *Respondent Understanding in Discrete Choice Experiments: A Scoping Review*, vol. 14, no. 1. Springer International Publishing, 2021.

[38] S. Bhatnagar and R. Kumra, "Understanding consumer motivation to share IoT products data," *J. Indian Bus. Res.*, vol. 12, no. 1, pp. 5–22, 2020.

[39] M. Turland and P. Slade, "Farmers' willingness to participate in a big data platform," *Agribusiness*, vol. 36, no. 1, pp. 20–36, 2020.

[40] E. M. Schomakers, C. Lidynia, and M. Ziefle, "All of me? Users' preferences for privacy-preserving data markets and the importance of anonymity," *Electron. Mark.*, vol. 30, no. 3, pp. 649–665, 2020.

[41] X. Wang, M.C. Lucic, H. Ghazzai, Y. Massoud, "Empowering Real-Time Traffic Reporting Systems With NLP-Processed Social Media Data", *IEEE Open Journal of Intell. Transp. Syst.,* vol. 1, pp. 159-175, 2020.

[42] M. Masoud, Y. Jaradat, A. Manasrah, and I. Jannoud, "Sensors of Smart Devices in the Internet of Everything (IoE) Era: Big Opportunities and Massive Doubts," *J. Sensors*, 2019.

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

[44] L. C. Klopfenstein, S. Delpriori, A. Aldini, and A. Bogliolo, "'Worth one minute': An anonymous rewarding platform for crowd-sensing systems," *J. Commun. Networks*, vol. 21, no. 5, pp. 509–520, 2019.

[45] P. Lorenzo, J. Padilla, and A. Requejo, "Consumer Preferences For

Personal Data Protection in Social Networks: A Choice Modelling Exercise," 2020.

[46] J. V. Johansson *et al.*, "Preferences of the public for sharing health data: Discrete choice experiment," *JMIR Med. Informatics*, vol. 9, no. 7, 2021.

[47] E. M. Schomakers, H. Biermann, and M. Ziefle, "Users' Preferences for Smart Home Automation – Investigating Aspects of Privacy and Trust," *Telemat. Informatics*, vol. 64, no. March, p. 101689, 2021.

[48] M. Aitken, G. McAteer, S. Davidson, C. Frostick, and S. Cunningham-Burley, "Public Preferences Regarding Data Linkage for Health Research: A Discrete Choice Experiment," *Int. J. Popul. Data Sci.*, vol. 3, no. 11, pp. 1–13, 2018

[49] N. Khoi, S. Casteleyn, M. Mehdi Moradi, and E. Pebesma, "Do monetary incentives influence users' behavior in participatory sensing?", *Sensors,* vol. 18, no. 5, pp. 1-29, 2018.

[50] P. Mariel *et al., "*Environmental evaluation with discrete choice experiments guidance on design, implementation and data analyis", 2021.

[51] A. Woodruff, V. Pihur, S. Consolvo, L. Brandimarte, and A. Acquisti, "Would a privacy fundamentalist sell their DNA for $1000... if nothing bad happened as a result? The Westin categories, behavioral intentions, and consequences," *SOUPS '14 Proc. Tenth Symp. Usable Priv. Secur.*, pp. 1–18, 2014.

[52] S. Derikx, M. de Reuver, and M. Kroesen, "Can privacy concerns for insurance of connected cars be compensated?," *Electron. Mark.*, vol. 26, no. 1, pp. 73–81, 2016.

[53] N. Gerber, P. Gerber, and M. Volkamer. "Explaining the privacy paradox: A systematic review of literature investigating privacy attitude and behavior," *Computers & Security*, 77, 2018.

[54] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Q.*, vol. 13, pp. 123–130, 1989.

[55] L. De Wilde, C. Macharis, and I. Keseru, "Technical requirements for organising campaigns in citizen observatories," *Transp. Res. Procedia*, vol. 48, no. 2019, pp. 1418–1429, 2020.

[56] K. Shilton and D. Estrin, "Ethical Issues in Participatory Sensing," *CORE Issues Prof. Res. Ethics*, vol. 1, no. 5, pp. 2160–8784, 2012.

Ria Johanna van den Boogert received her Master's degree from Delft University of Technology in Complex Systems Engineering and Management (2022). As part of her graduation project she investigated crowdsensing for smart mobility by taking a choice modelling perspective. Currently, she is working as a Financial Trainee at the Dutch Government, holding positions at the Ministry of Foreign Affairs and the Ministry of Infrastructure and Water Management.

Aaron Yi Ding leads the Cyber-Physical Intelligence (CPI) Lab as a Tenured Associate Professor at TU Delft. His research focuses on Edge AI solutions for cyber-physical systems in smart health, mobility and energy domains. He is an associate editor for ACM Transactions on Internet of Things (TIOT) and IEEE OJ-ITS. He received PhD (2015) from University of Helsinki. With over 16 years of R&D experience across EU, UK and USA, he has worked at TU Munich with Joerg Ott, at University of Cambridge with Jon Crowcroft, and at Columbia University with Henning Schulzrinne. He has 80+ peer reviewed publications, receiving best paper awards and recognition from ACM SIGCOMM, ACM EdgeSys, ACM SenSys CCIoT, IEEE INFOCOM, and the esteemed Nokia Foundation Scholarships.