Innovating Online Surveys using Experimental Methods

Inauguraldissertation zur Erlangung des akademischen Grades eines Doktors der Sozialwissenschaften der Universität Mannheim

Vorgelegt von

Christoph Beuthner

Dekan der Fakultät für Sozialwissenschaften
Prof. Dr. Michael Diehl

Betreuer
Prof. Dr. Florian Keusch
Dr. Henning Silber

Gutachter
Prof. Dr. Reinhard Pollak
Prof. Dr. Bella Struminskaya

Tag der Disputation
20.04.2023

Table of Contents

List of Figures

List of Tables

1 Introduction

This dissertation is dedicated to advance the application of experimental methods in survey methodology. In the first section of this dissertation I will give an introduction to experimental methods and casual inference. I will then reflect on the current state of experimental methods in the field of survey methodology, followed by the motivation of this dissertation. Finally, I give an overview of the main chapters.

1.1 Introduction to Experimental Methods

The experiment is one of the most important methods available to empirical research. It is the only way of directly observing causal relationships and thus, it is used in several disciplines like physics, medicine and the social sciences. While applying experimental designs in the natural sciences is rather straightforward, their use in social science disciplines is often problematic. Experiments can easily be confounded by a variety of effects such as the halo effect (Pohl, 2016) or biases caused by social desirability (Krumpal, 2013). In other situations conducting experiments is not possible due to the complexity of social situations or the inability to reproduce real life situations in a controlled environment.

In the realm of survey research, experiments are typically designed by randomly splitting respondents into a variety of groups and confronting them with different variations of a question or task (Sanders et al., 2002; Schuldt et al., 2011) or by manipulating question order and questionnaire length (Dillman et al., 1993). Common examples are incentive experiments where respondents receive a different amount of money for completing a task, or varying the position of two questions to measure recency effects.

1.2 Causal Inference and Experimental Methods

Besides problems arising from the implementation of specific experiments in social sciences, deriving casual inference from the gathered data is still problematic. Following Rubin (2005) all experiments are subject to the "fundamental prob-

lem of casual inference" (Holland, 1986, p. 947), meaning that in an experiment where subjects are assigned to various treatment and control groups, each individual can only be assigned to one treatment at the same time. However, to measure the causal effect of the treatment, it would be necessary to measure the effects of the other treatments on the same individual. Thus, causal inference on the unit level cannot be observed. Therefore, when conducting an experiment, participants are randomly assigned to the experimental groups which allows the estimation of causal inference on the population level (Sekhon, 2008). This so called average treatment effect (ATE) represents the difference in mean outcomes between a treatment and a control group (Rubin, 2005). However, if a random allocation of participation is not possible, the average treatment effect cannot be calculated.

As this dissertation is dedicated to the application of experimental methods in survey methodology, the next section will give an overview of experimental research in the field.

1.3 Experiments in Survey Methodology

The implementation of valid experiments is often limited or even impossible in social science disciplines, due to practical needs or confounding factors, e.g. self-selection into treatment. In many cases, data is gathered from already existing sources such as large-scale population surveys. In survey methodology however, the way in which surveys and questionnaires are designed, offers an ideal environment for experimental research (Mullinix et al., 2015; Schuldt et al., 2011). To start with, researchers often have complete control over the sampling procedure and thus, can randomly allocate respondents to different treatment groups, e.g. let respondents receive different incentives. Further, they can randomly allocate respondents to different versions of a questionnaire or a specific question. However, researchers have to consider that certain experimental conditions, e.g. when one group receives a very burdensome condition, may introduce bias or even reduce statistical power. Such bias can for example be created by forcing respondents to use a specific device to answer an online survey, which might lead to a large number of drop-outs amongst respondents allocated to a specific device, e.g. a

smartphone. These factors can drastically bias findings from experiments and pose a huge challenge to researchers.

A substantial amount of studies in the field of survey methodology already utilizes experimental methods. Typically, three aspects of surveys can be investigated. First, the general design of a survey can be researched. This encompasses all aspects of the study design that are related to sampling and fieldwork, e.g. the number of contacts, the survey mode or the amount of incentives given. Second, the questionnaire itself can be investigated. Typical manipulations in this area include survey length or the wording of specific questions. Finally, the analysis of survey data is an important field that can be studied. Studies focusing on this are mainly located in the field of statistics. This dissertation will focus on the first two aspects of experiments in survey methodology.

1.3.1 Survey Design Experiments

To date, different studies have investigated different subcategories of design choice effects, such as incentive structures (Becker & Glauser, 2018; McGonagle & Freedman, 2016; Mercer et al., 2015), contact mode effects (DeLeeuw, 2018; Millar & Dillman, 2011) or sampling frame effects (Blom et al., 2015; Scherpenzeel & Das, 2010). In all these applications, in the most optimal case, researchers can utilize an available and complete list of respondents and randomly split them into the respective experimental groups, e.g. a control group receiving no incentive and a treatment group receiving a 2€ incentive. However, in many cases conducting design experiments is more difficult. Most often recruiting respondents from a complete list is not possible and sampling is achieved through sophisticated procedures like the random route method (Bauer, 2014). In other cases respondents may refuse to participate in the design they are allocated to, for example when they are forced to participate using a specific device (Mavletova & Couper, 2015).

1.3.2 Questionnaire Design Experiments

Studies experimentally investigating the effects of questionnaire design, e.g. question placement (Sakshaug et al., 2013) or question wording (Schuldt et al., 2011;

Silber et al., 2018) are also common in the field of survey methodology. When such studies are conducted online, respondents can be randomly split in the respective groups shortly before the experiment, so that previous dropouts can not affect the random allocation of respondents. In other modes however, the implementation of such experiments is more complicated. For practical reasons, fully crossed experimental designs are often not implemented in mail surveys, as this would lead to a large number of different questionnaire versions and thus increase administrative costs. Furthermore, when respondents have to be allocated to an experimental group before the start of the survey, e.g. when receiving a paper questionnaire, experiments might be confounded by dropouts caused by differences between questionnaire versions. Additionally, especially in large scale surveys, researchers worry about the comparability of results and thus, are hesitant to implement experiments.

1.3.3 Analytical Experiments

Studies that focus on the analysis of experimental data to most parts examine different statistical methods used to analyze the data gathered beforehand. These studies often compare different algorithms for data preprocessing, e.g. imputation methods (Shao & Sitter, 1996; Shao & Steel, 1999) or different data analysis strategies. However, as these methods are not within the scope of this dissertation, I will not focus on them further.

In the next section I will explain the motivation for this dissertation and examine how it will help advance the field of survey methodology.

1.4 Motivation

The experiment is the only scientific method that allows to draw causal conclusions. While it is already widely used in the field of survey methodology, there are often issues arising from implementation of experiments that can make drawing casual inference impossible or prone to bias.

One reason for this can be that the intentionally induced difference between treat-

ment and control condition itself can cause drop-out or non response. A treatment can be perceived as more or less burdensome and thus, cause respondents to not participate in or drop-out of the survey or to skip questions. The consequences arising from this effect vary between experiments. A high drop-out or non-participation rate can lead to a drastically reduced sample size and thus, to an underpowered study, making it impossible to accurately estimate effects. However, it is also possible that a treatment such as smartphone participation causes a specific group, e.g. the elderly to not participate in an experiment at all, which then leads to non-response bias.

Similar issues can arise when implementing survey design experiments in small samples. A high number of experimental variations might lead to an underpowered study. This can be especially true when specific conditions are much more attractive for participants than others. Such effects are especially problematic in probability samples where compensating non-participation by drawing new participants from the sampling frame is not unproblematic or sometimes even impossible.

Another prominent issue when implementing experiments is the connection between stimulus and measurement. When respondents are subject to a stimulus, researchers have to measure the effect in close temporal proximity to it. The longer the time between treatment and measurement, the more likely the vanishing or distortion of the effect becomes. Put differently, experimental research is sensitive to the way in which stimulus and measurement are chained together.

When considering experiments in survey methodology, some phenomena such as incentives (Singer & Ye, 2013) received a lot of attention while others (e.g., respondents willingness to consent to data linkage (Sakshaug et al., 2013; Wenz et al., 2019) are less well researched so far. Therefore, this dissertation aims to further contribute on applying experimental methods in order to generate new knowledge in innovative areas of survey methodology. It will combine topics of survey research with with in-depth study of the experimental methods used. Overall, the studies presented in this dissertation will focus on the improvement

of survey and questionnaire design through experimental methods. In addition, from each applied method, conclusions are drawn in order to further advance the use of these experimental methods.

Each of the three upcoming chapters will introduce a empirical study utilizing experimental methods. While each of them introduces a new research design, they all share the commonality of using experimental designs while at the same time introducing novel procedures hopefully fruitful to other social scientists. Briefly, study 1 (chapter 2 of this dissertation) will focus on the effects of incentives as well as survey mode in a self-administered mixed-mode survey. It poses an innovative approach of delivering incentives in one treatment group, namely a 2€ delayed incentive. Study 2 (chapter 3 of this dissertation) introduces a study that experimentally investigates network name generators and recall aids on smartphones. The study in chapter 4 of this dissertation eventually uses experiments to investigate respondents willingness to consent to the linkage of additional data - a trend that has recently become more important in the social sciences. The final chapter will summarize the results and discuss the experimental methods in the light of new trends in survey methodology. Next, I will provide a detailed description of the following three chapters.

1.5 Summary of Chapters

Chapter 2 ("The interplay of mode sequence and incentives in self-administered mixed-mode surveys") introduces a study that focuses on optimizing the recruitment of respondents in a probabilistic self-administered mixed-mode survey (i.e., mail and online). Two of the most popular survey design options which influence response rates and survey costs are incentives and mode sequence. While incentives are known to increase response rates (Singer & Ye, 2013) they also directly translate into survey costs, thus it is necessary to find the right incentive height to optimize the balance between both factors. The same can be said about contact modes (DeLeeuw, 2018). While offering a mail option besides an online questionnaire in the initial contact can increase response rates, especially amongst older respondents, it also adds additional costs. Further, different combinations of

incentives and mode sequences might yield the best outcome for different demographic groups. In the study of Chapter 2, I randomly assigned 2,980 respondents to one of four incentive conditions (no incentive, 1€ prepaid, 2€ prepaid, 2€ delayed) and one of two mode sequences (concurrent, sequential). Results show that a 2€ delayed incentive combined with a sequential design worked best amongst respondents younger than 50 years, while a 2€ prepaid incentive combined with a concurrent design was favourable for respondents above 50 years. I also found these two designs to be cost effective for the respective groups. The chapter concludes with implications for the design of self-administered mixed-mode surveys. This chapter aims to improve the implementation of design experiments in register-based probability surveys. While it shows how self-administered mixed-mode surveys can be improved using experiments it also highlights issues arising from self-selection bias.

The study in chapter 3 ("Effects of smartphone use and recall aids on network name generator questions") focuses on the implementation of an experiment based on a network name generator in a smartphone survey. Network name generators are used to collect data on a respondents' social network. In many cases, respondents are asked to name their closest friends or relatives and specify their relationship with them. To lower response burden respondents can be confronted with a recall aid which triggers memories related to the network name generator and thus making information regarding the social network more accessible. Furthermore, the use of a smartphone is assumed to increase response burden due to worse presentation on a smaller screen or distraction by other applications. I used an online access panel to randomly allocate 3,891 respondents to either use a Smartphone or a PC to answer the question. Independently, respondents were randomly allocated to receive an open-ended recall aid question before or after the name generator. The results show no difference in the number of reported contacts between PC and Smartphone respondents. Additionally, the network generator did not lead to a higher number of contacts reported. However, the data generated by the recall question could be used as an indicator of satisficing. I conclude with implications for the design of survey experiments using smartphones and the use of open-ended questions as recall aids for network name generators. This chapter focuses on the

implementation of experiments to improve network name generators. It also aims at helping to further innovate questionnaire experiments by trying to include various devices used by respondents.

Chapter 4 ("Consent to Data Linkage for Different Data Domains – The Role of Question Order, Question Wording, and Incentives") introduces a study that uses an experimental design to generate knowledge on the willingness of respondents to consent to the linkage of data from non-survey sources. It is assumed that allowing researchers to access additional data sources comes with privacy costs for respondents and thus increases response burden. However, researchers can use known mechanisms such as incentives or a positive question wording to move respondents towards sharing data. This study sampled 3,374 respondents from an online access panel. Respondents were randomly allocated to one of three question wordings (i.e., time saving benefit, scientific benefit, no benefit) and one of two incentive conditions (i.e., promised incentive, no incentive). Additionally, they received 7 consent requests regarding the linkage of additional data (i.e., administrative data, smartphone usage data, bank data, biomarkers, Facebook data, health insurance data, and sensor data) in random order. The results show that respondents are more likely to share data from certain domains such as biomarkers, Facebook data and smartphone usage data then data from others, e.g. bank account data. Considering the experimental manipulations made, only the question order had a significant effect on consent rates. The chapter concludes with a reflection on how to design multiple requests for additional data to achieve the highest consent rates.

This chapter aims to improve experimental methodology by trying to estimate the effect of experimental treatments on several questions presented in a sequence. Thus, it helps to understand the effects that arise when treatment and measurement are separated in a questionnaire.

References

Bauer, J. J. (2014). Selection Errors of Random Route Samples. *Sociological Methods & Research*, *43*, 519–544.

Becker, R., & Glauser, D. (2018). Are prepaid monetary incentives sufficient for reducing panel attrition and optimizing the response rate? An experiment in the context of a multi-wave panel with a sequential mixed-mode design. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, *139*, 74–95.

Blom, A. G., Gathmann, C., & Krieger, U. (2015). Setting up an online panel representative of the general population. *Field Methods*, *27*, 391–408.

DeLeeuw, E. D. (2018). Mixed-mode: Past, present, and future. *Survey research methods*, *12*, 75–89.

Dillman, D. A., Sinclair, M. D., & Clark, J. R. (1993). Effects of questionnaire length, respondent-friendly design, and a difficult question on response rates for occupant-addressed census mail surveys. *Public Opinion Quarterly*, *57*, 289–304.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*, 945–960.

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality & Quantity*, *47*, 2025–2047.

Mavletova, A., & Couper, M. P. (2015). A meta-analysis of breakoff rates in mobile web surveys. In D. Toninelli, R. Pinter, & P. de Pedraza (Eds.), *Mobile research methods: Opportunities and challenges of mobile research methodologies* (pp. 81–98). Ubiquity Press.

McGonagle, K. A., & Freedman, V. A. (2016). The effects of a delayed incentive on response rates, response mode, data quality, and sample bias in a nationally representative mixed mode study. *Field Methods*, *29*, 221–237.

Mercer, A., Caporaso, A., Cantor, D., & Townsend, R. (2015). How much gets you how much? Monetary incentives and response rates in household surveys. *Public Opinion Quarterly*, *79*, 105–129.

Millar, M. M., & Dillman, D. A. (2011). Improving response to web and mixed-mode surveys. *Public Opinion Quarterly*, *75*, 249–269.

Mullinix, K. J., Leeper, T. J., Druckman, J. N., & Freese, J. (2015). The Generalizability of Survey Experiments. *Journal of Experimental Political Science*, *2*, 109–138.

Pohl, R. (2016). *Cognitive illusions: Intriguing phenomena in judgment, thinking and memory* (Second edition). Routledge, Taylor & Francis Group.

Rubin, D. B. (2005). Causal Inference Using Potential Outcomes: Design, Modeling, Decisions. *Journal of the American Statistical Association*, *100*, 322–331.

Sakshaug, J., Tutz, V., & Kreuter, F. (2013). Placement, Wording, and Interviewers: Identifying Correlates of Consent to Link Survey and Administrative Data. *Survey Research Methods*, *7*, 133–144.

Sanders, D., Burton, J., & Kneeshaw, J. (2002). Identifying the True Party Identifiers: A Question Wording Experiment. *Party Politics*, *8*, 193–205.

Scherpenzeel, A., & Das, J. (2010). 'true' longitudinal and probability-based internet panels: Evidence from the netherlands. In J. Das, P. Ester, & L. Kaczmirek (Eds.), *Social and behavioral research and the internet* (pp. 77–103). Taylor & Francis.

Schuldt, J. P., Konrath, S. H., & Schwarz, N. (2011). "global warming" or "climate change"?: Whether the planet is warming depends on question wording. *Public Opinion Quarterly*, *75*, 115–124.

Sekhon, J. S. (2008). The Neyman-Rubin model of causal inference and estimation via matching methods. In J. Box-Steffensmeier, H. Brady, & D. Collier (Eds.), *The Oxford Handbook of Political Methodology* (pp. 1–32).

Shao, J., & Sitter, R. R. (1996). Bootstrap for Imputed Survey Data. *Journal of the American Statistical Association*, *91*, 1278–1288.

Shao, J., & Steel, P. (1999). Variance Estimation for Survey Data with Composite Imputation and Nonnegligible Sampling Fractions. *Journal of the American Statistical Association*, *94*, 254–265.

Silber, H., Bernd, W., Florian, K., Christoph, B., & Jette, S. (2018). Framing consent questions in mobile surveys: Experiments on question wording. *BigSurv18 Conference, Barcelona, Spain, October*, *27th*.

Singer, E., & Ye, C. (2013). The use and effects of incentives in surveys. *The ANNALS of the American Academy of Political and Social Science*, *645*, 112–141.

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*, 1–22.

2 The interplay of mode sequence and incentives in self-administered mixed-mode surveys

Under review in:
Bulletin of Sociological Methodology

Abstract

Self-administered mixed-mode surveys are increasingly regarded as a viable alternative to in-person interviews to collect data from the general population. However, little is known on how decisions on the incentive scheme and the mode sequence jointly effect key survey outcomes, like survey response, sample balance and survey costs. To this end, we drew a probability sample of the residential population of the city of Mannheim, Germany (N = 2,980) and randomly assigned target persons to one of four incentive conditions (no incentive, 1€ or 2€ prepaid, and 2€ delayed) and one of two mode sequences (concurrent or sequential). Results show that a sequential design (web only in the first contact) with a 2€ delayed incentive primarily attracted younger target persons, whereas a concurrent design (online and mail from the beginning) with a 2€ incentive in the first contact was favoured by target persons aged 50 and above. These two designs also turned out to be most cost efficient for the respective age groups. Based on our results, we recommend using available sample frame information to address different age groups with different combinations of survey design features. This may help to maximize response rates, realize a balanced sample, and minimize survey costs.

Keywords

mixed-mode surveys, web-push surveys, contact mode sequence, prepaid incentives, survey costs

2.1 Introduction

As response rates are falling and costs are increasing, the search for cheaper alternatives to face-to-face surveys has been growing in importance in recent years. This development was further accelerated by the spread of COVID-19, with social distancing as one of the key policies to limit the impact of the pandemic. Web surveys comply with the needs to collect data cost-effectively and in physical distance to the respondents. However, they suffer from two important limitations: First, in most countries representative surveys for the residential population cannot solely rely on the online mode since it lacks an adequate sampling frame that would allow researchers to draw probability samples and recruit their target persons online (Blom et al., 2015; Scherpenzeel & Das, 2010). Hence, another mode for contact is necessary. Second, although in developed societies internet penetration is close to saturation, empirical evidence still suggests that the online mode predominantly attracts younger and well-educated persons. In contrast, older target persons prefer to fill in a paper questionnaire at higher rates and sometimes struggle with participating online (Olson, 2020). Thus, combining the online with the mail mode seems to be a promising avenue for collecting data cheaply while increasing sample balance by also including offline populations (Messer & Dillman, 2011).

Data collection experiments implemented in the European Values Study (EVS) 2017 suggest that self-administered mixed-mode (web, online) surveys yield comparable response rates to face-to-face surveys, even for long questionnaires (Luijkx et al., 2020). In the EVS-Germany, Wolf et al. (2021) implemented a survey mode experiment and found that the response rates in two self-administered mixed-mode surveys even exceeded the one in the face-to-face survey while costs were reduced by more than a half. At the same time, the study showed only minor differences in substantive answers to core items of the EVS questionnaire. While such findings suggest that self-administered mixed-mode surveys might be a viable alternative to face-to-face surveys, more research is required on the interplay of key survey design decisions (like on the incentive scheme and on the sequencing of survey modes) on response rates and sample composition. In our

view, such research is especially required in countries like Germany where the use of self-administered modes in general population surveys has only lately received elevated attention (Wolf et al., 2021). In addition, little is known (or published) on how key survey design elements effect survey costs and this shortcoming has recently led to a call to unpack this black box (Olson, 2020). Thus, we argue that not only is more research needed on the interaction between key survey design elements on survey response and sample balance but also on their implications for survey costs.

Our study addresses these research gaps by relying on a probability-based self-administered mixed-mode (online, mail) survey carried out in the German city of Mannheim (∼310,000 residents). We implemented an experimental design in which we varied the mode sequence (concurrent vs. sequential) and the amount of small prepaid incentives (no incentive, 1€, and 2€). Additionally, target persons from two experimental groups received a 2€ prepaid incentive only with the second contact (delayed incentive).

The main objective of our analyses is to uncover the effects of combinations of small prepaid monetary incentives and mode sequence on survey response, sample composition with regard to age and on survey costs. Therefore, in the following sections, we first review previous research on the effects of prepaid monetary incentives and mode sequencing on these three outcome variables. Based on this evidence, we discuss why incentive scheme and mode sequence might evoke combined effects with respect to survey response, sample balance and survey costs. After presenting our data, methods, and empirical results, we conclude with some recommendations on targeting strategies in self-administered mixed-mode surveys deduced from our study.

2.1.1 Prepaid Monetary Incentives

In many interviewer-mediated and self-administered surveys, researchers use prepaid monetary incentives to motivate their target persons to take part in the survey. Mostly, prepaid incentives are provided with the initial contact, yet some

researchers introduce them in subsequent contacts as a tool for refusal conversion (McGonagle & Freedman, 2016).

Empirical evidence clearly suggests that prepaid monetary incentives offered with the first contact attempt significantly improve response rates, and their effectiveness has been reported as relatively stable over the past several years (Mercer et al., 2015). This positive effect also holds for small prepaid incentives that seem to work particularly well in mail surveys (Edwards et al., 2005; Pforr et al., 2015). Small prepaid monetary incentives are supposed to work by triggering the reciprocity norm and establishing trust, meaning that the incentive evokes the feeling of a moral obligation for the recipient to comply with the survey request (Becker & Glauser, 2018; Dillman et al., 2014; Gouldner, 1960). Compared to prepaid monetary incentives offered with the initial contact, empirical evidence on delayed prepaid incentives is comparatively rare, albeit suggesting that they yield similar effects on survey response (Blom et al., 2015; McGonagle & Freedman, 2016).

Findings on the effects of prepaid incentives on sample composition are far less unambiguous than for survey response (McGonagle & Freedman, 2016; Petrolia & Bhattacharjee, 2009; Singer et al., 1999), with one study even reporting null findings and questioning a general relationship between the two (Groves & Peytcheva, 2008). For self-administered surveys, McGonagle and Freedman (2016) found delayed prepaid monetary incentives to be more effective for older than for younger adults, and also reported a reduction in nonresponse bias since the incentive elevated the participation rates of people with low education. Similarly, Petrolia and Bhattacharjee (2009) found prepaid incentives to predominantly attract people with low education. On the contrary, Sun et al. (2020) showed that a small prepaid incentive (2$) did not affect the survey response of groups that are usually underrepresented in self-administered surveys. However, older target persons were more inclined to respond when they received a prepaid incentive.

A potential drawback of prepaid monetary incentives in cross-sectional surveys are their associated costs. When researchers decide to offer prepaid incentives,

a part of their money is "wasted" in the sense that some target persons pocket the incentive but do not comply with the survey request. However, since prepaid incentives can strongly increase survey response and the speed of participation, the associated costs can often be partly compensated because a smaller amount of target persons needs to be contacted again (Mann et al., 2008).

2.1.2 Mode Sequence

In self-administered mixed-mode surveys, target persons are often offered a mail and a web questionnaire. The mode sequence simply refers to the time when the two modes are offered: While in concurrent self-administered mixed-mode surveys, target persons can choose to respond via both modes in each contact attempt, sequential self-administered mixed-mode surveys usually start with offering the less costly online mode and introduce the mail mode only in subsequent contact attempts (DeLeeuw, 2018; Hox et al., 2017).

Generally, including multiple modes for data collection can increase response rates, as different parts of the population favour different survey modes. Older target persons, for instance, prefer a paper-based over a web questionnaire at higher rates and are thus more likely to participate in a web-and-mail as opposed to a web-only survey (Olson et al., 2012). However, a meta-analysis by Medway and Fulton (2012) concluded that offering a concurrent web-based option in a mail survey results in a reduction of response rates by around 3.8 percent. In contrast, Millar and Dillman (2011) showed that when both modes are offered sequentially starting with the online mode, response rates do not differ from the mail-only condition. To explain this phenomenon, Dillman et al. (2014) and Tourangeau (2017) argue that offering target persons multiple options simultaneously makes things more complicated for them so that some target persons may be inclined to postpone their decision to participate. However, recent studies by Wolf et al. (2021) and Mauz et al. (2018) did not find the sequential design to outperform the concurrent one in terms of survey response in surveys of the German general population.

When it comes to sample composition, there seems to be substantial empirical evidence that the net samples of mixed-mode surveys are more balanced than those from single-mode surveys (Cornesse & Bosnjak, 2018). More precisely, some studies suggest a better demographic representation when a mail mode is offered in addition to the online mode, since a single-mode web survey excludes certain segments of the population, especially the elderly (Messer & Dillman, 2011). In a similar vein, Bandilla et al. (2014) reported that complementing a web survey with the mail mode brought their sample closer to a face-to-face reference sample. However, to our knowledge, empirical evidence on the effects of mode sequence on the sample composition in self-administered mixed-mode surveys is lacking. On the one hand, it seems plausible to not expect the composition of the final sample to differ strongly between a concurrent and a sequential self-administered mixed-mode survey since target persons are offered both survey modes, albeit at a different time. However, due to different mode preferences between the age groups, one might at least expect the sample composition to vary significantly between the different contact attempts depending on the mode sequence. This might, in turn, also result in differences in the final net sample for the two mode sequence designs with regard to sample composition.

When it comes to survey costs, sequential self-administered mixed-mode designs are very appealing: Target persons with a high willingness to participate will do so in the cheapest mode, while more resources are allocated to target persons with a lower response probability. While this presumably pays off in surveys with a small number of contact attempts targeting a web-prone population, it may turn out differently in the general population. Here, relying on a concurrent design may result in a substantially higher survey participation in the initial contact, as target persons with a strong preference for the mail mode are much more likely to respond. Consequently, fieldwork efforts in subsequent contacts and thus survey costs may be reduced. In essence, this front-loading effect in the concurrent design may at least partly compensate the expected cost savings in the sequential design.

2.1.3 The Interplay of Prepaid Monetary Incentives and Mode Sequence

In self-administered mixed-mode surveys, the interplay of prepaid monetary incentives and mode sequence may yield effects on survey response, sample balance, and survey costs that differ from those of each survey design element on its own. Since we do not have sufficient theoretical knowledge or empirical evidence to propose a full set of hypotheses, we rather elaborate on selected expectations. With respect to sample balance, we have initially focused on age and gender because we have been able to obtain this information also for non-participants from the cities' population registry. Since neither the literature nor our own analysis (see Appendix Table A2.3) suggest relevant design effects on gender, in the following we concentrate on age which is known to be highly related to mode preferences.

Our first expectation is that the positive effect of small prepaid monetary incentives on survey response is stronger in the concurrent than in the sequential design. In line with the psychological processes stated above, target persons feel morally obliged to comply with the survey request when they receive a small prepaid incentive (Becker & Glauser, 2018; Dillman et al., 2014; Gouldner, 1960). When target persons are offered two modes simultaneously, this limits their options for self-justification to dissolve cognitive dissonances evoked by their noncompliance despite of the incentive. In contrast, in the sequential design, strong mode preferences may counteract the desired impact of the incentive. However, when a more suitable mode is offered only in subsequent contact attempts some weeks later, the feeling of a moral obligation evoked by the prepaid incentive received in the first contact might have already disappeared. For the delayed incentive, we would, however, not expect different effects on survey response depending on mode sequence because the offered mode alternatives in the second contact are the same for both types of mode sequence at the time they receive the incentive.

With respect to sample composition, we expect older target persons to be more attracted by prepaid monetary incentives. Supposedly, this is due to prepaid incentives addressing the recipients' sense of trust and reciprocity with both, on

average, is known to increase with age (Dohmen et al., 2008; Sutter & Kocher, 2007). Moreover, older target persons are more inclined to fill in a mail rather than an online questionnaire (Olson et al., 2012). Thus, we expect substantial higher response rates for older target persons in the concurrent design after the first contact while the sequential design should succeed in bringing into the sample more older target persons after the second contact when the mail mode is introduced. Yet this may, again, depend on the incentive scheme: For older target persons with a pronounced preference for the mail mode, the prepaid incentive may not help to push them into the web mode, thus lowering their response probabilities in the initial contact. When the suitable mode is introduced subsequently, however, the effect of the prepaid incentive in the initial contact may have already disappeared. In contrast, a prepaid incentive in the concurrent design may particularly attract older target persons since this combination fits with their mode preference and successfully addresses their sense of trust and reciprocity. Hence, we expect the final sample to include the highest share of older target persons in the concurrent design with the prepaid incentive offered in the initial contact.

The assumed combined effects of prepaid incentives and mode sequence on survey response should directly translate into survey costs. If, for instance, small prepaid incentives offered in a concurrent mixed-mode survey yield the highest survey response in the initial contact, this combination may turn out to be the most cost-effective in terms of survey costs per realized case.

If we also take into account the assumed effects on sample composition, survey costs may additionally differ between target persons from different age groups with one combination being most cost-effective for younger target persons while another being most beneficial for older ones. This holds especially true if mode choice differs largely between the groups with varying incentive schemes and mode sequence, since mode choice is crucial for survey costs. Generally, we assume a sequential design to yield higher rates of web participation since respondents who see the online mode as a viable option will answer online already in the initial contact (Messer & Dillman, 2011). This effect should be especially pronounced when incentives are offered, as the moral obligation evoked by the

incentive might push target persons to answer online even when they prefer the mail mode. This, in turn, may result in a cost advantage for the sequential design with the prepaid incentive.

2.2 Methods

2.2.1 Data and Methods

The data for this study was collected with a self-administered mixed-mode survey (mail, online) carried out between November 3rd, 2019 and March 16th, 2020 in the city of Mannheim, Germany. A total of 2,980 residents (aged 18 and above) were invited to participate. The gross sample was drawn randomly from the city's population register.

We contacted our target persons by mail and informed them about the purpose and content of the survey. Respondents who did not participate after the initial contact were contacted again five weeks later. The survey was framed as a community survey asking about life quality in Mannheim, and also included questions on political attitudes and perception of surveys. The median response time was 27 minutes. For the web questionnaire, we used a responsive design optimized for the various devices.

Our survey included an experiment with two fully crossed factors: mode sequence and incentive scheme. We randomly assigned each target person to one of eight experimental groups (see Table 2.1). In the sequential design, we initially provided our target persons only with login information (URL and password) for the web questionnaire. In contrast, target persons assigned to the concurrent design additionally received a paper-based questionnaire and a return envelope with the first contact. In the second contact, a paper-based questionnaire and the login information for the web questionnaire were provided to all target persons who did not respond after the initial contact.
The second experimental factor was the incentive scheme. Before the initial contact, we assigned our target persons to one of three incentive conditions for the first contact (no incentive, 1€, 2€). Moreover, the four experimental groups with

Table 2.1: Overview of the Experimental Groups and the Design Elements

Group	Incentive First Contact	Incentive Second Contact	Mode Sequence	Sample Size
1	0€	0€	Concurrent	373
2	1€	0€	Concurrent	373
3	2€	0€	Concurrent	373
4	0€	2€	Concurrent	373
5	0€	0€	Sequential	372
6	1€	0€	Sequential	372
7	2€	0€	Sequential	372
8	0€	2€	Sequential	372

no incentive in the first contact were randomly split into two groups with either a 2€ prepaid incentive in the second contact or with no incentive again.

2.2.2 Measures

The main dependent variable of our study is *survey response*. We treated questionnaires with a rate of item nonresponse exceeding 50 percent as cases of nonresponse. Moreover, we created two dichotomous variables indicating whether target persons took part after the first or after the second contact.

Our main independent variable is the *experimental group* to which each target person was assigned. In order to estimate the effects of our experimental manipulation on sample balance, we rely on the target persons' age, provided by the registration office. For our analysis, we standardized age around the mean.[1] We also created a *mode variable* indicating which mode a respondent chose to participate (0 = online mode, 1 = mail mode). The *device* respondents used to participate online was measured relying on the user agent string information (Roßmann et al., 2020). We differentiate between desktop computer, smartphone, and tablet.

Survey costs were measured as costs per complete case for each of our eight experimental groups. As Olson (2020) pointed out, reporting costs per complete is one of the most generalizable methods of reporting costs across studies. For calculating costs per complete, we specified the numerator as group-specific overall

[1] The registration office also provided information on gender. The respective results are shown in Figure A2.1 in the Appendix A.

survey costs and the denominator as completes per experimental group. The overall survey costs are the sum of (1) costs for material (e.g., envelopes), printing and processing (e.g., folding) (2) costs for incentives (including costs for procuring and affixing the coins on the invitation letters) (3) costs for postage (including postage for returned questionnaires), and (4) costs for data entry of the returned paper questionnaires.

In addition, we also calculated group-specific costs per complete separately for two age groups, namely for target persons aged 50 and above, and for target persons younger than 50 years.

2.2.3 Analyses

Our analysis consists of five steps. First, we start with estimating the isolated effects of incentive scheme and mode sequence based on logistic regression models with survey response as the dependent variable.

Second, as we are interested in the interplay of prepaid monetary incentives and mode sequence, we then compare survey response across our eight experimental groups, again based on logistic regression models with survey response as the dependent and our experimental groups as independent variable. To gain further insights on the processes of survey response over the course of the field time, we estimate two additional models to predict participation in the first and second contact, respectively.

Third, we extend these models by including the age of our target persons. In these models, we allow for an interaction between age and the experimental groups, thus enabling us to investigate whether design effects differ between target persons of different ages. To illustrate the results, we plot predicted probabilities. We prefer this for the ease of interpretation but are fully aware that predicted probabilities represent the conditional effects of our experimental manipulations on age and do not correspond with the regression coefficient of the interaction term, especially in terms of statistical significance.

Fourth, in addition to overall survey response, we analyze the mode of participation across the experimental groups and further investigate the devices respondents used for online participation.

Fifth, we report survey costs (1) per experimental condition and (2) per condition and age groups to get a complete picture of the simultaneous effects of our experimental manipulations on survey response and sample composition. This approach allows us to deduce recommendations on how to design a self-administered mixed-mode survey that maximizes survey response, yields a solid sample composition with regard to age, and remains cost-effective.

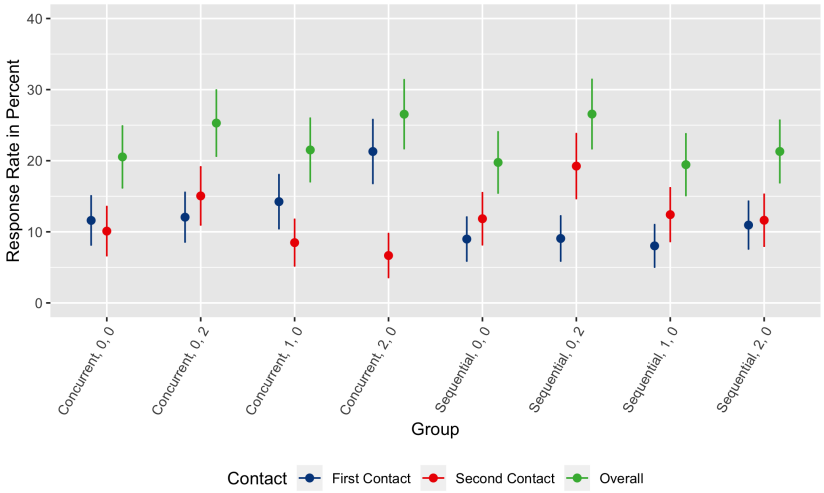
2.3 Results

Independent of the incentive scheme, the mode sequence had no effect on overall survey response (see Table A2.2 in the Appendix). However, the concurrent design yielded a significantly higher response rate in the first but a significantly lower response rate in the second contact. With regard to our prepaid incentives, we found the 2€ incentive to significantly increase response rates in both contacts, and the delayed incentive also improved the overall survey response.

2.3.1 The Effect of Incentive Scheme and Mode Sequence on Survey Response

Figure 2.1 displays the response rates after the first contact, after the second contact and overall by the experimental groups (the respective regression models on survey response by experimental group are provided in the Appendix Table A2.3). The blue lines represent survey response after the first contact and hint to a front-loading effect in the concurrent mixed-mode designs. However, survey response after the initial contact was only significantly higher for the concurrent design with the 2€ prepaid incentive provided with the first contact. In contrast, survey response after the second contact (red lines) was generally higher in the sequential design, thus compensating for its lower response in the initial contact. Again, we find the highest survey response for the sequential groups in the condition with

an enclosed monetary incentive (the delayed incentive). In terms of overall survey response (green lines), response rates substantially varied between 19.4 and 26.6 percent, with none of the conditions differed significantly from each other. However, the sequential design with a delayed incentive and the concurrent design with 2€ prepaid incentive performed best.



Note. Figure displays a 95% confidence interval.

Figure 2.1: Survey Response by Contact

2.3.2 The Effect of Incentive Scheme and Mode Sequence on Sample Composition

Although our previous analysis did not show any experimental group to perform significantly better than the others in terms of overall response, survey response may differ dependent on the age of the target persons. This is because target persons of different ages may be attracted differently by certain combinations of prepaid monetary incentives and mode sequence as outlined above.

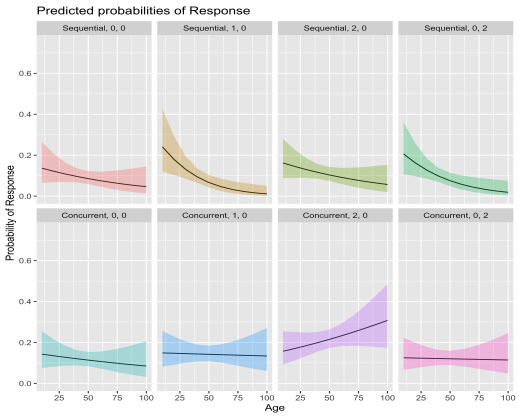
Figure 2.2 displays the predicted probabilities for survey response after the initial contact, after the second contact, and for overall survey response by the experimental groups and the age of the target persons.[2] For the first contact, the

[2]The results of these models can be found in the Appendix, Table A2.4.

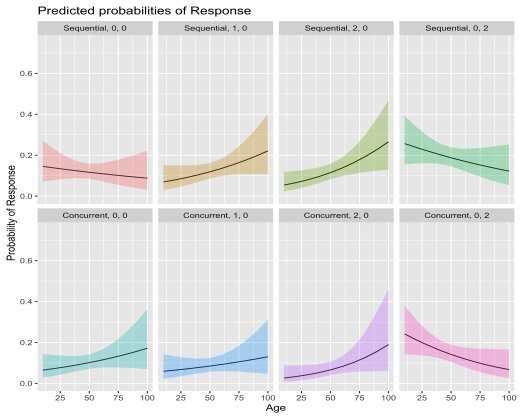
sequential design was successful in motivating younger target persons, as all four graphs on the top share a negative gradient (first panel of Figure 2.2). In contrast, survey response after the initial contact was far less affected by the target persons' age in the concurrent designs. However, the 2 € prepaid incentive group poses a remarkable exception since here we can see a strong positive gradient. This means that older target persons were particularly inclined to initially respond when they received a slightly more valuable prepaid incentive as well as a paper-based questionnaire already with the initial contact.

The predicted probabilities of survey response after the second contact (second panel in Figure 2.2) also offer some interesting insights into the combined effects of prepaid and delayed monetary incentives and mode sequence on survey response conditional on age. Although one would expect the sequential designs to now show a positive gradient (as older target persons may have welcomed the paper-based questionnaire more warmly as opposed to younger target persons), this was only the case in two out of four experimental groups. Interestingly, these were the groups in which the target persons received a prepaid incentive with the initial contact. In contrast, the delayed incentive rather attracted younger target persons, irrespective of mode sequence.

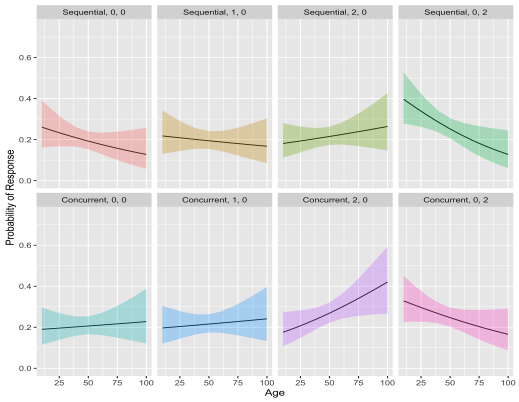
These differences in survey response by contact and age yielded in two designs emerging superior to the others for target persons of different ages (third panel in Figure 2.2): The first one was the sequential design with a delayed incentive since this combination yielded the highest response rate among younger target persons. The second one was the concurrent design with the 2€ prepaid incentive which realized a particularly high response rate among older target persons.



(a) First Contact



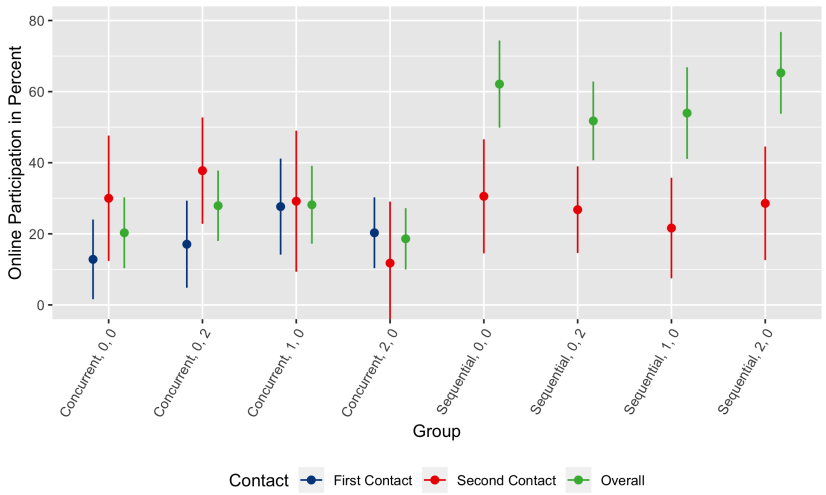
(b) Second Contact



(c) Overall

Figure 2.2: Predicted Probabilities of Survey Response by Age

As one of the sequential and one of the concurrent designs emerged superior for target persons from different age groups, we now have a closer look on respondents' mode choice. Figure 2.3 shows that the sequential designs yielded a remarkably higher overall online participation rate (51.8% to 65.3%) compared to the concurrent designs (18.6% to 28.2%). This resonates well with our previous findings that younger target persons were more likely to respond in the sequential design. As expected, respondents in the online mode were younger compared to respondents in the mail mode (mean age = 38.6 compared to 53.2 years for mail response).



Note. Figure displays a 95% confidence interval.

The online participation rate for the first contact of the sequential design was omitted because it is 100 percent since the mail mode was only offered in the second contact.

Figure 2.3: Online Participation by Mode Sequence

We also explored the device choice for our web respondents. The analysis revealed no significant differences between the sequential and the concurrent design, with smartphone and tablet each ranged around 10% (see Appendix Figure A2.2). These results further emphasize the importance of designing the questionnaire in a feasible way for all kinds of devices to motivate respondents to complete the questionnaire after starting on their device of choice. However, the results also suggest that desktops and laptops are still the predominant devices when people take part in an online survey.

2.3.3 Survey Costs

The survey costs per complete are displayed in Table 2.2. For the initial contact, not surprisingly, fixed costs were considerably lower for the sequential design groups and for experimental groups without prepaid incentives. For instance, costs per unit for material, printing, and processing were only 0.40€ in the sequential design but 1.08€ in the concurrent design. This is mainly due to the absence of costs for printing the questionnaire (0.48€ per unit) but also to cost-

savings for processing (folding), postage (since smaller envelopes were used for letters without a paper questionnaire) and because no envelopes for returning the questionnaire were needed in the sequential design. Costs related to the incentives do not simply equal their monetary value but also include costs for procuring and affixing the coins on the invitation letter as well as costs for additional efforts for folding (in total, 0.25€ per unit). In the sequential design groups, no variable costs incurred since these only comprise costs for the return postage (1.55€ per unit) and efforts for data entry for each returned mail questionnaire. For data entry, we calculated with 2.50€ per unit which roughly corresponds to a quarter of the hourly wage of a student research assistant, assuming that data entry for each mail questionnaire took approximately 15 minutes. When calculating overall survey costs by multiplying costs per case with the group-specific gross sample (i.e., 372 in the sequential groups and 373 in the concurrent groups, respectively), we found overall survey costs in the first contact to be lowest in the two sequential design groups without prepaid incentives while the concurrent design with the 2€ prepaid incentive yielded roughly fourfold higher costs.

Table 2.2: Survey Costs by Experimental Group and Age

	Con, 0, 0	**Con, 1, 0**	**Con, 2, 0**	**Con, 0, 2**	**Seq, 0, 0**	**Seq, 1, 0**	**Seq, 2, 0**	**Seq, 0, 2**
Fixed Costs: 1st contact								
Material	403.96	403.96	403.96	403.96	148.06	148.06	148.06	148.06
Incentives	0	466.62	839.62	0	0	465.37	837.37	0
Postage	623.28	623.28	623.28	623.28	310.25	368.65	368.65	310.25
Variable Costs: 1st contact								
Return postage	52.70	52.70	85.25	52.70	0	0	0	0
Data entry	85.00	85.00	137.50	85.00	0	0	0	0
Total Costs: 1st contact	**1164.94**	**1631.56**	**2089.61**	**1164.94**	**458.31**	**982.08**	**1354.08**	**458.31**
Fixed Costs: 2nd contact								
Material	318.09	303.09	273.11	320.23	325.58	319.16	322.37	311.66
Incentives	0	0	0	673.05	0	0	0	655.04
Postage	496.29	472.89	426.11	499.63	507.98	497.96	502.97	486.26
Variable Costs: 2nd contact								
Return postage	32.55	26.35	23.25	43.40	38.75	44.95	38.75	63.55
Data entry	52.50	42.50	37.50	70.00	62.50	72.50	62.50	102.50
Total Costs: 2nd contact	**899.43**	**844.83**	**759.97**	**1606.31**	**934.81**	**934.57**	**926.59**	**1619.01**
Overall Costs	**2064.37**	**2476.39**	**2849.58**	**2771.25**	**1393.12**	**1916.65**	**2280.67**	**2077.32**
n (respondents)	69	71	86	86	66	63	72	85
Costs per Complete (full sample)	**29.92**	**34.88**	**33.13**	**32.22**	**21.11**	**30.42**	**31.68**	**24.44**
Costs per Complete (under 50 yrs.)	**30.99**	**37.18**	**41.10**	**30.44**	**20.57**	**28.02**	**32.06**	**20.46**
Costs per Complete (50 yrs. and above)	**28.75**	**32.64**	**27.29**	**33.99**	**21.79**	**33.07**	**31.24**	**31.34**

For the second contact (rows 7 to 12 in Table 2.2), the fixed costs per unit in each group were multiplied with the gross sample minus the number of initial responses, the number of non-eligible target persons, and the number of explicit refusals. In this contact attempt, the fixed costs were highest for the two groups receiving a delayed incentive but also showed a higher level in the sequential design groups. This finding is the result of an, on average, lower initial response in the sequential designs leading to more cases that had to be contacted again. Moreover, the higher number of mail responses in the second contact also increased variable costs.

The overall survey costs (row 13 in Table 2.2) were lowest for the sequential design with neither a prepaid nor a delayed incentive, and roughly twice for the

concurrent design with the 2€ prepaid incentive. If we finally relate overall survey costs to the overall number of respondents and thus calculate costs per complete (row 13 in Table 2.2), results slightly changed with the concurrent design with a 1€ prepaid incentive now yielded the highest costs.

In the final step, we took the age-related differences in survey response by experimental group into account by calculating costs per complete for target persons aged below 50 and aged 50 and above. The last two rows in Table 2.2 show that group-specific costs per complete differed remarkably from costs per complete for the full sample. While the concurrent design with the 2€ prepaid incentive in the first contact was one of the most expensive designs for the entire sample, for older target persons it was one of the cheapest (27.29€). This is important since this group performed especially well in bringing older target persons into the sample. For younger target persons, the sequential design with the delayed incentive even turned out to be the cheapest (20.46€). Here, costs per complete were less than half compared to the most expensive design in this age group (the concurrent design with a prepaid incentive in the first contact).

Overall, our results clearly hint at promising targeting strategies for self-administered mixed-mode surveys. Thus, in a model calculation, we estimated overall survey response and overall survey costs if we had pursued such a strategy by implementing a concurrent design with a 2€ prepaid incentive for target persons aged 50 and above and a sequential design with a 2€ delayed incentive for target persons aged below 50 years. The results revealed that the response rate would have increased by almost 42 percent (actual response rate: 22.4 percent vs. predicted response rate: 31.8 percent), while overall survey costs would have only slightly increased by 11 percent.

2.4 Discussion

In this paper, we analysed the combined effects of mode sequence and prepaid monetary incentives offered in the initial and second contact on survey response, sample composition, and survey costs. With regard to the isolated effects of the

survey design elements on survey response, we were able to replicate empirical evidence, suggesting that even small prepaid monetary incentives increase response rates (Edwards et al., 2005; Mercer et al., 2015; Pforr et al., 2015). However, given that survey response in the 1€ condition hardly differed from the condition with no incentive, our study also suggests that incentives should exceed a minimum value in order to yield the desired effects on response rates.

Additionally, our study adds new evidence on the effectiveness of delayed, prepaid incentives in self-administered mixed-mode surveys. The 2€ offered in the second contact significantly improved survey response. With regard to mode sequence, we did not find any difference in overall response rates between both conditions. However, the concurrent condition led to a higher response rate in the first contact, while the sequential condition produced a higher response rate in the second contact. This indicates that for survey managers it might be worthwhile to relocate more resources to subsequent contacts after having collected responses from those target persons who do not need additional extrinsic motivational cues for their participation.

We set out to derive best practice recommendations on how to design a self-administered mixed-mode survey in a way that maximizes survey response, minimizes survey costs, and realizes a balanced sample composition. In our case, we were able to identify two optimal designs. The first one was a concurrent mixed-mode design with a prepaid incentive provided with the first contact while the second one was a sequential design with a delayed incentive. These two combinations of mode sequence and prepaid incentives can be regarded as optimal since they proved to be particularly cost-effective in two different age groups. These age-related differences in survey response are likely the result of varying combinations of design elements attracting different age cohorts.

In following our recommendations, researchers can expect a high share of online participants, especially among younger target persons. Therefore, we strongly suggest to use a responsive questionnaire design to allow respondents to easily answer the questionnaire on their device of choice and to avoid drop outs. How-

ever, if researchers aim to further increase the share of online participants by completely relying on a sequential mixed-mode design, they have to be aware of the potential exclusion of older target persons.

Although these recommendations for targeting in self-administered mixed-mode surveys appear promising, our study has limitations that offer avenues for future research. First, our survey was conducted in Germany and only comprises urban citizens. Hence, it is unclear whether our results are applicable to the residential population of Germany as a whole and generalize even further. Second, we varied mode sequence and incentive strategies, but there are numerous additional ways of varying the same or other design elements (e.g., higher prepaid incentives, more contact attempts, or different intervals between the contacts). Third, our community survey was framed in terms of local issues like the perceived quality of life in Mannheim. It seems interesting to replicate our findings in different settings with different survey topics (e.g., election studies, general population surveys, family research) and different target populations (e.g., older people, students). Fourth and foremost, our recommended strategies for targeting are only applicable if survey managers have information from their sampling frame regarding their target persons' age. Since self-administered mixed-mode surveys are likely to increase in importance in the survey landscape (e.g., Luijkx et al., 2020; Wolf et al., 2021), our results may stimulate further research that specifically addresses the question on how to design these surveys in a way to optimize various survey outcomes. In this sense, we hope to encourage researchers to use the information from their sampling frame to further elaborate targeting strategies for self-administered mixed-mode designs.

References

Bandilla, W., Couper, M. P., & Kaczmirek, L. (2014). The effectiveness of mailed invitations for web surveys and the representativeness of mixed-mode versus internet-only samples. *Survey Practice*, *7*, 1–9.

Becker, R., & Glauser, D. (2018). Are prepaid monetary incentives sufficient for reducing panel attrition and optimizing the response rate? An experiment in the context of a multi-wave panel with a sequential mixed-mode design. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, *139*, 74–95.

Blom, A. G., Gathmann, C., & Krieger, U. (2015). Setting up an online panel representative of the general population. *Field Methods*, *27*, 391–408.

Cornesse, C., & Bosnjak, M. (2018). Is there an association between survey characteristics and representativeness? a meta-analysis. *Survey Research Methods*, *12*, 1–13.

DeLeeuw, E. D. (2018). Mixed-mode: Past, present, and future. *Survey research methods*, *12*, 75–89.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys*. Wiley John + Sons.

Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2008). Representative trust and reciprocity: Prevalence and determinants. *Economic Inquiry*, *46*, 84–90.

Edwards, P., Cooper, R., Roberts, I., & Frost, C. (2005). Meta-analysis of randomised trials of monetary incentives and response to mailed questionnaires. *Journal of Epidemiology & Community Health*, *59*, 987–999.

Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review*, *25*, 161–178.

Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly*, *72*, 167–189.

Hox, J., de leeuw, E., & Klausch, T. (2017). Mixed-mode research: Issues in design and analysis. In P. P. Biemer, E. D. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. E. Lyberg, N. C. Tucker, & B. T. West (Eds.), *Total survey error in practice* (pp. 511–530). Wiley & Sons, Inc.

Luijkx, R., Jónsdóttir, G. A., Gummer, T., Stähli, M. E., Frederiksen, M., Ketola, K., Reeskens, T., Brislinger, E., Christmann, P., Gunnarsson, S. Þ., Hjaltason, Á. B., Joye, D., Lomazzi, V., Maineri, A. M., Milbert, P., Ochsner, M., Pollien, A., Sapin, M., Solanes, I., … Wolf, C. (2020). The European Values Study 2017: On the way to the future using mixed-modes. *European Sociological Review*, *37*, 330–346.

Mann, S. L., Lynn, D. J., & Peterson, A. V. (2008). The "downstream" effect of token prepaid cash incentives to parents on their young adult children's survey participation. *Public Opinion Quarterly*, *72*, 487–501.

Mauz, E., von der Lippe, E., Allen, J., Schilling, R., Müters, S., Hoebel, J., Schmich, P., Wetzstein, M., Kamtsiuris, P., & Lange, C. (2018). Mixing modes in a population-based interview survey: Comparison of a sequential and a concurrent mixed-mode design for public health research. *Archives of Public Health*, *76*, 8.

McGonagle, K. A., & Freedman, V. A. (2016). The effects of a delayed incentive on response rates, response mode, data quality, and sample bias in a nationally representative mixed mode study. *Field Methods*, *29*, 221–237.

Medway, R. L., & Fulton, J. (2012). When more gets you less: A meta-analysis of the effect of concurrent web options on mail survey response rates. *Public Opinion Quarterly*, *76*, 733–746.

Mercer, A., Caporaso, A., Cantor, D., & Townsend, R. (2015). How much gets you how much? Monetary incentives and response rates in household surveys. *Public Opinion Quarterly*, *79*, 105–129.

Messer, B. L., & Dillman, D. A. (2011). Surveying the General Public over the Internet Using Address-Based Sampling and Mail Contact Procedures. *Public Opinion Quarterly*, *75*, 429–457.

Millar, M. M., & Dillman, D. A. (2011). Improving response to web and mixed-mode surveys. *Public Opinion Quarterly*, *75*, 249–269.

Olson, K., Smyth, J. D., & Wood, H. M. (2012). Does giving people their preferred survey mode actually increase survey participation rates? an experimental examination. *Public Opinion Quarterly*, *76*, 611–635.

Olson, K. (2020). Unpacking the black box of survey costs. *Research in Social and Administrative Pharmacy*, *17*, 1342–1346.

Petrolia, D. R., & Bhattacharjee, S. (2009). Revisiting incentive effects. *Public Opinion Quarterly*, *73*, 537–550.

Pforr, K., Blohm, M., Blom, A. G., Erdel, B., Felderer, B., Fräßdorf, M., Hajek, K., Helmschrott, S., Kleinert, C., Koch, A., Krieger, U., Kroh, M., Martin, S., Saßenroth, D., Schmiedeberg, C., Trüdinger, E.-M., & Rammstedt, B. (2015). Are Incentive Effects on Response Rates and Nonresponse Bias in Large-scale, Face-to-face Surveys Generalizable to Germany? Evidence from Ten Experiments. *Public Opinion Quarterly*, *79*, 740–768.

Roßmann, J., Gummer, T., & Kaczmirek, L. (2020). Working with user agent strings in stata: The parseuas command. *Journal of Statistical Software*, *92*, 1–16.

Scherpenzeel, A., & Das, J. (2010). 'true' longitudinal and probability-based internet panels: Evidence from the netherlands. In J. Das, P. Ester, & L. Kaczmirek (Eds.), *Social and behavioral research and the internet* (pp. 77–103). Taylor & Francis.

Singer, E., Van Hoewyk, J., Gebler, N., & McGonagle, K. (1999). The effect of incentives on response rates in interviewer-mediated surveys. *Journal of Official Statistics*, *15*, 217–230.

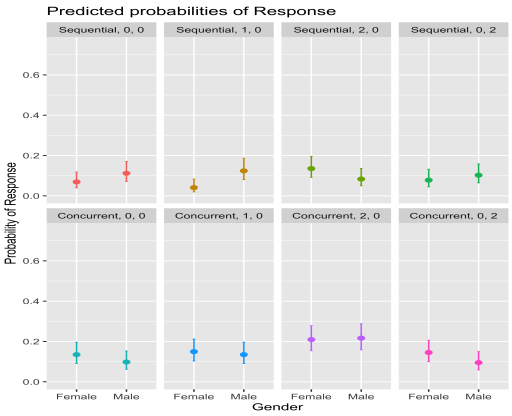
Sun, H., Newsome, J., McNulty, J., Levin, K., Langetieg, P., Schafer, B., & Guyton, J. (2020). What works, what doesn't? three studies designed to improve survey response. *Field Methods*, *32*, 235–252.

Sutter, M., & Kocher, M. G. (2007). Trust and trustworthiness across different age groups. *Games and Economic Behavior*, *59*, 364–382.

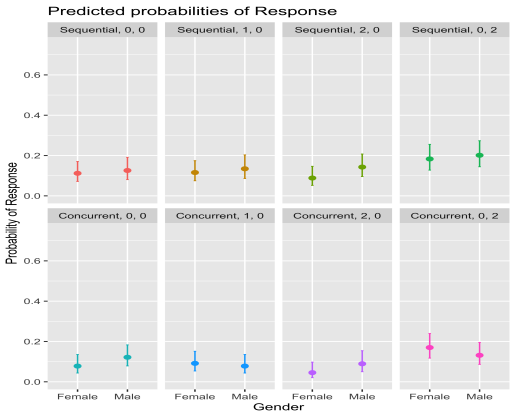
Tourangeau, R. (2017). Mixing modes. In P. P. Biemer, E. D. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. E. Lyberg, N. C. Tucker, & B. T. West (Eds.), *Total survey error in practice* (pp. 115–132). Wiley & Sons, Inc.

Wolf, C., Christmann, P., Gummer, T., Schnaudt, C., & Verhoeven, S. (2021). Conducting general social surveys as self-administered mixed-mode surveys. *Public Opinion Quaterly*, *85*, 623–648.

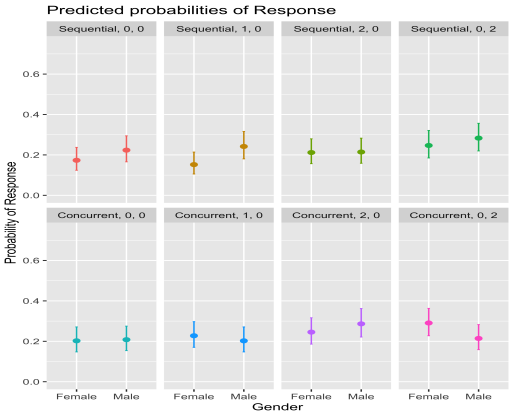
Appendix



(a) First Contact

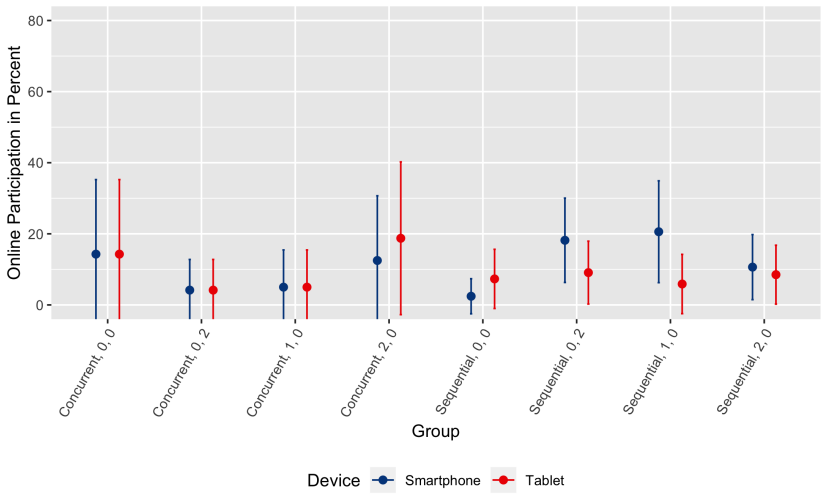


(b) Second Contact



(c) Overall

Figure A2.1: Predicted Probabilities of Survey Response by Gender



Note. Figure displays a 95% confidence interval.

Figure A2.2: Device Choice in the Online Mode

Table A2.1: Interactions between Experimental Manipulations and Gender

	First Contact	Second Contact	Overall
(Intercept)	-2.60^{***}	-2.07^{***}	-1.56^{***}
	(0.30)	(0.25)	(0.20)
Sequential, 1, 0	-0.56	0.04	-0.16
	(0.49)	(0.35)	(0.29)
Sequential, 2, 0	0.74^{*}	-0.26	0.25
	(0.37)	(0.38)	(0.27)
Sequential, 0, 2	0.13	0.58	0.45
	(0.42)	(0.33)	(0.27)
Concurrent, 0, 0	0.74	-0.40	0.19
	(0.38)	(0.40)	(0.28)
Concurrent, 1, 0	0.86^{*}	-0.22	0.34
	(0.37)	(0.38)	(0.27)
Concurrent, 2, 0	1.27^{***}	-0.97^{*}	0.44
	(0.35)	(0.49)	(0.27)
Concurrent, 0, 2	0.82^{*}	0.49	0.67^{*}
	(0.37)	(0.33)	(0.26)
Gender: Female	0.52	0.13	0.32
	(0.39)	(0.36)	(0.28)
Sequential, 1, 0*Gender: Female	0.68	0.03	0.26
	(0.60)	(0.50)	(0.40)
Sequential, 2, 0*Gender: Female	-1.07^{*}	0.41	-0.30
	(0.53)	(0.51)	(0.38)
Sequential, 0, 2*Gender: Female	-0.22	-0.02	-0.13
	(0.55)	(0.46)	(0.38)
Concurrent, 0, 0*Gender: Female	-0.88	0.36	-0.28
	(0.52)	(0.53)	(0.39)
Concurrent, 1, 0*Gender: Female	-0.64	-0.31	-0.47
	(0.50)	(0.56)	(0.38)
Concurrent, 2, 0*Gender: Female	-0.48	0.59	-0.11
	(0.48)	(0.63)	(0.37)
Concurrent, 0, 2*Gender: Female	-1.00	-0.44	-0.72
	(0.52)	(0.48)	(0.37)
AIC	1924.32	1707.79	2838.55
BIC	2018.42	1799.84	2932.64
Log Likelihood	-946.16	-837.90	-1403.28
Deviance	1892.32	1675.79	2806.55
Num. obs.	2646	2328	2646

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table A2.2: Isolated Effects of Experimental Manipulations

	First Contact	Second Contact	Overall
(Intercept)	-2.45^{***}	-2.01^{***}	-1.43^{***}
	(0.12)	(0.10)	(0.11)
Concurrent	0.54^{***}	-0.35^{**}	0.10
	(0.12)	(0.13)	(0.09)
1 € First Contact	0.08		0.02
	(0.15)		(0.14)
2 € First Contact	0.50^{***}		0.22
	(0.14)		(0.13)
2 € Second Contact		0.60^{***}	0.33^{*}
		(0.14)	(0.13)
AIC	1919.38	1691.43	2828.27
BIC	1942.90	1708.69	2857.68
Log Likelihood	-955.69	-842.72	-1409.14
Deviance	1911.38	1685.43	2818.27
Num. obs.	2646	2328	2646

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table A2.3: Effects of Experimental Manipulations on Response Rates

	First Contact	Second Contact	Overall
(Intercept)	-2.32^{***}	-2.01^{***}	-1.40^{***}
	(0.19)	(0.18)	(0.14)
Sequential, 1, 0	-0.12	0.05	-0.02
	(0.28)	(0.25)	(0.20)
Sequential, 2, 0	0.22	-0.02	0.09
	(0.26)	(0.25)	(0.19)
Sequential, 0, 2	0.01	0.57^{*}	0.38^{*}
	(0.27)	(0.23)	(0.19)
Concurrent, 0, 0	0.29	-0.18	0.05
	(0.26)	(0.26)	(0.19)
Concurrent, 1, 0	0.52^{*}	-0.37	0.11
	(0.25)	(0.28)	(0.19)
Concurrent, 2, 0	1.01^{***}	-0.63^{*}	0.38^{*}
	(0.23)	(0.31)	(0.19)
Concurrent, 0, 2	0.33	0.28	0.32
	(0.25)	(0.24)	(0.19)
AIC	1924.19	1699.13	2832.24
BIC	1971.24	1745.16	2879.29
Log Likelihood	-954.10	-841.57	-1408.12
Deviance	1908.19	1683.13	2816.24
Num. obs.	2646	2328	2646

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table A2.4: Interactions between Experimental Manipulations and Age

	First Contact	Second Contact	Overall
(Intercept)	-2.34^{***}	-2.01^{***}	-1.41^{***}
	(0.20)	(0.18)	(0.14)
Sequential, 1, 0	-0.22	-0.01	-0.00
	(0.30)	(0.26)	(0.20)
Sequential, 2, 0	0.21	-0.06	0.11
	(0.27)	(0.26)	(0.19)
Sequential, 0, 2	-0.10	0.57^{*}	0.36
	(0.30)	(0.23)	(0.19)
Concurrent, 0, 0	0.30	-0.19	0.06
	(0.26)	(0.26)	(0.19)
Concurrent, 1, 0	0.55^{*}	-0.38	0.12
	(0.25)	(0.28)	(0.19)
Concurrent, 2, 0	1.03^{***}	-0.68^{*}	0.38^{*}
	(0.24)	(0.32)	(0.19)
Concurrent, 0, 2	0.35	0.24	0.32
	(0.26)	(0.24)	(0.19)
Age	-0.25	-0.12	-0.19
	(0.21)	(0.19)	(0.15)
Sequential, 1, 0*Age	-0.46	0.41	0.12
	(0.32)	(0.26)	(0.20)
Sequential, 2, 0*Age	0.00	0.52^{*}	0.29
	(0.28)	(0.26)	(0.20)
Sequential, 0, 2*Age	-0.30	-0.07	-0.13
	(0.31)	(0.24)	(0.20)
Concurrent, 0, 0*Age	0.13	0.36	0.24
	(0.27)	(0.27)	(0.20)
Concurrent, 1, 0*Age	0.23	0.31	0.24
	(0.26)	(0.28)	(0.20)
Concurrent, 2, 0*Age	0.44	0.59	0.45^{*}
	(0.25)	(0.32)	(0.20)
Concurrent, 0, 2*Age	0.23	-0.20	-0.01
	(0.27)	(0.25)	(0.19)
AIC	1917.27	1695.67	2832.59
BIC	2011.36	1787.72	2926.68
Log Likelihood	-942.63	-831.84	-1400.30
Deviance	1885.27	1663.67	2800.59
Num. obs.	2646	2328	2646

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

3 Effects of smartphone use and recall aids on network name generator questions

Published in:
Beuthner, C., Silber, H., Stark, T. H. (2020). Effects of smartphone use and recall aids on network name generator questions. *Social Networks*, 69, 45-54.

Abstract

The increasing use of smartphones around the world provides new opportunities for network data collection with smartphone surveys. We investigated experimentally whether the use of smartphones and of a recall aid affects the number of reported names in a network name generator question. In a German online access panel (N = 3,327), respondents were randomly assigned to answer the survey on their PC or on their smartphone and were randomly assigned to receive an open-ended recall aid question before the name generator question or after. Results showed that respondents on PCs and smartphones reported the same number of network contacts. This suggests that smartphone surveys have no negative effect on the network sizes in ego-centered network studies. However, requiring people to answer on smartphones resulted in a selection bias due to non-compliance, which may have led to an overrepresentation of persons with larger network sizes. The recall aid question did not lead to more reported names, but it proved to be an indicator of respondents' motivation and response quality. In sum, the study suggests that smartphones can effectively be used for network research in tech-savvy populations or when respondents can choose to complete the survey on another device.

Keywords

ego-centered social networks, web surveys, smartphones, experiment, response quality, recall aid

3.1 Introduction

Past research has documented important survey design effects on the size of the network elicited from name generator questions in ego-centered network studies. Such name generator questions ask respondents to self-report the names of people in their personal network (Burt, 1984; Marsden, 2011). With respect to survey mode, there is mixed-evidence whether switching from traditional face-to-face interviews to online surveys reduces (Matzat & Snijders, 2010) or increases the network size (Fischer & Bayham, 2019), or has no effect on the number of names reported (Vriens & van Ingen, 2018). Within online surveys, the placement of a name generator question (Yousefi-Nooraie et al., 2019), the number of name boxes provided (Vehovar et al., 2008), and the use of name recall aids (Hsieh, 2015) can considerably affect network sizes. Regarding repeated measurement, Silber et al. (2019) showed that panel conditioning had only minor effects on the size of social networks in a German web survey.

Interestingly, the aforementioned study found that the network size of respondents who answered the questionnaire on their smartphones was slightly higher (3.48 friends) than that of respondents who answered on their PCs (3.32 friends, Silber et al. (2019)).[3] While the difference between the devices was statistically non-significant, this result is an encouraging signal for researchers who consider administering their entire survey on smartphones. Perhaps the use of smartphones has a positive effect on the size of ego-centered networks, despite the smaller display and keyboard of smartphones. Such a conclusion, however, cannot be drawn with confidence from that study because respondents were not randomly assigned to a device. Thus, there is no way of telling whether the results emerged because respondents who answer on smartphones have actually slightly more friends than respondents who answer on PCs or because respondents are motivated to report slightly more friends on smartphones. It could even be that respondents who answer on smartphones have more friends but tend to underreport their network size due to the device they use for participation.

[3] PC includes desktop computers as well as laptops, running windows, macOS, Linux, and other operating systems.

Previous methodological research showed that certain groups of respondents are more likely than others to use a smartphone to participate in an online survey when they can freely choose to do so. Specifically, younger (de Bruijne & Wijnant, 2013; Toepoel, 2017) and higher educated respondents (Fuchs & Busse, 2009) were overrepresented among smartphone participants. One study found women to be more likely than men to participate in smartphone surveys (de Bruijne & Wijnant, 2014) but other research could not replicate this finding (de Bruijne & Wijnant, 2013; Keusch & Yan, 2017).

Selective participation of certain groups in smartphone surveys may bias results of network studies if these groups differ in their sociability and thus have smaller or larger networks. For instance, women have been found to have larger social networks than men (Goodreau et al., 2009; Lewis & Kaufman, 2018; McLaughlin et al., 2010) and employed people have more social contacts than unemployed (Edin et al., 2003; Munshi, 2003). Likewise, people with a good health seem to have larger networks than those suffering from medical conditions (Michael et al., 1999; Schaefer et al., 1981). Moreover, some studies found that members of voluntary associations have larger social networks than people who are not members of such associations (Farkas & Lindberg, 2015; Putnam, 2000; Rotolo, 2000; but see Bekkers et al., 2008 who found no effect). Additional research has shown that inhabitants of urban areas are less likely to participate in voluntary association and show less community engagement (Oliver, 2000; Remmer, 2010). Hence, it can be assumed that inhabitants of urban areas are likely to have smaller social networks. If members of such groups are more or less likely to participate in network studies conducted on smartphones, the conclusions about the average network size in the population may be biased.

So far, little is known about the relationship between smartphone use to conduct ego-centered surveys and the elicited network size. To help close this research gap, the present research explored experimentally how the use of smartphones to answer online surveys affects the number of names elicited from name generator questions. Such insights are important because researchers will be inclined to col-

lect network data on smartphones in the near future given the wide-spread use of smartphones globally (Poushter et al., 2018) and the rapid improvements made in the development of new visual tools to collect ego-centered network data online such as GENSI (Stark & Krosnick, 2017), OpenEddi (Eddens & Fagan, 2018), or Network Canvas (Hogan et al., 2016).

This study reports results of an experimental online study in which respondents were randomly assigned to complete the same web survey either on a PC or a smartphone. Despite the encouraging findings of Silber et al. (2019) in their non-experimental study, we expected to elicit fewer names on smartphones than on PCs when respondents cannot self-select the device on which they answer. This is because, we hypothesized that respondents may be discouraged from thinking about and entering additional names because of smaller screen sizes of smartphones compared to PCs and the accompanying necessity to scroll down to see the entire name generator question. Additional technical issues, such as longer page loading times or the difficulty to click with the finger on answer boxes, may negatively affect the number of names reported on smartphones. Finally, people are much more likely to use their smartphones than their PC in a distracting environment outside of their home, such as while waiting for public transportation or in line at the grocery store (Couper et al., 2017). This may further reduce respondents' motivation to repeatedly enter contacts in a name generator question.

Respondents were also randomly assigned to answer a recall aid question before or after reporting names in a name generator question. A general challenge of ego-centered network studies is that people tend to forget to mention a substantive number of their personal contacts (Brewer, 2000). Previous work found that providing recall aids can increase the number of names reported in ego-centered network studies, both in face-to-face interviews (Marsden, 2011) and in online surveys on PCs (Hsieh, 2015). However, to our knowledge, no research has explored the effect of recall aids in smartphone surveys. We hypothesized that seeing a recall aid before the network prompt would counter the expected negative effect of the added difficulty of answering a smartphone survey because the names of network contacts are already available in the active memory.

3.2 Methods

The data for this study were collected with a web survey conducted between the 15th of July and the 31st of August 2018. Respondents were recruited from a German nonprobability online access panel using quotas for gender, education, age, and federal state. Before receiving the invitation via email, respondents were randomly allocated to either use a PC or a smartphone to complete the survey (device manipulation). All respondents were asked for their gender, age, and education at the start of the survey. After these questions, respondents violating the device assignment were screened-out.

Our two experiments were in the first part of a larger questionnaire and not preceded by any other experimental manipulation. Respondents had the possibility to proceed in the survey without answering a question but could not to go back in the questionnaire to change an answer they had already given. The questionnaire layout was optimized for smartphones and displayed similarly on both devices.

A total of 50,063 panel members were invited, from which 6,750 opened the invitation link. 538 (8.0%) broke off, and 2,838 (42.0%) were screened-out. The majority of those screened-out were respondents assigned to the smartphone condition but who tried to use a PC to complete the survey (2,563). The other 275 respondents were assigned to use a PC but were screened-out for trying to use a smartphone. Given these high numbers of screened-out respondents, we test below for a selection bias by comparing the characteristics of respondents who completed the survey with those who were screened-out using demographics and supplementary information obtained from the panel provider for all invited panel members. The final sample consisted of 3,327 respondents (completion rate 49.3%) of which 1,787 answered using a PC and 1,540 using a smartphone. The median response time for the survey was 29.6 minutes and the median response time for the network name generator question was 2.5 minutes.

Randomized Experiments

In the first experiment, half of the respondents were randomly assigned to use a smartphone for answering the survey and the other half was randomly assigned to use a PC. Both groups received the information about the device usage in the invitation email. In the second experiment, we randomly assigned respondents to one of two orders in which the name generator question and the recall aid question were asked. Half of the sample saw the open recall aid question first and answered then the name generator, and the other half saw first the name generator and then the recall aid question. This placement experiment allowed us to evaluate if the open question served as a memory trigger and helped respondents to recall more friends. The random assignments within the two experiments were independent of each other.

3.2.1 Measures

Name Generator

Number of Friends

The name generator question asked respondents to provide the first names of up to twenty friends: “Now we are interested in your closest circle of friends. Please enter the first names of your close friends here". Respondents could enter names in up to 20 vertically arranged separated answer boxes. A screenshot of this question is provided in the Appendix A3.1. No other instruction on how to complete the name generator was given. To avoid that a few respondents (10.1%) who named an exceptionally high number of friends have a high impact on the results, we truncated the measure and recoded all respondents that named 10 and more friends in a 10 category. Respondents who entered no names were considered to have a network size of 0 after our analysis revealed no significant differences between the PC and smartphone condition, suggesting that these were substantive answers (see Figure 3.1).

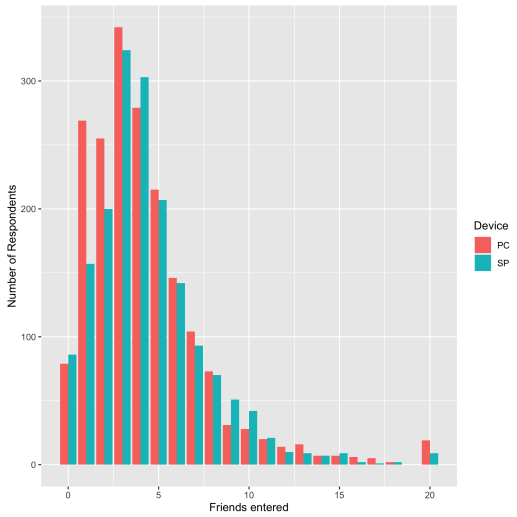


Figure 3.1: Number of Friends entered in the Network Generator Question by Device

Response Time

We calculated two measures for response time. Response time was measured as a relative timestamp in relation to the start of the survey for every question. Therefore, the total response time for the name generator question was calculated by subtracting the timestamp of the name generator question from the time stamp of the subsequent question. To account for response effort, the response time needed per name was calculated as a second indicator by dividing the total response time for the name generator question by the number of friends reported.

Recall Aid

The recall aid asked respondents: "When you think of your friends, what is important to you?". Respondents could type their answer in a single answer box (see Appendix Figure A3.2). This question is different from typically used recall aids that ask respondents to think about friends in certain foci, such as friends from school, from work, or from leisure activities (Belli et al., 2001; Glasner & Van der Vaart, 2009). We worried that explicitly mentioning certain foci might prime respondents to think only about friends from these foci, ignoring anyone else. The idea of our research aid was to activate general retrieval cues that might help respondents access memories of their friends (Yan & Tourangeau, 2008), without

priming them to a certain group of friends.

Quality of Answer

For the analysis, two coders independently coded the answers to the open recall aid question into ten categories dependent on their content. For further analysis, we combined the categories into two categories: (1) “no answer” or “answer not meaningful” and (2) “meaningful answer". The category “answer not meaningful” refers to an answer that is either not related to the question or not meaningful at all (e.g., ‘xxx’, ‘abc’, and ‘no idea’).[2] As an indicator of inter-coder reliability, we calculated Cohens Kappa for the binary coding and found high reliability (textitk = 0.81).

Control Variables

Because gender, age, and education were used for the quotas during the recruitment process, all respondents were asked about these characteristics as the very first questions of the survey, even before the screening by device took place. Skipping those three demographical questions was not possible. Therefore, these variables were also available for all respondents who were later screened out because they failed to use the device to which they were randomly assigned.

Gender

Respondents were asked for their gender. Possible answers were male or female.

Age

Each respondent was asked: “How old are you?”. Respondents had to give an open numeric answer.

[2] A total of 65.4% of respondents named values or emotions such as “trust,” “honesty,” or “happiness.” 18.2% of the answers given referred to support, such as “they help me solving my problems.” A further 10.8% were associated with activities like “going out together” or “meeting them to have fun.” All other categories were represented in less than 5% of the answers; those included: having regular contact with friends (“people I have regular contact with”), answers that were focused on the friends themselves (“they are nice”), answers which refer to common interests (“we have the same hobbies”), or such referring to the amount of friends (“there are only a few real friends”). Finally, non-meaningful answers (e.g., responses such as “.....” or “Xxx”) represented less than 5% of the cases.

Education

Education was asked as a closed-ended question: "What is your highest general school degree?". Respondents could select one of 9 response options representing German school degrees. It was also possible to enter a different school degree in an open format. For our analysis, we recoded education into three categories from low to high, in accordance with the tripartite school system of Germany (9-, 10-, and 12/13-year high school tracks). A screenshot of the question, showing all answer categories can be found in the Appendix A3.

Smartphone Skills

To measure how experienced respondents were with their smartphone, we asked all respondents, "How do you generally assess your ability to use your smartphone?". Respondents could answer on a 5-point rating scale with the end-points labeled as "Beginner" and "Expert" (Keusch & Yan, 2017).

Supplementary Information for Selection Bias Analysis

To test for a selection bias of groups that differ in their sociability, we requested supplementary information about the respondents from the panel provider. Those measures had been collected in a welcome survey, shortly after the registration of each new panel member. The requested indicators had been found to correlate with social network size in previous studies: employment status (Edin et al., 2003; Munshi, 2003), living in an urban or rural area (Oliver, 2000; Remmer, 2010), the number of medical conditions (Michael et al., 1999; Schaefer et al., 1981), and participating in voluntary organizations (Farkas & Lindberg, 2015; Putnam, 2000; Rotolo, 2000), which we tried to measure with the number of team sports respondents exercised.

Employment Status

Being unemployed was coded as 1 and employed as 0.

Urban Area

We divided panel members into two groups: those living in cities with more than

100,000 inhabitants were considered to live in an urban area (coded 1) and those living in cities or towns with less than 100,000 inhabitants were coded to live in a rural area (coded 0).

Number of Medical Conditions
The panel provider supplemented 15 dichotomous variables indicating whether the panel members had reported to suffer from each of 15 different medical conditions such as hearing problems or asthma (response categories: yes, no). These variables were summed to create an additive index ranging from 0 to 15. A complete list of medical conditions can be found in the Appendix B3.1.

Practiced Sports
The panel provider supplemented 25 variables describing whether or not a panel member participated in 25 different types of sports such as soccer, tennis, or volleyball (response categories: yes, no). We selected 18 sports, which are mainly practiced with others or in teams, as participation in such sports may be related to the network size. These variables were summed to create an additive index ranging from 0 to 18. A complete list of all 25 variables, including an indication of which variables were selected, can be found in Appendix B3.2.

Descriptive statistics for all variables are shown in Table 3.1. Bivariate correlations can be found in Table C3.1 in the Appendix.

Table 3.1: Descriptive Overview for all Measures

Variable	Range	Median	Mean	Standard Deviation	Missing Values
Metric					
Age		44	43.8	13.9	0
Smartphone skills	1-5	4	3.6	1.0	364
Number of medical conditions	0-14	1	1.2	1.6	339
Practiced sports	0-18	2	2.8	3.6	1,120
Number of friends	0-20	4	4.1	2.6	0
Ordinal					
		Low Education	Medium Education	High Education	
Education		15.0%	42.5%	42.5%	0
Nominal					
		Male	Female		
Gender		42.8%	57.2%		0
		Unemployed	Employed		
Employment Status		5.7%	94.3%		32
		Urban	Non-Urban		
Urban Area		65.2%	34.7%		157
		Desktop/Laptop	Smartphone		
Device		51.8%	48.2%		0
		Meaningful	Non-meaningful		
Quality of answer to recall aid		93.0%	7.0%		0

3.2.2 Analysis

We use multiple imputation (predictive mean matching with 10 samples) to replace missing values in all regression models (van Buuren & Groothuis-Oudshoorn, 2011). Analyses using listwise deletion instead of multiple imputation lead to very similar results (see Appendix Table C3.1 to Table C3.3). Since more people were screened-out in the smartphone sample than in the PC sample, we test for selection bias by comparing respondents who completed the survey with those who were screened-out separately for both devices. We then examine if the random allocation of respondents to the experimental conditions worked and investigate the influence of the device used to answer the survey on item nonresponse in the form of not entering any names. As a next step, we compare the distribution of the number of friends who were entered between the two devices. We also examine response times to uncover potential differences. Afterward, we focus on the effect of the recall aid experiment and on the combined effect of the device used

and the recall aid on the number of names entered. Finally, we conduct five OLS regression models, to examine possible interactions between the independent variables and to control for demographics, smartphone skills, and the supplementary measures. In these regression models, the number of friends named in the name generator question serves as the dependent variable.

3.3 Results

3.3.1 Selection Bias and Network Size

The results of the multivariate logistic regression model in Table 3.2 show that respondents who were older, male and those who were unemployed were more likely to be screened-out in the smartphone group because they tried to complete the survey on a computer. Also, those who suffered from more medical conditions were screened-out more often. However, we did not find significant differences regarding other factors, namely living in an urban area, the number of sports practiced, or education. In contrast to the smartphone screen-outs, respondents who were female, younger, and those with low or medium education were more likely to be screened-out in the PC group.

These results suggest that smartphone surveys are preferred over PC surveys by female respondents, those who are healthier, and those who are younger. Accordingly, a survey that is only conducted with smartphones could overestimate the average network size, as being female (Goodreau et al., 2009; Lewis & Kaufman, 2018; McLaughlin et al., 2010) and being healthy (Michael et al., 1999; Schaefer et al., 1981) are both associated with larger social networks. At the same time, allowing only the usage of PC's could lead to an underestimation of the network size, as men have been shown to have smaller social networks than women (Goodreau et al., 2009; Lewis & Kaufman, 2018; McLaughlin et al., 2010).

Table 3.2: Logistic Regression Analyses Predicting Being Screened-out or not in the Smartphone Condition and the PC Condition

	Screen-out smartphone	Screen-out PC
Gender female[a]	-0.48***	0.30*
	(0.06)	(0.14)
Low education[b]	-0. 16	0. 54*
	(0.10)	(0.23)
Medium education[b]	-0. 13	0. 60***
	(0.07)	(0.15)
Age	0.03***	-0.04***
	(0.00)	(0.01)
Employment status: unemployed[c]	-0.32*	0.07
	(0.14)	(0.29)
Urban area[d]	0.11	0.18
	(0.07)	(0.14)
Medical conditions	0.05*	0.00
	(0.02)	(0.03)
Practiced sports	0.02	0.04
	(0.02)	(0.03)
N	4,494	2,309
Pseudo-R^2 (McFadden)	0.34	0.34

Note: Logistic regression model with unstandardized coefficients and standard errors in parentheses.
*$p < .05$; **$p < .01$; ***$p < .001$.
[a] Reference category is male.
[b] Reference category is high education.
[c] Reference category is employed.
[d] Reference category is Non-Urban Area.

3.3.2 Randomization

To draw firm conclusions about the effects of the experimental manipulations, it is important that respondents were randomly allocated into experimental conditions. To check that the assignment worked, we used logistic regressions to predict the experimental group, separately for device and recall aid question placement for those who were not screened-out. The results in Table 3.3 show systematic differences between respondents who participated using a smartphone and those using

a PC. As can be expected based on the selection bias analyses, smartphone respondents were more likely to be female, younger, had a higher ability to use their smartphone, were less likely to have a medium education level, and less likely to suffer from a medical condition. It is thus crucial to keep these group differences in mind when interpreting the results of the device experiment. However, the assignment worked well for the recall aid question placement experiment as shown in Column 2 of Table 3.3. Specifically, there was no significant difference between respondents completing the survey under both placement conditions.

Table 3.3: Logistic Regression Analysis Testing Randomization to Device Assignment and Recall Aid Placement

	Assignment to Mobile	Question Placement Recall Aid first
Gender female[a]	0.39***	-0.08
	(0.07)	(0.07)
Age	-0.02***	0.00
	(0.00)	(0.00)
Education low	0.07	0.07
	(0.11)	(0.10)
Education medium	-0.26***	0.04
	(0.08)	(0.07)
Smartphone skills	0.04***	0.00
	(0.04)	(0.04)
Employment status: unemployed[b]	0.19	0.08
	(0.15)	(0.14)
Urban area[c]	-0.08	0.03
	(0.07)	(0.02)
Medical Condition	-0.04*	-0.03
	(0.02)	(0.02)
Practiced Sports	-0.01	0.02
	(0.02)	(0.01)

Note: Logistic regression model with unstandardized coefficients and standard errors in parentheses.
*$p < .05$; **$p < .01$; ***$p < .001$.
[a] Reference category is male.
[b] Reference category is high education.
[c] Reference category is employed.
[d] Reference category is non-urban area.

3.3.3 Item Nonresponse and Response Time

To evaluate how to treat respondents who did not enter a single name into the name generator, we tested for a potential item nonresponse bias caused by the device.

Table 3.4 shows that 166 respondents did not enter any friend. They were nearly equally distributed between PC respondents (n = 80, 0.02%), and smartphone respondents (n = 86, 0.02%). A Chi-squared test (χ^2 = 1.141, p = 0.285) did not reveal any relationship between the device and item nonresponse. This means that there is no device-related item nonresponse bias originating from the name generator. This can be seen as first evidence for smartphones being an equally feasible device to generate ego-centered networks – at least for the respondents who complied with the device assignment. Completing a name generator with a smartphone did not lead to increased nonresponse in the name generator.

Table 3.4: Amount of Item Nonresponse to the Name Generator Question

Device	Number of Item Nonresponse	Proportion of nonresponse	n total
PC	80	0.04	1,931
Smartphone	86	0.05	1,747
Total	166	0.04	3,681

Note. ChiSq-test: relationship item nonresponse and device: (^{2}= 1.134, *df* = 1, *p*= 0.29).

Besides the number of friends, the time a respondent needed to answer also poses an important element to evaluate the viability of smartphones for conducting ego-centered name generators. The median for answering the name generator was 9.0 seconds in total or 3.0 seconds per name across all respondents. A smartphone respondent needed on average 12.0 seconds in total or 3.5 seconds per name, compared to a PC respondent who needed on average 7.0 seconds in total or 2.7 seconds per name. A Mann-Whitney-U test revealed a significant difference for the overall time (W = 1,373,700; $p < 0.001$) and the time per name (W = 1,448,100; $p < 0.001$), indicating that answering the network name generator question on a smartphone was more time consuming. The greater amount of response time needed for smartphone respondents could have led to a greater response burden and thus potentially fewer reported names in the network name generator question.

3.3.4 Device Experiment: Number of Reported Names

Our next research question addresses whether the device affects the network size. Figure 3.1 shows the distributions of the number of friends entered for smartphones and PCs. Both distributions show a peak for respondents who entered 3 friends. The majority of respondents in both groups entered 1 to 5 friends. Figure 3.1 suggests that PC respondents entered 1 or 2 friends more often, while smartphone respondents were more likely to enter higher numbers of friends. However, a Kolmogorov-Smirnov test did not reveal a statistically significant difference between both distributions ($D = 0.15$; $p = 0.978$). These results suggest that smartphones are equally viable as PCs for collecting social network data – at least for groups that are willing to participate with a smartphone.

While the majority of respondents entered 10 or fewer friends, only a small group entered the maximum of 20 friends (0.9%). This group was two times larger in the PC condition than in the smartphone condition. When conducting the analysis without truncating at 10 names, we found that smartphone respondents entered on average 4.40 friends, which was not significantly more than the 4.25 friends' respondents entered when using a PC ($t = -1.49$; $p = 0.14$).[4] When truncating at 10, to avoid giving the few outliers a large impact on the results, smartphone respondents entered on average 4.25 names, while PC respondents entered significantly fewer (4.02) names on average ($t = -2.67$; $p = 0.01$).

In sum, smartphone respondents entered equally many or more friends than PC respondents. This result emerged despite the longer response time per name entered of smartphone respondents, indicating that the added difficulty of answering a smartphone survey did not have a negative impact on the number of friends that respondents reported. These preliminary results suggest that smartphones are indeed a viable option to use for ego-centered network name generators – at least

[4] Additionally, we ran a post-hoc power analysis using the distribution of our sample and assuming an alpha of 0.05 and a power level of 0.95. Under these assumptions the mean difference between both samples would need to be at least 0.12 names on average. To find an effect with an alpha of 0.01 and a power of 0.99 the mean difference would need to be at least 0.14 names.

for respondents who are willing to use a smartphone to fill in a survey.

3.3.5 Recall Aid Experiment: Number of Reported Friends and Quality of Answer

Our second research question focuses on the use of a recall aid, to help respondents complete the name generator question. Comparing respondents who saw the recall aid before completing the name generator to those who did not, did not reveal a significant difference in the number of friends entered (t = -0.73; p = 0.47). This means that there was no statistical evidence for the recall aid helping respondents to report a larger social network.

Since it is likely that the recall aid has no effect on respondents who did not seriously consider the recall aid question, we conducted a t-test to compare only those respondents who gave a meaningful answer and saw the recall aid first, to respondents who saw the name generator first. As can be seen in Table 3.5, respondents who saw the recall aid first and gave a meaningful answer entered on average more friends (4.41) than those who did not see the recall aid before answering the name generator question (4.10). This difference was statistically significant (t = 3.55; $p < 0.001$). This could be seen as evidence for the recall aid helping respondents who gave a serious answer to think of more friends when answering the network name generator question.

Table 3.5: Effect of the Quality of the Recall Aid Answers on the Mean Number of Friends Reported

Quality of answers to recall aid	Recall aid seen first	Recall aid seen second
Only non-meaningful answers	2.05 (142)	2.11 (109)
Only meaningful answers	4.41 (1,651)	4.30 (1,690)
All answers	4.16 (1,793)	4.10 (1,799)

Note. Mean values and total numbers in parentheses.
N = 3,592.
Number of friends entered was truncated at 10.

However, our fully crossed experimental design allows us to also exclude respondents who gave non-meaningful answers from the group who saw the recall aid

after completing the name generator. Interestingly, when comparing only respondents who gave at least one meaningful answer to the recall aid question and saw the recall aid question first to those respondents who gave at least one meaningful answer to the recall aid question and saw the name generator question first, the previously found significant difference vanishes (t = -1.27; p = 0.20). This suggests that the difference was not caused by the recall aid functioning as a memory trigger. Instead, respondents who gave meaningful answers to the recall aid question reported larger network sizes, independently of whether they first answered to the recall aid or first answered the name generator question. This means that the recall aid did not help respondents to remember their friends, but it functioned instead as a general indicator of response quality.

3.3.6 Combining Both Experiments: Number of Reported Friends and Quality of Answers

Table 3.6 present the combined analysis of both experiments. There was no significant difference in the number of names reported between respondents who completed the name generator before the recall aid. Independently of whether a meaningful or non-meaningful answer was given to the recall aid, respondents on PCs and smartphones reported similar numbers of names (see the first three rows in Table 3.6). However, differences emerged in the group of respondents who saw the recall aid first. Row 6 shows that smartphone respondents entered significantly more friends (4.33) than PC respondents (4.01), when seeing the recall aid first. This pattern emerged no matter if they gave a non-meaningful or a meaningful answer to the recall aid (see Rows 4 and 5), although the difference was only marginally significant in the latter group (t = -1.84; p = 0.07). This suggests, again, that the smaller screen of smartphones and the smaller digital keyboard did not negatively affect the number of network contacts elicited from the name generator question. In sum, the difference between devices was only significant when analyzing respondents who were asked the recall aid before entering their friends. This suggests that seeing the recall aid first has a slight positive effect on the number of friends entered on smartphones.

Table 3.6: Effect of the Recall Aid and Device on the Mean Number of Friends Reported

Seen first	Quality of answer to recall aid	PC	Smartphone	Overall	T-test Smartphone vs. PC (*p*-value)
Name Generator	Non-meaningful answers	2.09	2.10	2.11	0.98
	Meaningful answers	4.25	4.34	4.30	0.46
	Total	4.03	4.18	4.10	0.23
Recall Aid	Non-meaningful answers	1.74	2.52	2.05	0.01**
	Meaningful answers	4.29	4.53	4.41	0.07
	Total	4.01	4.33	4.16	0.01**
Overall		4.02	4.25	-	0.01**
T-test name generator vs. recall aid seen first (*p*-value)		0.96	0.24	0.47	-

Note. Mean values and total numbers in parentheses.
N = 3,592.
Number of friends entered was truncated at 10.

Finally, we ran five OLS regression models to control for socio-demographic variables, smartphone skills, and the supplementary variables obtained from the panel provider. This was necessary considering the fact that the random allocation of respondents to either use a smartphone or a PC did not work properly given the selection bias.

Model 1 in Table 3.7 shows a significant effect of device on the number of friends entered when controlling for the effect of the recall aid. In line with the previous results, there is no significant effect of the recall aid.

Table 3.7: OLS Regressions Predicting Number of Friends Reported

	Model 1	Model 2	Model 3	Model 4	Model 5
Device smartphone[a]	0.23**	0.09	0.10	0.07	-0.02
	(0.08)	(0.09)	(0.09)	(0.09)	(0.34)
Recall aid placement: first[b]	0.07	0.08	0.08	0.00	-0.109
	(0.09)	(0.08)	(0.09)	(0.32)	(0.08)
Gender female[c]		0.48***	0.49***	0.36***	0.36***
		(0.09)	(0.09)	(0.86)	(0.09)
Age		0.00	0.00	0.00	0.00
		(0.00)	(0.00)	(0.00)	(0.00)
Smartphone skills		0.18***	0.17***	0.12***	0.11*
		(0.05)	(0.05)	(0.05)	(0.06)
Low education[d]		-0.77***	-0.67***	-0.58***	-0.58***
		(0.13)	(0.13)	(0.13)	(0.13)
Medium education[d]		-0.36**	-0.30*	-0.25*	-0.24*
		(0.10)	(0.10)	(0.01)	(0.10)
Employment status: unemployed[e]			-0.34*	-0.30	-0.31
			(0.19)	(0.19)	(0.19)
Urban area[f]			0.10	0.12	0.19
			(0.09)	(0.09)	(0.09)
Medical conditions			-0.01	-0.01	-0.01
			(0.03)	(0.03)	(0.03)
Practiced sports			0.07**	0.08***	0.08***
			(0.02)	(0.02)	(0.02)
Quality of answer: meaningful[g]				2.08***	2.15***
				(0.25)	(0.16)
Smartphone*Recall aid placement: first					0.03
					(0.09)
Recall aid placement first*Quality of answer: meaningful				0.12	
				(0.33)	
Constant	3.99***	3.25***	3.06***	1.42***	1.27***
	(0.07)	(0.28)	(0.29)	(0.36)	(0.37)
N	3,891	3,891	3,891	3,891	3,891
R2	0.002	0.034	0.029	0.074	0.074
Adjusted R2	0.015	0.023	0.027	0.071	0.071

Note: Logistic regression model with unstandardized coefficients and standard errors in parentheses.
*$p < .05$; **$p < .01$; ***$p < .001$.
Number of friends entered was truncated at 10.
[a] Reference category is PC.
[b] Reference category is name generator seen first.
[c] Reference category is male.
[d] Reference category is high education.
[e] Reference category is employed.
[f] Reference category is non-urban area.
[g] Reference category is non-meaningful answer.

When adding demographic variables and smartphone skills in Model 2, the influence of the device becomes smaller and is no longer significant, while being female has a strong positive significant effect and low and medium education have negative significant effects. In addition, respondents' ability to use a smart-

phone shows a positive significant effect. This means that women and respondents who are experienced using their smartphone were more likely to report a greater number of friends compared to men and those with less smartphone experience. Moreover, low and medium educated respondents entered fewer friends than high educated respondents.

Thus, Model 2 provides clear evidence that the previously found positive effect of using a smartphone on the number of names reported is not caused by the device but by several factors that differentiate smartphone respondents from PC respondents in our sample. The selection bias discussed above prevented a successful random allocation but increased the number of women, higher educated, and respondents with better smartphone skills in the smartphone sample. The device effect thus appeared because these respondents reported more names than men, those who were lower educated, and those with fewer smartphone skills.

Model 3 additionally includes the variables we received from the panel provider. While the previously entered variables remain largely unchanged, unemployment has a significant negative effect on network size and the number of practiced sports a significant positive one. This means that employed respondents were more likely to report a larger network size than the unemployed, which is in line with results found earlier (Edin et al., 2003; Munshi, 2003). Similarly, respondents who practiced more sports were more likely to report more friends. This finding is supported by previous research that found larger social networks among people who were members of voluntary organizations (Farkas & Lindberg, 2015; Putnam, 2000; Rotolo, 2000).

Model 4 additionally includes a measure indicating whether a respondent gave a meaningful answer to the open question and an interaction of this variable with the recall aid experiment. The results show that whether a meaningful topic was named has a strong positive influence on the number of friends entered. In contrast, the interaction between the position of the recall aid and reporting a meaningful topic is non-significant. This confirms the previous conclusion (Table 3.6) that the recall aid did not lead to reports of more network contacts. Respondents

who gave a meaningful answer did not enter more friends because of the recall aid, but because they showed a higher motivation to complete the questionnaire effortfully.

Finally, in Model 5 the interaction is replaced by and interaction between the device used and the placement of the recall aid because Table 3.6 suggested that such an interaction effect may exist. Model 5 reveals no significant interaction effect between the placement of the recall aid and device used, showing that seeing the recall aid before the generator did not help smartphone respondents to complete the name generator to a greater extent than PC respondents. Again, our regression findings imply that the differences between the devices found in Table 3.6 were caused by respondent's characteristics, and not by the device used.

In sum, these results show that smartphones are an equally feasible option as PCs to conduct ego-centered social network research, as least for tech-savvy populations who are willing to answer online surveys on smartphones. Furthermore, the results suggest that respondents' likelihood to report a larger number of names is strongly related to their motivation and that this motivation can be measured by comparing meaningful and non-meaningful answers in an open-ended question.

3.4 Discussion

3.4.1 Selection Bias and Network Size

The study has several important implications for future network research using smartphones regarding selection effects, network sizes, usage of recall aids, and satisficing response behavior. Particularly noteworthy is our finding that 62.5% of participants who were randomly assigned to complete the survey on a smartphone and not on a PC did not comply with this request and were screened-out. As a consequence of this non-compliance, smartphone respondents were more likely to be female, younger, had a higher ability to use their smartphone, and were less likely to have a medium education level or to suffer from medical conditions compared to the PC group.

Such a selection bias can have severe consequences for social network research making use of smartphone surveys. In particular, many people who completed our survey on a smartphone belonged to demographic groups that are associated with having larger social networks. For instance, woman tend to report larger network sizes than men (Ajrouch et al., 2005; Goodreau et al., 2009; Lewis & Kaufman, 2018; McLaughlin et al., 2010) and younger people are more likely to report larger networks than older persons (Ajrouch et al., 2005). Moreover, previous research has shown that larger and more functional social networks are associated with better health (e.g., Michael et al., 1999; Schaefer et al., 1981). As people with fewer medical conditions, woman, and younger people were overrepresented in our smartphone sample, the average network size of our smartphone respondents may have been overestimated. This suggests that we had found a smaller average network size in the smartphone sample, if our experimental assignment would have worked as planned.

However, not all results point to an overestimation of the network size in the smartphone sample. We found that employed people were less likely than unemployed to complete the survey on a smartphone, whereas employed people have been found to report larger social networks than those without a job (Edin et al., 2003; Munshi, 2003; Rollins et al., 2011). Considering this finding, it is also possible that we would have found a larger average network size among smartphone respondents under fulfilled experimental conditions. However, a smaller network size seems more likely, as we found three factors that hint at an overestimation of the network size on smartphones and only one factor that hints at an underestimation. This "true" network size could be more similar to the one of PC respondents or even slightly smaller. In sum, it is thus not certain whether the selection bias led to an overestimation of the network size among smartphone respondents, and more research on the impact of selection effects on network sizes in smartphone surveys is needed.

It should be noted that we also found a selection bias in the PC group. Those who followed the instruction to complete the survey on a PC were more likely to be male, had a higher education, and were older. Since some of these character-

istics are associated with smaller networks (e.g., Goodreau et al., 2009; Lewis & Kaufman, 2018; McLaughlin et al., 2010), conducting ego-centered network surveys only with PCs may introduce a selection bias that leads to an underestimation of the average network size in a population. This selection bias illustrates that PCs cannot be seen as the gold standard to conduct ego-centered social network studies. Neither can responses of PC respondents be seen as entirely accurate. In line with previous studies on mode effects (Fischer & Bayham, 2019; Matzat & Snijders, 2010; Vriens & van Ingen, 2018), we have to conclude that none of the devices is clearly superior to the other when it comes to generating ego-centered social networks.

3.4.2 Device Effects

When comparing the network size in a PC survey to the one elicited on smartphones, we found that the use of smartphones to complete the name generator of an ego-centered network study did not negatively affect the reported network sizes. Previous work found mixed-evidence on whether respondents in an online (PC) survey name fewer or more network contacts than respondents in a face-to-face survey (Fischer & Bayham, 2019; Matzat & Snijders, 2010; Vriens & van Ingen, 2018). Our results initially suggested that more names were elicited in the smartphone condition than the PC condition. However, this was due to the overrepresentation of certain groups that tend to report larger networks, induced by the selection bias. Thus, this study concludes that moving from PC to smartphone does not increase the number of reported names but it also does not negatively affect it. This finding is quite remarkable considering the increased difficulty of answering a survey on smartphones that have considerably smaller displays and keyboards than PCs. Thus, at least among those respondents who are willing to complete a smartphone survey, using smartphones for data collection of ego-centered social network data seems to be an excellent opportunity compared to more traditional online methodologies. In fact, allowing smartphones as a response device could even be an option to reduce nonresponse of groups in PC surveys that may prefer to answer on a smartphone, such as tech-savvy populations.

A possible reason for the promising results regarding the implementation of network name generator question on smartphones could be that smartphones, as highly personalized devices used for private communication, help respondents to more easily recall their friends and therefore reduce cognitive effort. Respondents could also easily and quickly open their address books or recent conversations on the smartphone to recall important contacts, which can increase the number of names reported (Hsieh, 2015). While PC respondents may likewise access their contacts (e.g., recent email conversations), the majority of personal communication nowadays takes place via smartphones (e.g., via direct messenger apps). These factors might compensate for a longer response time that we found on smartphones, most likely due to the less comfortable input mechanics. It is also possible that a longer response time emerged from respondents leaving the survey to check their contact lists or social media apps for contacts. Another reason that could possibly have influenced the reported network size on smartphones is that some smartphone keyboards are able to autocomplete frequently written names, such as those of close friends. Future research could make use of para- and meta-data on respondents' behavior to investigate whether these factors are more relevant on smartphones than on PCs and test whether such behaviors increase the number of reported names.

3.4.3 Open Question as Data Quality Indicator

We did not find a positive effect of providing a recall aid before the name generator on the reported network size. This was the case for respondents who answered the questionnaire on their PC and for those who answered on their smartphones. Earlier studies suggest that reminding respondents of various social settings in which they could have interacted or of different types of relationships through asking multiple network generator questions leads to the reporting of more names (Brewer, 2000). A potential explanation for our null finding could thus be that the recall aid question was too general as we simply asked what respondents considered important with regard to their friends. Future research might thus be better advised to use more specific recall aids and probes that remind people of particu-

lar contexts and relationships.

When examining the responses to the open recall aid question more closely, we found their quality to be a strong predictor for reporting a larger number of friends to the name generator question. Specifically, whether respondents gave meaningful answers to the recall aid question or not was the strongest predictor of reported network size in our study. This suggests that the open question can serve as a proxy for a respondent's motivation to put effort into accurately engaging with a survey. The open question is therefore a potential tool to identify respondents who show satisficing response behavior. According to the theory of survey satisficing (Krosnick, 1991), some respondents are not willing to invest sufficient cognitive effort into answering a survey question adequately, but instead "satisfice" by providing an easily accessible answer such as selecting the first response option or saying "don't know" (Krosnick, 1991, 1999). This effect manifested itself in our survey in the form of respondents skipping the recall aid without answering at all or delivering a non-meaningful answer and also by providing no or only a few names to the name generator. Hence, this study suggests that including an open question previous or close to a network generator, can help to evaluate respondents' mindfulness (Vannette & Krosnick, 2014), which may be directly connected to the response quality.

In our study, when using the open-ended question as a proxy for response quality, only 7.1% of the respondents showed problematic response behavior, by giving non-meaningful answers (which compares to other studies on survey satisficing, e.g., Gummer et al., 2021). However, this group of respondents significantly reduced the average number of names reported in the network generator question on both devices, so that researchers cannot simply ignore respondents who try to shorten the response process and provide an answer that requires less consideration and thinking. Asking an open-ended question previous to a name generator can help identifying such respondents in future studies.

3.4.4 Limitations

While our results provide first evidence that ego-centered network studies on smartphones are feasible, further research is necessary for several reasons. Our survey was based on a non-probability sample, drawn from an online access panel. Such panels are prone to several biases, such as selection biases caused by the non-probability nature of the selection of panel members and the large number of surveys in which most panel respondents take part (Hillygus et al., 2014; Matthijsse et al., 2015). Further, we detected a large selection bias caused by the device assignment. While our analyses corrected for the bias to some extent by including relevant demographic and supplementary variables that were obtained from the panel provider, such an approach can never completely rule out the bias. Screened-out respondents may differ on other potentially important variables that are associated with people's sociability. For instance, research found that personality traits are associated with network size (Kalish & Robins, 2006; Tziner et al., 2004) and future studies could account for those.

In addition, the selection effects we found may imply a general reluctance of certain groups to complete surveys on smartphones (de Bruijne & Wijnant, 2014; Fuchs & Busse, 2009; Toepoel, 2017) but it may also be a consequence of the way in which people were invited to participate in the survey. The link to the survey was sent via email and many people may still read their email on their computer and not on a smartphone. Network researchers should thus consider inviting their respondents in ways that are more likely to be read on the device the survey is supposed to be completed such as, for instance, via text messages or scannable Quick Response (QR) codes.

Finally, it should be recognized that our experimental study used only a single network name generator question but did not employ a complete social network module. Further methodological research should examine effects of smartphone use on additional indicators, such as multiple name generator questions, name interpreter questions, and questions about the network structure. Repeatedly answering the same question about all alters and reporting on the existence of all

alter-alter ties can reduce respondents' willingness to effortfully answer all questions in a PC survey (Matzat & Snijders, 2010). This may be even worse on the small displays of smartphones. The small displays of smartphones also restrict the use of visual tools to collect ego-centered network data (Hogan et al., 2016; Stark & Krosnick, 2017), which have been found to increase respondents' motivation (Stark & Krosnick, 2017). Thus, while our results suggest that the network size is not lower in smartphone surveys in populations that are willing to complete such surveys, it remains unknown how smartphones affect the response quality of other network characteristics in ego-centered network studies.

3.4.5 Conclusion and Recommendations

Despite the limitations, our results suggest that smartphones are a feasible device to conduct ego-centered social network research and could help to increase response rates and measurement accuracy by including groups that are unlikely to participate in surveys on a regular computer. Thus, based on our results, we recommend allowing the usage of smartphones as an additional option to answer web surveys, as the free device choice is likely to reduce nonresponse bias and increase measurement quality. At the same time, our results also imply that forcing respondents to use smartphones to complete a survey can result in selection biases, as specific groups of people may be unwilling to use this device. Hence, forcing respondents to make use of a specific device to complete the survey is not recommendable. This is true for both smartphones and PCs since we also found evidence of a selection bias in the PC sample.

A potential strategy to deal with selection biases due to the selected device may be post-stratification weighting. However, we recommend using such weights only in cases were bias in the data was detected and not as a general data handling strategy. Our study shows that identifying appropriate weighting variables beyond demographics is not trivial and that information on such variables, such as the individual health status or personality traits, is often not available. Thus, it appears best to try to minimize selection bias by improving the study design in the survey planning phase and by that avoiding the necessity of post-stratification

weighting.

Nevertheless, our results also suggest that a smartphone-only survey is a feasible option for tech-savvy populations since tech-savvy respondents were particularly likely to take part in our survey on their smartphones. In addition, if researchers need their respondents to answer a survey on smartphones, for example, because the study includes additional measurements via an app, we recommend inviting respondents to the survey through a method that is likely read on smartphones, such as text messages or QR codes. This should prevent non-response caused by people's unwillingness to switch to a different device than the one on which they have read the invitation to the survey.

Lastly, our study suggests that an open-ended question about the network can be a valuable tool to identify respondents that are satisficing and not answering effortfully. Including such a question can help researchers to evaluate response quality efficiently. Given the wide-spread use of smartphones among people around the world (Poushter et al., 2018) and people's rapid adjustment to new technologies, researchers will soon be tempted to routinely collect network information on these devices. Our study suggests that this endeavor might be fruitful, but it also encourages more work on name interpreter questions as well as selection effects to uncover the full potential of this methodological avenue of social network research.

References

Ajrouch, K. J., Blandon, A. Y., & Antonucci, T. C. (2005). Social networks among men and women: The effects of age and socioeconomic status. *The Journals of Gerontology: Series B*, *60*, 311–317.

Bekkers, R., Völker, B., Van der Gaag, M., & Flap, H. (2008). Social networks of participants in voluntary associations. In N. Lin & B. Erickson (Eds.), *Social capital: An international research program* (pp. 185–205). Oxford University Press.

Belli, R. F., Shay, W. L., & Stafford, F. P. (2001). Event history calendars and question list surveys: A direct comparison of interviewing methods. *Public opinion quarterly*, *65*, 45–74.

Brewer, D. D. (2000). Forgetting in the recall-based elicitation of personal and social networks. *Social networks*, *22*, 29–43.

Burt, R. S. (1984). Network items and the general social survey. *Social networks*, *6*, 293–339.

Couper, M. P., Antoun, C., & Mavletova, A. (2017). Mobile web surveys. In P. P. Biemer, E. D. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. E. Lyberg, N. C. Tucker, & B. T. West (Eds.), *Total survey error in practice* (pp. 133–154). Wiley & Sons Ltd.

de Bruijne, M., & Wijnant, A. (2013). Comparing survey results obtained via mobile devices and computers: An experiment with a mobile web survey on a heterogeneous group of mobile devices versus a computer-assisted web survey. *Social Science Computer Review*, *31*, 482–504.

de Bruijne, M., & Wijnant, A. (2014). Mobile response in web panels. *Social Science Computer Review*, *32*, 728–742.

Eddens, K., & Fagan, J. M. (2018). Comparing nascent approaches for gathering alter-tie data for egocentric studies. *Social Networks*, *55*, 130–141.

Edin, P.-A., Fredriksson, P., & Åslund, O. (2003). Ethnic enclaves and the economic success of immigrants—evidence from a natural experiment. *The quarterly journal of economics*, *118*, 329–357.

Farkas, G. M., & Lindberg, E. (2015). Voluntary Associations' Impact on the Composition of Active Members' Social Networks: Not an Either/Or Matter. *Sociological Forum*, *30*, 1082–1105.

Fischer, C. S., & Bayham, L. (2019). Mode and Interviewer Effects in Egocentric Network Research. *Field methods*, *31*, 195–213.

Fuchs, M., & Busse, B. (2009). The coverage bias of mobile web surveys across European countries. *International Journal of Internet Science*, *4*, 21–33.

Glasner, T., & Van der Vaart, W. (2009). Applications of calendar instruments in social surveys: A review. *Quality and Quantity*, *43*, 333–349.

Goodreau, S. M., Kitts, J. A., & Morris, M. (2009). Birds of a feather, or friend of a friend? Using exponential random graph models to investigate adolescent social networks. *Demography*, *46*, 103–125.

Gummer, T., Roßmann, J., & Silber, H. (2021). Using instructed response items as attention checks in web surveys: Properties and implementation. *Sociological Methods & Research*, *50*, 238–264.

Hillygus, D. S., Jackson, N., & Young, M. (2014). Professional respondents in non-probability online panels. In M. Callegaro, R. Baker, J. Bethlehem, A. S. Goritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online panel research: A data quality perspective* (pp. 219–237). Wiley & Sons Ltd.

Hogan, B., Melville, J. R., Phillips II, G. L., Janulis, P., Contractor, N., Mustanski, B. S., & Birkett, M. (2016). Evaluating the paper-to-screen translation of participant-aided sociograms with high-risk participants. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5360–5371.

Hsieh, Y. P. (2015). Check the phone book: Testing information and communication technology (ICT) recall aids for personal network surveys. *Social Networks*, *41*, 101–112.

Kalish, Y., & Robins, G. (2006). Psychological predispositions and network structure: The relationship between individual predispositions, structural holes and network closure. *Social networks*, *28*, 56–84.

Keusch, F., & Yan, T. (2017). Web Versus Mobile Web: An Experimental Study of Device Effects and Self-Selection Effects. *Social Science Computer Review*, *35*, 751–769.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied cognitive psychology*, *5*, 213–236.

Krosnick, J. A. (1999). Survey research. *Annual review of psychology*, *50*, 537–567.

Lewis, K., & Kaufman, J. (2018). The conversion of cultural tastes into social network ties. *American journal of sociology*, *123*, 1684–1742.

Marsden, P. V. (2011). Survey methods for network data. In J. Scott & P. J. Carrington (Eds.), *The SAGE handbook of social network analysis* (pp. 370–388). SAGE.

Matthijsse, S. M., de Leeuw, E. D., & Hox, J. J. (2015). Internet panels, professional respondents, and data quality. *Methodology*, *11*, 81–88.

Matzat, U., & Snijders, C. (2010). Does the online collection of ego-centered network data reduce data quality? An experimental comparison. *Social Networks*, *32*, 105–111.

McLaughlin, D., Vagenas, D., Pachana, N. A., Begum, N., & Dobson, A. (2010). Gender differences in social network size and satisfaction in adults in their 70s. *Journal of Health Psychology*, *15*, 671–679.

Michael, Y. L., Colditz, G. A., Coakley, E., & Kawachi, I. (1999). Health behaviors, social networks, and healthy aging: Cross-sectional evidence from the Nurses' Health Study. *Quality of life research*, *8*, 711–722.

Munshi, K. (2003). Networks in the modern economy: Mexican migrants in the US labor market. *The Quarterly Journal of Economics*, *118*, 549–599.

Oliver, J. E. (2000). City size and civic involvement in metropolitan America. *American Political Science Review*, *94*, 361–373.

Poushter, J., Bishop, C., & Chwe, H. (2018). Social media use continues to rise in developing countries but plateaus across developed ones. *Pew Research Center*, *22*, 2–19.

Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. Simon & Schuster.

Remmer, K. L. (2010). Political scale and electoral turnout: Evidence from the less industrialized world. *Comparative Political Studies*, *43*, 275–303.

Rollins, A. L., Bond, G. R., Jones, A. M., Kukla, M., & Collins, L. A. (2011). Workplace social networks and their relationship with job outcomes and

other employment characteristics for people with severe mental illness. *Journal of Vocational Rehabilitation*, *35*, 243–252.

Rotolo, T. (2000). Town heterogeneity and affiliation: A multilevel analysis of voluntary association membership. *Sociological perspectives*, *43*, 271–289.

Schaefer, C., Coyne, J. C., & Lazarus, R. S. (1981). The health-related functions of social support. *Journal of behavioral medicine*, *4*, 381–406.

Silber, H., Schröder, J., Struminskaya, B., Stocké, V., & Bosnjak, M. (2019). Does panel conditioning affect data quality in ego-centered social network questions? *Social Networks*, *56*, 45–54.

Stark, T. H., & Krosnick, J. A. (2017). GENSI: A new graphical tool to collect ego-centered network data. *Social Networks*, *48*, 36–45.

Toepoel, V. (2017). Online Survey Design. In N. Fielding, R. M. Lee, & G. Blank (Eds.), *The SAGE handbook of online research methods* (pp. 184–202). SAGE.

Tziner, A., Vered, E., & Ophir, L. (2004). Predictors of job search intensity among college graduates. *Journal of Career Assessment*, *12*, 332–344.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*, 1–67.

Vannette, D. L., & Krosnick, J. A. (2014). A comparison of Survey Satisficing and Mindlessness. In A. Ie, C. T. Ngnoumen, & E. J. Langer (Eds.), *The Wiley Blackwell Handbook of Mindfullness* (pp. 312–327). Wiley & Sons Ltd.

Vehovar, V., Manfreda, K. L., Koren, G., & Hlebec, V. (2008). Measuring egocentered social networks on the web: Questionnaire design issues. *Social networks*, *30*, 213–222.

Vriens, E., & van Ingen, E. (2018). Does the rise of the Internet bring erosion of strong ties? Analyses of social media use and changes in core discussion networks. *new media & society*, *20*, 2432–2449.

Yan, T., & Tourangeau, R. (2008). Fast times and easy questions: The effects of age, experience and question complexity on web survey response times. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *22*, 51–68.

Yousefi-Nooraie, R., Marin, A., Hanneman, R., Pullenayegum, E., Lohfeld, L., & Dobbins, M. (2019). The relationship between the position of name gen-

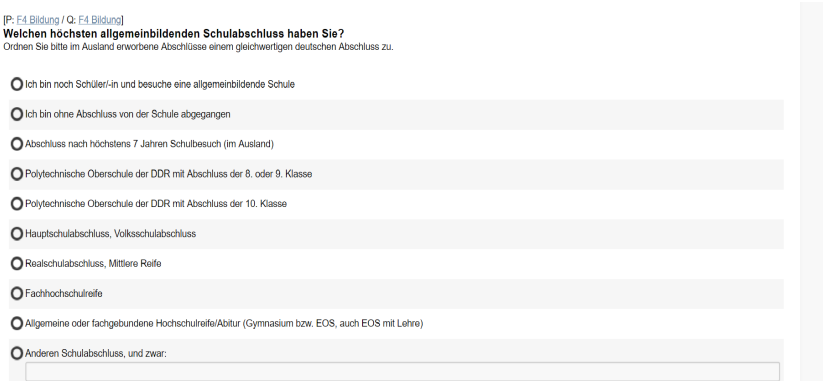
erator questions and responsiveness in multiple name generator surveys. *Sociological Methods & Research*, *48*, 243–262.

Appendix

Appendix A. Screenshots of the Online Questionnaire [in German]

Figure A3.1: Network Generator Question

Figure A3.2: Recall Aid Question


[P: F4 Bildung / Q: F4 Bildung]
Welchen höchsten allgemeinbildenden Schulabschluss haben Sie?
Ordnen Sie bitte im Ausland erworbene Abschlüsse einem gleichwertigen deutschen Abschluss zu.

- Ich bin noch Schüler/-in und besuche eine allgemeinbildende Schule
- Ich bin ohne Abschluss von der Schule abgegangen
- Abschluss nach höchstens 7 Jahren Schulbesuch (im Ausland)
- Polytechnische Oberschule der DDR mit Abschluss der 8. oder 9. Klasse
- Polytechnische Oberschule der DDR mit Abschluss der 10. Klasse
- Hauptschulabschluss, Volksschulabschluss
- Realschulabschluss, Mittlere Reife
- Fachhochschulreife
- Allgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS, auch EOS mit Lehre)
- Anderen Schulabschluss, und zwar:

Figure A3.3: Question on Education

Appendix B. Lists of Supplementary Information from the Panel Provider

B3.1 List of Medical Conditions (all medical conditions were included in the additive index)

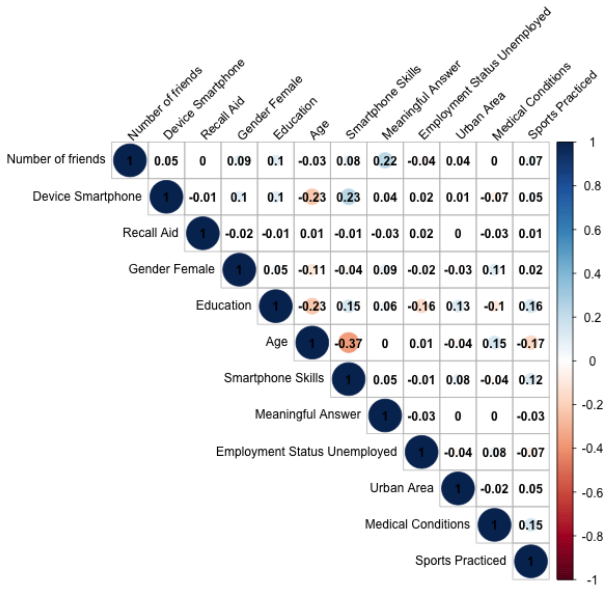
1. Asthma/chronic Bronchitis
2. Diabetes mellitus
3. Epilepsy
4. Erectile Dysfunction
5. Skin Complaints
6. Hearth Complaints
7. Hearing Problems
8. Incontinence
9. Migraine
10. Rheumatism
11. Back Problems
12. Thyroid (overactive/underactive)
13. Sleeplessness
14. Thrombosis
15. Stomach Complaints

B3.2 List of Sports Practiced (sports not included in the additive index are displayed in italic)

1. Aerobics
2. Badminton
3. Basketball
4. Dancing
5. Diving
6. Fitness
7. Football
8. Handball
9. Hockey
10. Jogging
11. Judo
12. Karate
13. Pilates
14. Horse riding
15. Swimming
16. Squash
17. Tennis
18. Volleyball
19. Yoga
20. Cycling

21. Golf
22. Sailing
23. Skiing
24. Surfing
25. Other sports

Appendix C. Correlation Plot and Tables showing Regression Analysis using Listwise Deletion



Note. The table displays Pearson's Product-Moment correlations between all variables used in our study. Correlations above .10 and below -.10 are highlighted. As the legend at the right of the figure shows, positive correlation coefficients are highlighted in blue, and negative correlation coefficients are highlighted in red. The size and the darkness of the circle for specific correlation coefficients display the strength of the correlation. For instance, the strongest correlation coefficients in this study appeared between smartphone skills and age (r = -.37). That circle is displayed darker and in a larger size as for instance, the correlation between sports practices and age, which was -.17.

Figure C3.1: Number of Friends entered in the Network Generator Question by Device

Table C3.1: Logistic Regression Analyses Predicting Being Screened-out or not in the Smartphone Condition and the PC Condition when using Listwise Deletion

	Screen-out Smartphone	Screen-out PC
Gender female [a]	-0.50***	0.24
	(0.08)	(0.16)
Low education[b]	-0.01	0.43
	(0.12)	(0.27)
Medium education[b]	-0.06	0.78***
	(0.09)	(0.19)
Age	0.04***	-0.03***
	(0.00)	(0.01)
Employment status: unemployed[c]	-0.34**	0.06
	(0.17)	(0.36)
Urban area[d]	0.08	0.18
	(0.08)	(0.17)
Medical conditions	0.07***	0.04
	(0.02)	(0.05)
Practiced sports	0.01	0.04
	(0.02)	(0.03)
N	3,249	1,660
Pseudo-R^2 (McFadden)	0.34	0.34

Note: Logistic regression model with unstandardized coefficients and standard errors in parentheses.
*$p < .05$; **$p < .01$; ***$p < .001$.
Number of friends entered was truncated at 10.
[a] Reference category is male.
[b] Reference category is high education.
[c] Reference category is employed.
[d] Reference category is non-urban area.

Table C3.2: Logistic Regression Analysis Testing Randomization to Device Assignment and Recall Aid Placement when using Listwise Deletion

	Assignment to Mobile	Question Placement Recall Aid first
Gender female[a]	0.39***	-0.01
	(0.27)	(0.06)
Age	-0.02***	0.00
	(0.00)	(0.00)
Education low	-0.13	0.05
	(0.13)	(0.13)
Education middle	-0.26***	0.09
	(0.09)	(0.09)
Smartphones skills	0.0***	-0.2
	(0.05)	(0.04)
Employment status: unemployed[b]	0.27	0.14
	(0.19)	(0.18)
Urban area[c]	-0.06	0.00
	(0.02)	(0.08)
Medical condition	-0.06**	-0.05*
	(0.03)	(0.03)
Practiced sports	0.00	0.02
	(0.02)	(0.02)
N	3,249	1,660
Pseudo-R^2 (McFadden)	0.34	0.34

Note: Logistic regression model with unstandardized coefficients and standard errors in parentheses.
*$p < .05$; **$p < .01$; ***$p < .001$.
Number of friends entered was truncated at 10.
[a] Reference category is male.
[b] Reference category is employed.
[c] Reference category is non-urban area.
[d] Reference category is non-urban area.

Table C3.3: OLS Regressions Predicting Reported Number of Friends using List-wise Deletion

test	Model 1	Model 2	Model 3	Model 4	Model 5
Device smartphone[a]	0.26**	0.12	0.13	0.11	0.03
	(0.10)	(0.11)	(0.11)	(0.10)	(0.14)
Recall aid placement: first[b]	0.01	0.03	0.02	-0.25	-0.02
	(0.10)	(0.10)	(0.10)	(0.38)	(0.13)
Gender female[c]		0.48***	0.49***	0.39***	0.39***
		(0.10)	(0.10)	(0.10)	(0.10)
Age		0.01	0.01*	0.01	0.01
		(0.00)	(0.00)	(0.00)	(0.00)
Smartphone skills		0.19***	0.18***	0.15***	0.15***
		(0.05)	(0.05)	(0.05)	(0.05)
Low education[d]		-0.72***	-0.62***	-0.54***	-0.54***
		(0.16)	(0.16)	(0.16)	(0.16)
Middle education[d]		-0.26**	-0.19*	-0.14	-0.14
		(0.11)	(0.11)	(0.11)	(0.11)
Employment status: unemployed[e]			-0.28	-0.22	-0.22
			(0.23)	(0.23)	(0.23)
Urban area[f]			0.15	0.15	0.15
			(0.11)	(0.10)	(0.104
Medical conditions			-0.02	-0.02	-0.02
			(0.03)	(0.03)	(0.03)
Practiced sports			0.06***	0.07***	0.07***
			(0.02)	(0.02)	(0.02)
Quality of answer: meaningful[g]				1.90***	2.08***
				(0.29)	(0.20)
Smartphone*Recall aid placement: first					0.16
					(0.20)
Recall aid placement first*Quality of answer: meaningful				0.32	
				(0.39)	
Constant	4.12***	3.14***	2.96***	1.41***	1.27***
	(0.09)	(0.33)	(0.34)	(0.42)	(0.37)
N	2,547	2,547	2,547	2,547	2,547
R^2	0.003	0.025	0.029	0.071	0.071
Adjusted R^2	0.002	0.022	0.025	0.066	0.066

Note: Logistic regression model with unstandardized coefficients and standard errors in parentheses.
*$p < .05$; **$p < .01$; ***$p < .001$.
Number of friends entered was truncated at 10.
[a] Reference category is PC.
[b] Reference category is name generator seen first.
[c] Reference category is male.
[d] Reference category is high education.
[e] Reference category is employed.
[f] Reference category is non-urban area.
[g] Reference category is non-meaningful answer.

4 Consent to Data Linkage for Different Data Domains – The Role of Question Order, Question Wording, and Incentives

Under review in:
International Journal of Social Research Methodology

Abstract

As our modern world has become increasingly digitized, various types of data from different data domains are available that can enrich survey data. To link survey data to other sources, consent from the survey respondents is required. This article compares consent to data linkage requests for seven data domains: administrative data, smartphone usage data, bank data, biomarkers, Facebook data, health insurance data, and sensor data. We experimentally explore three factors of interest to survey designers seeking to maximize consent rates: consent question order, consent question wording, and incentives. The results of the study using a German online sample (N = 3,374) show that survey respondents have a relatively high probability of consent to share smartphone usage data, Facebook data, and biomarkers, while they are least likely to share their bank data in a survey. Of the three experimental factors, only the consent question order affected consent rates significantly. Additionally, the study investigated the interactions between the three experimental manipulations and the seven data domains of which only the interaction between the data domains and the consent question order showed a consistent significant effect.

Keywords

consent, data linkage, online surveys, experiment

4.1 Introduction

With the help of emerging digital technologies, the ability to easily record and process information on people's everyday life offers new possibilities for researchers (Link et al., 2014). Linking additional data with traditional survey data provides an opportunity for survey methodology, a field that permanently tries to reduce survey costs and measurement error. Data linkage could help increase data quality by delivering more reliable data and substituting missing data and therefore increase the value of datasets. As surveys are also used in an increasingly interdisciplinary environment and across different fields of science, including public health, economics, and education, the linkage of additional data might help researchers to generate new knowledge by enriching datasets with information not obtainable from surveys (e.g., linking medical records to self-assessments). Some studies already exist that link survey data to additional data sources, such as biomarkers (Avendano, 2018; McFall et al., 2014) and administrative data (Baker et al., 2000; Christoph et al., 2008).

However, scientists cannot simply link additional data but have an ethical and legal obligation to obtain respondents' consent before implementing data linkage procedures. As some data domains (e.g., financial data) are often seen as sensitive by respondents (Walzenbach et al., 2022), asking for and obtaining consent to link these data can be rather difficult. While various factors can influence consent rates, only some can be manipulated by researchers, including incentives, question wording, and the position of the consent request in the questionnaire (Keusch et al., 2019; Revilla et al., 2019; Sakshaug et al., 2012; Sakshaug et al., 2013). So far, only a few studies have compared consent decisions with respect to different data domains (Revilla et al., 2019; Wenz et al., 2019). By investigating contextual factors and different data domains in one study, we will be able to give practical advice to practitioners on how to increase consent rates for data linkage requests in surveys.

In this study, we investigate respondents' willingness to consent to the linkage of data from various data domains, and we experimentally test the influence of con-

textual factors that can be optimized to achieve higher consent rates. In particular, we are interested in answering the following research questions: (1) Do consent rates for data linkage differ by data domain? (2) How do contextual factors (i.e., question wording, incentives, and question order) influence consent rates? (3) How do data domains and contextual factors interact with respect to willingness to share additional data?

In the next section of this paper, we discuss previous research on data linkage and develop hypotheses. Afterward, we describe the methods used in our study and present our results. Finally, we discuss the theoretical and practical implications of our findings.

4.2 Background

With their consent to data linkage, survey respondents allow the researcher further access to additional personal information. Thus, the act of consenting to data linkage is always associated with a certain amount of previously agreed privacy intrusions (Martin & Shilton, 2016; Nissenbaum, 2018, 2019). As the perceived sensitivity of data domains might differ, consent rates between data domains might differ. However, researchers can try to increase consent rates by varying certain methodological elements within the consent request(s). In the following, we provide an overview of how the consent decision can differ by data domain and which measures researchers can implement to potentially increase consent rates.

4.2.1 Do consent rates for data linkage differ by data domain?

As data domains might vary in sensitivity, allowing access to them might be associated to higher or lower costs for respondents. Only a few studies confront survey respondents with consent requests regarding different tasks or data domains, which can largely vary regarding the task difficulty and related burden for respondents (Silber et al., 2021). With respect to comparing different data domains, Revilla et al. (2019) conducted a study where Spanish respondents in a nonprobability online panel were given a list of 20 different tasks that went beyond responding to a web survey. Consent rates ranged from 74% for receiving

a product at home and testing it to 6% for letting respondents' children wear a small device that delivers real-time information about the child's stress levels. The results show that respondents were more likely to accept additional tasks if they were able to report the information themselves (e.g., provide a self-report of blood cholesterol levels) compared to when data was shared automatically (e.g., share GPS location via smartphone automatically). Respondents also were less likely to consent when they were asked for measures that allow drawing conclusions about their behavior (e.g., share all information on Facebook profile) compared to other tasks. Wenz et al. (2019) used the Understanding Society Innovation Panel in the UK to ask smartphone respondents for their willingness to perform additional tasks with their smartphones. 65% of the participants stated that they were willing to take photos or scan barcodes, 61% would use the built-in accelerometer to record movements, 39% would share their location via GPS, and 28% would install an app that tracks their phone usage anonymously.

While no study compared the different data domains that we include in our study, previous studies showed that different consent requests indeed yielded different consent rates. This might be because survey respondents consider data from some domains more sensitive than others and thus are less willing to share this information. Therefore, we assume that a respondent's willingness to consent might differ between different data domains.

H1. Consent rates differ between data domains.

4.2.2 How do contextual factors (i.e., question order, consent question wording and incentives) influence consent rates?

Consent Question Order Effect

Linking several data domains to survey data in one study can help to explore research questions from various angles. This can be necessary in interdisciplinary contexts, for example, when assessing health risks using both information from biomarkers and fitness apps. Nevertheless, research on order effects deriving from

multiple requests is rare.

To our knowledge, only one study has yet explicitly investigated consent question order effects. Walzenbach et al. (2022) conducted two experiments in two surveys using an access panel in the UK comparing five different data domains. In the first experiment, they tested two question orders (starting with the most sensitive data domain (health data) vs. starting with the least sensitive data domain (tax data)). They found that asking a less sensitive consent request first yields higher average consent rates. Additionally, they could not find an increase or decrease in consent rates associated with question order of the following questions. However, the effects found in the first study could not be replicated in the second study.

Research conducted by Keusch et al. (2019) can deliver further evidence on consent question order effects. In a vignette experiment, web survey respondents had to rate their willingness to participate in studies that involved installing a research app to their smartphone that passively collected data about smartphone usage and geolocation. Respondents received eight study descriptions that experimentally varied several features of the study (e.g., sponsor, incentive, length of data collection period). The stated willingness of respondents to participate in the described study was significantly higher for the first vignette seen compared to all the other seven vignettes regardless of the study description.

Another study by Sakshaug et al. (2019) focused on the position of a single consent request within the questionnaire. The authors found a consent request placed at the beginning of the survey to increase consent rates by 15.5 percentage points in a telephone sample and by 11.6 percentage points in a web sample, compared to when the question was placed at the end.

Based on these earlier findings, we assume that every additional consent request asked in a survey is associated with additional costs. Thus, respondents should be more likely to consent to earlier consent requests and be less likely to consent to later ones. We hypothesize that a ceiling effect could decrease consent probability for later consent requests drastically (Wang et al., 2008).

H2.1 The earlier a respondent is confronted with a data linkage request in the sequence of consent questions, the higher the likelihood to consent.

Consent Question Wording

Going beyond research that compared consent rates for different data domains, there is a growing body of literature that investigates how consent rates to data linkage can be increased. In particular for cases when linkage requests pertain to data domains that are perceived as sensitive by many respondents, for example, financial information or information about social behavior, the wording of the request seems especially important (Tourangeau & Yan, 2007). However, the empirical evidence on how much influence question wording actually has on the consent decision is mixed. Pascale (2011) found no significant differences in data linkage consent rates for administrative records in a telephone survey between the stated benefits of reduced costs, reduced time, and better data accuracy. Similarly, Sakshaug et al. (2013) found no differences in consent rates between a neutral and a time savings framing when conducting an experiment in a telephone survey. However, in another experimental study, Sakshaug and Kreuter (2014) found a 6 percentage points higher consent rate for linking the web survey responses to administrative records if a time saving framing was used compared to a neutral framing. In a telephone survey, Kreuter et al. (2015) tested a gain against a loss framing using an experimental setup. The gain framing stated that the information provided by the respondent would gain value if the respondent consented to data linkage, whereas the loss framing stated that the information would lose value. They found consent rates to be 10 percentage points higher for the loss than for the gain framing. In another study by Struminskaya et al. (2020), the authors compared a neutral consent request against one emphasizing time savings arising from the linkage of sensor data using the Dutch LISS panel. However, they did not find a significant effect of the framing.

Using beneficial wording in a consent request can help to balance out privacy costs by associating a benefit (e.g., time savings or added scientific value) with the act of data sharing. In that way, a beneficial statement can help reduce costs

arising from sharing information. Therefore, we assume that consent rates can be increased by accompanying the request with a statement that emphasizes specific benefits when consenting.

H2.2 When presented with a stated benefit in the context of a consent request, respondents are more likely to give consent than when the benefit is not explicitly stated.

Incentive

An incentive might be an efficient way to motivate respondents to give linkage consent. The most straightforward way to do so is to link the consent decision to a financial incentive. With respect to survey participation, incentives have been successfully used to increase response rates and data quality (Singer & Ye, 2013). The influence of consent-related incentives, however, has not been studied exhaustively. Jäckle et al. (2017) found no significant difference between incentives of two and six British Pounds for downloading a spending app in the UK Understanding Society Innovation Panel. In their experimental vignette study, Keusch et al. (2019) found that the willingness to download a research app that would passively collect sensor and log file data from participants' smartphones increased by 18, 19, or 26 percentage points, respectively, when a 10 Euro incentive for downloading the app, a 10 Euro incentive for leaving the app installed until the end of the field period, or when both an incentive for downloading and at the end of the field period were promised compared to when no incentives were announced. We expect to find a similar beneficial effect by providing a financial incentive to respondents for making a consent decision.

H2.3 When promised a financial incentive respondent are more likely to give consent, then when no incentive is given.

4.2.3 How do data domains and contextual factors interact with respect to willingness to share additional data?

We assume that the mechanisms of question order, question wording, and incentive influence consent to all seven data domains. However, it is possible that the size of the effect differs between domains. For some data domains, where the perceived costs of data linkage for the respondents are relatively high, the beneficial effects of the contextual factors might be relatively low, as question order, question wording, and incentive might not provide enough of a push for people to give consent. At the same time, for data domains where the perceived costs of data linkage are rather low, and the baseline consent rates are relatively high, the contextual factors might not have a strong additional effect either. As we do not know which consent decisions are more or less costly for respondents, we will not formulate any interaction hypotheses but instead, conduct exploratory analyses.

4.3 Methods

To study our research questions and test the stated hypotheses, we gathered data from a web survey conducted between July 15 and August 31, 2018. Respondents were recruited from a German nonprobability online access panel. Quotas according to the general population of Germany were set for gender, education, age, and federal state.

In addition to the experimental set-up that allows us to test our hypotheses, the questionnaire also included questions and experiments on other topics (e.g., concerning the device used to respond to the survey, misreporting, attentiveness, and social networks). For example, before receiving the invitation, panel members were randomly allocated to a desktop/laptop computer group and a smartphone group, and in the invitation, panel members were instructed to complete the survey on the assigned device. Violations of the requirement were checked by asking the respondent for the device they used and by analyzing the user agent string. Respondents violating the device assignment were screened out. While this experiment is not the main focus of this study, we do control for the device as part of our analyses. Additionally, questions regarding trust, attitudes, smartphone us-

age, socio-demographics, and other variables were included. Respondents had the possibility to proceed without answering a survey question but could not to go back in the questionnaire. The questionnaire was optimized for smartphones, meaning that questions and question formats were optimized to be displayed on smaller screens. For example, larger grids were split and presented as individual questions on subsequent screens so that the questionnaire was shown in a similar design on both desktop/laptop computers and mobile devices.

A total of 50,063 panel members were invited of which 6,750 opened the online questionnaire by clicking on the invitation link. 2,838 or 42% of the panel members who started the survey were screened out because they used a device to which they were not assigned, and 538 (8%) broke off the survey before the last question. The final sample consists of 3,374 completed interviews, of which 1,826 completed the survey on a desktop/laptop computer and 1,548 using a smartphone. The median response time for completing the questionnaire was 29 minutes and 36 seconds.

In this paper, we focus on the consent to data linkage module of our questionnaire (for an overview of the question sequence in this module, see Figure 4.1). The module started with an introductory page, providing information about data protection, and preparing respondents for the consent to data linkage part. At this point, respondents were randomly assigned to one of three different framing conditions. Independent of the framing manipulation, all respondents were also randomly assigned to receive information about an additional incentive or not (the exact wording of the introductory page can be found in Appendix A). One the next seven pages, we asked a sequence of seven consecutive consent requests referring to data linkage regarding the seven data domains, which were shown on separate pages and ordered randomly for each respondent to allow experimental comparison. All respondents who consented to the linkage of Facebook data received a page delivering detailed instructions for the actual data linkage after replying to all seven requests. The remaining respondents were shown a debriefing page, thanking them for their willingness to share data and informing them that no additional data would be gathered. If respondents declined the consent to certain domains,

they were asked to explain their decision. Similarly, if respondents consented to the linkage of Facebook data, they were asked to provide a reason for why they did so. Finally, all respondents were asked if they participated in a survey asking for data linkage of one of our domains before. The wording of all instructions and questions can be found in Appendix A.

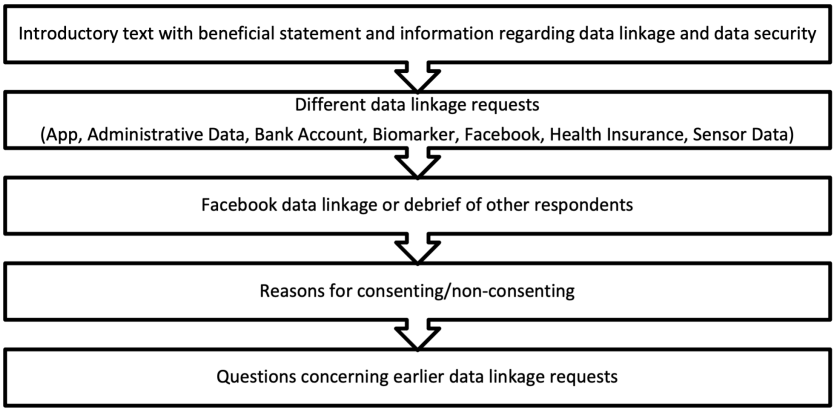


Figure 4.1: Steps of the Consent for Data Linkage Module

4.3.1 Measures

Summary statistics for all variables can be found in Table 4.1.

Experimental Manipulations

Consent Requests

Respondents were confronted with seven consent requests regarding different data domains in random order resulting in 5,040 unique possible question sequences. All of the requests named the type of data which would be collected and gave examples of why these data are of interest (e.g., the number of friends for Facebook data to investigate social interactions on social media platforms). These seven requests, which were placed on separate survey pages, were (1) the installation of an app that tracks smartphone usage behavior, (2) the linkage of administrative data (employment records), (3) the collection of biomarkers, (4) the linkage of

bank account data, (5) the linkage of data from a respondent's Facebook account, (6) the linkage of health insurance data, and (7) the gathering of data measured by smartphone sensors (e.g., GPS and barometer). The exact wording of the requests can be found in Appendix A. For each of the seven requests, respondents could select "Yes, I agree" (coded 1) or "No, I do not agree" (coded 0).

Incentives

Respondents were randomly assigned to one of two incentives groups (no incentive vs. incentive). When answering the consent requests, respondents were only informed that they would receive an incentive or that they would not, but they were not informed about the concrete incentive amount. For our analyses, we coded the no incentive group with 0 and the incentive group with 1.

Question Wording

Respondents were randomly assigned to one of three question wording groups. Based on the experimental groups, respondents were coded with 0 if they received no beneficial framing, with 1 for the scientific benefit framing (emphasizing the scientific use of the data) and 2 for the time saving benefit framing (emphasizing the possibility to shorten the questionnaire by consenting).

Control Variables

Gender

At the beginning of the questionnaire, we asked respondents if they were male or female. It was mandatory to answer the question since it was used for screening to meet the defined quotas. Female was coded as 1, and male was coded as 0.

Age

Respondents had to provide their age in years, and skipping this question was not possible.

Education

Education was also used for screening, and thus, another mandatory question that

was asked in a closed-ended format. For our analysis, we recoded education into three categories from low to high, in accordance with the German school system (9-, 10-, and 12/13-year school tracks).

Device

The device variable was coded as 0 for respondents using a desktop/laptop computer to complete the survey and as 1 for respondents using a smartphone.

Table 4.1: Overview of Measures

Variable					Missing Values
Age	Range	Median	Mean	Standard Deviation	
	18 - 70	44	43.8	13.9	16
Education		Low Education	Medium Education	High Education	
		14.3%	41.7%	44.0%	0
Gender		Male	Female		
		42.8%	57.2%		16
Incentives		No Incentive	Incentive		
		41.8%	58.2%		0
Question Wording		Neutral	Scientific Benefit	Time Saving	
		33.5%	33.0%	33.5%	0
Device		Desktop/Laptop	Smartphone		
		53.7%	46.3%		0

Note. Mean values and total numbers in parentheses.
n = 3,327.

4.4 Analyses Plan

All analyses were conducted in R version 3.6.2 (R Core Team, 2018). For data analysis, we created a dataset consisting of all respondents who gave an answer to all seven consent requests, reducing the dataset to 3,327 cases to ensure comparability across data domains. We then transformed the dataset into the long format using each consent decision as an observation that is nested within a respondent. To answer Research Questions 1 and 2, we specified a logistic multilevel model predicting the probability of consent, including data domains and experimental manipulations nested in the respondent (Model 1). These manipulations are question order, incentive, and question wording. The regression model allows us to estimate the main effects of the data domain and our experimental manipulations

on linkage consent. To answer Research Question 3, we added interactions between the data domains and the experimental manipulations to our model (Model 2). This allows us to identify differences in the effects caused by a manipulation based on the data domain. In both models, control for device, gender, and education. To estimate the multilevel regression models, we used the lme4 library (Bates et al., 2015).

4.5 Results

4.5.1 Does consent rates for data linkage differ by data domain?

In Model 1 (see Table B4.1 in the Appendix), where administrative data is the reference category, we see that consent rates significantly differ for app data (p = .01), biomarker ($p < .001$), Facebook (p = .002), and bank (p <.001) data, thus supporting H1. Figure 4.2 displays the predicted probabilities of consent by data domain. Except for the linkage of bank account data (19%), 95% CI [14.1, 24.4], all consent probabilities ranged between 45% for sensor data, 95% CI [39.0, 55.4] and 54% for biomarkers, 95% CI [46.0, 62.0]. It should be noted that predicted probabilities do not resemble regression coefficients but result out of a specific configuration of the regression model where reference categories are held constant for the calculation. Thus, significant differences as displayed in Table B4.1 might not be visible in the graph.

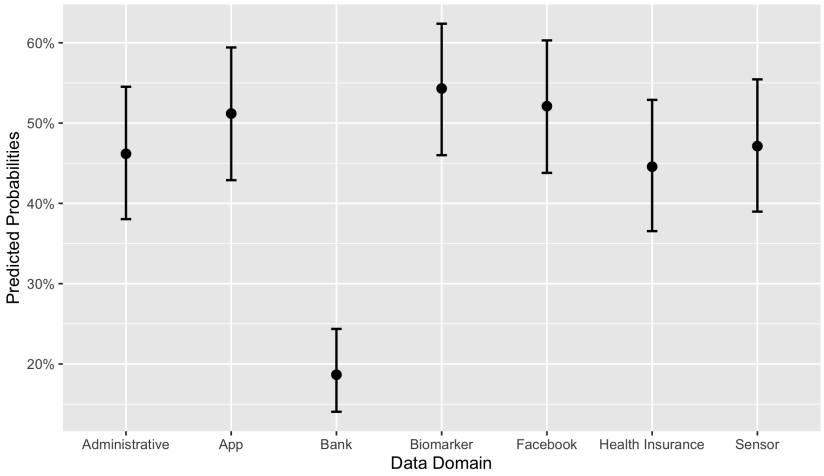


Figure 4.2: Predicted Probabilities by Data Domain (see Model 1 in Table B4.1)

4.5.2 How do contextual factors (i.e., question order, consent question wording, and incentives) influence consent rates?

To answer Research Question 2, we again use Model 1, including main effects only. Our regression model shows significant effects comparing position 1 to all other positions, thus supporting H2.1. As shown by the predicted probabilities of consent displayed in Figure 4.3, the consent probability drops when a question is asked in the second (22%, 95% CI [16.6%, 28.2%]) instead of the first position (46%, 95% CI [38.0%, 54.5%]). This trend continuous throughout all 7 positions with a declining strength. Eventually, consent probabilities stay nearly consistent between the fourth (10%, 95% CI [7.3%, 13.4%]) and the seventh position (6%, 95% CI [4.5%, 8.5%]).

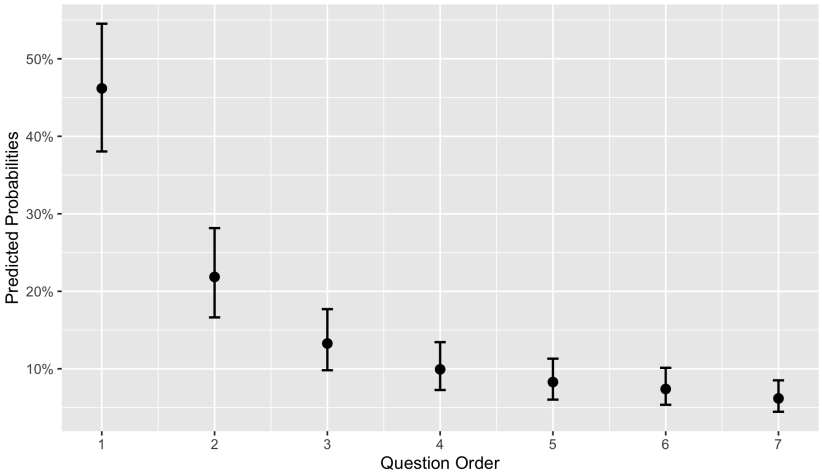


Figure 4.3: Predicted probabilities of consent to data domain by consent question position (see Model 1 in Table B4.1)

Concerning the influence of consent question wording, we do not find a significant effect for any of the variations ($p > .1$, see Table B4.1). Therefore, H2.2 is not supported. Similarly, the incentive condition did not significantly influence consent probability ($p > .1$, see Table B4.1), leading to a lack of support for H2.3.

4.5.3 How do data domains and contextual factors interact with respect to willingness to share additional data?

To examine how the effect of our experimental manipulations differ by data domains, we specified a multilevel logistic regression model including interaction terms between domains and manipulations (Model 2 in the Appendix Table B4.2).

When looking at the interaction between data domain and question order, we find several significant effects. Figure 4.4 shows the predicted probabilities of consent by question order by data domain. All seven data domains show a common pattern of the probability of consent continuously declining between the first and seventh positions. However, for some domains, the effect between the first and the second position is more pronounced (administrative data, bank data, biomarkers, Facebook data, health insurance data) than for others (app data, sensor data) where the 95% confidence intervals overlap.

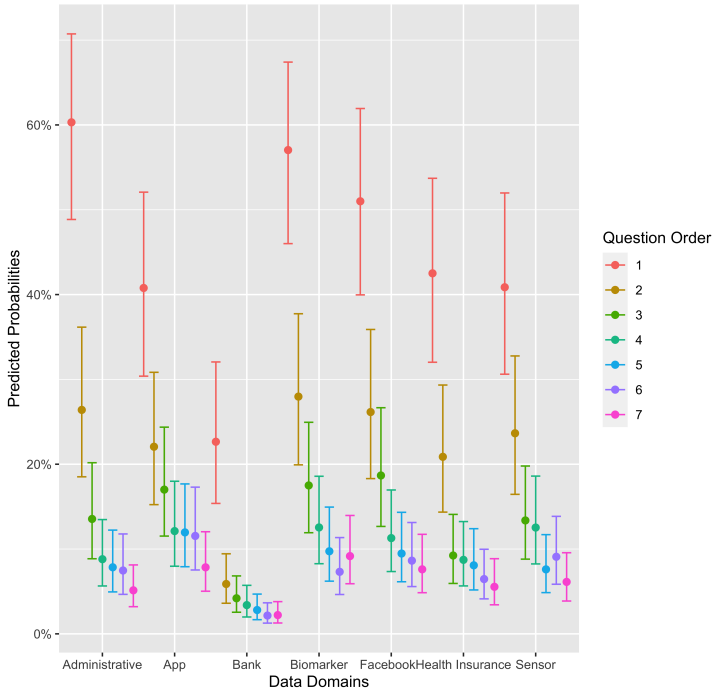


Figure 4.4: Predicted probabilities between Data Domain and Question Order (see Model 2 in Table B4.2)

None of the interactions between data domain and question wording are significant ($p > .1$, see Table B4.2). Concerning the interaction between data domain and incentive, we find significant interactions between Facebook data and incentive group ($p = .014$, see Figure 4.5 and Table B4.2) and between app data and incentive ($p = .031$, see Figure 4.5 and Table B4.2).

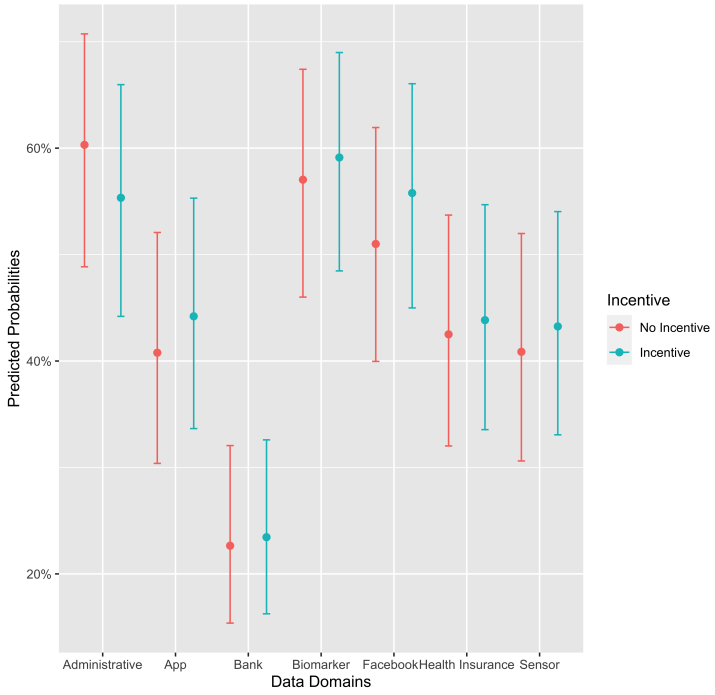


Figure 4.5: Predicted probabilities between Data Domain and Incentive (see Model 2 in Table B4.2)

4.6 Discussion

In this paper, we investigated respondents' willingness to consent to data linkage for seven different data domains in a web survey. We found variation in the likelihood to consent depending on the data domain and the position at which the data linking request was asked. We found that asking for consent to link app data, biomarker data, and Facebook data to the survey responses created higher consent rates (all larger than 50%) compared to administrative data (63.0%), which did not significantly differ from consent likelihood to health insurance (46.5%) and sensor data (46.3%). By far the lowest likelihood to consent was observed for bank account data (33.8%). This finding is in line with research on the sensitivity of certain survey topics (Singh & Hill, 2003; Tourangeau & Yan, 2007), e.g., that financial information, in general, is considered to be more private and sensitive than for other data domains.

Furthermore, the results showed that the likelihood to consent to data linkage is negatively associated with question position. All data domains achieved significantly higher consent rates when the request came first in a sequence compared to when the same question was asked at a later position. Our result is in line with the findings by Keusch et al. (2019). The likelihood to consent further decreased with each additional request, showing that respondents may have the feeling of increasing privacy costs with every additional linkage of data they allow. This might be caused by the feeling that researchers can get a more complete picture of the respondent with every additional data domain linked. Interestingly, the decline in the consent rate with every additional request was not consistent across data domains, with some domains showing a steeper decline from the first to the second request than others. Together with findings from Walzenbach et al. (2022), we see a first indication that starting with a less sensitive data domain might be advantageous.

We further analyzed the influence of consent question wording and incentive on the likelihood to consent to data linkage. When considering the effect of consent question wording, we found that neither outlining the scientific benefit of the data linkage nor the time savings for respondents increased the consent rate compared to not providing any beneficial argument. We also found no interaction between question wording and data domain. Similar findings were reported by Sakshaug et al. (2013) and Pascale (2011), while other research found variations of question wording to increase consent rates Kreuter et al. (2015) and Sakshaug and Kreuter (2014). A possible explanation for this null effect could have been the separation of stimulus and question in our study, as the beneficial phrase was written on the introductory page of the consent module before the specific consent request(s) and not on the same page as the request(s). Additionally, we found no effect of an incentive, contradicting our assumptions. A similar null finding was reported by Jäckle et al. (2017), while Keusch et al. (2019) reported a positive effect of providing an incentive. Similar to the null effect of question wording, we assume that the separation between stimulus and consent question could have weakened the effect of the incentive. Additionally, the information about the incentive might not have

been specific enough, as we did not state the exact amount that participants would receive upon data linkage, which may have further reduced its effect. However, while the main effect of the incentive was non-significant, we found significant interactions between the incentive treatment and data domain. The willingness to consent to the linkage of Facebook data and app data was positively affected by the incentive.

Our study is not without limitations. First, we conducted our experiments with members of a non-probability online access panel. Even though quotas for sociodemographic characteristics were used, we do have to acknowledge that our respondents may differ from the general population in Germany in other characteristics, including their likelihood to consent to data linkage requests. While we might overestimate general consent rates, because members of a nonprobability online panel who have volunteered to regularly respond to surveys and are thus used to share a lot of information might be more willing to consent to a data linkage request, our goal was not to produce population estimates with this study. In contrast, we were mainly interested in identifying casual relationships through our experimental variations, and we assume that the effects of data domain and question order we found here may also hold for the general population. Second, the questionnaire we used was about 30 minutes long, with the consent and data linkage part being located toward the end of the questionnaire. This could have resulted in an underestimation of respondents' willingness to consent and may have had a negative effect on their response behavior in general due to respondents' fatigue (Sakshaug et al., 2013).

Nevertheless, we think that our study provides valuable insights for researchers who want to implement data linkage requests in a web survey. Understanding how respondents react to different consent requests and how different factors such as the data domain or question wording can affect consent decisions will help researchers to design the consent process in a way that maximizes consent rates. Our study is one of the first to compare consent to data linkage requests for a variety of different data domains. Considering our results, we recommend confronting respondents with a single consent request or, if necessary, to sort the consent re-

quest by sensitivity and/or importance for the researcher, starting with the ones that are least sensitive and in which the researcher is most interested. We further recommend, to continue the research on the effect of specific question wording and incentives, but to put this information closer to the consent request, as this design improvement will likely increase the effectiveness compared to our study.

References

Avendano, M. (2018). Can Biomarkers be collected in an Internet Survey? A Pilot Study in the LISS Panel 1. In M. Das, P. Ester, & L. Kaczmirek (Eds.), *Social and behavioral research and the internet* (pp. 371–412). Routledge.

Baker, R., Shiels, C., Stevenson, K., Fraser, R., & Stone, M. (2000). What proportion of patients refuse consent to data collection from their records for research purposes? *British Journal of General Practice*, *50*, 655–656.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48.

Christoph, B., Müller, G., Gebhardt, D., Wenzig, C., Trappmann, M., Achatz, J., Tisch, A., & Gayer, C. (2008). Codebook and documentation of the panel study'Labour Market and Social Security'(PASS): Vol. 1: Introduction and overview, wave 1 (2006/2007). *Institut für Arbeitsmarkt-und Berufsforschung (IAB), Nürnberg*.

Jäckle, A., Burton, J., Couper, M. P., & Lessof, C. (2017). Participation in a mobile app survey to collect expenditure data as part of a large-scale probability household panel: Response rates and response biases. *Institute for Social and Economic Research, University of Essex: Understanding Society Working Paper Series No*, *9*.

Keusch, F., Struminskaya, B., Antoun, C., Couper, M. P., & Kreuter, F. (2019). Willingness to participate in passive mobile data collection. *Public opinion quarterly*, *83*, 210–235.

Kreuter, F., Sakshaug, J. W., & Tourangeau, R. (2015). The framing of the record linkage consent question. *International journal of public opinion research*, *28*, 142–152.

Link, M. W., Murphy, J., Schober, M. F., Buskirk, T. D., Hunter Childs, J., & Langer Tesfaye, C. (2014). Mobile technologies for conducting, augmenting and potentially replacing surveys: Executive summary of the AAPOR task force on emerging technologies in public opinion research. *Public Opinion Quarterly*, *78*, 779–787.

Martin, K., & Shilton, K. (2016). Putting mobile application privacy in context: An empirical study of user privacy expectations for mobile devices. *The Information Society*, *32*, 200–216.

McFall, S. L., Conolly, A., & Burton, J. (2014). Collecting biomarkers using trained interviewers. Lessons learned from a pilot study. *Survey research methods*, *8*, 57–66.

Nissenbaum, H. (2018). Respecting Context to Protect Privacy: Why Meaning Matters. *Science and Engineering Ethics*, *24*, 831–852.

Nissenbaum, H. (2019). Contextual Integrity Up and Down the Data Food Chain. *Theoretical Inquiries in Law*, *20*, 221–256.

Pascale, J. (2011). Requesting Consent to Link Survey Data to Administrative Records: Results from a Split-Ballot Experiment in the Survey of Health Insurance and Program Participation (SHIPP). *Survey Methodology*, *3*.

R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Revilla, M., Couper, M. P., & Ochoa, C. (2019). Willingness of online panelists to perform additional tasks. *Methods, data, analyses: a journal for quantitative methods and survey methodology*, *13*, 223–252.

Sakshaug, J. W., & Kreuter, F. (2014). The effect of benefit wording on consent to link survey and administrative records in a web survey. *Public Opinion Quarterly*, *78*, 166–176.

Sakshaug, J. W., Couper, M. P., Ofstedal, M. B., & Weir, D. R. (2012). Linking survey and administrative records. *Sociological Methods & Research*, *41*, 535–569.

Sakshaug, J. W., Schmucker, A., Kreuter, F., Couper, M. P., & Singer, E. (2019). The effect of framing and placement on linkage consent. *Public opinion quarterly*, *83*, 289–308.

Sakshaug, J. W., Tutz, V., & Kreuter, F. (2013). Placement, wording, and interviewers: Identifying correlates of consent to link survey and administrative data. *Survey Research Methods*, *7*, 133–144.

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*.

Singer, E., & Ye, C. (2013). The use and effects of incentives in surveys. *The ANNALS of the American Academy of Political and Social Science*, *645*, 112–141.

Singh, T., & Hill, M. E. (2003). Consumer privacy and the Internet in Europe: A view from Germany. *Journal of consumer marketing*, *20*, 634–651.

Struminskaya, B., Toepoel, V., Lugtig, P., Haan, M., Luiten, A., & Schouten, B. (2020). Understanding willingness to share smartphone-sensor data. *Public Opinion Quarterly*, *84*, 725–759.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological bulletin*, *133*, 859–883.

Walzenbach, S., Burton, J., Couper, M. P., Crossley, T. F., & Jäckle, A. (2022). Experiments on multiple requests for consent to data linkage in surveys. *Journal of Survey Statistics and Methodology*, Advance online publication.

Wang, L., Zhang, Z., McArdle, J. J., & Salthouse, T. A. (2008). Investigating Ceiling Effects in Longitudinal Data Analysis. *Multivariate Behavioral Research*, *43*, 476–496.

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*, 1–22.

Appendix

Appendix A: Questionnaire Wording

Introductory Text

[In addition to your responses/ For scientific purposes/ In order to shorten the survey duration and save you some questions,] we would like to collect some data in addition to your answers that are of interest for our evaluation. In this way, you support our research and make a valuable contribution to scientific progress.

In order to combine these data with your survey data, we will ask for your consent in each case. When evaluating the data, we absolutely ensure that all data protection regulations are complied with and that the data are not passed on to third parties.

Your consent is, of course, voluntary. You can revoke it at any time at onlinestudie2018@gesis.org.

In the following, you will be asked in each case for your consent to the use of this data. You will then be randomly selected for certain data and will receive detailed information on the further course of the study. [As compensation for the additional effort, you will receive additional mingle points from us.]

Consent Requests

Administrative Data

We would like to collect some data which we can obtain from the competent authorities. These include, for example, information on previous employment relationships, periods of unemployment and participation in measures during unemployment, and characteristics of your employer. We use these data to explore the increasing complexity of work in our society.
Do you agree?

Yes, I agree
No, I don't agree.

App Data

We would like to collect some data that we collect with a program (an "app") on your smartphone. For example, we may collect information about the frequency of smartphone use, the number of apps used, or other aspects of usage behavior. With the help of this data, we investigate human behavior in an increasingly digitalized world.
Do you agree?
Yes, I agree
No, I don't agree.

Bank Data

We would like to collect some data that we request from your bank with your consent. This includes, for example, information on consumption and savings behavior as well as income levels. With the help of this data, we investigate consumption and saving behavior.
Do you agree?
Yes, I agree
No, I don't agree.

Biomarker

We would like to collect some data, which we determine with the help of saliva and blood samples. This includes, for example, information on environmental pollution as well as data on hormonal values. We use these data to investigate the relationship between environmental conditions and health status.
Do you agree?
Yes, I agree
No, I don't agree.

Facebook Data

We would like to collect some data from your Facebook account. This includes key data such as the number of friends or the number of posts on your own or on other people's walls. With the help of this data, we investigate the changes in interaction through digital media and the Internet.
Do you agree?
Yes, I agree
No, I don't agree.

Health Insurance Data

We would like to collect some data, which we request from your health insurance company with your consent. This includes, for example, information such as prescribed medication, examinations performed or the frequency of visits to the doctor. With the help of this data, we investigate differences between statutory and private health insurance companies.
Do you agree?
Yes, I agree
No, I don't agree.

Sensor Data

We would like to collect some data that using the sensors on your smartphone. These include GPS location data, ambient brightness information, and air pressure measurements. With the help of these data we investigate the influence of your environment on your behavior.
Do you agree?
Yes, I agree
No, I don't agree.

Appendix B: Supplementary Tables

Table B4.1: Estimates and standard errors (in parentheses) from multilevel regression predicting consent to data linkage (Model 1)

	Estimate	(S.E.)
H1: Data Domain (Reference: Administrative Data)		
App Data	0.201***	(0.078)
Bank Data	-1.319***	(0.086)
Biomarker Data	0.326***	(0.077)
Facebook Data	0.237***	(0.078)
Health Insurance Data	-0.065	(0.079)
Sensor Data	0.038	(0.078)
H2.1: Question Position: (Reference: Position 1)		
Position 2	-1.121***	(0.072)
Position 3	-1.724***	(0.075)
Position 4	-2.052***	(0.077)
Position 5	-2.250***	(0.079)
Position 6	-2.374***	(0.080)
Position 7	-2.567***	(0.081)
H2.2: Question Wording (Reference: No Benefit)		
Science Benefit	0.081	(0.141)
Time Benefit	-0.078	(0.141)
H2.3: Incentive (Reference: No Incentive)	0.058	(0.118)
Female	0.741***	(0.117)
Education (Reference: High)		
Medium	0.234*	(0.126)
Low	0.416**	(0.176)
Device Mobile (Reference: Desktop/Laptop)	-0.228*	(0.117)
Constant	-0.153	(0.171)
Observations	23,177	
Log Likelihood	-9,589.202	
Akaike Inf. Crit.	19,220.400	
Bayesian Inf. Crit.	19,389.470	

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table B4.2: Estimates and standard errors (in parentheses) from multilevel regression predicting consent to data linkage including interactions (Model 2)

	Estimate	(S.E.)
H1: Data Domain (Reference: Administrative Data)		
App Data	-0.791***	(0.257)
Bank Data	-1.647***	(0.265)
Biomarker Data	-0.135	(0.252)
Facebook Data	-0.378	(0.252)
Health Insurance Data	-0.720***	(0.255)
Sensor Data	-0.788***	(0.254)
H2.1: Question Position: (Reference: Position 1)		
Position 2	-1.443***	(0.211)
Position 3	-2.272***	(0.221)
Position 4	-2.756***	(0.222)
Position 5	-2.882***	(0.232)
Position 6	-2.935***	(0.235)
Position 7	-3.335***	(0.233)
H2.2: Question Wording (Reference: No Benefit)		
Science Benefit	0.134	(0.192)
Time Benefit	0.086	(0.193)
H2.3: Incentive (Reference: No Incentive)	-0.204	(0.160)
Female	0.739***	(0.117)
Education (Reference: High)		
Medium	0.228*	(0.126)
Low	0.415**	(0.176)
Device Mobile (Reference: Desktop/Laptop)	-0.235**	(0.117)
Interaction Effect: Data Domain * Question Position		
App Data * Position 2	0.554*	(0.302)
Bank Data * Position 2	-0.101	(0.314)
Biomarker Data * Position 2	0.214	(0.295)
Facebook Data * Position 2	0.365	(0.301)

Continued on next page

Table B4.2 – *Continued from previous page*

	Estimate	(S.E.)
Health Insurance Data * Position 2	0.413	(0.297)
Sensor Data * Position 2	0.641**	(0.300)
App Data * Position 3	1.060***	(0.306)
Bank Data * Position 3	0.372	(0.325)
Biomarker Data * Position 3	0.438	(0.305)
Facebook Data * Position 3	0.760**	(0.308)
Health Insurance Data * Position 3	0.289	(0.313)
Sensor Data * Position 3	0.773**	(0.310)
App Data * Position 4	1.148***	(0.313)
Bank Data * Position 4	0.634*	(0.335)
Biomarker Data * Position 4	0.531*	(0.310)
Facebook Data * Position 4	0.655**	(0.311)
Health Insurance Data * Position 4	0.711**	(0.311)
Sensor Data * Position 4	1.183***	(0.309)
App Data * Position 5	1.259***	(0.319)
Bank Data * Position 5	0.566*	(0.338)
Biomarker Data * Position 5	0.372	(0.323)
Facebook Data * Position 5	0.584*	(0.317)
Health Insurance Data * Position 5	0.752**	(0.319)
Sensor Data * Position 5	0.753**	(0.321)
App Data * Position 6	1.272***	(0.326)
Bank Data * Position 6	0.349	(0.341)
Biomarker Data * Position 6	0.112	(0.325)
Facebook Data * Position 6	0.535*	(0.320)
Health Insurance * Position 6	0.564*	(0.322)
Sensor Data * Position 6	1.001***	(0.321)
App Data * Position 7	1.244***	(0.323)
Bank Data * Position 7	0.775**	(0.343)
Biomarker Data * Position 7	0.758**	(0.322)
Facebook Data * Position 7	0.798**	(0.321)

Continued on next page

Table B4.2 – *Continued from previous page*

	Estimate	(S.E.)
Health Insurance Data * Position 7	0.802**	(0.330)
Sensor Data * Position 7	0.975***	(0.323)
Interaction Effect: Data Domain * Question Wording		
App Data * Science Benefit	-0.094	(0.191)
Bank Data * Science Benefit	-0.354*	(0.214)
Biomarker Data * Science Benefit	0.047	(0.192)
Facebook Data * Science Benefit	-0.137	(0.193)
Health Insurance * Science Benefit	0.178	(0.194)
Sensor Data * Science Benefit	-0.115	(0.193)
App Data * Time Benefit	-0.242	(0.193)
Bank Data * Time Benefit	-0.156	(0.213)
Biomarker Data * Time Benefit	-0.177	(0.194)
Facebook Data * Time Benefit	-0.238	(0.195)
Health Insurance Data * Time Benefit	-0.118	(0.197)
Sensor Data * Time Benefit	-0.143	(0.195)
Interaction Effect: Data Domain* Incentive		
App Data * Incentive	0.344**	(0.160)
Bank Data * Incentive	0.249	(0.177)
Biomarker Data * Incentive	0.290*	(0.160)
Facebook Data * Incentive	0.396**	(0.161)
Health Insurance Data * Incentive	0.259	(0.162)
Sensor Data * Incentive	0.302*	(0.161)
Constant	0.418*	(0.237)
Observations	23,177	
Log Likelihood	-9,546.980	
Akaike Inf. Crit.	19,243.960	
Bayesian Inf. Crit.	19,847.780	

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

5 Conclusion and Discussion

This dissertation aimed to apply experimental methods to novel areas of survey methodology. I showed how experiments can help to advance surveys, and how the evidence gathered can be used to derive concrete recommendations for practitioners. In this final chapter, I will summarize and discuss the findings and conclusions of the studies described before. I will also highlight some of the limitations of the studies and their implications for experimental research.

The study in chapter 2 focused on a fully crossed experiment designed to investigate the recruitment of respondents drawn from a probability sample in a self-administered mixed-mode survey. Respondents were randomly assigned to one of two modes (i.e, concurrent and sequential) and one of four incentive conditions (i.e, no incentive, 1€ prepaid, 2€ prepaid, 2€ delayed). The results showed that different combinations of mode sequence and incentive worked best for different age groups. Younger respondents were most likely to participate when a delayed incentive was combined with a sequential mode sequence, while a 2€ prepaid incentive combined with a concurrent mode sequence turned out to be the best design for respondents aged above 50. Using these designs for the respective groups also lead to a minimization of survey costs. However, the findings of this study are based on a small survey conducted amongst the population of a minor German city which can lead to issues with regards to generalization. This limitation highlights a general problem with experimental research. Experiments can be underpowered when the sample drawn by researchers is too small. Further sample size can shrink due to self-selection or due to a treatment effect. In the study in chapter 2 specific combinations of treatments led to higher survey participation and thus to larger sample sizes in the respective groups.

The second study of this dissertation focused on the effects of smartphone use in surveys and recall aids on network name generators. 3,374 respondents were recruited from a German online access panel and independently randomly allocated to one of two device groups (i.e, smartphone and PC) and to one of two recall aid conditions (i.e., recall aid question before the network name generator,

recall aid question before the network name generator). The results did not show a significant difference between the number of friends reported on smartphones and on PCs. Additionally, no significant effect of the recall aid question on the number of friends reported could be found. However, I was able to use the data generated by the recall aid question as an indicator of satisficing. The study was severely limited by the fact that the smartphone condition caused a high number of screen-outs and dropouts in the beginning of the survey. This highlights another aspect of experiments that researchers have to be aware of: some experimental treatments might be unacceptable to a large number of participants, and thus lead to increased dropout rates. In extreme cases this can produce biases between experimental groups and confound the experiment. In the experiment at hand in chapter 3, the need to switch to a smartphone to fill in the survey, was considered as too burdensome by many respondents, so that consequently they decided to not fill in the survey.

The final study in chapter 4 described an experiment on the willingness of respondents to share additional data from non-survey sources. Respondents were recruited from a German online access panel and randomly allocated to one of three question wordings (i.e., time saving benefit, scientific benefit, no benefit) and one of two incentive groups (i.e., promised incentive, no incentive). Further they were confronted with consent requests for 7 data domains in random order. Those encompass administrative data, smartphone usage data, bank data, biomarkers, Facebook data, health insurance data, and sensor data. The results showed that data domain had a significant impact on whether a respondent was willing to share data or not. Additionally, I found only question order to have a significant impact on consent decisions. This specifically means that the first question asked, resulted in higher consent rates then later consent requests. Regarding experimental research this study highlighted the importance of the relationship between treatment and measurement. I found neither the question wording nor the incentive condition to have an effect. I assume that this in part might be caused by the fact that both treatments were presented to respondents on questionnaire pages previous to the consent requests. Consequently, respondents might have already forgotten the treatment effect.

Overall, this dissertation has shown that experimental methods are an invaluable way to generate empirical evidence and derive knowledge on causal effects. However, implementing ideal experimental designs in practice often comes with challenges eventually leading to limitations. First, researchers have to plan ahead and consider the effect size of their treatments, when calculating sample sizes. The implementation of certain treatments can lead to non-participation and leave researchers with underpowered studies incapable of conducting valid statistical testing. Second, researchers have to consider if certain treatments are feasible. Respondents can refuse participation at any given time in a study. Hence, treatments with a high associated burden can cause large numbers of dropouts and bias the data resulting from the experiment. Finally, the treatment has to be strongly associated to the point of measurement, keeping it salient enough in respondents minds to have a potential effect. If treatment and measurement are separated to far, the effect of the treatment on respondents might vanish, thus making the effect size inestimable.

This dissertation was able to deliver evidence from three different studies on the use of experimental methods in survey research. While I addressed challenges of the implementation of experiments within questionnaires and social science studies, the ability of generating causal evidence is indispensable for survey research and science in general. When researchers design their experiments carefully and implement them while keeping the practical shortcomings in mind they pose a valuable tool to further advance science.

Further research should focus on improving experimental designs by addressing the issues investigated in this dissertation. This especially means finding innovative ways of conducting experiments including decisions about the usage of devices (e.g., smartphone studies) and if not possible, trying to use novel statistical approaches, e.g. matching to draw casual inference. Additionally, a variety of improvements can be made to experiments regarding consent decisions, e.g. using different wording for consent requests or simplifying the linkage process. Focusing on improving willingness-to-consent to the linkage of additional non-survey

data might be invaluable to social science disciplines. Finally, researchers should always review their experiments and report unintended effects of their designs, to allow the further development of experimental methods in general.