MASTER THESIS

IS MORE ALWAYS BETTER? THE EFFECT OF MONETARY INCENTIVES ON PRIVACY CONCERNS AND THEIR IMPACTS ON WILLINGNESS TO PARTICIPATE IN PASSIVE MOBILE DATA COLLECTION.

August 25, 2022

Author: Wai Tak Tung

Student ID: 1713671

University of Mannheim

Faculty of Social Sciences

Supervisors: Prof. Annelies Blom and Prof. Florian Keusch

Examiner: Prof. Florian Keusch

Author Personal Info: 017636515654. waitak.tung@uni-mannheim.de

Rm. 214, Haus 3, Ulmenweg 55,

Friedrich Traumann Straße, Mannheim, 68167, Germany.

ABSTRACT

Passive mobile data collection (PMDC) via a smartphone app provides unprecedented opportunities and data for researchers to study interactions between various kinds of social behaviour. Given its advantages, researchers would like to increase the participation rate in PMDC studies. Although offering monetary incentive to participants is a common practice to entice participation in conventional self-reporting surveys, the effect of this practice is theoretically and empirically under-explored in the context of PMDC. This research answers to the theoretic and empirical research gap by first developing a theoretical framework explaining the causal mechanism between monetary incentive and willingness to participate in PMDC based on Nissenbaum (2019) Theory of Contextual Integrity. The theory argues the effect of monetary incentive on willingness to participate in PMDC is mediated by participants' level of privacy concern about PMDC. Subsequently, Keusch (2019) incentive experimental data collected from a German non-probability based panel is leveraged to empirically test the validity of theory. 1214 respondents above 18 who owned smartphones are surveyed with vignettes describing a hypothetical PMDC sponsored by researchers from University of Mannheim with different incentive schemes. Participants have to rate their willingness to participate in PMDC after reading the vignette varied by different monetary incentive amount. The results show that 1) Willingness to participate is not influenced by the amount of maximum incentive. 2) Privacy concern about PMDC is not influenced by the amount of maximum incentive. 3) Participants' privacy concern about PMDC is not a mediator between maximum incentive and willingness to participate. 4) Willingness to participate is influenced by participants' level of privacy concern for PMDC independently.

Keywords: Passive Mobile Data Collection. Monetary Incentive. Privacy Concern. Data Security Concern. Contextual Integrity.

Word Count: 9477

1 INTRODUCTION

In recent years, researchers are increasingly interested in passively collecting survey respondents' personal data on smartphone via an app. With an app that runs in the background of participants' smartphones, personal data can be *passively* collected without any active action from participants once they have given their consent to researchers (Wenz et al, 2019). This increasing interest towards *passive mobile data collection* (PMDC) can be explained by some of its advantages compared to conventional self-reporting surveys. PMDC reduces respondents' burden as the data are collected from an app that runs in the background of participants' smartphones (Keusch et al, 2019; Keusch et al, 2020; Struminskaya et al, 2020); allows researchers to collect data that cannot be collected from conventional self-reporting surveys, such as continuous GPS locations, device usage and chat-logs (Struminskaya et al, 2020; Wenz et al 2019) and reduces bias caused by self-reporting (Keusch et al, 2020; Struminskaya et al, 2020).

Given all these advantages of PMDC, it is natural to think researchers want to increase participation rate in PMDC. One conventional method to entice participation in self -reporting survey is to provide monetary incentive to participants (Singer & Ye, 2013). Generally speaking, monetary incentive comes in three modes in PMDC studies (Haas et al, 2020; Jäckle et al, 2019; Keusch et al, 2019; Singer & Ye, 2013). Participants can be paid immediately for installing the app - *prepaid incentive*, or paid conditionally for retaining on the PMDC app - *promised incentive*. Some studies also mention the sum of both the prepaid and promised incentives participants can obtain - *maximum incentive*, in a vignette that describes a hypothetical scenario of PMDC (Keusch, 2019; Keusch et al, 2019); or a flyer that describes the

incentive scheme in an actual PMDC study (Haas et al, 2020). In PMDC incentive studies, participants earn points for their participation in PMDC and once their points have reached a certain threshold, they can redeem these points into digital gift vouchers (Haas et al, 2020; Jäckle et al, 2019, Keusch, 2019).

However, the effect of monetary incentive on enticing willingness to participate in PMDC is theoretically unclear with mixed empirical results. Theoretically, many PMDC incentive studies do not provide a rigorous theoretic account explaining the casual mechanism between monetary incentive and willingness to participate in PMDC (Haas et al, 2020; Jäckle et al, 2019; Keusch et al, 2019; Silber et al, 2021). Without a clear theoretic framework in the literature, this somewhat explains the low external validity of previous PMDC incentive studies. Empirically, positive significant effect (Keusch et al, 2019; Silber et al, 2021); non-significant effect (Jäckle et al, 2019); and an inverted U-shaped effect (Haas et al, 2020) of different modes of monetary incentive on participants' hypothetical and actual participation in PMDC are observed. It is therefore elusive whether monetary incentive can entice participation in PMDC both theoretically and empirically.

These studies prompts the following research question. What is the casual mechanism between monetary incentive and willingness to participate in PMDC? This paper argues monetary incentive influences how participants judge the appropriateness of the data transmission of PMDC that influences their willingness to participate in PMDC. In particular, how monetary incentive influences participants' level of privacy concern about PMDC is crucial in understanding their willingness to participate in PMDC. This research comes in five major parts. First, it provides a brief literature review on the existing literature explaining

the causal mechanism between monetary incentive and willingness to participate in PMDC. Second, it develops a theory explaining the casual mechanism between monetary incentive and willingness to participate in PMDC. Three hypotheses are proposed under the theoretic framework. Third, it outlines the data, empirical setup, operationalization and analysis plan for hypothesis testing in this study. Fourth, the empirical analysis results are presented. Finally, it discusses the results as well the limitations of this study.

2 LITERATURE REVIEW

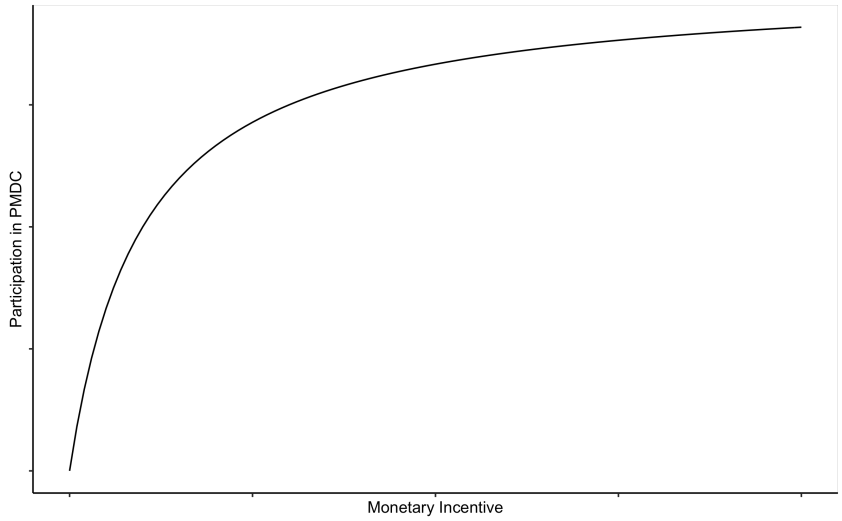

Figure 1: Haas et al (2020) diminishing marginal effect of monetary incentive on participation in PMDC.

While a handful of studies investigate the effect of monetary incentives on hypothetical and actual participation in PMDC conducted by university researchers, most of them do not provide an extensive elaboration on the casual mechanism between monetary incentive and participation in PMDC (Jäckle et al, 2019; Keusch et al, 2019; Silber et al, 2021). To the best of my knowledge, only Haas et al (2020) provide a brief theoretic account.

Haas et al (2020) expect a diminishing marginal effect of monetary incentive on participation to participate in PMDC. Drawing insights from existing PMDC studies, they argue people in general are concerned about their privacy when participating in PMDC (Haas et al, 2020, pp.390-392). Assuming privacy concern about PMDC is correlated to trust towards the data collector, their theory implies participants have little trust towards researchers who conduct PMDC. Subsequently, monetary incentive is used by researchers as a token to earn trust from participants (Haas et al, 2020, p. 392). Once the critical threshold of trust is reached between the researchers and participants, the effect of monetary incentive diminishes as the incentive amount increases (Haas et al, 2020, p.392). Figure 1 visualises the diminishing marginal utility in Haas et al (2020) theory.

In the empirical analysis, Haas et al (2020) find partial support for their theory. On one hand, they do find a critical threshold of trust might exist between €60 and €70 maximum incentive as participants who are offered €70 or €90 statistically significantly stay on average 10 days longer on the PMDC app than those offered €60 (Haas et al, p.402). Yet, instead of a diminishing marginal effect of monetary incentive on actual participation in PMDC, Haas et al (2020) find an inverted U-shaped relationship between maximum incentive and days of retention on the PMDC app. The days of retention on the PMDC app is not statistically significantly different between those who are offered €60 and the €100 maximum incentives (Haas et al, 2020).

Two shortcomings in Haas et al (2020) study are worth mentioning. On a theoretical standpoint, Haas et al (2020) do not provide rigorous elaborations nor empirical evidences showing why privacy concern and trust towards the data collector is correlated. Without much

elaboration, Haas et al (2020) stipulate "(t)hose concerns (privacy concern for PMDC) may be tied to trust issues: individuals do not trust researchers to protect their data adequately." (Haas et al, 2020, p. 392). It is therefore unclear why two concepts are correlated theoretically. On an empirical standpoint, neither privacy concern for PMDC nor trust towards the data collector is operationalized in the analysis. Instead, the empirical analysis investigates the effects of different modes of incentive on downloading a smartphone app for PMDC, uninstalling the PMDC app, and days of retention on the app with participants having different welfare status (Haas et al, 2020). It is therefore unclear from their empirical analysis whether privacy concern for PMDC or trust towards the data collector statistically significantly mediates between maximum incentive and participation in PMDC as proposed in their theory.

Against this background, this study answers to both the theoretic and empirical research gap. Based on Nissenbaum (2019) Theory of Contextual Integrity, this paper proposes a theory arguing there exists an inverted U-shaped relationship between monetary incentive and willingness to participate in PMDC. The theory also argues that this inverted U-shaped relationship is mediated by participants privacy concern about PMDC without making the assumption that participants level of privacy concern about PMDC is correlated with their level of trust to researchers. After the theory is constructed, privacy concern about PMDC and monetary incentive are operationalized empirically to test whether monetary incentive and privacy concern about PMDC influence the appropriateness of PMDC as reflected by their willingness to participate in a PMDC using Keusch (2019) incentive experimental data.

3 THEORY

3.1 Nissenbaum (2019) Theory of Contextual Integrity

Apart from being a token to earn participants' trust proposed by Haas et al (2020), studies from survey methodology (Groves et al, 2000); and e-commerce (Li et al, 2014; Nget et al, 2017; Staiano et al, 2014), suggest a variety of normative arguments that could explain the effect of monetary incentive on enticing participation in PMDC. Some suggest personal data are monetized commodities and data collectors need to offer money to purchase such commodities from the data owners (Li et al, 2014; Nget et al, 2017; Staiano et al, 2014). Whereas survey methodology study suggests monetary incentive can be considered a benefit that outweighs the negative consequences of participating in a survey such as loss of privacy or lack of interest (Groves et al, 2000, p.302).

In fact, these normative arguments point to the notion that monetary incentive in fact influences how participants perceive the appropriateness of the data transmission in PMDC. This is clearly visible once we put this notion into context. For example, once trust is built between researchers and participants, participants consider the data transmission in PMDC appropriate and become willing to share their data with researchers; or if researchers offer a price below the expectation of participants in exchange of their personal data, or insufficiently compensate their lack of interest, they do not consider the data transmission in PMDC appropriate and become unwilling to participate in PMDC.

To assess the appropriateness of the data transmission in PMDC, Nissenbaum (2019) Theory Contextual Integrity (TCI) becomes relevant. This is because TCI is precisely designed

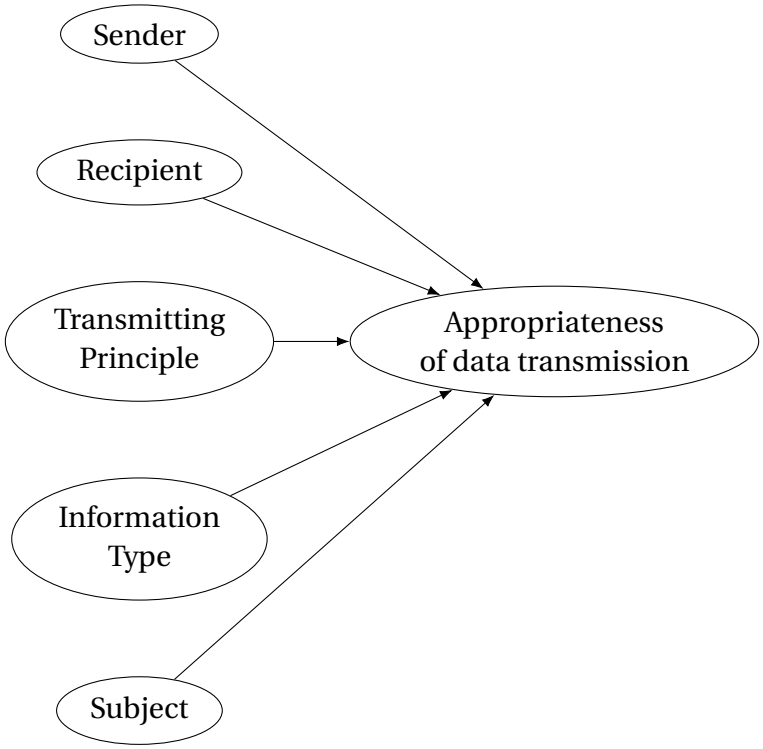

Figure 2: Nissenbaum (2019) stipulates five contextual parameters determining the appropriateness of a data transmission.

to assess the appropriateness of data transmission based on five contextual parameters: 1) the sender of the data 2) the recipient of the data 3) transmitting principle 4) the information type and 5) the subject of the data (Nissenbaum, 2019, p.228). Figure 2 visualises Nissenbaum (2019) theoretic framework. If the data transmission defined by the five contextual parameters conforms to the privacy norm, this data transmission is then considered appropriate and respects privacy (Nissenbaum, 2019).

Nissenbaum (2019) framework can be easily understood once it is put into context. A simple example is subject consent in academic researches (Nissenbaum, 2019). Informed consent is an ethical requirement in modern academic researches (Couper & Singer, 2013; Nissenbaum, 2019). Therefore, it will be considered inappropriate and unethical when

academic researchers (*recipient)* collect information about people's (*sender & subject)* attitude or opinion (*information type)* without their stated consent (*transmitting principle).*

3.2 Putting Nissenbaum (2019) TCI into the Context of PMDC

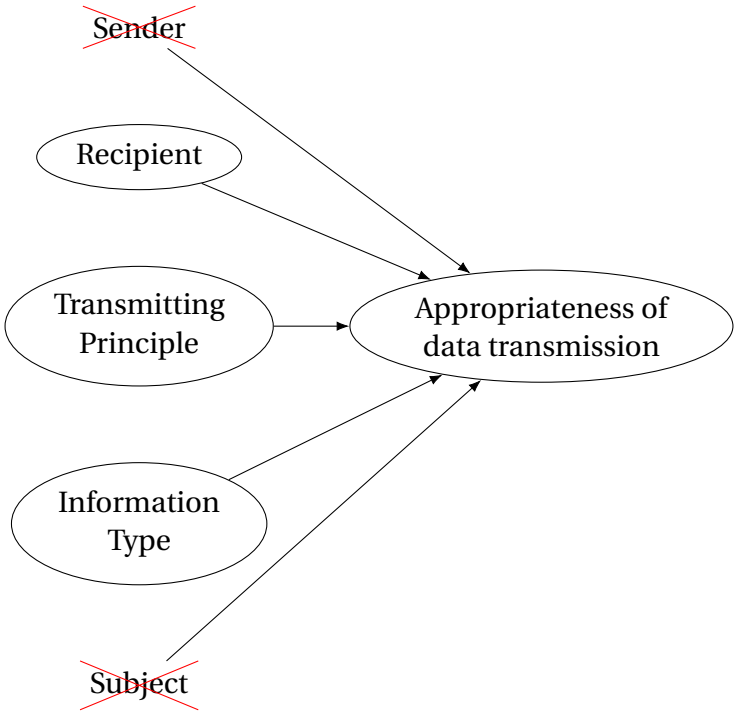

Figure 3: Nissenbaum (2019) Contextual Integrity framework from the participants' perspective when determining whether the the data transmission in PMDC is appropriate.

In order to put TCI into the context of PMDC, some modifications and clarifications are needed. First, changing any contextual parameter can already change how people perceive the appropriateness of a data transmission, holding other contextual parameters constant. This argument can be understood easily when we look at the previous subject consent example. When subject consent is involved and holding other contextual parameters constant, the privacy norm in academic research induces us to consider such data transmission to be appropriate and ethical. Alternatively, it will be considered inappropriate and unethical for

researchers to collect the same type of data, from the same group of participants but without subject consent.

Second, participants are only considering three contextual parameters when determining the appropriateness of the data transmission in PMDC. This is because PMDC participants are both the *sender* and *subject* of the data in this context. Therefore, participants are only judging the appropriateness of PMDC by who is receiving the data (*recipient*), under what circumstances and conditions the data are transmitted (*transmitting principle*), and what types of data are being transmitted (*information type*).

3.2.1 Empirical Evidences

Previous PMDC studies support the notion that we can apply the modified TCI framework into understanding the mechanism of willingness to participate in PMDC. The contextual factors in the modified framework do influence how participants assess the appropriateness of PMDC. The appropriateness of PMDC can be reflected by respondents' stated willingness to participate in participation in PMDC. First, a statistically significant sponsor effect on willingness to participate in PMDC is observed using both probability-based panel data (Struminskaya et al, 2020); and non-probability-based panel data (Keusch et al, 2019). Participants are significantly more likely to be willing to participate in PMDC sponsored by university than a PMDC sponsored by marketing agency or statistic agency (Keusch et al, 2019; Struminskaya et al, 2020). Another example is that willingness to participate in PMDC differs by the information type being collected by researchers. Struminskaya et al (2020) show that participants are significantly less likely to be willing to share their geo-locations compared to

data collected from wearing a fitness bracelet to university researchers.

These empirical evidences are informative because they show participants do judge the appropriateness of PMDC differently upon different contextual parameters (change of recipient and information type in the above examples) as predicted by the modified TCI framework. These studies also justify the choice of applying TCI framework into the context of this study. As the next section unfolds, it will explain monetary incentive is in fact one of the transmitting principles that influence participants' perceptions on the appropriateness of PMDC and subsequently their willingness to participate in PMDC.

3.3 Monetary Incentive, Transmitting Principle & Willingness to Participate in PMDC

How can we explain the causal mechanism between monetary incentive and willingness to participate in PMDC under the modified TCI framework? Theoretically, monetary incentive can be understood as one of the transmitting principles that stipulates the condition and circumstance the data is transmitted in PMDC. As the theory predicts, once we change the amount of monetary incentive offered to participants, we are in fact changing one of the transmitting principles that defines the appropriateness of PMDC. From a theoretical standpoint, we should therefore anticipate participants to judge the appropriateness of PMDC differently when they are offered a different amount of monetary incentive. Such difference of appropriateness is translated into different kinds of data sharing behaviour in PMDC. Hence, we can reasonably expect when researchers offer different amount of monetary incentives to participants in PMDC, they will exhibit different kinds of data sharing

behaviour.

3.3.1 Empirical Evidences

Previous PMDC incentive studies show that when participants are offered a different amount of incentive, they do exhibit different kinds of data sharing behaviour (Haas et al, 2020; Keusch et al, 2019; Silber, 2021). Generally speaking, previous PMDC incentive studies suggest a positive relationship between a small sum of monetary incentive and willingness to participate in PMDC (Haas et al, 2020; Keusch et al, 2019; Silber et al, 2021). Sampling from a German non-probability panel, Keusch et al (2019, p.224) show that the predicted probabilities of willingness to participate in PMDC increases by 20 percentage point for participants who are offered €10 prepaid incentive, 19 percentage point for those offered €10 promised incentive and 26 percentage point for those offered €20, over participants who are offered no incentive. Haas et al (2020) find similar results from a German annual household panel showing that those offered €20 prepaid incentive have a significantly 3 percentage point higher installation rate of the PMDC app than those offered €10 prepaid incentive holding other factors constant.

These findings suggest that when offered a small sum of monetary incentive, participants tend to view the data transmission in PMDC more appropriate and more willing to participate in PMDC than their counterpart who receive less or no monetary incentive. If participants consider the data transmission with a small sum of incentive offered appropriate, we can also reasonably assume that they are not concerned about data security when engaging in PMDC or their privacy concern about PMDC is adequately compensated by the monetary incentive.

This argument is also in line with the existing literature explaining why monetary incentive might entice participation in PMDC (Groves et al, 2000; Haas et al, 2020; Li et al, 2014; Nget et al, 2017; Staiano et al, 2014).

3.4 Is More Always Better? Not Really.

The previous section shows that monetary incentive and willingness to participate in PMDC is correlated theoretically and empirically. While the majority of the existing literature suggests those who are offered a small sum of monetary incentive are more likely to participate in PMDC than their counterparts with no or less monetary incentive, it is intuitive to think researchers should offer as much monetary incentive as they can to entice a higher participation rate in PMDC. However, a large sum of monetary incentive can have a negative effect on willingness to participate in PMDC. This is because a large sum of monetary incentive might make participants' privacy concern about PMDC salient. Once participants privacy concern about PMDC is made salient by incentive, participants might consider the data transmission in PMDC inappropriate and become unwilling to participate in PMDC.

3.4.1 Can Participants' Privacy Concern about PMDC be Primed?

First, participants privacy concern about PMDC can be primed by a sensory input. Priming occurs when a sensory input - *the prime*, increases the salience and accessibility of a concept -*the target*, that ultimately affects a person's other cognitive processes (Cassino & Erisen, 2010, p.375). In particular, semantic priming refers to priming when the prime and the target belongs to the same semantic category (eg. iPhone & Andriod Phone) (Cassino & Erisen, 2010). Previous survey methodology studies have shown that participants' privacy concern can be

primed by a sensory input that affects their willingness to share their personal information in online self-reporting surveys (Couper & Singer, 2013; Joinson et al, 2008; Marrerios et al, 2017). Marrerios et al (2017) show that participants who have read either positive or negative privacy-related news prior to their participation are statistically significantly less likely to disclose their given name and email in an online survey compared to their counterparts who did not read any privacy-related news. This shows that once the topic of privacy is made salient by a sensory input to participants, participants are less likely to share their personal data to researchers.

Extending this logic to PMDC incentive studies, these findings suggest that if the topic of privacy is made salient by a sensory input in PMDC, it might also influence participants' willingness to share their personal data with researchers. Existing PMDC studies already know that privacy concern for PMDC and data security concern are both negatively significantly correlated to willingness to participate in PMDC (Keusch et al, 2019; Struminskaya et al, 2020; Wenz et al, 2019). Hence, the remaining question is whether a large sum of monetary incentive makes the topic of privacy salient and reduces willingness to participate in PMDC.

3.4.2 There Is No Free Lunch In The World.

Literature from survey methodology studies seem to support the notion that a large sum of incentive can make the topic of privacy salient. This is because previous studies have shown that people positively correlate the data sensitivity of their personal data to the amount of monetary incentive or monetary compensation they anticipate from data collectors (Carrascal et al, 2013; Staiano et al, 2014). Staiano et al (2014) show that Italian participants value

the most sensitive personal data with the highest monetary value in a 6-week mobile data collection study. Carrascal et al (2013) find similar results which Spanish respondents value data about their age and address (€25) approximately 3.6 times more than their online browsing history (€7). Because participants associate a higher monetary value to personal data that are more sensitive, we can reasonably suspect that as the amount of monetary incentive or compensation offered by data collectors increases, participants will be aware that their personal data being collected are increasingly sensitive too. This means when a large sum of monetary incentive is offered in exchange of participants' personal data, participants will be aware that the data they are providing are sensitive information. In other words, a large sum of monetary incentive can make the topic of data sensitivity salient. When the topic of data sensitivity is made salient by a large sum of monetary incentive, it can cause semantic priming that triggers participants to think about other semantically related topics such as privacy concern about the PMDC and whether they will incur a large degree of privacy loss if they decide to participate in this PMDC.

Putting these findings into the context of PMDC, it is therefore possible that participants privacy concern about PMDC can be made salient through priming when offered a large sum of monetary incentive. Once participants have a high privacy concern about PMDC, they become unwilling to participate in PMDC. This might explain why an inverted U-shaped relationship between maximum incentive and days of retention on the PMDC app is observed in Haas et al (2020) study because the critical threshold that makes privacy concern salient exist between €60 and €70.

3.5 Theoretic Framework

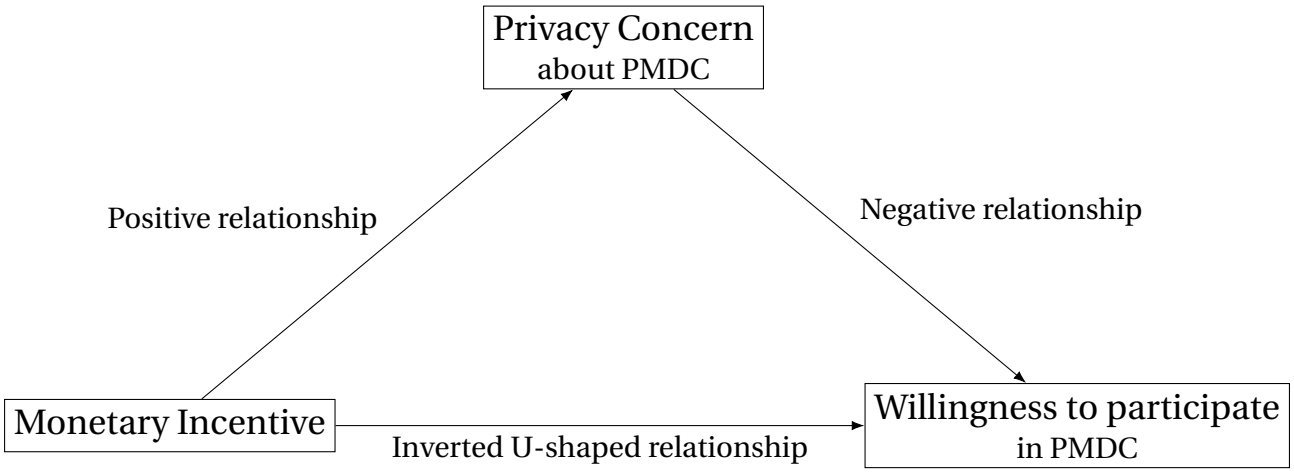

Figure 4: Theoretic framework explaining the mechanism between monetary incentive and participation in PMDC.

All together, the literature induces the following theory explaining the mechanism on how monetary incentive influences willingness to participate in PMDC mediated by their privacy concern about PMDC. When a small amount of monetary incentive is offered to participants, participants increasingly consider the data transmission in PMDC appropriate and become willing to participate in PMDC. However, when a large sum of monetary incentive is offered to participants, it makes participants' privacy concern about PMDC salient. As privacy concern about PMDC increases as the incentive amount increases, participants become increasingly unwilling to participate in PMDC. This means in between a small sum of incentive and a large sum of incentive, there exists a critical threshold of monetary incentive that will make participants privacy concern about PMDC salient. This also means that there will be an inverted U-shaped relationship between monetary incentive and willingness to participate in PMDC. Figure 4 visualises the underlying casual mechanism the theory this study. Therefore, the theory induces the following three hypotheses:

H1: There exists an inverted U-shaped relationship between monetary incentive and willingness to participate in PMDC.

H.2: Participants who are offered more monetary incentive will be more likely to have a high level of privacy concern about PMDC than participants who are offered less monetary incentive.

H.3: The relationship between monetary incentive and willingness to participate in PMDC is mediated by participants' level of privacy concern about PMDC.

4 METHODOLOGY

4.1 Data

Keusch (2019) incentive experimental data on passive mobile data collection is leveraged to test the three hypotheses. In Keusch (2019) experiment, 1398 German smartphone users who are 18 or above are recruited through Survey Sampling International – a German non-probability online panel in December 2017. 87 per cent (N=1214) respondents participated in the survey. Excluding incomplete cases, 1204 respondents are analysed in the main analysis. The median time required to complete the survey is 9 minutes and 15 seconds (Keusch, 2019).

4.2 Empirical Setup

In Keusch (2019) experiment, participants are first prompted with a vignette describing a hypothetical scenario where researchers from University of Mannheim would like to collect their smartphone data via a research app[1]. The vignettes also include information about the

[1] In Keusch (2019) original experimental setup, the vignettes varies by three characteristics: 1) prepaid incentive 2) promised incentive and 3) learning about one's smartphone behaviour, with two dimension levels

duration of the hypothetical PMDC study, the amount of monetary incentive participants will receive, data protection regulation and the types of data to be collected passively.

After reading the vignette, participants have to rate their willingness to participate in this PMDC study on a 11-point bipolar scale (0: Would definitely not participate, 10 would definitely participate). After answering their willingness to participate in PMDC, participants are prompted with a number of questions on different survey topics such as general privacy concern, data security concerns in different situations, survey attitudes, and their socio-demographics (Keusch, 2019).

Leveraging on Keusch (2019) experimental data, this study designs a 2x2 empirical setup on prepaid and promised incentives. The empirical setup is visualised in figure 5. For prepaid incentive, participants are randomly given either €10 or €20; for promised incentive, participants are randomly given either €50 or €80. Combining both the prepaid and promised incentives, participants receive one of the four possible amounts of maximum monetary incentive: €60, €70, €90 or €100. Essentially, participants are randomly assigned to one of the four maximum incentive treatment groups.

Two key features of the empirical design in this study are worth mentioning. First, questions on participants' privacy concern about PMDC have to be asked after the vignette describing the amount of monetary incentive participants receive. Such question order is important because it is only possible to observe whether participants' privacy concern about PMDC is primed by the incentive amount they are offered. Serendipitously, the question

for each characteristic. This makes the original experimental setup a 2x2x2 setup with 8 vignettes. However, since the theory in this study does not explain nor predict the effect of non-monetary incentive on enticing participation in PMDC, the third dimension on self-learning is removed here.

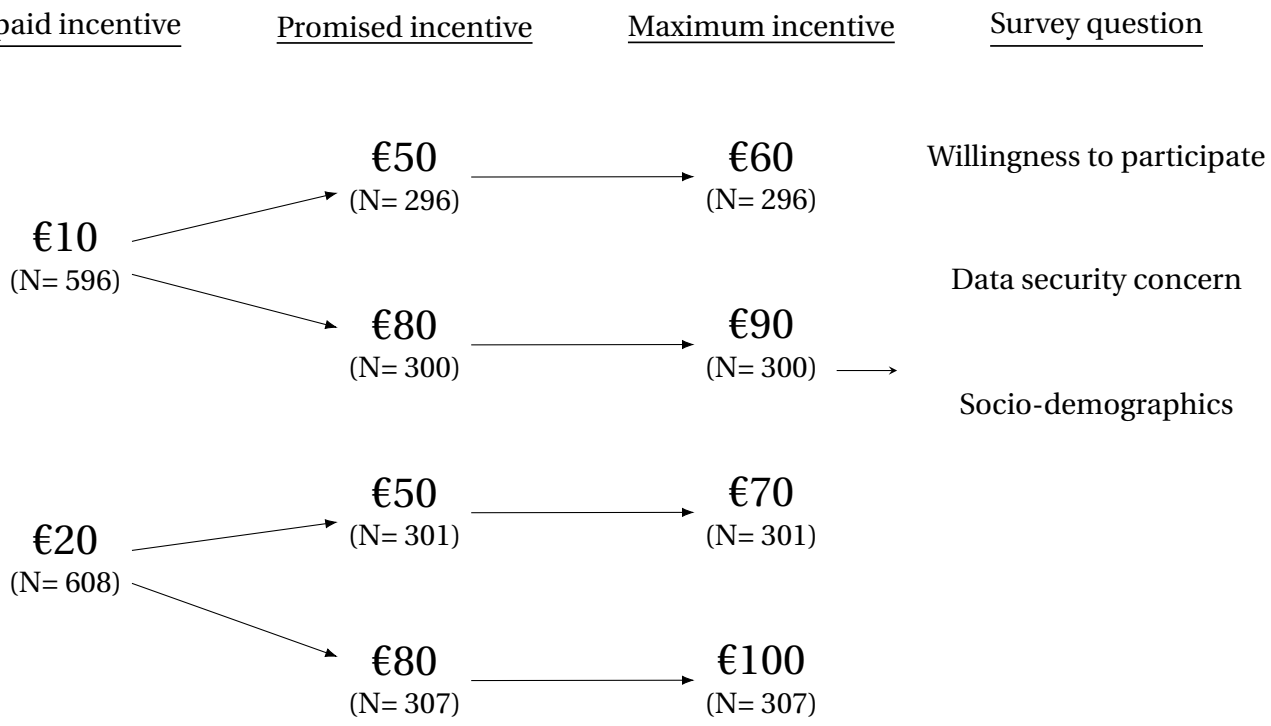

Figure 5: Empirical design based on Keusch (2019) incentive experiment study (N=1204).

operationalizing the theoretic construct of "privacy concern about PMDC" is asked after the item asking participants willingness to participate in the study in Keusch (2019) study. This allows us to leverage on the data to test for H.2 & H.3.

Second, all contextual parameters except monetary incentive are held constant in Keusch (2019) vignettes. Mapping Keusch (2019) incentive experiment data to the modified Nissenbaum (2019) theoretic framework, researchers from University of Mannheim are the *recipient* of the data. The amount of monetary incentive, and the data protection conditions are the *transmitting principles* that define the conditions and circumstances the PMDC that took place. In the vignettes, the *information types* to be collected are participants' network quality, geo-locations, interaction history, social network characteristics, activity data and smartphone usages. Participants in Keusch (2019) study are therefore judging whether this

PMDC defined by these contextual parameters is appropriate for research purposes - *privacy norm*. All these contextual parameters are held constant except the amount of monetary incentive in the vignettes. This empirical setup therefore allows us to claim that if different kinds of data sharing behaviour between the four maximum incentive treatment groups are observed, such difference is induced by the amount of monetary incentive participants received but not other contextual parameters.

4.3 Operationalisation

4.3.1 Willingness to Participate in PMDC

Willingness to participate in PMDC is operationalized with question (Q6) asking participants' stated willingness to participate in PMDC. In the original questionnaire, participants' willingness to participate in PMDC is measured on a 11-point bipolar scale (0: Would definitely not participate, 10 would definitely participate). This variable is re-coded in two ways. First, participants' willingness to participate in PMDC is mapped to the maximum incentive they receive - *WTP_GRP* (1: €60, 2: €70, 3: €90 and 4: €100). This variable allows further coding to observe the distribution of participants' privacy concern about PMDC across the four maximum incentive groups for answering H.2. This variable also allows us to observe whether participants are equally distributed across the four maximum incentive groups as shown in the empirical setup. Second, participants are dichotomized into a binary variable - *WTP_dummy*. Those who score from 0 to 5 are categorised into *unwilling to participate* and coded 0; and those who scored from 6 to 10 are categorised into *willing to participate* and coded 1.

There are three reasons why *WTP_dummy* is dichotomized in such fashion. First, dichotomizing willingness to participate in PMDC is a common practice in PMDC studies (Keusch et al, 2019; Struminskaya et al, 2020). Second, the dichotomization of responses allows us to run logistic regressions and estimates the predicted probabilities of willingness to participate in PMDC. Third, Keusch et al (2019) show that participants' responses to willingness to participate in PMDC cluster at the extreme points on the scale (0 & 10 in this 11-point bipolar scale). Using a different cutoff point to dichotomize the data or restricting responses to extreme points on the scale (0 & 10) will produce stable estimates of key predictors (Keusch et al, 2019). The same procedures are repeated in this study and the results corroborate with Keusch et al (2019) arguments[2].

4.3.2 Privacy concern about PMDC

Participants privacy concern about PMDC is operationalized with the question (Q21c) that asks participants how concerned they are about data security when downloading an app which collects data about how they use their smartphones for research purposes on a 4-point uni-polar scale (1:"Not concerned at all" to 4: "Very concerned"). This variable is operationalized in two ways. First, the original 4-point scaled variable - *security concern*, is kept for creating jitter plots showing the distribution of participants security concern across the four incentive treatment groups.

Another operationalization of this variable is a dichotomized variable – *security_dummy*.

[2]For distribution of willingness to participate on the 11-point scale measurement in aggregate and across 4 maximum incentive groups, see appendix figure 1 to 5; for the logistic coefficient estimates with different cutoff (0-4: unwilling to participate & 5-10: willing to participate and just extreme value (0& 10) for testing H1 and H.3, see appendix table 9.

For this dichotomized variable, "Not concerned at all" and "A little concerned" are categorised into *low concern* ; and "Somewhat concerned" and "A lot concerned" are categorised into *high concern. Low concern* is coded 0, whereas *high concern* is coded 1. This binary variable serves two functions. First, it allow us to estimate the predicted probabilities for having high security concern in H.2 and whether security concern is a mediator between monetary incentive and willingness to participate in PMDC in H.3. Second, Struminskaya et al (2020) has shown that using a dichotomized predictor and a original 4-point scale predictor in PMDC study produce stable estimates of key predictors. This procedure is repeated in this study and the results corroborate with Struminskaya et al (2020) finding[3]. For consistency sake, the dichotomized security concern variable - *security_dummy* is used in the main analysis.

There are two reasons why this measurement is chosen to operationalize "privacy concern about PMDC". First, there is no question asking participants' privacy concern about PMDC specifically in Keusch (2019) questionnaire. An attitude will be a better predictor of an action when it is measured at the same level of specificity of the behaviour (Ajzen, 2005). Hence, although there is a question about participants' general privacy concern (Q19) in Keusch (2019) questionnaire, it is not the appropriate measurement to be operationalized because it does not match the specific data sharing action this study is interested in - willingness to participate in PMDC sponsored by university researchers. Among all the data security concern questions, (Q21c) has the closest proximity to the theoretic construct of "privacy concern about PMDC", which makes (Q21c) the most optimal measurement available in Keusch (2019) dataset for operationalizing "privacy concern about PMDC". From here onward, pri-

[3]For logistic coefficients with a 4-point scale security concern predicting the likelihood of willingness to participate in PMDC for H.3, see Appendix table 10.

vacy concern about PMDC and security concern about PMDC are used symonymously unless further specified.

4.3.3 Monetary Incentive

Since there are 4 maximum incentive treatment groups , three monetary incentive dummy variables (being offered €70, €90 and €100) are created. For simplicity, they are labelled as *70_dummy*, *90_dummy* and *100_dummy*. The reference group for the three incentive dummy variables is the €60 maximum incentive group. The main reason to use maximum incentive to operationalize "monetary incentive" is because it is the only incentive mode with more than 2 dimensional levels available from Keusch (2019) dataset. This feature is essential to observe a non-linear relationship between monetary incentive and willingness to participate in PMDC as predicted in H.1. Indeed, Haas et al (2020) observed an inverted U-shaped effect of monetary incentive on PMDC within the same range of incentive. This suggests that the critical threshold that make privacy concern salient within this range of maximum incentive might also lie within the same range in this study.

4.3.4 Control Variables

Participants age, gender and education level are controlled for H.3. For age, participants are first categorized into three age groups, 18-29, 30-49 and 50 or above, where the 18-29 group is the reference group. For gender, a female dummy variable is created using the male gender as the reference group. Finally, participants' education level is dichotomized into whether they have a high school degree using having a high school degree as the reference group.

4.4 Analysis Plan

4.4.1 First Hypothesis

The first hypothesis would like to understand the effect of monetary incentive on willingness to participate in PMDC. To answer this question, the log of the odds of participants' willingness to participate, $\pi_{WTP\_dummy} = Pr(WTP\_dummy = 1)$, is regressed on the three maximum incentive dummies. The betas (β_1 to β_3) denote the corresponding logistic coefficients to each dummy variable. ϵ indicates the residual error term.

$$log(\frac{\pi_{WTP\_dummy}}{1-\pi_{WTP\_dummy}}) = \beta_0 + \beta_1 dummy\_70 + \beta_2 dummy\_90 + \beta_3 dummy\_100 + \epsilon \quad (1)$$

4.4.2 Second Hypothesis

The second hypothesis would like to understand whether an increase of monetary incentive will increase participants privacy concern about PMDC. This hypothesis is answered in two ways. First, participants distributions of security concern (on a 4-point scale) across 4 incentive treatment groups are visualised using jitter plots. Second, leveraging on the binary variable - *security_dummy*, the log of the odds of participants security concern, $\pi_{security\_dummy} = Pr(security\_dummy = 1)$, is regressed on the three maximum incentive dummies. The betas (β_1 to β_3) denote the corresponding logistic coefficients to each dummy variable. ϵ indicates the residual error term.

$$log(\frac{\pi_{security\_dummy}}{1-\pi_{security\_dummy}}) = \beta_0 + \beta_1 dummy\_70 + \beta_2 dummy\_90 + \beta_3 dummy\_100 + \epsilon \quad (2)$$

4.4.3 Third Hypothesis

The third hypothesis would like to understand whether privacy concern about PMDC is a mediating variable between monetary incentive and willingness to participate in PMDC. This hypothesis is answered with Baron and Kenny (1986) test for mediation method. According to Baron and Kenny (1986), three conditions have to be met simultaneously to demonstrate a complete mediation effect of privacy concern about PMDC between monetary incentive and willingness to participate in PMDC. This is visualised in figure 6.

$$\text{Monetary incentive} \longrightarrow \text{Willingness to participate in PMDC} \qquad \text{(R1)}$$

$$\text{Monetary incentive} \longrightarrow \text{Security concern} \qquad \text{(R2)}$$

$$\text{Monetary incentive} \quad + \quad \text{Security concern} \longrightarrow \text{Willingness to participate in PMDC} \qquad \text{(R3)}$$

Figure 6: The three regressions models required to demonstrate a complete mediation of security concern between monetary incentive and willingness to participate in PMDC according to Baron and Kenny (1986) mediation testing method.

First, monetary incentive has to be significantly correlated to willingness to participate in PMDC in the first regression model - *R1*. Second, monetary incentive has to be significantly correlated to data security concern about PDMC in the second regression model - *R2*. Third, security concern about PMDC is significantly correlated to willingness to participate and monetary incentive is no longer significantly correlated to willingness to participate in the third regression model- *R3*. If three conditions are met, complete mediation of security concern about PMDC between incentive and willingness to participate in PMDC is present. Since the *R1* and *R2* are already tested in the analysis for H.1 and H.2, only *R3* is run for parsimony reason. Two regressions models are built to test for *R3*. In the base model, the log of the odds

of participants willingness to participate in PMDC, $\pi_{WTP\_dummy} = Pr(WTP\_dummy = 1)$, is regressed on the three maximum incentive dummies and security concern dummy. The betas (β_1 to β_4) denote the corresponding logistic coefficients to each dummy variable. ϵ indicates the residual error term.

$$log(\frac{\pi_{WTP\_dummy}}{1-\pi_{WTP\_dummy}}) = \beta_0+\beta_1 dummy\_70+\beta_2 dummy\_90+\beta_3 dummy\_100+\beta_4 security\_dummy+\epsilon \tag{3}$$

In the model with participants' socio-demographics controlled, the log of the odds of participants willingness to participate in PMDC, $\pi_{WTP\_dummy} = Pr(WTP\_dummy = 1)$, is regressed on the three maximum incentive dummies, security concern dummy and the socio-demographics control. The betas (β_1 to β_8) denote the corresponding logistic coefficients to each dummy variable. ϵ indicates the residual error term.

$$\begin{aligned} log(\frac{\pi_{WTP\_dummy}}{1-\pi_{WTP\_dummy}}) = \beta_0 + \beta_1 dummy\_70 + \beta_2 dummy\_90 + \beta_3 dummy\_100 + \beta_4 security\_dummy \\ +\beta_5 Age\_30\_49 + \beta_6 Age\_50\_or\_above + \beta_7 Female + \beta_8 Without\_HS + \epsilon \end{aligned} \tag{4}$$

All the logistic regressions are run and produced using RStudio[4] (RStudio Team, 2021). Logistic regressions are run with the library(mass) (Venables & Ripley, 2002). In order to manipulate and relabel variables, library(dplyr) is used (Wickham et al, 2022). To export well formatted regression and summary statistics tables from RStudio to the LaTeX environment, library(stargazer) is used (Hlavac, 2022). In order to visualize the distribution of data across different categories aesthetically, library(viridis) is used to create jitter plots (Garnier et al, 2021). For easy interpretation of the multiple logistic regressions' results, average marginal

[4]The R script for all the coding and analysis can be found here: https://github.com/waitaktung1/Thesis

effects of each predictor with standard error with 95 percent of confidence is visualized using library(margins) (Leeper, 2021); and library(dotwhisker) (Solt & Yu, 2015). Frequency distributions and Haas et al (2020) theory are visualised using library(ggplot2) (Wickham, 2016). The summary statistics of all the variables operated in this study can be found in appendix table 1.

5 Main Analysis

5.1 Is There an Inverted U-shaped Relationship between Incentive and Willingness to Participate in PMDC?

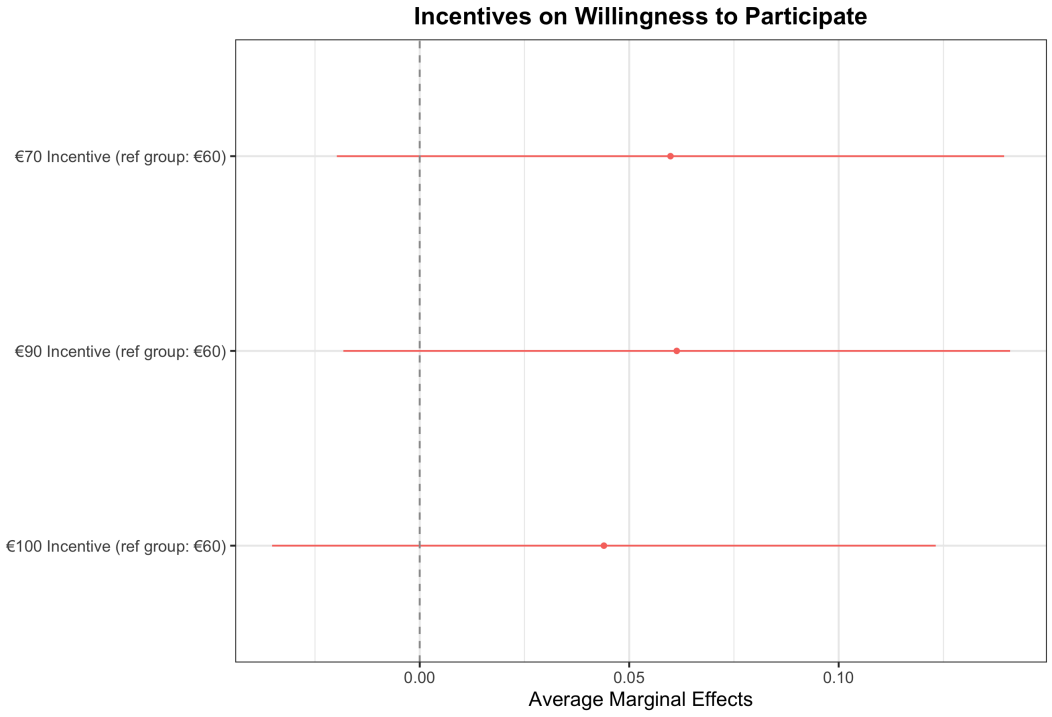

Figure 7: Average marginal effects with 95 per cent confidence interval from multiple logistic regression predicting likelihood of willingness to participate in PMDC (N=1204).

Monetary incentive is not significantly correlated to willingness to participate in PMDC

(p>0.05), showing no support for H1. Figure 7 shows the average marginal effect of maximum incentive on willingness to participate in PMDC [5]. Figure 7 shows participants' willingness to participate in PMDC with €70, €90 and €100 are not statistically significantly higher than those who receive €60 maximum incentive (p>0.05). Therefore, H1 is rejected.

5.2 Can The Amount of Monetary Incentive Prime Participants' Level of Security Concern about PMDC?

Additional monetary incentive does not influence participants' level of security concern, showing no support for H.2. Figure 8 shows the hypothetical distribution of participants' security concern across 4 maximum incentive groups if monetary incentive is positively correlated to participants' level of security concern. If large amount of incentive can make participants' security concern salient, participants level of security concern should increasingly clustered in the "Very Concerned" and "Somewhat Concerned" responses as the amount of maximum incentive increases.

However, maximum incentive does not seem to be positively correlated to participants' level of security concern as shown in figure 9. Figure 9 shows the actual distribution of participants' security concern across 4 maximum incentive groups. Instead of a positive relationship between incentive and security concern, the level of security concern across 4 treatment groups are rather consistent. This suggests that participants' level of security concern might not be influenced by the amount of incentive they are offered.

The results of logistic regression corroborate figure 8 & 9 findings. Figure 10 shows

[5] The logistic regression model and average marginal effects with standard errors figure 7 based on can be found in appendix table 2 & 3.

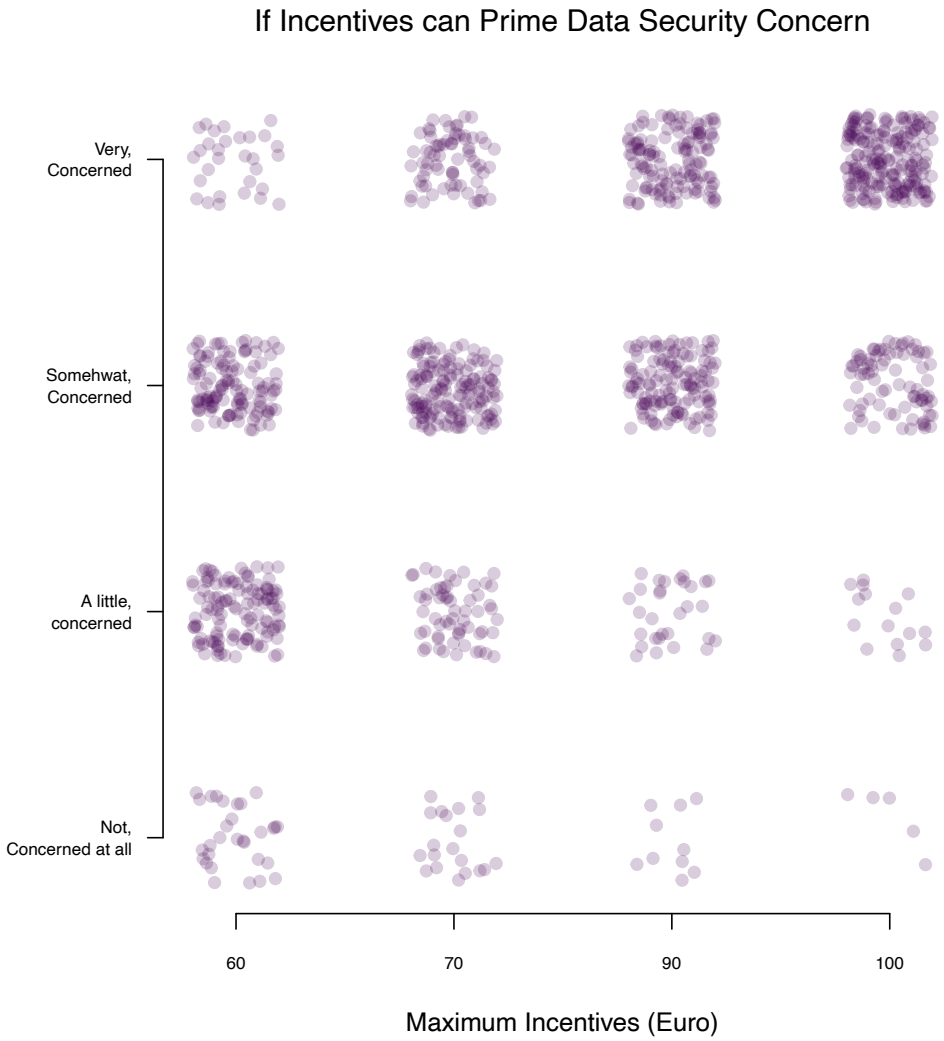

Figure 8: Hypothetical distribution of participants' security concern about downloading an app that collect their smartphone usage data across 4 maximum incentive groups if monetary incentive successfully prime their level of security concern.

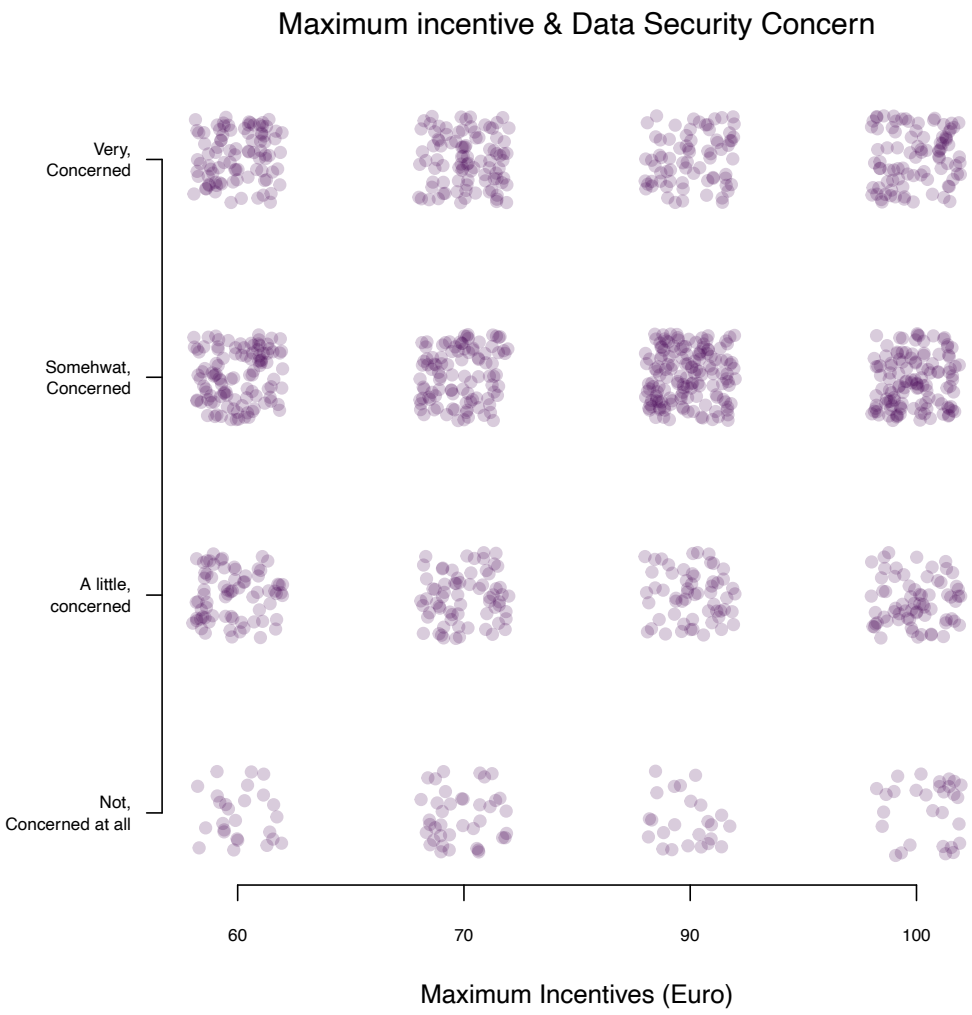

Figure 9: Actual distribution of participants' security concern about downloading an app that collect data about their smartphone usage across 4 maximum incentive groups (N=1204).

the average marginal effect of monetary incentive with 95 per cent confidence interval on predicting the likelihood of having high security concern[6]. Figure 10 shows that monetary incentive is not significantly correlated to participants' level of security concern ($p > 0.05$). The results show that participants' level of security concern is not statistically significantly influenced by the amount of incentive they are offered. Together with the results from figure

[6]The logistic regression model and average marginal effects with standard errors figure 10 based on can be found in appendix table 4 & 5.

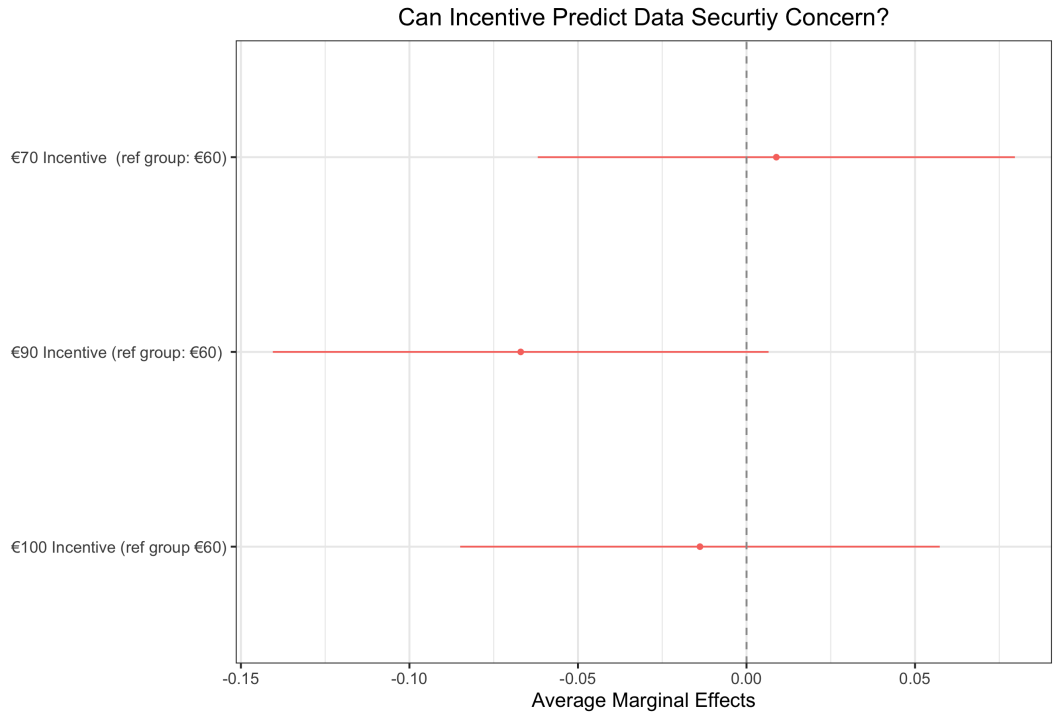

Figure 10: Average marginal effects with 95 per cent confidence interval from multiple logistic regressions predicting the likelihood of having high security concern. (N=1204).

8 and 9, H.2 is rejected.

5.3 Is Security Concern a Mediator between Monetary Incentive and Willingness to Participate in PMDC?

Because H.1 and H.2 is rejected, this means the first two conditions for complete mediation of security concern about PMDC are not met based on Baron and Kenny (1986) method. Hence, security concern about PMDC cannot be a mediator between monetary incentive and willingness to participate in PMDC. Therefore H.3 is rejected.

For completeness sake, the results of the third mediation testing regressions are also presented here. Figure 11 shows the average marginal effects of security concern about PMDC and monetary incentive on predicting the likelihood of willingness to participate in

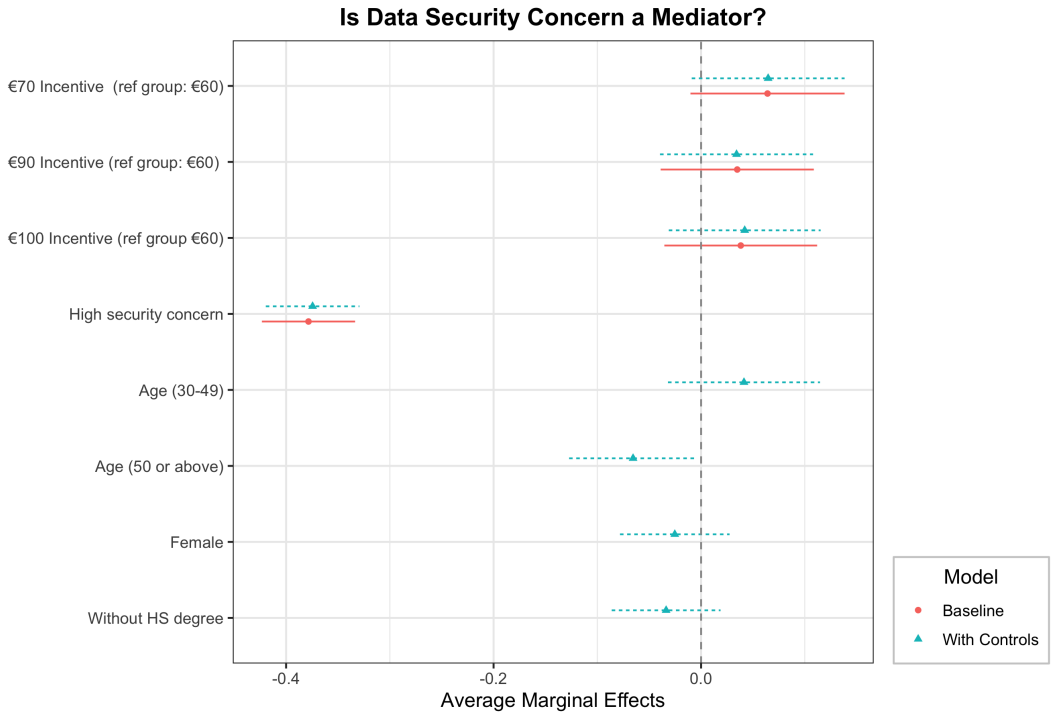

Figure 11: Average marginal effects with 95 per cent confidence interval from multiple logistic regressions predicting likelihood of willingness to participate in PMDC.(N=1204).

PMDC in a base model and a model with socio-demographics controlled [7]. Figure 11 shows two major findings. First, consistent with the H1 analysis results, monetary incentive is not significantly correlated to willingness to participate in PMDC ($p>0.05$), holding other variables constant in both models. Second, security concern is negatively significantly correlated to willingness to participate in PMDC ($p<0.01$), holding other variables constant in both models. In the base model, predicted probabilities of willingness to participate in PMDC decrease with high security concern by 38 percentage points (AME:- 38 per cent,$p<0.01$) over low security concern. In the model with controls, predicted probabilities decreases with high security concern by 37 percentage points (AME: -37 per cent, $p<0.01$) over low security concern. These

[7] The logistic regression models and average marginal effects with standard errors figure 11 based on can be found in Appendix table 6, 7 & 8.

findings suggest security concern does not have a mediating effect between incentive and willingness to participate in PMDC albeit having an independent negative significant effect on willingness to participate in PMDC.

6 SUMMARY AND DISCUSSION

This research questions the fundamental casual mechanism between monetary incentive and willingness to participate in PMDC. It builds a theoretic framework based on a collective literature from e-commerce (Carraascal et al, 2013; Li et al, 2014; Nget et al, 2017; Staiano et al, 2014); survey methodology (Couper & Singer, 2013; Groves et al, 2000; Haas et al, 2020; Jäckle et al, 2019; Joinson et al, 2008; Keusch et al, 2019; Keusch et al, 2020; Marrerios et al, 2017; Silber et al, 2021; Singer & Ye, 2013; Struminskaya et al, 2020; Wenz et al, 2019); socio-psychology (Cassino and Erisen, 2010); and privacy (Nissenbaum, 2019). The theory hypothesizes 1) an inverted U-shaped relationship between monetary incentive and willingness to participate in PMDC 2) a positive relationship between monetary incentive and level of privacy concern for PMDC 3) privacy concern for PMDC being a mediator between monetary incentive and willingness to participate in PMDC. Contrary to what the theory predicts, the empirical analysis finds no support to all three hypotheses.

6.1 Monetary Incentive & Willingness to participate

Monetary incentive is not significantly correlated to willingness to participate in PMDC. Willingness to participate in PMDC is not significantly different among participants who are offered more monetary incentive and participants who are offered less. Contrary to what

the theory predicts, how participants consider the appropriateness of the data transmission in PMDC does not change as the incentive amount increases. The results are inconsistent with Haas et al (2020) findings. This inconsistency is not surprising because incentive studies in PMDC often have low external validity and this study measures stated willingness to participate in PMDC, whereas Haas et al (2020) measure actual participation in PMDC.

There are three possible explanations why no significant relationship is found between monetary incentive and willingness to participate. First, it could be that participants value their data more than €100. A sub-group analysis on participants who are unwilling to participate and offered less than €100 corroborates this speculation[8]. However, offering incentives above €100 to participants might not be pragmatic due to budgetary reason. If researchers want to entice participation of those who can be converted from unwilling to willing to participate in PMDC in Keusch (2019) study with €100 or above €100 incentive (N=278), the cost will soar up to €39650 [9]. Researchers do not have unlimited budget to collect data. Further studies should investigate whether offering incentive above €100 to entice participation in PMDC is both effective and pragmatic.

Second, it could be that participants are not motivated by monetary but non-monetary motives to participate in PMDC. The null effect of additional monetary incentive and negative significant effect of security concern in this study are good evidences to support this argument. A second sub-group analysis on participants who are willing to participate and offered more

[8]A sub-group analysis is further conducted on those who are unwilling to participate and offered less than €100 to observe how much extra money is needed for their participation. Overall, 278 participants (66 percent of this sub-group) can be swayed by more monetary incentive. Among these 278 participants, 199 participants are willing to participate with €100; and 79 participants need incentive above €100 (on average €250) to be willing to participate in PMDC. For data flowcharts and further details, see Appendix figure 6 & 7.

[9]Since 199 participants wanted €100 and 79 participants on average want €250 to change their mind from unwilling to willing to participate in PMDC. €199*100 + 250* 79= €39650.

than €60 further corroborates this argument[10]. The result shows that 83 per cent of them (N=416) are willing to participate with just €60 incentive. In a broader sense, these findings also answer to the most recent mobile data collection study where non-monetary motives such as whether participants get feedback from their participation in mobile data collection plays a significant role in influencing their willingness to participate (Wenz et al, 2022). Further studies should continue to explore the effect of non-monetary incentives that can entice participation in PMDC.

Third, additional monetary incentive might be useful for a particular group of participants. There are signs that additional monetary incentives might have a differential effect on enticing participation in PMDC. The two sub-group analyses suggest that participants who are willing to participate do not seem to be motivated by additional monetary incentive when considering their participation in PMDC; and participants who are unwilling to participate seem to be motivated by additional monetary incentive when considering their participation in PMDC. Among those who are unwilling to participate and offered less than €100, additional monetary incentive can convert 66 per cent of them (N=278) to be willing to participate; Among those who are willing to participate in PMDC and offered more than €60, 83 per cent of them (N=416) are willing to participate with just €60. What factor or characteristic of participants that distinguishes participants into those who can be motivated by additional monetary incentive is unfortunately not identified in this study.

More broadly speaking, the potential differential effect of monetary incentive has pro-

[10]A second sub- group analysis is conducted on those who are willing to participate and offered more than €60 to understand whether they are willing to participate with just €60 (N=503). Up to 83 per cent of these participants (N= 416) are willing to participate with just €60. See Appendix figure 8 for frequency distribution.

found implications because it raises questions to the ethical and methodological concern of using monetary incentive to entice participation in PMDC. A PMDC with high response rate does not equate to a PMDC that is representative. From a methodological standpoint, if additional monetary incentive is effective on enticing participation of a certain sub-group in the population in PMDC, this might perpetuate the existing non-response bias in PMDC by recruiting a particular sub-group from the population to participate and excluding a certain sub-group from participation (Singer & Ye, 2013). Even if additional monetary incentive can entice a sub-group that will otherwise not participate and reduce the non-response bias, monetary incentive might induce undue influence and coerce their participation in PMDC, making the use of incentive unethical (Singer & Couper, 2008). Further studies not only should explore whether a particular sub-group in the population is prone to be enticed by monetary incentive to participate in PMDC, but perhaps more importantly to question the fundamental ethical and methodological grounds of using monetary incentive to entice participation in PMDC.

6.2 Privacy concern about PMDC, Priming & Willingness to participate.

There is also no support showing the level of participants security concern is positively significantly correlated to the amount of monetary incentive they are offered. The results show the incentive does not have a priming effect that makes participants level of security concern salient. Although security concern is not a mediator between incentive and willingness to participate as hypothesized, security concern is found serendipitously to be negatively significantly correlated to willingness to participate in both the base model and the model

with socio-demographics controlled when testing for H.3. This finding is also consistent with the existing literature (Keusch et al, 2019; Struminskaya et al, 2020; Wenz et al, 2019).

There are two answers to why participants level of security concern is not positively correlated to the amount of maximum incentive they are offered. The intuitive answer is that the critical threshold that make participants' security concern salient does not fall within the range of maximum incentive (€60 - €100) in this study. Hence, we did not observe any significant different level of security concern across the four incentive treatment groups. Further PMDC studies can design an empirical setup with a higher upper limit of monetary incentive to observe whether a critical threshold of security concern or privacy concern exists as proposed in this research.

The methodological answer is that the theoretic construct of "privacy concern about PMDC" in this study is not operationalized with the most ideal measurement. In this research, theoretic construct of "privacy concern about PMDC" is operationalized with (Q21c) in Keusch (2019) questionnaire that measures participants' concern about data security when downloading an app that collect their smartphone data for research purposes. Strictly speaking, privacy concern is not conceptually equivalent to data security concern [11]. Further studies can explore whether a priming effect of privacy concern about PMDC is present with a more accurate operationalization of "privacy concern about PMDC".

For researchers who want to increase participation rate in PMDC, the null effect of incentive on participants' privacy concern about PMDC can be a good news since additional monetary incentive does not increase participants' privacy concern about PMDC that ulti-

[11] Please kindly refer the operationalization section where an argument is made to explain why data security concern (Q21c) is chosen to operate the theoretic construct of "privacy concern about PMDC".

mately reduces their willingness to participate in PMDC. However, it can also be a bad news for researchers as this finding suggests that participants' privacy concern for PMDC cannot be easily mitigated by factors researchers can control (eg. how much incentive is given, whether participants get feedback from their participation). This can pose a challenge for researchers who want to increase participation rate in PMDC because high privacy concern about PMDC significantly reduce willingness to participate in PMDC. Apart from incentive, how survey respondents are recruited (Jäckle et al, 2022); and offering participants controls over what data they are giving to researchers (Keusch et al, 2019; Struminskaya et al, 2020) are two known external factors under researchers control that can reduce participants concern for PMDC (as reflected with increasing hypothetical and actual participation rate in the mobile data collection in these studies). How to mitigate participants' privacy concern about PMDC will remain an important topic for future PMDC researches.

7 LIMITATIONS

This study has five limitations. First, the data analysed in this study is collected in 2017, five years prior to this paper is written. Within 2017-2022, there might be a paradigm shift of privacy norm which people assess the appropriateness of PMDC completely differently in 2022. For example, downloading a health app that tracks people personal data (eg. geo-locations) has become a norm since the outbreak of the Covid-19 global pandemic in 2019 (Munzert et al, 2021). With increasing familiarity of a mobile app that collects personal data, people in post Covid-19 era might exhibit different kinds of data sharing behaviour compared to participants who were recruited in 2017 in Keusch (2019) study. Post Covid-19 survey

respondents either become more willing to participate in PMDC because they are more familiar with the mobile data collection process, or unwilling to participate because previous exposure to a health app that collects their personal data makes their data security concern salient. More recent incentive studies are needed to understand how monetary incentive influences participation in PMDC in post Covid-19 era.

Second, maximum incentive is used to operationalized the theoretic construct of "Monetary Incentive" in this study due to data availability. It is possible that the mode of incentive can influence how participants perceive the appropriateness of PMDC, holding the incentive amount constant (Haas et al, 2020). Further studies should explore on this topic by introducing more levels for each mode of incentive and observe whether participants exhibit the same pattern of data sharing behaviour in PMDC found in this study.

Third, the hypothetical PMDC experiment in Keusch (2019) study is conducted by academic researchers. As explained in the theory section, changing one contextual parameter can influence how people perceive the appropriateness of the data transmission in PMDC. Sponsor effects are also observed in existing PMDC study (Keusch et al, 2019; Struminskaya et al, 2020). Therefore, the results in this study cannot be generalised to PMDC sponsored by commercial data brokers or governmental agency holding other variables constant. Further studies should explore whether monetary incentive is effective in enticing participation in PMDC sponsored by non-academic personnel.

Fourth, the personal data to be collected passively as described in Keusch (2019) vignettes are collected in aggregate. Previous studies have shown that participants assess the willingness to participate in a PMDC differently upon different data types being collected

(Struminskaya et al, 2020). Struminskaya et al (2020) findings also align with the theory proposed in this study which changing one contextual parameter- *information type* can already make a difference in how people assess the appropriateness of the data transmission in PMDC. Further studies should explore whether monetary incentive is useful at enticing participants' willingness to share a specific type of personal mobile data passively.

Fifth, the results in this paper cannot be generalized to the general population because respondents in this study are recruited from a non-probability panel. Respondents in Keusch (2019) dataset are recruited from a German non-probability panel. Silber et al (2021) has argued that survey respondents recruited from a non-probability panel might be more willing to share their personal data with researchers compared to respondents recruited via a probability panel. This is because survey respondents from a non-probability panel are already voluntarily participating in online surveys and more familiarised with the survey process (Silber et al, 2021). If the analysis conducted in this paper is based on a probability-based sample, it is reasonable to suspect that people will be more concerned about privacy and data security regarding PMDC because they are not familiarised with the survey process. With a higher privacy concern about PMDC among survey respondents recruited from a probability based sample, the negative effect of data security concern on willingness to participate might be even stronger compared to what is found in this research. Further studies should explore whether the findings in this study can be extended to sample collected from probability based panels.

8 CONCLUSION

Although no statistically significant relationship between incentive and willingness to participate is found, this study still makes some novel contributions to the PMDC literature. This study answers to the existing theoretic gap by building a theory explaining the casual mechanism between monetary incentive and willingness to participate in PMDC. Apart from that, the theory also maps different theoretic arguments and empirical studies under one overarching framework with clearly defined contextual parameters that influence how people perceive the appropriateness of PMDC. If future PMDC incentive studies adopt this framework, this might address the existing low external validity issue in PMDC incentive studies (Lucas, 2003). More broadly, the framework developed here can also be extended to understand how non-monetary contextual factors that influence participation in PMDC as long as researchers map the factors in interested to their corresponding contextual parameters in the modified TCI framework.

The empirical results also raise many meaningful questions about the effects of monetary incentive on enticing participation in PMDC under different circumstances, and the methodological & ethical grounds of using monetary incentive to entice participation in PMDC. The null effect of monetary incentive on participants' level of privacy concern about PMDC and the serendipitous discovery of an independent negative effect of privacy concern about PMDC on willingness to participate in PMDC also open door for new research topics on how to mitigate participants' privacy concern about PMDC. In summary, it is safe to say there are still many un-answered questions regarding the effect of monetary incentive on enticing participation in PMDC. Given the increasing popularity of smartphone app that

collects people data passively in today's society, more researches are needed to unveil not just the mechanism between incentive and willingness to participate in PMDC in different contexts and circumstances, but different mechanisms of willingness to participate in PMDC as well.

9 REFERENCE

Ajzen, I. (2005). *Attitudes, Personality and Behavior*. Maidenhead: McGraw-Hill International (UK) Ltd.

Baron, R.M. & Kenny, D.A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6), 1173-1182. https://doi.org/10.1037/0022-3514.51.6.1173

Carrascal, J.P., Riederer, C., Erramilli, V., Cherubini, M., & De Oliveira, R. (2013, May). Your browsing behavior for a big mac: Economics of personal information online. In *Proceedings of the 22nd international conference on World Wide Web* (pp.189-200). https://doi.org/10.1145/2488388.2488406

Cassino, D. & Erisen, C. (2010). Priming Bush and Iraq in 2008: A Survey Experiment. *American Politics Research*, 38(2), 372-394. https://doi.org/10.1177/1532673X09346799

Couper, M.P. & Singer, E. (2013). Informed Consent for Web Paradata Use. *Survey Research Methods*, 7(1), 57-67. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3790572/

Garnier, S., Ross, N., Rudis, R., Camargo, P.A., Sciaini, M. & Scherer, C. (2021). *viridis - Colorblind-Friendly Color Maps for R*. https://doi.org/10.5281/zenodo.4679424, R package

version 0.6.2. 10.5281/zenodo.4679424.

Groves, R.M., Singer, E., & Corning, A. (2000). Leverage-Saliency theory of survey participation: description and an illustration. *Public Opinion Quarterly*, 64 (3), 299–308. https://doi.org/10.1086/317990

Haas, G.C., Kreuter, F., Keusch, F., Trappmann, M. & Bähr, S. (2020). Effects of Incentives in Smartphone Data Collection. In *Big Data Meets Survey Science: A Collection of Innovative Methods* (pp.387-413). First Edition. Hoboken, NJ: John Wiley & Sons. https://doi.org/10.1002/9781118976357.ch13

Hlavac, M. (2022). *stargazer: Well-Formatted Regression and Summary Statistics Tables.* Social Policy Institute, Bratislava, Slovakia. R package version 5.2.3,https://CRAN.R-project.org/package=stargazer.

Jäckle, A., Burton, J., Wenz, A., Couper, M.P. & Lessof, C. (2019). Participation in a mobile app survey to collect expenditure data as part of a large-scale probability household panel: coverage and participation rates and biases. *Survey Research Methods*, 13(1), 23-44. https://doi.org/10.18148/srm/2019.v1i1.7297

Jäckle, A., Wenz, A., Burton, J. & Couper, M.P. (2022). Increasing Participation in a Mobile App Study: The Effects of a Sequential Mixed-Mode Design and In-Interview Invitation. *Journal of Survey Statistics and Methodology*, 1-25. https://doi.org/10.1093/jssam/smac006

Joinson, A.N., Paine, C., Buchanan, T. & Reips, U. (2008). Measuring self-disclosure online: Blurring and non-response to sensitive items in web-based surveys. *Computers in Human Behavior*, 24(5), 2158-2171. https://doi.org/10.1016/j.chb.2007.10.005

Keusch, F. (2019). Mobile Data Collection - Incentive Experiment. *GESIS Data Archive,*

Cologne. ZA6978 Data file Version 1.0.0, https://doi.org/10.4232/1.13247.

Keusch, F., Struminskaya, B., Antoun, C. Couper, M.P. & Kreuter, F. (2019). Willingness to Participate in Passive Mobile Data Collection. *Public Opinion Quarterly*, 83(S1), 210-235. https://doi.org/10.1093/poq/nfz007

Keusch,F., Struminskaya,B., Kreuter,F. & Weichbold, M. (2020). Combining Active and Passive Mobile Data Collection: A Survey of Concerns. In *Big Data Meets Survey Science: A Collection of Innovative Methods* (pp. 657-681). First Edition. Hoboken, NJ: John Wiley & Sons. https://doi.org/10.1002/9781118976357.cH.22

Leeper, T.J. (2021). *margins: Marginal Effects for Model Objects*. R package version 0.3.26. https://cran.r-project.org/web/packages/margins/citation.html

Li, C., Li, D.Y., Miklau, G., & Suciu,D. (2014). A Theory of Pricing Private Data. *ACM Transactions on Database Systems (TODS)*, 39(4), 1-28. https://doi.org/10.1145/2691190.2691191

Lucas, J.W. (2003). Theory-Testing, Generalization, and the Problem of External Validity. *Sociological Theory*, 21 (3), 236-253. https://doi.org/10.1111/1467-9558.00187

Marreiros, H., Tonin, M., Vlassopoulos,M. & Schraefel, M.C. (2017). "Now that you mention it": A survey experiment on information, inattention and online privacy. *Journal of Economic Behavior & Organization*, 140, 1-17. https://doi.org/10.1016/j.jebo.2017.03.024

Munzert, S., Selb, P., Gohdes, A., Stoetzer, L. & Lowe, W. (2021). Tracking and promoting the usage of a COVID-19 contact tracing app. *Nature Human Behaviour*, 5, 247-255. https://doi.org/10.1038/s41562-020-01044-x

Nissenbaum, H. (2019). Contextual Integrity Up and Down the Data Food Chain. *Theoretical Inquiries in Law*, 20(1), 221-256. https://doi.org/10.1515/til-2019-0008

Nget,R., Yang,C. & Masatoshi,Y. (2017). How to balance privacy and money through pricing mechanism in personal data market. *arXiv preprint arXiv:1705.02982*. https://doi.org/10.48550/arXiv.1705.02982

RStudio Team. (2021). *RStudio: Integrated Development Environment for R*. Boston, MA. Available at: http://www.rstudio.com/

Singer, E. & Couper, M.P. (2008). Do Incentives Exert Undue Influence on Survey Participation? Experimental Evidence. *Journal of Empirical Research on Human Research Ethics*, 3 (3), 49-56. https://doi.org/10.1525/jer.2008.3.3.49

Singer, E. & Ye, C. (2013). The Use and Effects of Incentives in Surveys. *Annals of the American Academy of Political and Social Science*, 13(1), 112-141. https://doi.org/10.1177/0002716212458082

Silber, H., Breuer, J., Beuthner, C., Gummer, T., Keusch, F., Siegers, P., Stier, S. & Weiss, B. (2021). Linking surveys and digital trace data: Insights from two studies on determinants of data sharing behavior. *SocArXiv Papers*. https://doi.org/10.31235/osf.io/dz93u

Solt, F. & Hu, Y. (2015). *dotwhisker: Dot-and-whisker plots of regression results*. R pacakge version 0.7.4. https://cran.r-project.org/web/packages/dotwhisker/index.html

Staiano, J., Oliver, N., Lepri, B., de Oliveira, R., Caraviello, M. & Sebe, N. (2014, September). Money walks: a human-centric study on the economics of personal mobile data. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp.583-594). https://doi.org/10.1145/2632048.2632074

Struminskaya, B., Toepoel, V., Lugtig, P., Haan, M., Luiten, A. & Schouten, B. (2020). Understanding Willingness to Share Smartphone-Sensor Data. *Public Opinion Quarterly*,

84(3), 725-759. https://doi.org/10.1093/poq/nfaa044

Venables, W.N. & Ripley, B.D. (2002). *Modern Applied Statistics with S*. Fourth edition. Springer, New York. https://www.stats.ox.ac.uk/pub/MASS4/

Wenz, A., Jäckle, A. & Couper, M.P. (2019). Willingness to use mobile technologies for data collection in a probability household panel. *Survey Research Methods*, 13(1), 1-22. https://doi.org/10.18148/srm/2019.v1i1.7298

Wenz, A., Jäckle, A., Burton, J. & Couper, M.P. (2022). The Effects of Personalized Feedback on Participation and Reporting in Mobile App Data Collection. *Social Science Computer Review*, 40 (1), 165-178. https://doi.org/10.1177/0894439320914261

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wickham, H., François, R., Henry, L. & Müller, K. (2022). *dplyr: A Grammar of Data Manipulation*. https://dplyr.tidyverse.org, https://github.com/tidyverse/dplyr

10 APPENDIX

Table 1: Summary statistics

Statistic	N	Mean	St. Dev.	Min	Max
dummy_70	1,204	0.249	0.433	0	1
dummy_90	1,204	0.250	0.433	0	1
dummy_100	1,204	0.255	0.436	0	1
WTP	1,204	5.591	3.897	0	10
WTP_GRP	1,204	2.514	1.119	1	4
WTP_dummy	1,204	0.535	0.499	0	1
security_concern	1,204	2.868	0.940	1	4
security_dummy	1,204	0.287	0.452	0	1
female	1,204	0.500	0.500	0	1
Age_30_49	1,204	0.230	0.421	0	1
Age_50_or_above	1,204	0.478	0.500	0	1
without_HS	1,204	0.477	0.500	0	1

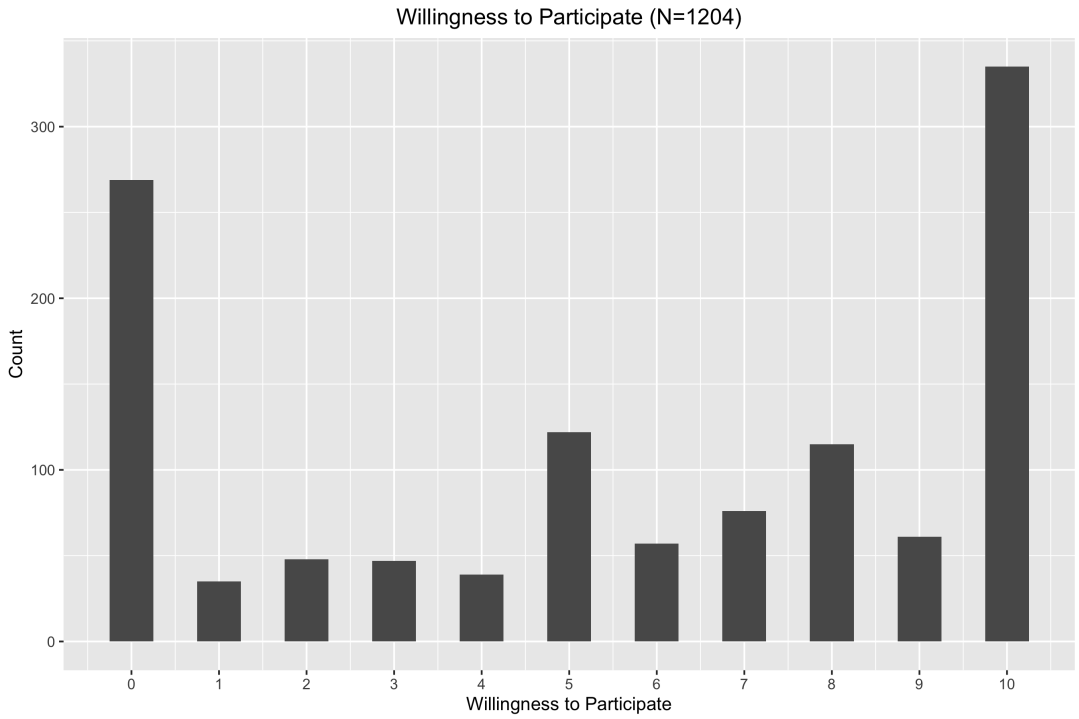

Figure 1: Distribution of willingness to participate on the 11-point scale at aggregate (N=1204).

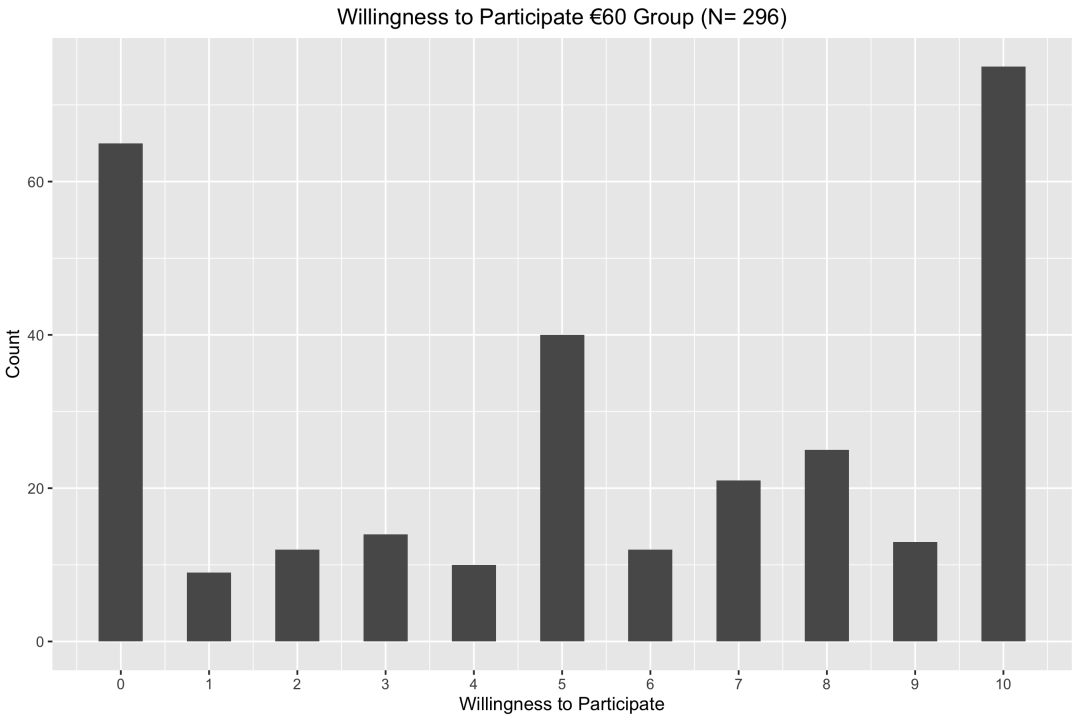

Figure 2: Distribution of willingness to participate among participants on the 11-point scale in €60 incentive group (N=296).

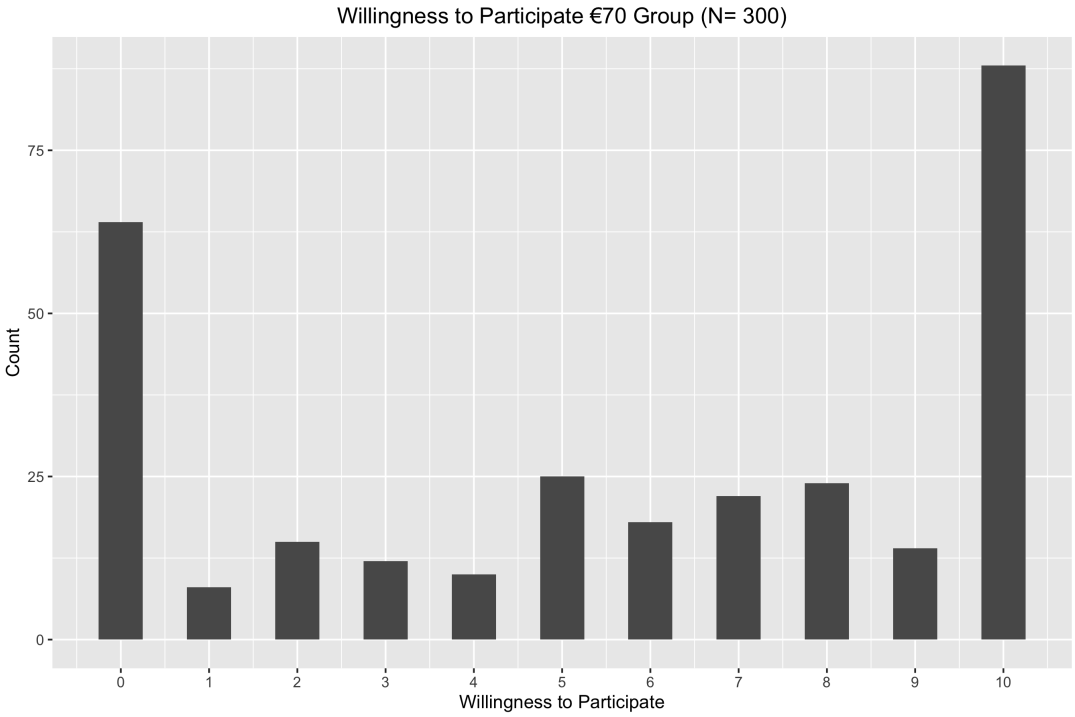

Figure 3: Distribution of willingness to participate among participants on the 11-point scale in €70 incentive group (N=300).

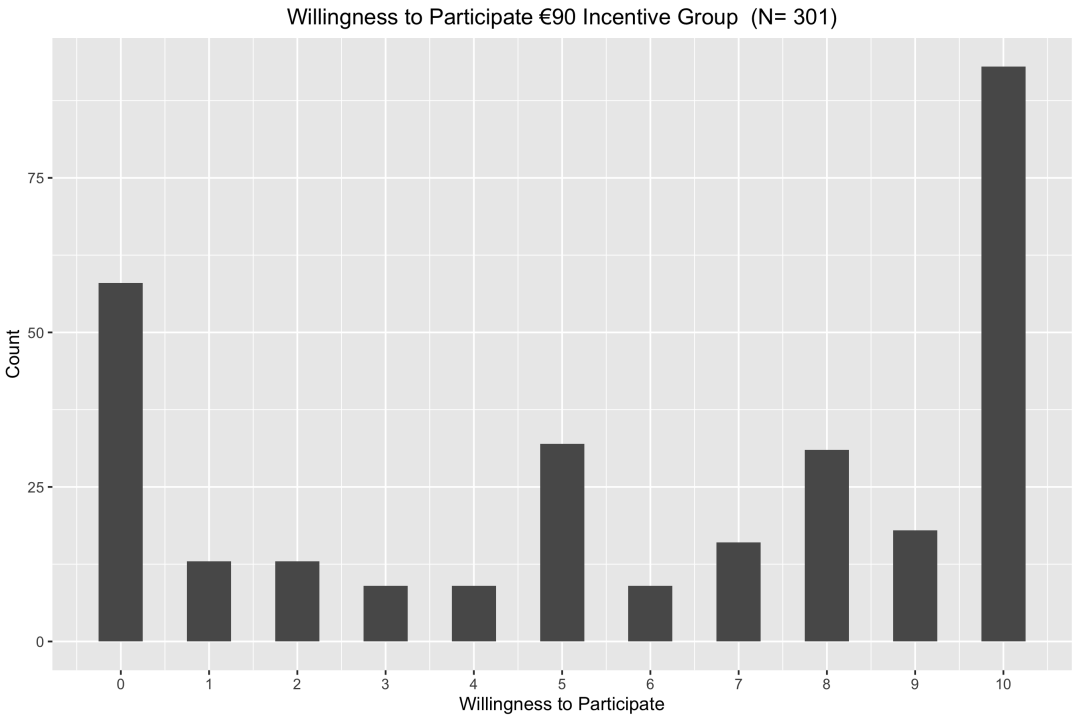

Figure 4: Distribution of willingness to participate on the 11-point scale among participants in €90 incentive group (N=301).

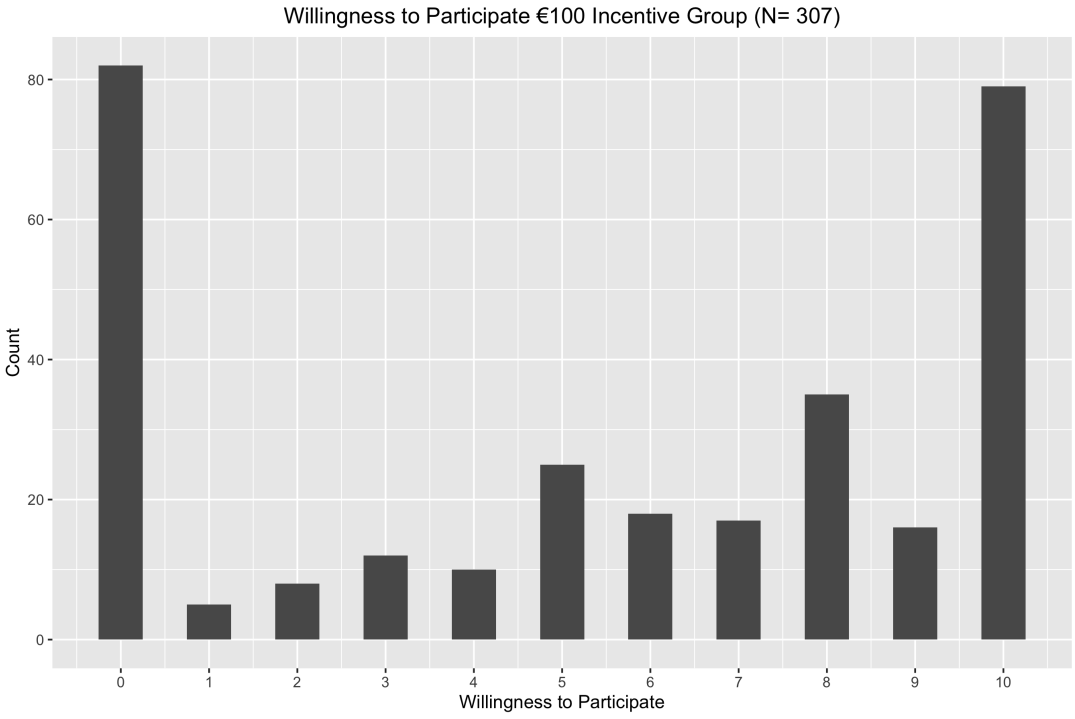

Figure 5: Distribution of willingness to participate on the 11-point scale among participants in €100 incentive group (N=307).

Table 2: Logit coefficients with standard errors (in parenthesis) predicting the likelihood of willingness to participate in PMDC for hypothesis 1

	Dependent variable:Willingness to participate
	Logit Coef (s.e.)
€70 (ref:€60)	0.241
	(0.164)
€90 (ref:€60)	0.247
	(0.164)
€100 (ref:€60)	0.177
	(0.163)
Constant	−0.027
	(0.116)
Observations	1,204
Log Likelihood	−830.137
Akaike Inf. Crit.	1,668.275
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 3: Average Marginal Effect (AME) with standard error (SE) predicting the likelihood of willingness to participate in PMDC for hypothesis 1

	factor	AME	SE	z	p	lower	upper
1	€100	0.044	0.040	1.088	0.277	-0.035	0.123
2	€70	0.060	0.041	1.473	0.141	-0.020	0.139
3	€90	0.061	0.041	1.511	0.131	-0.018	0.141

Table 4: Logit coefficients with standard errors (in parenthesis) predicting the likelihood of having high level of security concern about downloading an app which collects data about how participants use their smartphones with different incentive scheme in hypothesis 2.

	Dependent variable:High security concern
	Logit Coef (s.e.)
€70 (ref:€60)	0.043
	(0.177)
€90 (ref:€60)	−0.329*
	(0.185)
€100 (ref:€60)	−0.068
	(0.178)
Constant	−0.828***
	(0.126)
Observations	1,204
Log Likelihood	−718.775
Akaike Inf. Crit.	1,445.550
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5: Average Marginal Effect (AME) with standard errors (SE) predicting the likelihood of having high concern about privacy concern about download an app which collects data about how participants use their smartphones in hypothesis 2

	factor	AME	SE	z	p	lower	upper
1	€100	-0.014	0.036	-0.380	0.704	-0.085	0.057
2	€70	0.009	0.036	0.245	0.806	-0.062	0.080
3	€90	-0.067	0.038	-1.785	0.074	-0.141	0.007

Table 6: Logit coefficients with standard errors (in parenthesis) predicting the likelihood of willingness to participate in PMDC for hypothesis 3 in a base model and a model with participants' socio-demographics controlled.

	Dependent variable:Willingness to participate	
	Logit Coef (SE)	Logit Coef (SE)
€70 (ref:€60)	0.299*	0.305*
	(0.178)	(0.178)
€90 (ref:€60)	0.163	0.162
	(0.176)	(0.178)
€100 (ref: €60)	0.179	0.199
	(0.176)	(0.177)
High security concern	−1.767***	−1.769***
	(0.145)	(0.146)
Age(30-49)		0.195
		(0.177)
Age (50 or above)		−0.310**
		(0.150)
Female		−0.120
		(0.128)
without HS		−0.159
		(0.127)
Constant	0.469***	0.703***
	(0.130)	(0.181)
Observations	1,204	1,204
Log Likelihood	−745.182	−738.317
Akaike Inf. Crit.	1,500.363	1,494.635
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 7: Average Marginal Effect (AME) with standard error (SE) predicting the likelihood of willingness to participate in PMDC for hypothesis 3 (base model)

	factor	AME	SE	z	p	lower	upper
1	€100	0.038	0.038	1.019	0.308	-0.035	0.112
2	€70	0.064	0.038	1.691	0.091	-0.010	0.138
3	€90	0.035	0.038	0.926	0.355	-0.039	0.109
4	High security concern	-0.378	0.023	-16.506	0	-0.423	-0.333

Table 8: Average Marginal Effect (AME) with standard error (SE) predicting the likelihood of willingness to participate in PMDC for hypothesis 3 (with socio-demographic controls)

	factor	AME	SE	z	p	lower	upper
1	Age (30-49)	0.041	0.037	1.107	0.268	-0.032	0.115
2	Age (50 or above)	-0.066	0.031	-2.083	0.037	-0.127	-0.004
3	€100	0.042	0.037	1.126	0.260	-0.031	0.115
4	€70	0.065	0.038	1.720	0.085	-0.009	0.138
5	€90	0.034	0.038	0.909	0.363	-0.040	0.108
6	Female	-0.025	0.027	-0.939	0.348	-0.078	0.028
7	High security concern	-0.374	0.023	-16.256	0	-0.420	-0.329
8	without HS	-0.034	0.027	-1.258	0.208	-0.086	0.019

Table 9: Logit coefficients with standard errors (in parenthesis) predicting the likelihood of willingness to participate in PMDC for hypothesis 1 & 3 with different cutoff point (0-4: unwilling to participate, 5-10: willing to participate) or just extreme responses (0 and 10).

	Dependent variable:Willingness to participate					
	Different cutoff			Just extreme point		
	(1)	(2)	(3)	(4)	(5)	(6)
€70 (ref: €60)	0.036	0.060	0.065	0.175	0.101	0.084
	(0.170)	(0.184)	(0.184)	(0.236)	(0.269)	(0.272)
€90 (ref: €60)	0.143	0.035	0.038	0.329	0.086	0.121
	(0.171)	(0.184)	(0.187)	(0.238)	(0.270)	(0.275)
€100 (ref: €60)	−0.040	−0.075	−0.053	−0.180	−0.205	−0.207
	(0.168)	(0.182)	(0.183)	(0.231)	(0.264)	(0.269)
High security concern		−1.720***	−1.719***		−2.289***	−2.324***
		(0.138)	(0.139)		(0.216)	(0.221)
Age (30-49)			0.052			−0.094
			(0.188)			(0.294)
Age (50 or above)			−0.469***			−0.848***
			(0.157)			(0.245)
Female			−0.108			−0.291
			(0.133)			(0.194)
Without HS			−0.035			−0.038
			(0.132)			(0.192)
Constant	0.525***	1.099***	1.383***	0.143	0.900***	1.556***
	(0.120)	(0.139)	(0.194)	(0.169)	(0.204)	(0.306)
Observations	1,204	1,204	1,204	604	604	604
Log Likelihood	−788.657	−705.426	−698.168	−412.279	−343.259	−334.155
Akaike Inf. Crit.	1,585.314	1,420.852	1,414.335	832.559	696.517	686.311

Note: *p<0.1; **p<0.05; ***p<0.01

Table 10: Logit coefficients with standard errors (in parenthesis) predicting the likelihood of willingness to participate in PMDC with a different operationalization of data security concern using the original 4-point uni-polar scale.

	Dependent variable:Willingness to participate	
	Logit Coef (s.e.)	Logit Coef (s.e.)
€70 (ref:€60)	0.259	0.264
	(0.186)	(0.187)
€90 (ref:€60)	0.306*	0.294
	(0.182)	(0.184)
€100 (ref:€60)	0.220	0.241
	(0.183)	(0.185)
Security concern	−1.158***	−1.165***
	(0.082)	(0.083)
Age (30-49)		0.146
		(0.183)
Age (50 or above)		−0.354**
		(0.155)
Female		−0.057
		(0.133)
without HS		−0.219*
		(0.132)
Constant	3.333***	3.617***
	(0.274)	(0.306)
Observations	1,204	1,204
Log Likelihood	−698.931	−691.545
Akaike Inf. Crit.	1,407.863	1,401.090
Note:		*p<0.1; **p<0.05; ***p<0.01

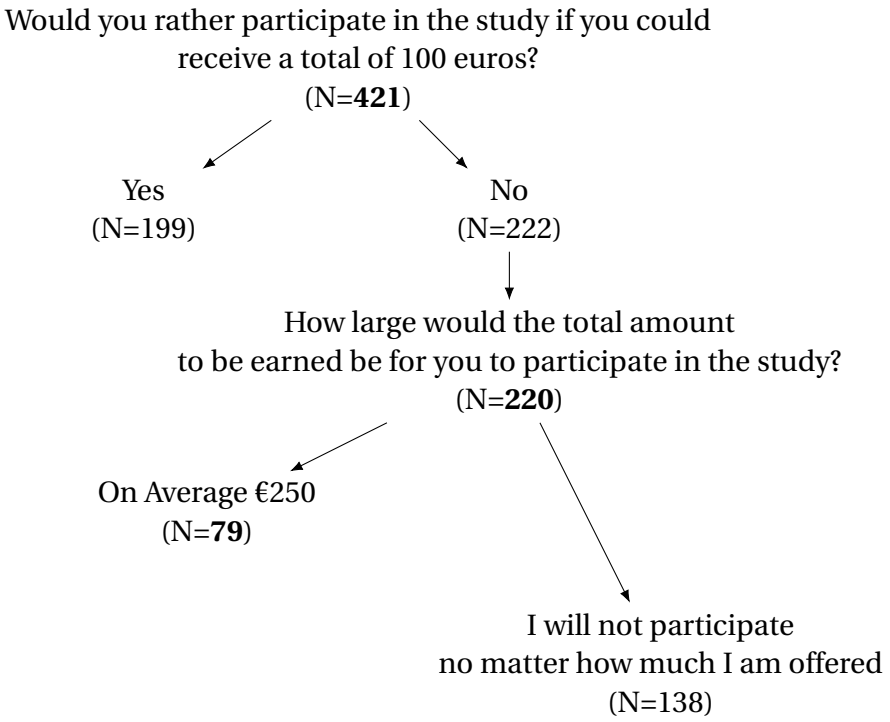

Figure 6: Data flowchart for sub-group analysis 1 (N=421).4 responses selecting both wanting more monetary incentive and never participate are excluded in the first question. 2 non-responses when answering how large the amount participants need are excluded in the second question; and 3 non-meaningful answers for extra amount they need are excluded in the third question.

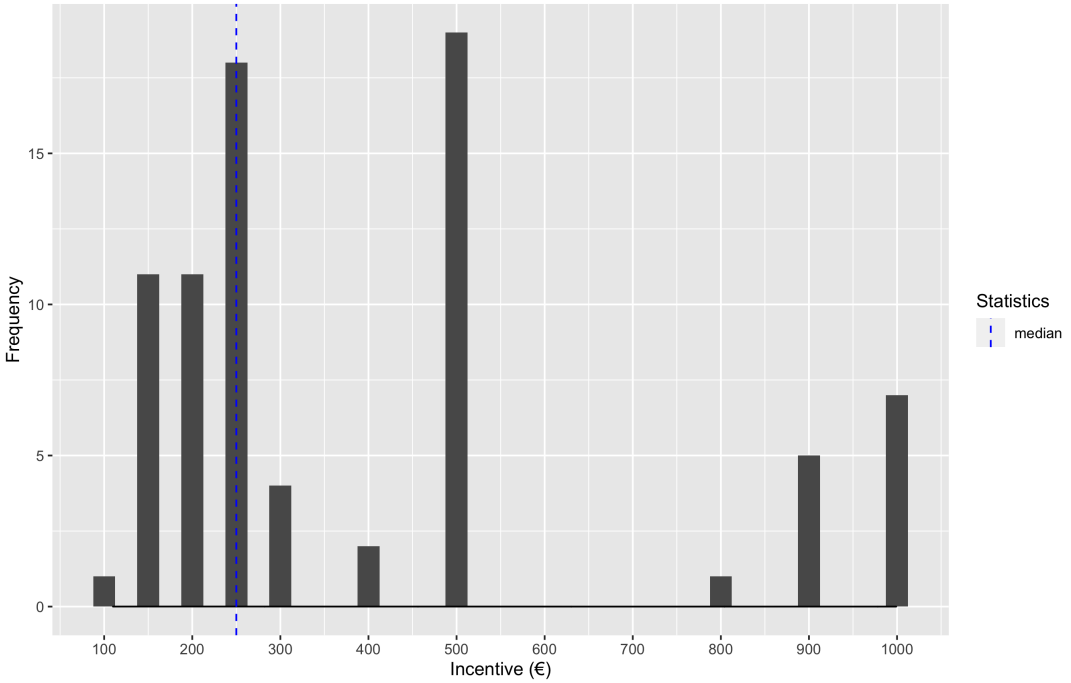

Figure 7: Histogram on how much additional incentive above €100 is needed for participants who are unwilling to participate and originally offered with less than €100 to entice their participation in PMDC (N=79).

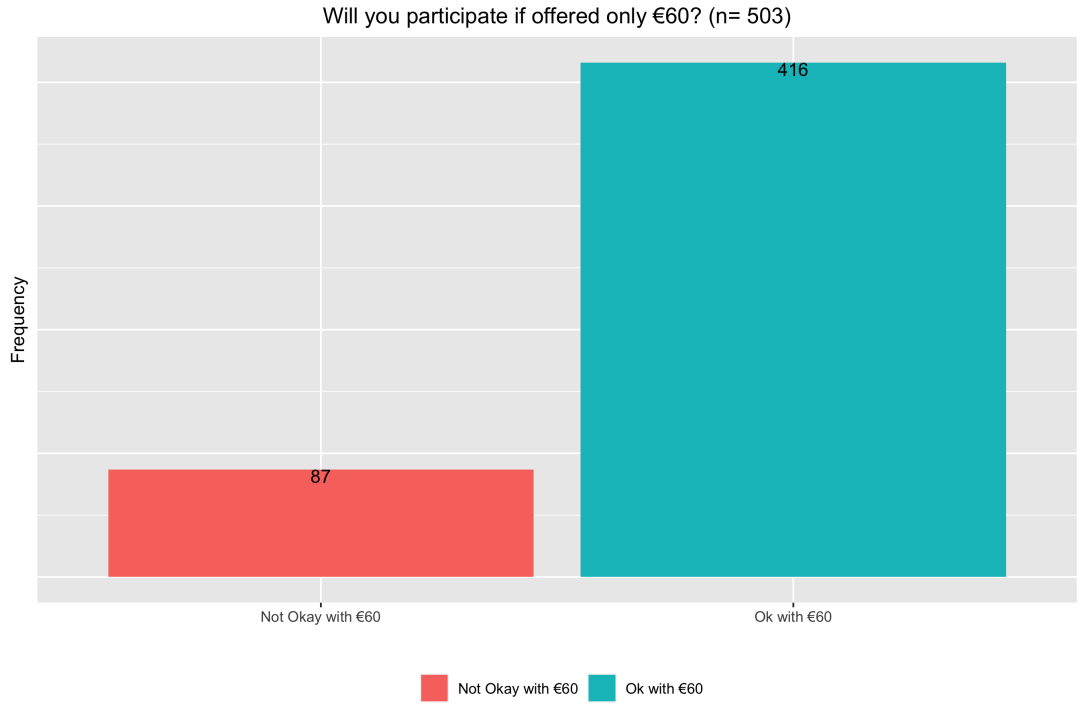

Figure 8: Distribution on the question whether participants who are willing to participate and offered more than €60 are willing to participate with just €60 (N=503).

11 Statutory Declaration

Hiermit versichere ich, dass diese Arbeit von mir persönlich verfasst ist und dass ich keinerlei fremde Hilfe in Anspruch genommen habe. Ebenso versichere ich, dass diese Arbeit oder Teile daraus weder von mir selbst noch von anderen als Leistungsnachweise andernorts eingereicht wurden. Wörtliche oder sinngemäße Übernahmen aus anderen Schriften und Sekundärliteratur und sonstige Quellen sind nachgewiesen und in der Bibliographie aufgeführt. Das Gleiche gilt für graphische Darstellungen und Bilder sowie für alle Internet-Quellen. Ich bin ferner damit einverstanden, dass meine Arbeit zum Zwecke eines Plagiatabgleichs in elektronischer Form anonymisiert versendet und gespeichert werden kann. Mir ist bekannt, dass von der Korrektur der Arbeit abgesehen und die Prüfungsleistung mit „nicht ausre-

ichend" bewertet werden kann, wenn die Erklärung nicht erteilt wird.

I hereby declare that the paper presented is my own work and that I have not called upon the help of a third party. In addition, I affirm that neither I nor anybody else has submitted this paper or parts of it to obtain credits elsewhere before. I have clearly marked and acknowledged all quotations or references that have been taken from the works of other. All secondary literature and other sources are marked and listed in the bibliography. The same applies to all charts, diagrams and illustrations as well as to all Internet sources. Moreover, I consent to my paper being electronically stores and sent anonymously in order to be checked for plagiarism. I am aware that the paper cannot be evaluated and may be graded “failed” (“nicht ausreichend”) if the declaration is not made.