A Penny for your Thoughts
The Use of Cash Incentives in Face-to-Face Surveys

BY

Ulrich Krieger

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

Dekan der Fakultät für Sozialwissenschaften
Prof. Dr. Michel Diehl

Betreuerin
Prof. Dr. Annelies G. Blom

Gutachterin und Gutachter
Prof. Dr. Annelies G. Blom
dr. Peter Lugtig
Prof. Dr. Christof Wolf

Tag der Disputation: 27. Februar 2018

Contents

List of Figures

List of Tables

To Katrin. And Mika, Zora and Karla.

Acknowledgments

This dissertation has been years in the making. Accordingly, there are many people to thank for their input in this work.

First and foremost I express my gratitude towards my advisor Annelies G. Blom. She has been a constant force pushing for improvements, continuation and completion of the dissertation. She has been there for me constantly with kind words of encouragement, helpful hints, and clever ideas. After all my years in academia she taught me the skills I lacked to tackle this project: structured writing and tenacity. While I haven't completely mastered those skills, I am thankful for what I have learned and for the companionship along the path to completion of the thesis.

I am thankful to my whole team at the German Internet Panel. We have created a unique environment over the past five years producing a high quality data product and enabling insightful survey research. This dissertation is one of four dissertations completed during our time in this project. This is a great achievement considering our jobs come with a high workload doing the actual work that keeps the GIP running. I thank Annette Holthausen, Franziska Gebhard, Dayana Sieger, Daniela Ackermann-Piek, Carina Cornesse, Susanne Helmschrott, Jessica Herzing, Barbara Felderer and Christian Bruch (characters in order of their appearance).

And I must not forget to thank the one and only boss I ever called 'der Chef", Axel Börsch-Supan. Our cooperation was rocky due to outside influences (move to Munich, the "annus horribilis" in German survey fieldwork and the incentive troubles). However, he was a great boss, supportive, loyal and fair. Thank you for all you have done to support me and my work.

Many thanks to the SHARE Team in Munich. While running an excellent study, they also make fine colleagues and friends. Thanks to Sabrina Zuber,

Stephanie Stuck, Johanna Bristle, Marina Brand, Frederik Malter, Barbara Schaan and Christina Diemand. And a very special thank you to Julie Korbmacher.

My student assistants have worked countless hours formatting text and tables, going over Stata code, and fight for the right bibliography format. I tip my hat to you, Ina Holdik, Johannes Bär, Marina Fikel, Peter Derheim, Melvin John and Felix Schmidt.

There have been many researchers inspiring me and my work. The community meeting on the "survey methods conference circus" is a very friendly, supportive and inclusive group and I am happy to be a part of that. Meeting Jürgen Schupp, Oliver Lipps, Peter Lugtig, Mick Couper, Stephanie Eckman, Frauke Kreuter, Mark Trappmann (in no particular order) and many more on RC 33 conferences, ESRA conferences, Nonresponse Workshops, and Panel Survey Methods Workshops was inspirational and helpful for my work as a whole. Thank you all. And thanks to those who organize those conferences, that can't be expressed too often.

Thank you to everyone in the collaborative research center 884. Without you, neither the GIP nor this work would exist. I especially thank Thomas König, Margarita Maklakova, Marina Jesse, Martin Eichler and Jason Eichorst.

Special thanks to Klaus Pforr, the best colleague I ever had. And a true friend. You have always been supportive and positive about my work. It is good to know that you're there and helping me talk over survey topics and beyond. And, he even hunts down that odd reference just because he can. Vielen lieben Dank!

Leaving the professional field, I am supported by a great family. I am thankful for the loving support of my wife Katrin, and for not mentioning the thesis for a while when that was what I needed. Thanks for choosing me and being with me, I could not have done it without you.

How could I not thank my parents? They're just great parents. They never let you down and are always there for you. Being a parent now myself, I think I know what I am talking about: You did a heck of a job.

In addition, I would like to thank my kids, Mika, Zora and Karla, what a fine combo. Thanks for being in my life and making my life less boring and less about survey methods.

And as I have gone so far, thanks to my in-laws and brothers in law.

Finally, thank you to all those, I forgot to mention.

I've got ninety thousand pounds in my pyjama.
I've got forty thousand French francs in the fridge.
I've got lots of lovely lira,
Now the deutschmark's getting dearer,
and my dollar bills could buy the Brooklyn Bridge.

Eric Idle, The Money Song

1 Introduction

It seems the scientific survey is in a bad shape these days. Polls are deemed unreliable to predict election outcomes (Skibba, 2016), Big Data analysis promises to deliver insights more cheaply and reliably than survey results (Japec et al., 2015) and respondents are just unable to be convinced to participate in surveys (Meyer et al., 2015).

At the same time there is an increasing number of large scale surveys in Germany (Nationale Kohorte, the National Education Panel Study and the Programme for the International Assessment of Adult Competencies - Longitudinal to name just a few) (Rat für Sozial- und WirtschaftsDaten, 2017), election polls are still the best predictor for election outcomes (Kennedy et al., 2017) and while response rates are still low we seem to have reached the bottom (Keeter et al., 2017) some panel surveys still register very high retention rates (Schoeni et al., 2013).

I conclude that there still is a demand for the kind of knowledge only old fashioned probability based surveys can generate: Insights on how the public feels about a topic, capturing trends and developments in societies and generate data

of high quality on these for scientific research (Couper, 2013).

While there is a demand for surveys and their performance being much better than how they are currently perceived there are undoubtedly challenges for survey research, of which survey nonresponse "is increasingly becoming the most pressing issue"(Couper, 2013, p. 153).

Nonresponse increases the likelihood of not observing important information about the population in a survey. The resulting data can then be biased resulting in nonresponse bias (see Groves & Couper, 2012) (see section 2.1).

One measure to counter nonresponse in surveys is the use of respondent incentives (Pforr et al., 2015). Respondent incentives are items, gifts or money presented to respondents as encouragement to take part in a survey (Boulianne, 2008). Incentives can take on many forms, however my dissertation will focus on cash incentives as I have experimental data on this incentive form.

When I entered the world of survey research in 2006 and visited conferences I became fascinated with presentations on incentive use by international colleagues (e.g. at the MOLS conference, (Laurie & Lynn, 2009)).

Research on how incentives improve response rates have a long history in survey methodology. In an early meta-study by Armstrong (1975), which collects studies published as early as 1931 (Shuttleworth, 1931) two of the fundamental questions on the use of incentives are examined. First, what is the optimal value of an incentive? Researchers want to pay as little as possible to keep the survey budget small but they want the best possible results on outcomes like response rates. And second, should the researcher offer an incentive before or after the interview? Offering the incentive prior to the interview may be more efficient but also more expensive as more sample members receive an incentive. Armstrong concludes that "The only type of monetary incentive that has an impact on the nonresponse rate is the prepaid monetary incentive"(Armstrong, 1975, p. 116).

However, the research on incentive use continues until the present day and the questions of "how much" and "when" are still being examined. In chapter 3 I introduce a comprehensive overview on types or modes of respondent incentives. As will be shown in chapter 3, there are many different kinds of incentive designs and effects can differ between these incentive designs.

Whereas incentives where not used regularly in Germany (Pforr et al., 2015),

reports on incentive success by scholars and practitioners from the United States or Great Britain were very encouraging (for examples and citations see section 3). However, German surveys were late to adopt incentive use (Pforr et al., 2015).

Later in my career, I was fortunate to work on two projects evaluating the use of respondent incentives in the Survey of Health, Ageing and Retirement in Europe (SHARE) (Börsch-Supan, Krieger, & Schröder, 2013) and the German Internet Panel (GIP) (Blom et al., 2015). Both studies are being described as data sources in chapter 5.

The main body of this dissertation are three experiments on respondent incentives.

The incentive experiments have similarities but also cover different aspects of incentive use. All experiments are conducted in the first wave of a longitudinal study with the goal to motivate respondents not only to participate in one survey, but to stay for the long haul. All experiments include at least one condition, offering cash incentives upfront without any strings attached (unconditional cash incentives).

The first experiment I present in chapter 5 is comparing unconditional cash incentives to conditional cash incentives in the GIP recruitment survey. I present effects on response rate and sample composition. The added twist to this experiment is that the values of both types of incentive are chosen in a way that their total cost is the same: paying € 10 offered conditionally to about half the (responding) sample members compared to € 5 sent unconditionally to all sample members. Incentive effects in this experiment are thus measured under the realistic assumption of a fixed survey budged, that needs to be used most effectively on an incentive scheme.

The cost aspect is expanded on chapter 6 when discussing the GIP reminder incentive experiment. Here I compare registration rates to the GIP online panel conditional on receiving an unconditional cash incentive with a reminder letter to just receiving the reminder. In addition to reporting on registration rates and sample composition I examine if the use of the cash incentive was cost effective for the GIP project.

Finally, I included the results of the SHARE Germany wave 4 incentive experiment form 2011. Parts of the results presented in chapter 7 are taken from a working paper on this topic (Börsch-Supan, Krieger, & Schröder, 2013). However, the

results fit nicely into the present volume and the presentation has been improved and the findings are still relevant for todays survey researchers:Testing unconditional cash incentives of high value in a setting of a large, national face-to-face study.

There is nothing quite as wonderful as money!
There is nothing like a newly minted pound!

Eric Idle, The Money Song

2

Theory on incentives and survey response

In this chapter I will explain why incentives are being used in surveys. The purpose of respondent incentives is to influence respondents' behavior, to encourage them to take part in the survey. In this process, as (supposedly) more sample members take the survey, survey participation increases and survey nonresponse decreases.

In a meta-study on the impact of nonresponse rates on nonresponse bias, Groves & Peytcheva (2008) conclude that there is no automatic link between response rates and nonresponse bias. While high response rates reduce the risk of nonresponse bias, "some surveys with low nonresponse rates have estimates with high relative nonresponse bias"(Groves & Peytcheva, 2008, p. 183). In section 2.1 I will show why survey participation is important for survey quality by discussing the relationship between nonresponse and nonresponse bias, a form of survey error.

While there is no direct link between non-participation and survey error, I will show that large numbers of nonparticipants increase the likelihood of the occurrence of such an error.

After establishing the importance of minimizing nonresponse (or more precisely the nonresponse bias) I will focus my attention on why incentives can help counter non-participation in surveys. For this purpose I look at theories of response behavior (see section 2.2.4) in order to examine why sample members take part in a survey at all. Theories can then help explain the mechanisms of how and why survey incentives can be used to nudge respondents into participating.

In section 3 I will discuss the relevant literature on how incentives have been used in the past. Different forms of incentive use and their effects will be described. Past research will also help us discover successful incentive use in the past.

2.1 Nonresponse and nonresponse bias

It can be assumed that with the first surveys in history came the first nonrespondents. Whenever surveys are conducted, some members of the sample will refuse to cooperate. This even holds true for mandatory surveys such as the German Mikrozensus, where refusing is considered a misdemeanor. For example, in 1996, 2,356 units out of 58,084 refused to cooperate in the State of Hesse alone, from which 91 ended up being ordered to pay a fine. Encouragingly, in a report to the Hesse parliament the administrators state that they are not seeking jail sentences in these cases (Al-Wazir & Staatskanzlei Hessen, 1997). This example is just to illustrate that there will always be nonresponse. The task at hand is to minimize it as much as possible.

In this section the term nonresponse will be defined and the potential negative effects on data quality will be explained.

2.1.1 (Unit) nonresponse

I distinguish several types of Nonresponse (Goyder, 2008). If just some answers to questions are not gathered I speak of item nonresponse. If the sampled unit drops out altogether, this is called unit nonresponse. Attrition is a special case of unit nonresponse, where a sampled unit that once participated in a survey drops out from subsequent panel waves. In this chapter, I concentrate on unit nonresponse. The terms nonresponse and unit nonresponse are used interchangeably.

Nonresponse rates are defined as the proportion of cases in the samle that are invited to participate but do not participate due to refusal, absence, illness and so on.

Next, I examine the consequences of this non-participation, the error induced by nonrespondents.

As straightforward as this may sound, the definition of what constitutes an eligible sample member can be debated and has lead to establishing common definitions (AAPOR, 2016). The sampling frame sometimes contains sample members that are ruled ineligible during fieldwork. If, for example, private households are sampled and an interviewer finds a business at a sampled address the unit is no longer followed and ruled ineligible for the survey. Ineligible sample units are then taken out of the equation when calculating response rates. Generally, response rates can be written as

$$RR1 = \frac{I}{(I+P)+(R+NC+O)+(UH+UO)} \tag{2.1}$$

where I are complete interviews which are divided by the sum of all eligible sample members: those that have been interviewed (complete and partial interviews(P), those that have not been interviewed (refusals and break-offs R, non-contacts NC, and other outcomes O) and those sample members with unknown eligibility (Unknown if household/occupied housing unit UH and Unknown, other UO) (AAPOR, 2016)).

Missing from the equation are ineligible cases, they are not used in the calculation of nonresponse rates. It becomes clear why coding as many units from refusals and known eligibility into the non-eligible cases category will increase response rates: The cases are removed from the denominator of the equation.

To avoid confusion and give data producers a guideline, agreed upon definitions are being used. The most widely used definitions are the standard rates published by the American Association for Public Opinion Research (AAPOR) (2016).

2.2 How nonresponse may affect data quality

"The basis for unbiased inference from relatively small observed samples to large unobserved samples is probability sampling"(Singer, 2006, p. 637). If we find systematic nonresponse in a survey, the principles of probability sampling can be violated (Stoop, 2012). Nonresponse is systematic when response is correlated with survey variables. For example if sample members in a face-to-face study on fear of crime refuse to participate because they are afraid to open the door due to fear of crime. In this case respondents differ systematically from nonrespondents.

2.2.1 Nonresponse bias

We have learned so far how unit nonresponse is defined and how a bias introduced by nonresponse can affect data quality. I will now focus on a formalization of this nonresponse bias.

2.2.2 Measuring nonresponse bias

There are two ways of formalizing nonresponse rates, the deterministic and the probabilistic formula. In earlier work (see Groves & Couper, 2012), nonresponse bias was estimated with the deterministic formula. It is thereby assumed, that there are fixed, disjoint parts of samples, the respondents and the nonrespondents. They either participate or do not participate in surveys. The nonresponse bias of the unadjusted respondent mean $\overline{Y}_r$ can be formalized as

$$B(\bar{y}_r) = (\frac{M}{N})(\overline{Y}_r - \overline{Y}_m), \tag{2.2}$$

where M is the number of nonrespondents in the population and N the size of the target population, thus $(\frac{M}{N})$ is the nonresponse rate in the population. According to Groves & Couper (2012, p.3) following the idea of a fixed percentage of nonrespondents in the population and a fixed difference between respondents and nonrespondents implies a view that ensures the reduction of nonresponse bias only by increasing response rates.

Dropping the assumption of fixed sets of respondents and nonrespondents, the

probabilistic model of nonresponse was developed (Bethlehem, 2002; Bethlehem et al., 2011). Here the participation propensity ρ_k is introduced. All sample members have an unknown propensity to participate. Nonresponse is the result of a random selection "... when an element ... is selected in the sample, a random mechanism is activated that results with the probability ρ_k in response and $1 - \rho_k$ in nonresponse ..." (Bethlehem et al., 2011, pp. 43). The process is assumed to be random, as many factors determine the outcome, just a fraction of which can be observed. The nonresponse bias of the unadjusted respondents mean can then be written as

$$\overline{Y}_r = \overline{Y}_m + \frac{\sigma_{y\rho}}{\overline{\rho}}, \tag{2.3}$$

where $\sigma_{y\rho}$ is the covariance between y, the variable of interest in the survey statistic and ρ, the propensity to respond among units of the gross sample; and $\overline{\rho}$, the mean of all response probabilities in the population. The covariance measures, the joint variation of the variable of interest and the participation propensity and is given by

$$\sum_{i=1}^{N}(y_i - \bar{y})(\rho_i - \overline{\rho})/(N-1), \tag{2.4}$$

where y_i is the variable of interest for element i, $\bar{y}$ the mean of the variable of interest in the population and ρ_i the propensity to respond for element i. This "implies that when the likelihood of responding is strongly related to the variable of interest in the survey, then nonresponse bias for the respondent mean will be large" (Groves et al., 2013, p. 189). For nonresponse bias the size of nonresponse as well as the correlation between the variables of interest are important. A large nonresponse rate does not necessarily correspond to large nonresponse bias. Instead, the correlation between the nonresponse propensity and the variables of interest has to be examined (Groves, 2006).

For my research into the effect of incentives, the consequence of this research is twofold. First, a reduction in nonresponse rates through incentive use is maybe helpful in reducing the likelihood of nonresponse bias. Secondly, increasing survey participation does not necessarily correspond with a reduction in bias.

The effect of incentives has to be measured not only as its impact on the nonre-

sponse rate but also on how it affects nonresponse bias.

2.2.3 The continuum of resistance: describing nonrespondents

Who are the sample members which are not taking part in the survey? And, most importantly, how do we find out about their preferences and characteristics? As seen above we need information about both respondents and nonrespondents to get information on the size of the nonresponse bias.

Information on nonrespondents can sometimes be obtained from the sampling frame, i.e. from information provided when the sample was drawn. This information is valuable as it is available for both respondents and nonrespondents.

In the German survey context frame information is most often very limited as samples are commonly drawn from registers or random route procedures are applied (e.g see chapter 5).

Sometimes samples are drawn from groups that allow access to rich data sources like in the PASS survey (Trappmann et al., 2013) where an employment history of sample members is available. But these are rare circumstances.

A different approach to find out information on nonrespondents are nonresponse follow up studies. Here, in the event of noncooperation and noncontact, sample members are approached with a shortened questionnaire to gain some information on the cases that will not be in the survey (Stoop, 2005). The information from these short questionnaires can then be used to assess the size of the bias. While the idea is intriguing there are many practical issues. In some scenarios it might be unlawful to re-approach nonrespondents. In addition, interviewers might be tempted to field the shorter nonresponse questionnaire instead of the long form and thereby increase nonresponse.

The main problem with nonresponse studies is that there is nonresponse in nonresponse studies. Willingness to participate in such nonresponse surveys is often very low as most sample members have already refused the earlier survey requests. Thus there might be nonresponse bias within the nonresponse survey itself.

A different approach to find out about nonrespondents is proposed with the idea of a continuum of resistance to survey requests (Lin & Schaeffer, 1995). Assuming every sample member has a fixed, determined, unobserved resistance to

surveys. The higher the resistance, the harder it is to convince a respondent to take part in a survey. It will take more effort, e.g. more calls, more visits and a longer field time to get higher resistance sample members into the survey. Also it is assumed that this resistance continuum correlates with variables of interest.

If these assumptions hold true, nonrespondents should have similar outcome characteristics on said variables of interest as hard-to-reach respondents. Hard-to-reach respondents are characterized by requiring lots of visits or calls, requiring convincing strategies, maybe refusal conversion. Following the continuum theory they rank relatively high on the resistance variable. Nonrespondents have even higher values on this dimension, the increased field efforts that managed to reach difficult cases was simply not enough to realize interviews with these cases. In identifying these hard to reach cases we can isolate a subset of respondents that is closest in characteristics to nonrespondents.

The continuum approach is criticized for postulating heavy assumptions to hold true (Stoop, 2005). The most notable assumption is that interviewer effort is standardized and constant among all interviewers. This assumption will not hold in practice given the high differences in effort and success between interviewers. The assumption of a fixed continuum of resistance towards all surveys under all circumstances is also hard to justify (Schnell, 1997). As we will see in the next section, external factors play an important role in the decision to participate.

2.2.4 So, why do they respond? A very brief look at theories of survey participation

Explaining why sample members refuse to participate requires one to look at why respondents participate at all. Therefore, I will very briefly describe theories used to explain survey response in this section. The aim of this exercise is not to give a full and exhaustive rundown of theories. Other authors have done that already (Albaum et al., 1998; Schnell, 1997). Rather, in this section I want to outline theories and frameworks often used to explain survey response especially in conjunction with the use of incentives. These include social exchange theory, rational choice theory, and leverage-salience theory.

Social exchange theory was proposed by Homans (1958) explaining the dy-

namics of human behavior as exchanges between actors. These actors analyze costs and benefits of actions as well of expectations about certainty of returns to guide their behavior. Applying the theory to survey research (Dillman, 1978), the costs of answering a questionnaire were compared to the rewards, both non-monetary and monetary. In the application of the theory, the aspect of human interaction in the survey situation is emphasized. While costs and benefits are important, benefits and rewards can be drawn from the interaction itself. Dillman (1978) proposes that the interaction of the survey response itself, for example the conversation with an interviewer or expressing one's opinion can be rewarding in itself and outweigh the costs of the interview. Survey incentives can play a role in the evaluation of respondents costs and benefits. Nonetheless, exchange theory emphasizes the importance of human interaction.

In an economic rational choice theory framework, actors choose alternative behaviors such that it maximizes their subjective expected utility. In the classic theory, each and every alternative behavior option can be ranked by the actor according to the costs and benefits expected. From these benefits costs of the chosen alternative have to be deducted. Those can be direct costs, such as transportation costs to reach a test center or (more likely) indirect costs or opportunity cost, such as the opportunity costs of participating in a 90 minute household interview. Finally, the option maximizing the actors utility is chosen.

A direct consequence of this strict, classical formulation of this theory of action is that a rational actor rarely participates in a survey. The direct or indirect benefits of a typical survey are minimal to nonexistent and the (opportunity) costs of taking a survey will always be present. Only if the opportunity costs are virtually nonexistent (for example in the case of retirees, or the unemployed) survey participation is possible.

If the strict assumptions are relaxed and the cognitive effort reducing framing theory is assumed (Schnell, 1997), survey participation can be explained much more easily. In this approach, actors can choose actions from a reduced set of actions instead of performing extensive cost/benefit- calculations. If a doorstep approach by a SHARE interviewer is framed in terms of "helping out strangers"(e.g. interviewers), the actor just performs according to a script and helps out the person at their door.

The reciprocity norm, in contrast, emphasizes the moral duty to help others in return for a favor that has been provided. Small gifts or kind gestures evoke a feeling of duty to reciprocate the favor towards the giver. These reciprocity norms are believed to be universal and thus work across contexts, times and places (Gouldner, 1960). A survey request can be viewed in this context. Complying with the request is most often a low cost situation to the respondent, a favor provided to an interviewer or survey organization. Once the organization or the interviewer starts the interaction with a kind gesture (e.g. an unconditional incentive), reciprocity norm theory states that compliance with the following request is more likely.

The reciprocity norm is often cited when explanations for the effect of incentives are being sought for (Singer et al., 2000). There are two major predictions from this theory: one, incentives must be given unconditionally in advance to encourage respondents' cooperation and two, it is essential that incentives have the properties of a gift, a token of appreciation or a kind gesture. Only then can a respondent mirror the gesture with her survey participation. In a test of the reciprocity norm, Diekmann & Jann (2001) explicitly tested reciprocity theory. Prepaid telephone cards bearing the study design and photos of researchers were used as incentives. The individual design and the use of real photos of the researchers instead of stock photos enhanced the gift character of the incentive. Sending this gift out in advance increased the participation rate and was interpreted as supporting evidence for the validity of the reciprocity theory.

In an attempt to find a framework tailored to explaining survey participation Groves et al. (2000) proposed Leverage-Salience Theory. At the core of this concept is a person's underlying propensity to participate in a survey. On this propensity continuum a threshold is assumed that has to be cleared in order for an actor to participate in the survey. In order for the actor to make her decision regarding survey participation, two additional concepts come into play. Leverage describes the importance the sample person puts towards certain survey attributes. For example, helping research could be a very positive experience for this person, while not wasting her time is of prime importance.

The second concept, salience, is not influenced by the sample person but the survey organization or the interviewer. It denotes how salient certain aspects of the survey will be made to the respondent. An interviewer might tailor a interview

in a way to make the top notch research conducted with the data very salient and downplay the dreaded length of the interview.

By tailoring the survey approach in this way, the interviewer is more likely to realize an interview with the sampled person: Factors that make the interview more appealing are being made more salient and the reverse is true for aspects with negative leverage. Survey incentives can help in this framework if high leverage is attached to them and they are made salient to the respondents when the interviewer is making the survey pitch.

To sum up I showed how survey participation can be explained and what role survey incentives play in these theories. All theories predict respondent incentives to increase survey participation rate, only the mechanisms of how incentives improve response rates differ.

Exchange theory and reciprocity norm focuses on the social interaction of exchange as such. The more a situation is built like an exchange the more likely the success, meaning survey participation. Rational choice theory and leverage salience theory emphasize the weighting of benefits and costs when deciding about survey participation. In leverage salience theory there is an emphasis on the role of the design features and how important these are to respondents.

In the next section past research on incentives will be presented.

Everyone must hanker for the butchness of a banker,
it's accountancy that makes the world go round!

Eric Idle, The Money Song

3

Previous research on incentives in survey research

In this chapter I will discuss earlier research on incentives.

Generally, the term incentive is used to describe a benefit given to an actor in order to nudge him towards a desired behavior. In economics, incentives are a central concept when explaining behavior and options of actors (see e.g. Mankiw, 2004; Sowell, 2007). Incentives can be the result of existing structures (e.g. norms) or take on the form of deliberate planned action like laws, fines or benefits. Assuming that actors behave to maximize their utility, the description of an incentive structure of said actors is the foundation to explaining their behavior.

When manipulating the incentive structure of actors, cost efficiency is a guiding principle. Providing changes to the incentive structure should come with a return to the actor paying for the incentive.

There are many parallels to the way survey researchers use the term incentive. By providing an incentive a change or reassurance in a certain behavior (compliance

with the survey request) is intended and the incentive gift should be beneficial to the survey sponsor, meaning survey incentives should be cost efficient. When paying for incentives, survey organizations aim to gain a return they otherwise would not have obtained given budget restrictions.

Incentives used in survey research are used in many forms, modes and values. In this section, I provide an overview of incentive use in survey research. When considering incentives, these are defined as a gift to a sample person in a survey somehow connected to his or her survey participation: a pen offered at an interview, a $5 note included in an advance letter or a personalized ticket to a charitable lottery handed out by the interviewer after the interview. Many studies have examined the effects of the administration of different types of incentives in surveys.

In this dissertation, I will focus on respondent incentives and not consider incentives to interviewers or other survey organization staff. In addition, most incentives are not a payment or an expense allowance, the incentives described here have a gift character. That does not mean that they are of low value to the respondent. Just that the framing of the transaction is that of a gift or a token of appreciation rather than a business transaction. Also, I will primarily discuss findings of experiments in interviewer-mediated studies as this is most relevant to my own research.

To categorize the diverse research findings on incentive effects in surveys, I choose 8 categories to order the findings. A summary of studies and categories can be found in table 3.0.1.

In the first category I group studies investigating the effect of offering incentives compared to no incentives.

The second category contains studies comparing monetary incentives to non-monetary, in-kind incentives.

Studies on the optimal value of incentives make up the third category.

Studies in the forth category report effects of the payout form of monetary incentives such as cash, checks or vouchers.

In the fifth category I grouped studies discussing the payout timing of incentives, comparing prepaid or unconditional incentives to postpaid or conditional incentives.

Studies of incentives targeted towards a specific subgroup of respondents are

Table 3.0.1: Overview of studies discussed by incentive mode

No	Effects tested	Studies
1	Main effect of incentives	Blohm & Koch (2013), Castiglioni et al. (2008), Church (1993), Pforr et al. (2015), Schröder et al. (2013), Singer, Hoewyk, et al. (1999), Singer & Ye (2013)
2	In-kind and monetary incentives	Church (1993), Diekmann & Jann (2001), Gendall & Healey (2008), Ryu et al. (2005), Singer, Hoewyk, et al. (1999)
3	Optimal incentive value	Börsch-Supan, Krieger, & Schröder (2013), Scherpenzeel & Toepoel (2012), Schröder et al. (2013), Singer, Hoewyk, et al. (1999), Yu & Cooper (1983), Martin et al. (2014), Zagorsky & Rhoton (2008),Hsu et al. (2017), Mercer et al. (2015), Hanly et al. (2014)
4	Payment transaction: cash, check, voucher	Birnholtz et al. (2004), Börsch-Supan, Krieger, & Schröder (2013), Gajic et al. (2010), Hogan (2009), Schröder et al. (2013), Singer & Ye (2013), Warriner et al. (1996)
5	Prepaid versus promised incentives	Börsch-Supan, Krieger, & Schröder (2013), Castiglioni et al. (2008), Church (1993), Gelman et al. (2003), Scherpenzeel & Toepoel (2012), Singer, Hoewyk, et al. (1999), Kretschmer & Müller (2017),Hsu et al. (2017)
6	Targeted incentives	Abreu & Winters (1999), Cyffka et al. (2012), Juster & Suzman (1995), Kay et al. (2001), Laurie & Lynn (2009), Olsen (2005), Rodgers (2011), Singer, Groves, & Corning (1999)
7	Incentives in longitudinal surveys	Booker et al. (2011), Castiglioni et al. (2008), Goldenberg et al. (2009), Jäckle & Lynn (2008), Laurie & Lynn (2009), Singer et al. (1998), Pforr et al. (2015); Singer & Ye (2013)
8	Incentive use in Germany	Pforr et al. (2015), Schröder et al. (2013)

grouped in category six. Category seven summarizes the incentive use in longitudinal surveys.

Finally, I look into the use of incentives in Germany.

3.1 Using incentives compared to no incentives

Incentives are widely used in survey research to facilitate contact and cooperation with a survey request (Church, 1993). When studying their effectiveness the basic question is whether these incentives have any effect on survey outcomes in terms of response rates and resulting sample composition (Church, 1993; Pforr et al., 2015; Singer & Bossarte, 2006). Experimental research in the use of incentives has a long tradition (Armstrong, 1975). The effect of incentives to increase participation rates is well replicated across numerous surveys, countries and survey modes (Singer, Hoewyk, et al., 1999). The overall finding is that incentives tend to improve response rates in surveys (Singer & Ye, 2013).

In a meta-study Church (1993) finds that offering an incentive increases response rates on average by 13.2 percentage points across the 38 mail studies that were examined. Studies were identified using two databases (Sociological Abstracts, Psychological Abstracts) as well as two Journals (Public Opinion Quarterly and Journal of Marketing Research) and included references to response rates and incentives. In addition, theses studies contained a control group. As a caveat he notes that the overall effect of increased response rates due to incentive is largely mediated by the type of incentive treatment (See subsections 3.2 and 3.5).

Translating these findings from mail to interviewer mediated surveys, Singer, Hoewyk, et al. (1999) conducted a meta-analysis of 39 experiments in face-to-face and telephone studies. Using an incentive increased response rates. "On average, each dollar of incentive paid results in about a third of percentage point difference between incentive and the zero-incentive condition"(Singer, Hoewyk, et al., 1999, p. 223). The findings are robust when controlling for the characteristics of the incentive and burden of the study. This effect is also reported in a review article by Singer & Ye (2013) for numerous studies. Note that most of the studies were conducted by mail or by telephone and the evidence on large scale face-to-face studies is more limited (Pforr et al., 2015).

In a face-to-face study, the German Allbus survey, Blohm & Koch (2013) report a significant 3.6 percentage point difference when testing no incentive versus a promised € 10 cash incentive.

Contrary to these findings, the pilot study of the German Family Panel (pairfam), the Minipanel, did not find an incentive effect (Castiglioni et al., 2008). 600 Respondents were approached in four cities with a survey request to partake in the Minipanel study. Shopping vouchers valued at € 10 were offered prepaid or promised to random subsets of the sample and can be compared to a no incentive group. In the first wave, no significant difference between the incentive and the no incentive conditions were observed.

In some experiments the no incentive condition is not even tested as no incentive is not seen as a viable option for survey research (e.g. Schröder et al., 2013). Overall, the literature finds that incentives improve response rates in surveys.

3.2 Monetary incentives and in-kind incentives

There are monetary and non-monetary, in-kind incentives. Monetary incentives have a clear monetary value attached such as cash or a check. But incentives can also come in the form of gifts or presents. These objects like a pen or a stuffed animal do not come with a clear monetary value attached. An intermediate form of incentives are incentives such as store vouchers, stamps or phone cards. As store vouchers can be used like cash I will discuss them as monetary incentives. The usability of stamps and phone cards is limited to the postal service or a phone company, therefore I will regard them as in-kind incentives, though this might depend on the cultural context. In many African countries phone cards are used for transferring money across long distances (see e.g. Meuleman et al. (2017))

In-kind incentives are often used at the beginning of the interaction with respondents. The German Socio-Economic Panel for example used these kinds of incentives called "door opener" (Schröder et al., 2013). Here interviewers have a budget to pay for these presents and it is at their discretion if and which incentives they use.

In other studies, gifts like a pen or an elaborated mix of children's toys are the only incentives (De Luca & Lipps, 2005) (N. Biedinger, personal communication,

14.2.2014). The rationale behind the decision not to use cash can be justified with norms of reciprocity. Offering a small, symbolic present to the respondent is thought to improve the likeliness to return a favor. This interaction is perceived not to be an economic exchange but a mere gesture of kindness (Diekmann & Jann, 2001).

In a rare controlled experiment testing of in-kind against monetary incentives, Ryu et al. (2005) confirmed cash incentives to be more effective in raising response rates than in-kind incentives. Accordingly, the meta-studies by Church (1993) and Singer, Hoewyk, et al. (1999) report that monetary incentives are more effective in raising response rates than in-kind incentives.

When faced with new regulations of the postal service which forbid sending cash via mail, Gendall & Healey (2008) experimented with the use of in-kind incentives to replace their former $1 incentive. In-kind incentives improved response rates but not with the consistency and to the same level as cash incentives. In todays research practice in the social sciences, cash incentives can be observed much more commonly than in-kind incentives (Pforr et al., 2015; Singer, Hoewyk, et al., 1999; Singer & Ye, 2013). In addition to the greater effect on response rates they are easier to obtain, store, ship and return.

3.3 What is the optimal value of incentives?

The effect of different values of incentives on the response rate has also been researched, especially in face-to-face and telephone modes. Finding the optimal incentive amount is a crucial question when trying to optimize a research budget.

The primary research question when studying the effect of the incentive value is thus: What is the relationship between incentive value and response rate and nonresponse bias? And is there a monotonous relationship between incentive value and response rate? Or, conversely, will high incentive values diminish the effect of incentives as respondents may feel pressured, coerced or bribed?

In a meta-analysis, Church (1993) finds a strong relationship between the cash value of incentives and response rates. He does not find indications of a diminishing return of higher incentive values. This finding is replicated in the Singer, Hoewyk, et al. (1999) study on findings from interviewer-mediated surveys. They

find that "each dollar of incentive paid results in about a third of a percentage point difference between the incentive and the zero-incentive condition" (Singer, Hoewyk, et al., 1999, p. 223). This finding also holds when examining the effect of the incentive value by comparing the highest versus the lowest incentive condition. Similar linear increases in response rates by incentive value can also be found in Yu & Cooper (1983).

However, other studies have found nonlinear effects. When experimenting with the recruitment procedures of the Longitudinal Internet Studies for the Social sciences (LISS) panel, Scherpenzeel & Toepoel (2012) found a € 20 prepaid incentive to yield a similar effect size as a € 10 prepaid incentive. The odds ratio of registering versus not registering for the LISS panel was even 0.21 points lower for the high € 50 prepaid incentives than for the € 10 prepaid incentives (1.76 vs. 1.97). These findings lead to the conclusion that there might be incentives that are too large, having negative effects on respondents motivation to participate. The analysis by Scherpenzeel & Toepoel (2012) also includes a report on the cost effectiveness of recruitment incentives in terms of the response rate. This analysis indicates that € 10 are more cost efficient than no incentive, € 20 or € 50. Experimenting with even lower incentive values, € 5 respondent incentives performed slightly better than offering € 10 in the German Socio-Economic Panel (SOEP) (Schröder et al., 2013). When increasing incentives incrementally from $0 to $10 in an RDD survey, Trussell & Lavrakas (2004) found no significant increase in response rate between $1 and $5, but significant increases above $5. A similar result was achieved in the Household, Income and Labour Dynamics in Australia (HILDA) pilot study incentive experiment, where a € 25 incentive resulted in a 27 percentage points higher response rate compared to a € 10 incentive (Hanly et al., 2014). In a € 20 vs. € 40 experiment during an ongoing panel study, Zagorsky & Rhoton (2008) find the high incentive group to perform better in terms of completion rate. The survey was conducted to ensure cooperation in a longitudinal survey among those that previously refused to participate. A monotonous effect of incentive value was also found by Börsch-Supan, Krieger, & Schröder (2013) where a € 40 condition performed better than a € 20 or € 10 incentive condition (all prepaid, see chapter 7). This finding is also supported by the 2010 PIAAC Germany field test experiment, where response rates increased with incentive value.

The € 50 condition provided the highest response rate (Martin et al., 2014).

Hsu et al. (2017) also find a monotonously increasing relationship between incentive value and response rate, comparing a $150 dollar postpaid incentive compared to $50 and $100 conditions. Nonetheless in a meta-analysis of incentive experiments, Mercer et al. (2015) find a "strong nonlinear effect of incentives". While response rates improve with higher incentive values, there are diminishing returns with high incentives. Important moderating factors in the analysis are survey mode and time of payment.

The results of this overview are mixed. While most individual studies and meta-studies report that larger incentive values result in larger response rates, some studies report the largest response rates in median incentive conditions or no effect of incentive value. The effect of incentive value seems to be moderated by many factors in fielded surveys. A pretest experiment is advised when drafting a new survey. Overall the literature supports the notion that there is a monotonously positive relationship between incentive value and response rates (with some exceptions). There is support for the notion that there is a diminishing return for high incentive values. Special consideration should be given to cost effectiveness of the incentive value as evidence is sparse on this question (Gelman et al., 2003; Scherpenzeel & Toepoel, 2012).

3.4 Different forms of monetary incentives: cash, check, voucher and lotteries

Money can change hands in different ways: cash, written checks, use of a credit card or vouchers, and in many other ways. In the same way monetary incentives can be handed over to a respondent in many different forms. Different modes of payment can feel different to the respondent and invoke different reactions. A check or a voucher is not as direct and universal as a bank note. A prepaid credit card imprinted with a pleasant picture and the surveys' logo however may potentially be remembered better than a wire transfer. Even though there might be reasons to suspect different effects of different transfer modes of incentives, there has been little research into the matter.

In a direct test of a cash incentive versus a gift voucher both worth $5, Birnholtz

et al. (2004) found that cash incentives yield the higher response rate for new and repeat participants in a web survey. In field practice, however, the choice of payment mode seems to be guided more by issues of handling and available methods than by the effectiveness of the mode.

Using checks or prepaid credit cards also provides the practical advantage that funds form uncashed checks or unused prepaid credit cards will be paid back to the survey organization as reported by Hogan (2009) (provided legal, ethical and budgetary regulations do not interfere). But as the effect of payment mode is unclear, more experimental research in this area is necessary.

A special case of monetary incentives are lotteries, as a monetary value can be attached to the prize but that value might be misidentified by the respondents. In spite of early evidence against the use of lotteries published under the the programmatic title "Charities, no; Lotteries, no; Cash, yes" (Warriner et al., 1996), lotteries are still being used in survey research. Gajic et al. (2010) confirmed a modest prepaid incentive to perform better than two lottery conditions in a sequential mixed mode survey (mail/online). When comparing three conditions to each other, offering a high $250 lottery win worked better than the $25 lottery win. However, a third condition, a $2 prepaid cash incentive was the most cost efficient. Lotteries remain in web surveys as they tend to be the least expensive incentive option. In their review Singer & Ye (2013) conclude that lotteries show small positive effects in comparison to no incentive groups and that "lotteries (prize draws) are no more effective in Web surveys than they are in other kinds of surveys"(Singer & Ye, 2013, p. 125). Lottery incentive schemes should therefore be used with caution. In the German survey context, large scale surveys have moved away from the established use of charitable lotteries as an incentive in recent years (Börsch-Supan, Krieger, & Schröder, 2013; Schröder et al., 2013).

3.5 When to pay out incentives: in advance or after the survey?

When to offer the incentives is a continuous debate: right with the survey request or upon completion of the survey. Offering an incentive with no strings attached with the survey request is usually called an unconditional or prepaid incentive or

very rarely pre-incentive. When the handout of incentives is conditional on survey completion they are typically called conditional, postpaid or promised incentives.

The reason for choosing between those payment moments can be a practical one. It may simply be impossible to offer an incentive on first contact (for example, in telephone surveys without address frame) or legal regulations may not allow sending a gift in advance. In addition, the net cost of incentives may be lower with conditional incentives, as they are paid out only to the smaller subset of the sample that responded to the survey request. A final advantage of conditional incentives is that the logic of the interaction (an effort is followed by a reward) is familiar to respondents from everyday life whereas unconditional incentives reverse that logic and may therefore be regarded as unconventional.

When going back to the theory of action underlying the response process, support for both approaches can be found. An unconditional incentive emphasizes the role of the incentive as a gift, a token of appreciation. The rationale behind this mode is to evoke the reciprocity norm in the respondent and encourage her/him to participate. In contrast, an incentive paid on condition of survey participation plays more into the general exchange frame: The respondent is receiving a good in exchange for a service.

In experiments testing the effectiveness of unconditional incentives against conditional incentives most studies find unconditional incentives to perform better than conditional incentives, but the reverse result can also be found in the literature. The positive effect of prepayment is reported in meta-studies. Church (1993) reports an average increase in response rates of 19.1 percentage points for unconditional monetary incentives whereas that average increase was only 4.5 percentage points when using postpaid incentives (t=3.67). In the review study of Singer, Hoewyk, et al. (1999) the effect of prepayment is positive when controlling for incentive size and other characteristics but fails to reach significance.

In a reanalysis of the metadata collected by Singer, Hoewyk, et al. (1999), Gelman et al. (2003) refine the model used and include interactions of time of payment and other characteristics in the analysis (burden, value, mode, form). As a result they find an interaction of value and incentive. Unconditional incentives have "a higher effect per dollar compared to postpaying" (Gelman et al., 2003, p. 217).

Support for the positive effect of unconditional incentives can also be found in review articles by Booker et al. (2011) and Edwards et al. (2005). Looking at individual studies, I can illustrate those mixed reviews. When setting up the Dutch LISS panel, only prepaid incentive conditions had a significant (positive) effect on panel registration compared to promised incentives (Scherpenzeel & Toepoel, 2012). When experimenting to find the incentive scheme for the pairfam Study conditional incentives performed better than unconditional incentives, both in the first interview and in subsequent follow-up waves (Castiglioni et al., 2008). Contrary to this, in the face-to-face recruitment for the German Internet Panel, unconditional incentives returned an about 9 percentage points higher response rate than conditional incentives (t=5.14) in spite of the value of the conditional incentive being twice as large (€ 10 vs. € 5, see chapter 5).

As mentioned above, the net cost of prepaid incentives is higher than that of promised incentives. Therefore, for survey practitioners, the question of cost efficiency will be raised. Do the improved results justify the higher net costs? The question is not an easy one to answer as the value of additional survey participants is hard to measure. When analyzing the total cost of registering participants for the LISS panel Scherpenzeel & Toepoel (2012) conclude that the € 10 value and offering the incentive unconditionally is the most cost efficient. Gelman et al. (2003) conclude that postpaid incentives will be more cost efficient for their survey situation. The question of cost efficiency is far from being solved, especially when going beyond response rates and examining questions of data quality and fieldwork effort.

Unconditional incentives can be viewed as an unconventional approach. This is especially the case for Germany, where prepaid incentives have not been used until recently. This can lead to a higher number of calls to the survey hotline as experienced in the SHARE Germany wave 4 survey (Börsch-Supan, Krieger, & Schröder, 2013). As a result of an experiment conducted with unconditional incentives in this panel wave the number of calls to the hotline was about 3 percentage points higher with prepaid incentive compared to the conditional incentive. There is no indication that overall surveys suffer from these complaints, but practitioners must be aware of this consequence of unconditional incentives and plan accordingly with advising hotline staff and sponsors.

Prepaid and postpaid incentives can also be combined. In the NEPS Wave 6 incentive experiment a random half of sample members were switched from a € 20 postpaid to a € 10 prepaid combined with a € 10 postpaid incentive (Kretschmer & Müller, 2017). The combined post- and prepaid incentive treatment yielded higher response rates, a more balanced sample, and reduced fieldwork effort. Also, Hsu et al. (2017) report higher response rates when combining promised incentives with a notification postcard and a $ 5 prepaid incentive.

In summary, most studies find that prepaid incentives result in higher response rates than postpaid incentives. The theoretical explanation that this is due to evoking the reciprocity norm which works better than a payment frame highlighted by postpaid incentives is compelling. But survey research lacks theory to explain the circumstances of why prepaid incentives sometimes perform worse than postpaid incentives.

3.6 Targeted incentives

In previous sections, incentives were paid out to all respondents of a survey in the same way. However, survey organizations can also use different incentive schemes to target different subgroups of respondents.

Higher incentives may be paid out to a hard to reach population or an additional gift could be offered to initial nonrespondents or sleepers in a panel survey or used to aid with the refusal conversion of reluctant respondents. This is called targeted or differential incentives. In this section, I will discuss evidence on the use of targeted incentives and discuss the ethics or fairness of such incentive use.

When discussing targeted incentives ethical questions are often examined. Respondents are treated unequally (Laurie & Lynn, 2009). For example late or reluctant respondents may receive higher incentives or gifts. Inadvertently, surveys thus reward unwanted behavior like a late response or skipping a panel wave. Managers of both German large survey organizations carrying out face-to-face surveys have therefore repeatedly expressed their unwillingness to use targeted incentives, citing ethical concerns.

In a paper by Singer, Groves, & Corning (1999) perceptions on targeted incentives were examined. In a two-stage survey differential incentives were used

to convert reluctant respondents from the first stage. In the second stage, an experiment about disclosing this practice to respondents was carried out. It became clear that most people think targeted incentives are being used and the majority of respondents feel that these targeted incentives are unfair. In practice, the disclosure of differential incentives made no difference to response propensity. "We conclude that respondents are sensitive to issues of fairness in the distribution of incentives, but that these issues are not especially salient, nor are they prominent among the factors that motivate survey participation" (Singer, Groves, & Corning, 1999, p. 259).

There is evidence about the success of differential incentives paid to increase the likelihood of reluctant respondents in a survey. In the first wave of the Health and Retirement Study in the USA (HRS), nonrespondents were targeted with increased incentives. The initial incentive was $10 for individuals and $30 if both partners of a couple participated. Dividing nonrespondents into two experimental groups and a control group, different treatments were tested: Incentives were doubled for the first group ($20 for individuals /$60 for couples) and raised to $100 for individuals ($200, for couples) for the second group. The control group received no additional incentives. The interventions were successful in converting households into HRS participants. The sample composition analysis showed that the treatments were successful in getting important respondents into the survey to balance out the sample (Juster & Suzman, 1995).

In the LISS panel study, incentives were used to target so called sleepers, panellists that did not complete questionnaires for three consecutive months (Scherpenzeel, 2014). Offering a € 30 unconditional incentive increased the percentage of re-activated sleepers by 12 percentage points compared to a no incentive control. group (44% versus 32%). The € 10 conditional incentive group was almost as successful (43% versus 32%, %d=11) in increasing re-activation rates but at a much lower cost than the € 30 condition.

In the Survey of Program Dynamics 2000 data collection debit cards valued $40 were successfully used. Respondents that showed signs of an initial refusal during the doorstep approach were immediately offered the debit card to avoid refusal (Kay et al., 2001). Unfortunately, the direct effect of the treatment is not provided in the paper, but evidence reported in the paper indicates the success of the treat-

ment: My own calculations show a 34 percentage points increase in response rates among reluctant respondents when using the incentive compared to not using an incentive. However, much is left unclear in the paper, for example which part of the assignments were at random and which on condition of previous participation in the study. Results from this study should thus be treated with caution.

For successful examples of refusal conversion see also the studies by Abreu & Winters (1999) using the Survey on Income and ProgrammParticipation and the experimental study by Cyffka et al. (2012), where targeting reluctant respondents with cash incentives increased participation.

In summary, differential incentives are being used mainly for refusal conversion. These interventions are successful in that they bring nonrespondents, sleepers or reluctant respondents (back) into the survey. Taken together with the findings by Rodgers (2011) that increased targeted incentives do not create an expectation or disappointment when incentive amounts converge back to normal, differential incentives can be regarded as a valid and cost efficient tool to counter decreasing response rates (Olsen, 2005). It is noteworthy that experimental evidence is mostly from the United States; the cross-cultural compatibility of such an intervention needs to be tested.

3.7 Incentives in panel surveys

Panel surveys are surveys that approach the same sample unit longitudinally with survey requests. In contrast to cross-sectional surveys, the survey panel request is thus not a unique event but a repeated one. When faced with the survey request, respondents can therefore use past experiences with the survey organization as a reference point for their decision to take part in a survey or not. Some respondents stay loyal to a panel survey, they have formed a close relationship with the survey enterprise. Furthermore, the use of incentives can be put in reference to past incentive treatments by the respondents. It is therefore not unlikely that incentive effects may differ in longitudinal surveys as their effect may accumulate over time or may be diminished by the personal relationship of respondents to the survey that has grown over time. Researchers examining incentives in longitudinal surveys are typically interested in the effect of incentive change over time and the

effect of continuous incentive offerings.

Increasing the incentive amount by £3 from wave 13 to wave 14 of the British Household Panel Study (BHPS) significantly increased the cooperation rate in wave 14. This effect was especially strong among those that did not respond to wave 13 (Laurie & Lynn, 2009). As retaining participants and motivating reluctant respondents are very important challenges to longitudinal surveys, a small increase in incentive value may help master these challenges.

Keeping respondent incentives and treatment groups constant over the three-waves preliminary study of the pairfam, Castiglioni et al. (2008) found offering a € 10 gift voucher to significantly increase participation over time compared with offering no incentive: the conditional incentive outperformed the unconditional incentive. Carrying out a rather complex incentive experiment on the Youth Cohort Study of England and Wales, Jäckle & Lynn (2008) found incentives to be a valid tool for maintaining sample size and data quality over time. The positive effects on retention did not differ across subgroups of the sample. Incentive treatment groups however, showed larger levels of item nonresponse.

An experimental factor unique to longitudinal incentives is the possibility to change the value of incentives over time. In a test of the effect of decreasing incentives Singer et al. (1998) found no negative effects - respondents did not react negatively when the expected incentives were no longer offered.

Evidence from the Consumer Expenditure Survey, a rolling panel of consumer behavior in the USA, Goldenberg et al. (2009) found $40 to result in higher initial response rates than $20 or no incentive. The incentive was payed out unconditionally prior to the first wave of the panel. Throughout the five waves of the panel the $40 group showed the highest response rates. In addition, the $40 group performed best on indicators of fieldwork effort, for example contact attempts or number of personal visits needed to complete an interview.

In a systematic comparison of 28 population-based studies covering health outcomes Booker et al. (2011) found incentive effects to be reported in ten studies. The authors find evidence for increased participation rates with higher incentive values.

When summarizing the state of the literature on the effect of incentives in lon-

gitudinal surveys, a well cited statement [1] is "…limited knowledge of what the optimum strategies are for any given design and whether or how incentive strategies translate into improvements in the accuracy of estimation over the longer term" (Laurie & Lynn, 2009). I, however, conclude that the evidence supports the positive effects of incentives in longitudinal surveys. It is true that the effects are even more inconsistent than in cross-sectional surveys. This is most likely due to longitudinal surveys being very diverse in terms of fieldwork efforts, population and survey goals, even more so than surveys carried out at a single point in time.

3.8 The use of incentives in Germany

While many incentive studies have been conducted in the United States, there has been little research on the use of incentives in Germany. This is especially true for face-to-face surveys (see Pforr et al., 2015, for a comprehensive discussion). Historically, incentives were not used or discussed as widely in Germany as in some other countries: Incentives are mentioned only in passing in one of the two leading German survey methods text books (Diekmann, 2006; Schnell et al., 2013). Against the backdrop of decreasing response rates in Germany in the past years survey managers began to systematically research and think about incentives. A typical example of this survey context is the case of the SOEP. The incentives scheme was changed from using a charitable lottery in combination with doorstep incentives at the discretion of the interviewer to postpaid cash incentives. Also experimental studies with the SOEP have been carried out (for an overview see Schröder et al., 2013).

Germany being late to the game when it comes to survey incentives poses practical challenges and promotes a research question. The practical challenges arise when making the case for incentive use in communication with survey sponsors, university administrators, commercial survey agencies or respondents. As incentives are relatively new and unconventional there is often an element of surprise when faced with survey incentives. In addition, problems and potential risks and drawbacks of the method are brought forward. Administrators feel respondents

[1] Cited in Pforr et al. (2015); Singer & Ye (2013).

have to sign receipts for incentives or refuse the notion of incentive payments altogether. Survey sponsors have to come up with additional funds for incentives and an increased budget. When dealing with commercial survey organizations I have encountered a general skepticism about the usefulness and practical administration of incentives. Last but not least, there is a considerable opposition to (especially) cash incentives coming from respondents. A small but vocal minority of the sample members complain about receiving cash incentives. This is especially prominent with prepaid incentives. In their complaints, these respondents predominately challenge the legality of the cash offering or complain about their tax money being wasted or a combination of both. It is hard to quantify the number of these complaints but I will take a closer look at this issue in chapter 7.

The research question then is whether the results from studies in other countries can be generalized and transferred to Germany. As cross cultural differences in field methods are well documented (e.g. Blom et al., 2010) the effects of incentives might be different in Germany than in other countries. In a comparison of known evidence of research on incentives in large face-to-face studies in Germany, Pforr et al. (2015) reported the results of incentive experiments on eight surveys in Germany, including two studies presented in this volume. The experiments examined different aspects of incentive treatments: prepaid vs. postpaid, lotteries, and incentive value. The consensus in the literature is that incentives increase response rates in the surveys examined. The authors find some support for larger cash values performing better, especially large sums. Overall, prepaid incentives have a stronger effect than conditional incentives. Finally, the authors identify the area of nonresponse bias as an area that has not been studied well enough yet.

Existing research carefully suggests that incentives affect survey results in a similar way as has been shown in the literature on surveys based in the United States. Nonetheless, researchers should proceed with caution when using incentives as acceptance with respondents can't be generally assumed.

3.9 Lessons from past research

In summary, the literature finds that incentives are an able tool to improve survey participation. Across survey and incentive modes, incentive conditions tend

to outperform control groups when it comes to survey response. In general, larger sums of prepaid cash incentives have the strongest effects on response rates. But experiments on incentives use very different surveys as vehicles, with different target populations, topics and survey environments. As all of these factors influence the outcomes of the surveys, comparing results from incentive experiments is difficult. To a survey practitioner it is difficult to give general recommendations. I suggest including an experiment in the pretest to find out how incentives work in a particular context and to address issues specific to the respective survey (Scherpenzeel & Toepoel, 2012).

You can keep your Marxist ways
for it's only just a phase.
Money, money, money makes the world go round!

Eric Idle, The Money Song

4

Data used in the dissertation

4.1 Survey of Health, Ageing and Retirement in Europe

The Survey of Health, Ageing and Retirement in Europe (SHARE) is a multidisciplinary and cross-national European research infrastructure collecting longitudinal micro data in 27 European countries and Israel (Börsch-Supan, Brandt, et al., 2013). The project is mainly funded by the European Union. The project runs a biannually computer assisted face-to-face survey of the population of participating countries aged 50 years and over. The main topics covered in SHARE are health, socioeconomic status and social networks (see Börsch-Supan & Jürges, 2005). Six panel waves with more than 120,000 participating individuals have been completed so far, with a seventh wave starting in 2018 (SHARE - Survey of Health, Ageing and Retirement in Europe, 2017).

4.1.1 Sampling

SHARE consists of country samples and sampling procedures can vary between countries. For most countries a population register is being used, but in some countries those are not available (Lynn et al., 2013). As national deviations in sampling have to be accounted for, the central coordination assures that comparable probability samples are being drawn in all SHARE countries.

Excluded from the target population aged 50 plus are those in institutions or those that cannot speak the country's language(s)(Börsch-Supan, Brandt, et al., 2013). In addition to the target population, the SHARE survey has a household component. Current partners living in the household of sample members are eligible for a SHARE interview, irregardless of their own age. All those that have been interviewed in past SHARE waves are followed and interviewed in following SHARE waves if they do not move abroad (Lynn et al., 2013).

To maintain sample size over time, SHARE countries can add refreshment samples to their panel sample. The data used in this dissertation are from the SHARE Germany wave 4 refreshment sample. This refreshment sample had a target size of 4,000 individuals based on a register sample. The sample was drawn in two stages. In the first sampling stage, 200 municipalities were selected with a selection probability proportional to their population size of persons aged 50 years or older. This resulted in 219 primary sampling units (PSUs) as large cities were included with more than one PSU (Lynn et al., 2013, p. 98). In the second stage, 44 individuals were sampled per PSU resulting in a gross sample of 9,636 units. The age cohorts born 1957 to 1960 were oversampled to better match the existing baseline sample of SHARE Germany (Lynn et al., 2013).

4.1.2 Interviewing

SHARE fieldwork is done in person by interviewers face-to-face. Interviewing is contracted to local survey agencies.

The SHARE questionnaire is programmed by the SHARE coordination team to harmonize the fieldwork instrument for SHARE countries. From a common generic questionnaire localized versions are translated with the help of software tools and translation guidelines (Wijnant et al., 2013).

Fieldwork of the German wave 4 SHARE was contracted to the Infas institute (MEA - Max Planck Institute for Social Law and Social Policy, 2017). Survey interviewers are trained following a program provided by the SHARE core team. Fieldwork is coordinated, recorded and controlled by the use of the SHARE sample management system. In this way SHARE coordination assures interviewing standards are met and information on fieldwork progress is collected in a timely fashion.

SHARE data is generally made available to the scientific community as a research infrastructure free of charge via the website share-project.org. Note however that the data used in this publication is not included in the public releases but is made available on site at SHARE in Munich upon request (MEA - Max Planck Institute for Social Law and Social Policy, 2017).

4.2 The German Internet Panel (GIP)

The GIP is an infrastructure project of the Collaborative Research Center SFB 884 "Political Economy of Reforms" at the University of Mannheim funded by the German Research Fund (DFG). The project runs an ongoing online panel study based on a strict probability sample of the German population. Its aim is to provide a vessel for researchers of the center to collect data on individuals' preferences and opinions on reforms in Germany and the change of theses opinions over time. There have been two recruitment phases in 2012 and 2014 (Blom et al., 2015). The 2012 recruitment of the GIP, which I will use in my analyses, implemented two incentive experiments that are ideally suited to research the question of effects of incentives on sample composition. In the following, I will only discuss the 2012 recruitment as only this procedure is relevant for the work described in this dissertation.

4.2.1 Sampling

The GIP 2012 study is based on a probability sample of the general population born between 1937 and 1996. The sampling design was composed in three steps: A stratified random selection of PSUs, a listing of households from a random starting address in each sample point, and a random selection of listed households.

In a first step, 250 sample points were drawn from the ADM sampling frame. This sample frame covers all populated areas of Germany and distinguishes approximately 53,000 geographical areas with 300-700 households each (Heckel & Hofmann, 2014). The GIP made use of the stratification categories provided on this sampling frame by selecting sample points such that the selection probabilities of households were proportional to their occurrence in the population regarding federal states, region and a classification of size and structure of municipalities (BIK Strukturtyp, a classification of populated areas based on the properties of population size and urbanity). 250 Primary Sampling Units (PSUs) were selected representing 208 local administrative units.

In a second step, households were listed within these PSUs by a separate random route procedure. The listing was conducted prior to fieldwork and carried out by interviewers not involved in the fieldwork in this specific PSU. Starting from a randomly selected starting address, 100 households were listed with the address and last name found on door bells and letter boxes. Note that in Germany, it is common to have the inhabitant's last name displayed at an address: Less than 3 percent of addresses had to be listed without a resident's name. Many of these "unnamed" households are likely to be uninhabited. All households were listed, no households were skipped.[1]

In a third step, 22 addresses were drawn from the address list in each PSU: starting from a random number between 1 and 5, each 5th address was selected to ensure sufficient physical distance between sample members. The first 16 addresses drawn were defined as the first address batch, the other six addresses were kept as a second reserve address batch. With this random division of the addresses into two working samples it was easier for the the GIP team to steer the progress of fieldwork and for the survey agency to manage the workload of their interviewers. These two address batches are of special significance to the invitation incentive experiment, as the experiment was only conducted with the addresses from the first

[1] It is important to note that this listing of households without interval is unique in the German fieldwork setting. Typically, surveys list and interview every 5^{th} household only (e.g. German Socio-Economic Panel). The listing without interval was chosen for quality control purposes, because a) we assumed that interviewers would be less likely to skip unattractive households and b) it is easier to cross-check an uninterrupted list.

address batch. The final sample consisted of 5,500 listed households.

Once sampling was completed the recruitment process began. In order to invite respondents to the online panel, recruitment was carried out in two steps. First, a face-to-face interview was conducted to list eligible individual sample members in each household and to obtain consent from sample members to invite household members to the online survey. Then, in a second step, consenting eligible sample members were invited to the online panel.

4.2.2 Face-to-face recruitment

Face-to-face recruitment as well as sampling was contracted to TNS Sozialforschung. 135 interviewers were jointly trained by TNS and the GIP team to conduct the interviews. About a week prior to interviewer contact, listed households received an advance letter. In cases where no family last name was listed, letters were addressed to "tenant's of".

The letter briefly outlined the face-to-face survey without mentioning the panel study. With the advance letter a project leaflet and a data protection statement was sent. In all materials, only the face-to-face study was mentioned. We did not specify that the purpose of the interview was recruiting panel members, nor that the interview mode for the panel would be online. This approach was used to not evoke selective nonresponse due to respondents being averse to a long-term commitment.

As part of this letter, incentives were sent or announced. When sending the invitation letters to the 4,000 respondents of the starting address batch, a controlled random incentive experiment was conducted testing a € 10 promised cash incentive against a € 5 unconditional cash incentive. For more details on this experiment see chapter 5 in this dissertation.

The main purpose of the recruitment interview was to build a detailed household grid and assess the internet and computer equipment in the household. In addition, some selected substantive questions were asked to show respondents, what further survey waves may look like.

At the end of the interview, consent for a recontact was asked for all household members aged 16 to 75 (born 1937 to 1996). Consenting sample members moved

on to the second recruitment stage, the online recruitment.

4.2.3 Offline households

Many online panels suffer from coverage bias, as persons without access to a computer and the internet are not covered (Yeager et al., 2011). To overcome such a coverage bias the GIP provides internet access and computers to such offline households. In the face-to-face interview, interviewers recorded whether households were equipped with an internet connection and / or computers. If households lacked a computer, fast internet connection or both the GIP offered to set up the equipment necessary to participate in the study. The inclusion of previous offline households are an important feature to increase representativeness of the GIP (Blom et al., 2017). Offline households are not part of the experiments discussed in this dissertation, because their recruitment involved different steps, which may not be affected by the incentives experiments.

4.2.4 Invitation to the online panel

Where consent was obtained in the face-to-face recruitment, all age-eligible household members were sent a personally-addressed invitation letter to register for the online panel.

The invitation letter contained information about the study, the URL of the study page and login credentials. Using this information the invited household members could register online for the GIP and thus became panel members. If respondents did not register within two weeks, a reminder letter was sent. With this first reminder letter, a second incentive experiment was conducted. A random subset of respondents received a € 5 unconditional cash incentive. This incentive experiment is discussed in chapter 6. In case respondents needed additional reminders, households were reminded via telephone and finally a second reminder letter was sent when needed.

Once respondents registered with the panel, the GIP confirmed their email address and invited respondents to a first online interview containing questions on basic demographics, their internet usage, citizenship and place of birth. The first wave of the online panel was fielded in September 2012, concluding the main re-

cruitment process. Registration into the panel was open for latecomers until early 2013.

4.2.5 Online interviewing

Regular panel waves are fielded every other month. Panel management and questionnaire hosting and programming is conducted by the LINK Institute.[2] Questionnaire content is delivered by researchers of the Collaborative Research Center 884. Those questionnaire proposals are compiled into a complete questionnaire by the GIP team in consultation with the questionnaire design experts of GESIS - Leibniz - Institute for the Social Sciences.

Once questionnaires are programmed and tested the recruited sample members are invited to participate in a GIP online wave every other month. Field duration is one calendar month. Reluctant respondents are reminded via email twice each wave. In addition, during the final days of the field period, respondents are reminded via telephone.

When data collection is completed, the survey data are delivered to the GIP for testing, documentation and anonymization. A Scientific Use File (SUF) of each wave is published to the research community via the GESIS - Data Archive for the Social Sciences six months after the fieldwork of that wave has finished.

[2] After a merger the panel is now run by forsa marplan.

There is nothing quite as wonderful as money!

Eric Idle, The Money Song

5

Optimal use of incentives under a fixed budget: The effect of conditional and unconditional incentives on response rates and sample composition

[1]

5.1 Introduction

In this chapter, I discuss survey budgets and incentives by focusing on a different aspect: How do we best allocate money to offer incentives given a fixed budget?

[1] This chapter uses data from the German Internet Panel. A study description can be found in (Blom et al., 2015). The German Internet Panel is funded by the German Research Foundation through the Collaborative Research Center 884 "Political Economy of Reforms" (SFB 884).

Is it better to promise money to the responsive part of the sample as a conditional incentive or to give money to all sampled units in advance as an unconditional incentive?

Previously, I discussed that incentives are best payed out in advance (see 3.5), in cash (see 3.4) and that incentives of greater value increase the response rate more (see 3.3). Given an unlimited budget, the recommendation would thus be to incentivize with a considerable unconditional cash incentive. In real life however, budgets are limited and survey managers have to optimize results within their financial restrictions.

In this situation, unconditional incentives' inherent property of being paid out to all sample members becomes a problem: It makes survey designs with unconditional incentives much more expensive than designs with conditional incentives of the same value. This is especially true in a survey environment like Germany with a (comparatively large) proportion of nonrespondents (Pforr et al., 2015). Furthermore this leads to a trade-off between incentive amount and payment option: promising incentives to respondents leads to a reduced group of recipients, allowing for larger incentive amounts. If the incentive budget has to be distributed among a larger group of recipients, as is the case with unconditional incentives, the funds available per head will be lower.

The GIP recruitment incentive experiment addresses this trade-off and allows one to answer the question of how to best allocate an incentive budget: promising an incentive, knowing that the conditional mode is sub-optimal, or paying an unconditional incentive with a lower monetary value, knowing that incentives of larger values perform better with respect to response rates.

This chapter will be structured as follows: first, I will provide an overview of what previous studies have found regarding the use of conditional versus unconditional incentives. I will then describe the experimental design of the 2014 recruitment incentive experiment and which data and methods will be used to analyze potential bias resulting from different incentive modes.

I will subsequently present and discuss results of the recruitment process and the experimental variation and its effects on sample composition of the resulting sample in the GIP study.

5.2 Conditional versus unconditional incentives: Previous findings and research outline

Previous research has shown that response rates tend to be higher when sample units receive unconditional incentives rather than conditional incentives (Singer, Hoewyk, et al., 1999; Singer & Ye, 2013). This robust finding has been discussed in section 3.5.

A prominent discussion of incentive effectiveness between conditional and unconditional incentives can be found in Scherpenzeel & Toepoel (2012). The authors included cost calculations for the recruitment sample experiment in the LISS online panel. For the personal interviews, an unconditional € 20 incentive was the most cost-efficient solution, as for this option the total cost for a registered household was the lowest. It remains an open question whether the different incentive treatments influence not only the sample size, but also the sample composition.

When examining sample composition effects of incentives, other studies have found no effect of unconditional incentives on nonresponse bias (Jäckle & Lynn, 2008) or bias-reducing effects, meaning those with lower response propensities participate more when offered an incentive (Singer, Hoewyk, et al., 1999). Mixed results are reported by Schröder et al. (2013): Whereas incentives increase the participation of immigrants, a group typically underrepresented in surveys, it also increases participation rates in rural areas, a demographic already well represented. In a summary of findings from nine German surveys, Pforr et al. (2015) state that the results are inconclusive regarding nonresponse bias.

In the GIP incentive experiment, unconditional incentives were used in order to achieve higher response rates. But as the use of unconditional cash incentives is uncommon in survey research in Germany (Pforr et al., 2015), we conducted an experiment to test for the presence of an effect of unconditional vs. conditional incentives on response rates. In addition, we tested for the effect on sample composition.

There were many reasons for running this experiment. First and foremost, the effect of unconditional incentives on response rates was tested in the specific context of the GIP. Secondly, the attitudes of respondents towards receiving unconditional cash incentives was tested. As I will mention when discussing the

SHARE survey in this dissertation (see chapter 7), there had previously been problems with unconditional cash incentives in Germany (Börsch-Supan, Krieger, & Schröder, 2013). And finally, the experiment was designed to test the effect of the unconditional incentives on the sample composition.

Even if unconditional cash incentives yielded higher response rates than conditional incentives, this increase in participation might come at the expense of increased bias in the sample. Unconditional incentives might motivate a unique subset of respondents more strongly than other groups.

To expand on this proposed mechanism let me give an illustrative example. Let us assume that low income target households in the sample were highly motivated by an unconditional incentive. Further, we found that affluent target households were not motivated by an unconditional incentive but rather by conditional incentives. If we now examined the resulting sample from this example, we would find it biased towards low income households in the unconditional cash incentive condition compared to the sample resulting from conditional incentive condition.

With this example I illustrate that higher response rates realized with an unconditional incentive do not necessarily result in a favorable outcome with respect to sample composition. The resulting sample might be larger but more biased than a sample realized with the more conservative approach of conditional incentives.

When deciding between unconditional and conditional incentives the issue of sample composition influenced by the timing of the incentive payment needs to be examined. If unconditional incentives lead to higher participation rates but this group is of a select demographic, the effect may be an increased overall bias. If, however, the additional respondents drawn to the survey by unconditional incentives closely resemble those who would also participate with an conditional incentive, the resulting sample would be more balanced in addition to containing an increased number of observations available for analysis and, in the case of the German Internet Panel, for subsequent panel participation. This chapter will provide evidence on the effect of incentive mode on sample composition.

As mentioned earlier, there are many different ways of administering survey incentives. Due to this variety there is no one incentive effect on the sample composition but the effects depend on how incentives are administered. A potential effect of incentives on the sample composition will not be universal but will depend on

how exactly incentives are used.

When using unconditional incentives the issue of survey costs has to be considered, as implementing unconditional incentives is the more costly option compared to conditional incentives of a given incentive value.

In the GIP we expected a 50 percent response rate at the recruitment interview. The GIP allocated the same part of the budget per respondent for conditional and unconditional incentives. Thus, unconditional incentives had to be smaller. They were discounted by the expected proportion of nonrespondents. As a result, the incentive mode was expected to be cost neutral to the GIP budget.

Conditional and unconditional incentives are used in the GIP in the form of a controlled experiment. This chapter will answer two questions: Do unconditional incentives have a larger effect on nonresponse than conditional incentives and is there an effect on the sample composition when comparing unconditional and conditional incentives while keeping the incentive costs equal across the two incentives groups.

5.3 Data and methods

I use data from the 2012 German Internet Panel incentive experiment. The survey is described in chapter 4.2, variables and analytical steps taken are described in this section.

5.3.1 Design of the recruitment incentive experiment

In the GIP recruitment experiment the incentive budget was fixed, i.e. that conditional and unconditional incentives were of identical cost per sample member. In the case of the GIP, a fixed budget for incentives was calculated for a € 10 promised incentive to respondents. The response rate was expected to be around 50 percent. The budget therefore was € 10 times 50 percent of 5,500 sampled addresses, i.e. € 27500.

At this fixed budget, offering a conditional incentive meant that the incentive value had to be discounted by the proportion of sample members not participating in the survey (i.e. by 50%). The prepaid incentive value thus had to be half of the € 10 promised incentive. To test the effectiveness of a prepaid incentive versus a

conditional incentive at a fixed budget, we thus decided to conduct an experiment of a € 5 unconditional versus a € 10 conditional incentive.

Prior to the GIP fieldwork, the total sample was randomly split into a first and second batch to be fielded at different times during fieldwork. This split was done to better steer fieldwork processes and to be able to achieve the target sample size, if response rates fell behind expectations. For the experiment, only the 4,000 addresses of the first batch were used. Exactly 100 cases had to be excluded from the experiment as no family name was collected during household listing procedure. The remaining 3,900 addresses are the base sample for the experiment.

Addresses were randomly selected into the conditional incentive control group and the unconditional incentive treatment group. The experimental treatment consisted of a variation in the incentive presentation and an advance letter adapted accordingly. The advance letter announced an interviewer contact in the near future, assured confidentiality and the voluntary nature of the survey. At the very end of the letter the experimental variation was introduced. In the conditional incentive control group, the letter announced an incentive of € 10 handed out by the interviewer as a thank you after the interview. Letters to sampled households in the unconditional incentive group closed with referring to an enclosed € 5 bank note as a thank you (see figures 9.0.1 and 9.0.2 in the Appendix). An overview of the distribution of the random allocation of sampled households into experimental conditions is displayed in figure 5.3.1.

5.3.2 Methods

In this subsection, I lay out the analytical steps taken to investigate the effects of the two incentive conditions. There are four steps in this process: describing the effect of the experimental conditions on response rates, a manipulation check, a nonresponse bias analysis by incentive conditions and a comparison of sample compositions between incentive groups. Steps one to three provide supporting evidence and manipulation checks for the sample composition analysis.

The results section opens with a simple percent difference analysis in response rates.

Next, I conduct a manipulation check for the randomization in the recruitment

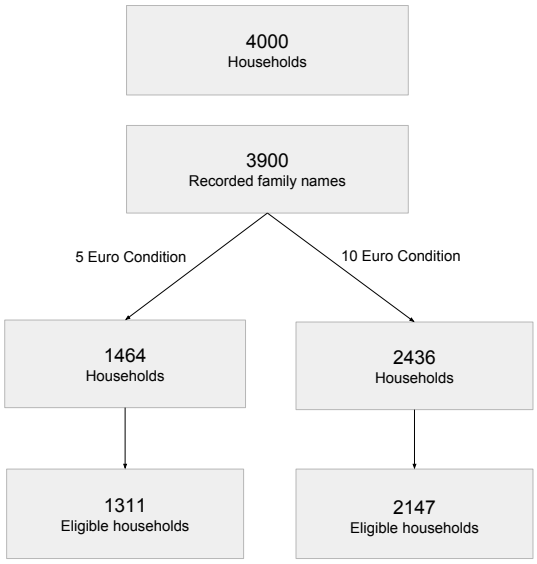


Figure 5.3.1: Overview of the 2012 GIP recruitment incentive experiment

experiment, because, while I ensured that the allocation to control and treatment groups occurred randomly, even a random allocation may lead to an unequal distribution of key characteristics. Therefore, I compare data available for all sample members, respondents and nonrespondents, according to their assignments to experimental conditions.

As a third step I include a nonresponse bias analysis. Whereas the experiment is set up to be randomized over sample households, a selective nonresponse effect of the incentive condition could lead to differences in the composition of the achieved sample. Nonresponse bias analysis is conducted by means of a fixed effects logistic regression (Allison, 2009; Bell & Jones, 2015) on survey participation with PSU as a group variable. This method controls for interviewer and regional effects by keeping them constant. Interaction effects of individual characteristics with the experimental incentive conditions are tested in the model to check for selective nonresponse effects across experimental groups.

The main analysis answering the research question itself is a comparison of the sample composition between recipients of unconditional and conditional incentives. To examine the sample compositions, I use sample member characteristics

from the survey and survey paradata but also using external data sources like commercial and public area information. A detailed data description can be found in the next subsection. Differences in means between the two experimental groups are compared using a two-sample t-tests. As recommended by Gelman & Hill (2007) I include no controls for post-treatment variables for the casual interpretation of the main experimental effect. Following Brüderl & Ludwig (2015) no controls for the group variable are added.

5.3.3 Data

The data I use in these analyses are from the recruitment interview and the interviewer observation data of the 2012 GIP recruitment. These data are enriched with area information data from public sources (Bundesinstitut für Bau- Stadt- und Rauminformation, 2015) and a commercial data vendor (Microm, 2011).

5.3.3.1 Recruitment experiment

Recruitment interview data encompass all data directly derived from conducting the experiment. As described, 3,900 households were part of the experiment. These households were randomly allocated to treatment groups with 1,464 households in the unconditional and 2,436 households in the conditional incentive group.

During fieldwork, about 12 percent of all households were discovered to be ineligible for the interview as they were not part of the GIP population, for example because no private household was found at the specified address or all household members were outside the GIP age range. Excluding ineligible households resulted in a sample of 1,311 households in the € 5 group and 2,147 households in the € 10 condition (see figure 5.3.1).

5.3.3.2 Commercial area information

The GIP Data were enriched with data bought from Microm GmbH. These data can be linked to street addresses in Germany. Microm data are collected from a variety of data sources from federal agencies, as well as private enterprises. This information is geocoded and processed. Microm data are provided at the level

of a so-called "Microzelle", a group of about five houses (that can include multiple households). Information linked to a street address is thus part of such a Microzelle, an aggregate of multiple households. The data are aimed at marketing clients to gain insight on (prospective) customers. Microm data have been used in the past to examine nonresponse patterns (e.g. Sinibaldi et al., 2014).

Commercial area information is used for the manipulation checks and sample composition analyzes. I include information on economic prosperity of sample points, an index of the purchasing power in a region (in thousand euros). An indicator of persons with a university degree living at a given address is used as a measure of education. Also I use dummy indicators marking adresses with a high probability of single households, singles and foreigners. These indicators must not be confused with a real observations of these groups, it is rather a high likelihood that, for example, single households are living at a given address.

5.3.3.3 Public area information

In order to assess regional effects on respondents' response behavior, information on the PSUs is used. I use statistics collected and aggregated by the federal statistical agencies. This information can be matched at the level of communities ("Gemeinden"), the smallest level of political aggregation in Germany. In my analysis, I use an indicator for low, medium or high unemployment rates.

5.3.3.4 Interviewer observations

Another set of variables is the interviewer observations. Interviewers were required to record four observations at each sampled household: whether the household had an intercom, the type of building the household lives in, the maintenance status of the building and the perceived socioeconomic class of the household.

In a paper by Sinibaldi et al. (2014) the authors demonstrate the value of such interviewer observations for nonresponse adjustment. These variables are therefore used as control variables for neighborhood and income related effects on sample composition. Having an intercom system is viewed as an additional burden for contacting household members and might affect respondents cooperation negatively. Only one interviewer observation per address is recorded at the first visit.

Some interviewer observations were not recorded by the interviewer, about 12% household information on intercom use, building type and building maintenance and about 20% of class assessment by the interviewer are missing. Missing interviewer observations are imputed using multiple imputation (with chained equations)(Rubin, 1987). Variables are summarized in table 5.3.1.

5.3.3.5 Recruitment survey data

The recruitment interview data is used in the final step of the analysis comparing the resulting sample of the unconditional and conditional incentive groups. Recruitment data is only available for those respondents who participated in the recruitment interview (n=1,410).

A first set of variables is chosen to mirror data used in the manipulation checks. The data used in the manipulation checks originate from outside data sources and might be less precise than the answers of the respondents themselves. I use self-reported health assessment recorded on a 5-point scale transformed into a dummy indicator of very bad or bad health. The frequency of home internet usage is recorded into an indicator of the daily occurrence of such usage. Further indicators of respondent characteristics include age (65 years or older coded as one), gender, full-time employment, being a foreigner and having a college education. Household information is captured with the indicators of household size, a separate indicator for single households.

In addition, indicators of substantive interest to the GIP are included: These are variables political interest and political attitudes. The variables chosen here are respondents reporting high interest in political affairs and the results of a recall question of a respondent's vote choice at the previous general election (Bundestag) held on September 27th 2009. I generated indicators for the five major parties as well as non-participation in the election ("other"-category has been excluded).

An important characteristic of the GIP recruitment is the inclusion of so-called offline households, i.e. households without high speed internet and/or computer access. Therefore, I included an indicator for such households to see if the incentive type is influencing the recruitment of offline households (Blom et al., 2017).

Table 5.3.1: Summary statistics GIP recruitment incentive experiment, recruitment paradata

	mean	min	max
Recruitment experiment data			
Participation in recruitment interview	0.45	0	1
Unconditional Incentive	0.38	0	1
Commercial area information			
Purchasing power			
Low	0.33	0	1
Medium	0.33	0	1
High	0.34	0	1
High probability of:			
College graduates	0.22	0	1
Singles	0.17	0	1
Foreigners	0.21	0	1
Single households	0.16	0	1
Public area information			
Unployment rate			
Low	0.32	0	1
Medium	0.35	0	1
High	0.29	0	1
Interviewer observations			
Intercom installed?	0.60	0	1
Building type			
Farm	0.02	0	1
Familiy home	0.27	0	1
Semi detached house	0.16	0	1
3-4 apartments	0.13	0	1
5-8 apartments	0.23	0	1
9+ apartments, no high rise	0.18	0	1
High rise	0.03	0	1
Building maintenance			
Good	0.19	0	1
Medium	0.74	0	1
Poor	0.06	0	1
Perceived class			
Poor	0.03	0	1
Working	0.41	0	1
Middle	0.48	0	1
Upper middle	0.07	0	1
Upper	0.01	0	1
Observations	3,458		

Note: GIP 2012 recruitment

An overview of the recruitment survey data can be found in table 5.3.2.

Table 5.3.2: Summary statistics GIP recruitment incentive experiment, recruitment interview variables , face-to-face respondents

	mean	min	max
Bad health	0.12	0	1
Daily Internet use	0.15	0	1
High political interest	0.46	0	1
Bundestag election			
Abstained	0.17	0	1
Voted CDU	0.20	0	1
Voted SPD	0.19	0	1
Voted FDP	0.06	0	1
Voted Greens	0.10	0	1
Voted Left	0.03	0	1
65 and older	0.24	1	1
Household size	2.33	1	8
Single household	0.26	0	1
Female respondent	0.39	0	1
Full-time employed	0.37	0	1
University degree	0.19	0	1
Foreigner	0.05	0	1
No Internet in household	0.13	0	1
Observations	1,410		

Note: GIP 2012 recruitment, face-to-face recruitment interview respondents only; standard deviation (household size) = 1.19

5.4 Results

5.4.1 Main incentive effect

For the cases in the experimental condition, the message of the main effect is clear and simple: unconditional incentives outperformed conditional incentives in terms of the response rate realized. The difference between the two groups in response rates is 8.4 percentage points (50.8% for the unconditional incentive group

compared to 41.9% for the conditional incentive group; t=-5.1; RR1; (AAPOR, 2016)).

As the experiment was designed to be neutral in terms of the cost to the researcher, meaning both conditions have the same upfront per unit cost, unconditional incentives are more cost-effective, because the resulting sample is larger when unconditional incentives are used.

5.4.2 Manipulation check

The GIP incentive experiment is a randomized controlled experiment. Households were randomly assigned to either the conditional or unconditional experimental group. Due to the randomization and the large number of observations, the gross sample should be evenly distributed across the two experimental conditions. To make sure that the randomization resulted in an evenly distributed sample across experimental groups, I checked the distribution of key characteristics available from the gross sample for the conditional and unconditional incentive groups.

As a manipulation check, I perform a comparison of distribution means. The results are shown in table 5.4.1. As expected, there are no significant differences in means between the two experimental groups. This indicates that the sample is evenly distributed across experimental groups with regards to interviewer observations, PSU characteristics and commercial paradata.

5.4.3 Nonresponse bias

To check for effects selected recruitment had on the treatment groups I conduct a nonresponse bias analysis. The focus of this analysis is to detect if there are different nonresponse effects of the conditional and unconditional incentive treatment groups on participation in the survey, while controlling for frame variables and survey paradata.

The nonresponse bias will be modeled using a fixed effects regression on recruitment survey participation of the household informant. The group variable is the PSU. By selecting this group variable the model controls for all PSU and interviewer effects, that are constant across households in the sample as interviewers

Table 5.4.1: Comparison of means between experimental groups, recruitment paradata

	Cond. incentive	Uncond. incentive	t-stat.
Commercial area information			
Purchasing power			
Low	0.33	0.33	−0.29
Medium	0.34	0.33	0.67
High	0.34	0.34	−0.32
High probability of:			
College graduates	0.22	0.22	0.00
Singles	0.17	0.16	0.08
Foreigners	0.21	0.21	−0.36
Single households	0.16	0.17	−0.22
Public area information			
Unemployment rate			
Low	0.31	0.32	−0.43
Medium	0.35	0.34	0.52
High	0.29	0.29	0.09
Interviewer observations			
Intercom installed?	0.61	0.59	0.72
Building maintenance			
Good	0.19	0.19	0.30
Medium	0.74	0.75	−0.50
Poor	0.07	0.06	0.42
Building type			
Farm	0.01	0.02	−0.93
Familiy home	0.26	0.27	−0.79
Semi-detached house	0.17	0.16	0.58
3-4 apartments	0.13	0.13	0.02
5-8 apartments	0.23	0.22	0.68
9+ apartments, no high rise	0.18	0.17	0.38
High rise	0.02	0.03	−1.12
Perceived class			
Poor	0.03	0.04	−1.14
Working	0.41	0.41	−0.10
Middle	0.49	0.48	0.80
Upper middle	0.06	0.07	−0.64
Upper	0.01	0.01	0.09
Observations	2,147	1,311	

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

Note: t-stat.: two-sample t-test

could work in multiple PSUs but only one interviewer was assigned to work in a given PSU.

In the model, all commercial area information and interviewer observations variables as seen in table 5.3.1 are introduced. Public area information variables are constant within PSU and therefore not included in the model as main effects but rather the effects are controlled for by the fixed effects regression model.

To detect whether nonresponse patters differ between experimental conditions I include interaction effects with the unconditional incentive treatment for each independent variable of the commercial area information and interviewer observations. In addition, I include the interaction effects with public area information variables. In contrast to the main effect of public area information, the interaction terms are not constant within PSUs and can be included in the model. If interaction effects are significant in the model, I can conclude that the effect of a given independent variable on participation differs between incentive conditions.

Figure 5.4.1 displays all interaction odds ratios and their confidence intervals. One interaction odds ratio yields a significant result: The interaction between high probability of single households and incentive condition. The odds of participating in the survey are about two times larger for the unconditional incentive treatment group in PSUs with a high proportion of single households compared to the conditional incentive group while controlling for other model effects. The other interaction effects do not reach statistical significance.

The full nonresponse model can be found in the appendix in table 9.0.1. Here it can be shown that three main effects have significant effects on survey participation: living in a house in need of some (medium) maintenance, having an intercom installed and living in an area with higher purchasing power as measured by marketing data all decreases the odds of participating in the survey compared to relevant reference categories. These findings are not uncommon in nonresponse research. In the context of this section they are not relevant to the research as the effect of these independent variables is the same over both experimental conditions.

Overall, one out of 21 tested interaction coefficients is significant, indicating a different effect of the independent variable conditional on incentives: That of PSUs with a high percentage of single households in the incentive treatment group.

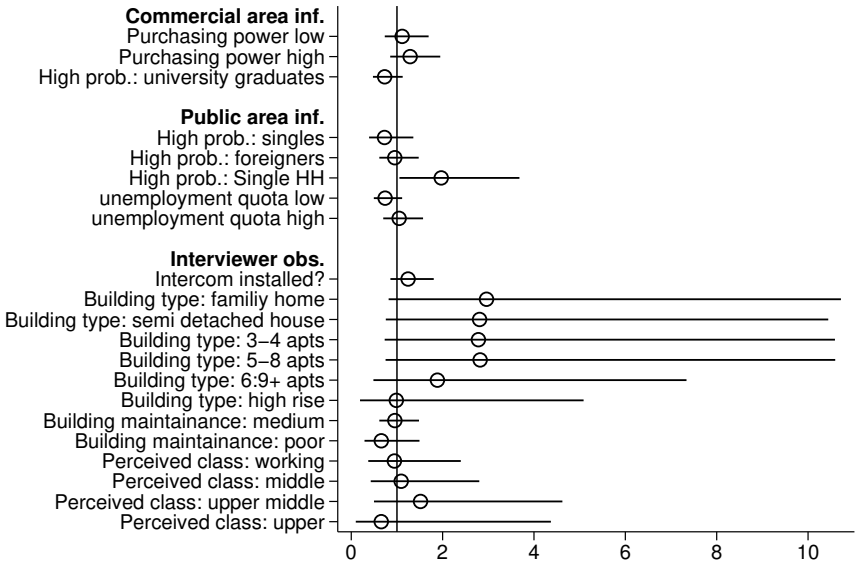


Figure 5.4.1: Logistic fixed effects regression on survey participation, interaction effects with incentive experiment, odds ratio coefficients and 95%KI

This finding has to be taken into consideration when interpreting sample composition effects of this variable, as selective participation by incentive treatment group may confound the sample composition analysis. Summing up the evidence so far, we can conclude that unconditional incentives lead to higher response rates than conditional incentives and there is little evidence for nonresponse bias when using unconditional incentives.

5.4.4 Comparison of distributions

As just presented, I find few indications that incentives influence the sample composition in the GIP via a selective participation propensity. Next I focus on the realized sample. Does the incentive treatment deferentially affect the GIP sample composition? That is, are there differences in key characteristics of sample members that were recruited using unconditional incentives compared to conditional incentives? To answer these questions I compare the distributions of variables in the GIP online panel sample. As part of the incentive experiment, 1410 respondents completed the first interview of the online panel and are the basis for this

analysis.

First I compare the GIP online panel sample with respect to those variables that I used in the manipulation check. The manipulation check analyses have shown no significant differences regarding the experimental conditions in the gross sample. Now I perform the same comparison of means for the GIP online panel sample. There are multiple ways to test for the significance of the mean difference. I chose to use t-tests (Student, 1908).

The t-tests displayed in table 5.4.2 show no significant differences for the interviewer observations, block level data, and sample point characteristics. Therefore I can conclude that based on these variables the sample resulting from the face-to-face recruitment interview does not differ in terms of the experimental conditions.

In addition to variables available for all sample members, variables from the interview itself can be used for an analysis of the sample composition. I included variables that are important indicators of the sample quality given the topic of the survey, i.e. political reform making. The variables thus examined are self reported health, daily internet use, political interest, voting behavior, age, gender, household composition and education. Results of mean comparison t-tests are displayed in table 5.4.3.

The results indicate no significant differences between the unconditional incentive and the conditional incentive group. The means of household characteristics, age, health, voting record, sex, citizenship and education are equal between the two groups.

5.4.5 Discussion

This chapter presents results of the German Internet Panel recruitment experiment. Unconditional cash incentives were tested against conditional cash incentives. As incentive expenses per target household were held constant, conditional incentives had double the value of unconditional incentives. I have shown that unconditional incentives outperform conditional incentives in terms of response rates. Unconditional incentives proved a more successful strategy to increase the sample size than conditional incentives. Furthermore, when checking the sample distribution, no differences between the two experimental groups could be de-

Table 5.4.2: Comparison of means between experimental groups, recruitment paradata, face-to-face respondents

	Cond. incentive	Uncond. incentive	t-stat.
Commercial area information			
Purchasing power			
Low	0.26	0.28	−0.72
Medium	0.38	0.36	0.65
High	0.37	0.36	0.08
High probability of:			
College graduates	0.24	0.22	0.76
Singles	0.15	0.17	−0.96
Foreigners	0.19	0.22	−1.21
Single households	0.15	0.19	−1.53
Public area information			
Umployment rate			
Low	0.38	0.35	1.37
Medium	0.33	0.35	−0.75
High	0.25	0.26	−0.66
Interviewer observations			
Intercom installed?	0.56	0.58	−0.90
Building type			
Farm	0.02	0.01	0.55
Familiy home	0.29	0.31	−0.57
Semi detached house	0.17	0.18	−0.32
3-4 apartment	0.12	0.12	−0.09
5-8 apartment	0.20	0.21	−0.30
9+ apartment, no high rise	0.16	0.14	1.07
High rise	0.03	0.02	0.51
Building maintenance			
Good	0.21	0.21	−0.11
Medium	0.74	0.75	−0.60
Poor	0.05	0.03	1.49
Perceived class			
Poor	0.03	0.03	−0.33
Working	0.38	0.35	1.08
Middle	0.50	0.51	−0.56
Upper middle	0.08	0.10	−0.81
Upper	0.01	0.01	0.45
Observations	803	607	

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

Note: GIP 2012 recruitment, face-to-face recruitment interview respondents only; t-stat.: two-sample t-test

Table 5.4.3: Comparison of means between experimental groups, recruitment interview variables , face-to-face respondents

	Cond. incentive	Uncond. incentive	t-stat.
Bad health	0.11	0.13	−0.78
Daily Internet use	0.15	0.15	−0.01
High political interest	0.48	0.43	1.87
Bundestag election			
Abstained	0.17	0.17	−0.08
Voted CDU	0.21	0.18	1.33
Voted SPD	0.18	0.19	−0.04
Voted FDP	0.07	0.06	0.46
Voted Greens	0.10	0.09	0.84
Voted Left	0.03	0.03	−0.07
65 and over	0.23	0.25	−1.13
Household size	2.36	2.29	1.02
Single household	0.25	0.27	−1.16
Female respondent	0.40	0.38	0.78
Full-time employed	0.37	0.36	0.45
University degree	0.18	0.20	−1.03
Foreigner	0.04	0.06	−1.10
No Internet in household	0.13	0.12	0.79
Observations	803	607	

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

Note: GIP 2012 recruitment, face-to-face recruitment interview respondents only; t-stat.: two-sample t-test

tected. Incentive treatment thus is not a factor influencing the sample composition in the 2012 GIP recruitment survey. Only in the nonresponse analysis one interaction coefficient showed an influence on response, conditional on incentive treatment. This effect is arguably in a desired direction, as single households are often an underrepresented demographic in survey samples. However, given the number of effects tested this one significant effect may well be by chance. In addition, no effect was detectable when examining the sample composition of the GIP online sample.

Given the findings of this experiment, the use of unconditional cash incentives should be encouraged. Larger sample sizes are the result and I can detect no effect on sample composition of this incentive treatment compared to the more common use of conditional incentives.

However, there are some caveats to the study. First, the results of the sample comparisons are restricted to the variables available for both respondents and nonrespondents.

Second, The experimental design is no full factorial design of incentive treatments, thus it does not allow me to contrast paying incentives to not paying incentives (see 3.1), to compare the effect of unconditional and conditional incentives of the same size (see 3.5), or to investigate the optimal incentive value (see 3.3).

The goal of this experiment is a different one. The research question that can be answered with this experiment is a rather practical one. Given a fixed budget, how can the money be utilized for the maximum benefit of the survey: a prepaid incentive to all sample members or a promised incentive upon completion to respondents only? From the GIP recruitment incentive experiment, I can conclude that given budget constraints, unconditional incentives are a more effective use of available funds than conditional incentives.

There is nothing quite as beautiful as cash!

Eric Idle, The Money Song

6

Using incentives in reminders to facilitate recruitment

[1]

6.1 Introduction

In the previous chapter I discussed how to improve response rates in face-to-face recruitment interviews for the German Internet Panel (GIP) using different types of incentives under a fixed budget. In this chapter I examine the next recruitment step, the invitation to the online panel. I describe the results of an experiment on unconditional cash incentives when sending reminder letters to sample members that did not already register with the study.

---

[1] This chapter uses data from the German Internet Panel. A study description can be found in Blom et al. (2015). The German Internet Panel is funded by the German Research Foundation through the Collaborative Research Center 884 "Political Economy of Reforms" (SFB 884).

High registration rates are crucial for the success of the GIP. Registration in the GIP is defined as the mode switch from the face-to-face interview to participation in the online panel.

6.1.1 Push to web

In the literature encouraging respondents to switch to the online mode is referred to as push-to-web. This is a subset of research on mixed modes (Mavletova & Couper, 2016). The main lines of research in this area are studies examining the effects of switches on either measurement error or mode effects (e.g. Klausch et al., 2015) and studies examining the effects on response rates (e.g. Millar & Dillman, 2011). As the GIP is not a mixed mode study and face-to-face results are not compared to online results, measurement error and mode effects are of less interest in this context. Instead, I concentrate on push-to-web studies examining the effects on response rates. In some mixed mode studies the results indicate that offering an additional choice of survey mode in a mail survey decreases the overall response rate (Medway & Fulton, 2012). Cernat & Lynn (2014), however, find no effect on overall response rates when offering a web mode in addition to a CAPI mode. They were able to increase web pickup by e-mail reminders and thus reduce fieldwork costs. Again contrary to both findings, Millar & Dillman (2011) report an effective sequential mixed mode strategy using mail advance letters, a token unconditional cash incentive and e-mail and a final mail follow up reminder. This strategy maximized response rates compared to single mode strategies. Recent reports suggest a sequential mixed mode strategy, where respondents are first offered one (cheap) survey mode and are switched only if they refuse to participate in this mode (Tourangeau, 2017).

6.1.2 Reminders in surveys

Recruiting sample members in person by interviewers is costly and time consuming (see chapter 5). If these sample members cannot be convinced to join the panel, this recruitment effort will not generate any return on investment. In addition, a high sign up rate reduces the potential for systematic bias due to non-sign up (see 2.1).

A standard measure in survey research are reminders (Dillman, 2011; Groves et al., 2013). These can come in different modes (phone, mail, in person, e-mail). Using reminders, cooperation can be increased, as hesitant respondents get an additional chance to comply with the survey request.

Reminders can also be combined with incentives targeted towards nonparticipating sample members. In the GIP, such a procedure was tested: A reminder letter was mailed to encourage sample members to go online and register with the GIP. With this reminder letter, a € 5 unconditional cash incentive was tested against no incentive.

Including incentives when reminding nonparticipants is a form of targeting incentives towards a specific sub-population of the sample. This is a rather uncommon procedure in surveys, where usually all sampled persons receive an incentive. (see 3.6).

I discussed targeted incentives in chapter 3.6. Therefore, I will just briefly recap the most important aspects of the method in the research literature.

The ethical aspect of targeted incentives, for example the unfairness in rewarding uncooperative response, is often discussed in literature. However, research finds that respondents do not perceive such measures as unfair (Singer, Groves, & Corning, 1999). Most studies report the use of targeted incentives to aid refusal conversion (see Abreu & Winters, 1999; Cyffka et al., 2012; Juster & Suzman, 1995; Kay et al., 2001). As with other forms of incentives, there is little evidence of an influence on sample composition, although Juster & Suzman (1995) report a more balanced sample with the incentive intervention.

A comprehensive study on the effects of reminder incentives not just on response rates but also on data quality and sample composition is reported by McGonagle & Freedman (2017). The authors report on a controlled experiment using targeted incentives on the 2013 fieldwork of the Panel Study of Income Dynamics (PSID). Nonrespondents were randomly allocated into a $5 or $10 unconditional cash incentive treatment group in addition to a no incentive control group. Response rates increased significantly in the cash incentive conditions without there being a difference between the $5 and $10 groups. Looking at different demographic subgroups, cash incentives were found to have a larger effect on increasing response rates for older respondents. There was no effect

of the treatments on item nonresponse, a finding the authors interpret as there being no effect on data quality. Comparing the sample composition of the target sample with the realized sample in treatment and control groups, the study concludes that the reminder incentive "brought in a disproportionate number of less-educated respondents, who were less likely to have responded before the…mailing" (McGonagle & Freedman, 2017, p. 16).

From these findings I derive the first two research questions for this chapter. First, I examine the effect of reminder incentives on panel participation: Can an unconditional cash incentive, included in a reminder letter increase the response rate in the GIP? Expanding on the research of effects of incentives, I examine whether including an unconditional cash incentive influences the composition of the realized sample.

In addition to examining the effects of incentives on sample composition and response rates, incentive cost are examined in this chapter. Spending part of the budged on incentives has direct costs. But how do those costs compare to the total survey costs? What is the overall balance of incentives, are the direct costs outbalanced by savings in other areas of the survey? The data from the GIP provides some evidence on the question as to whether including a reminder incentive is cost efficient or not.

The chapter is structured as follows: The reminder experiment sample and data used are described in section 6.2. Then results regarding the effect on response rates and on the sample composition are presented, followed by an overview of incentive cost in the GIP reminder experiment. The chapter closes with a discussion of the results.

6.2 Data and methods

6.2.1 Sample

The recruitment incentive experiment was conducted during the 2012 GIP online recruitment. The full recruitment procedures are described in section 4.2, especially subsection 4.2.4. Two GIP sample subgroups are not available for the experiment due to diverging fieldwork procedures: sample members interviewed using

the short face-to-face doorstep interview and those without computer and or Internet access classified as offliners. All other GIP sample members that failed to register within two weeks of their initial invitation were part of the reminder experiment sample. Figure 6.2.1 gives an overview of the online recruitment procedures after completion of face-to-face interviewing.

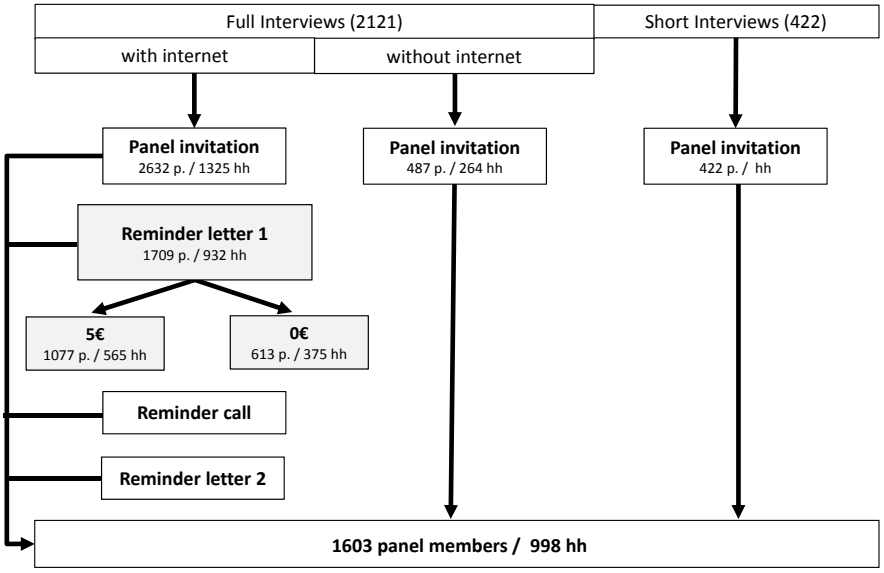


Figure 6.2.1: Push-to-web process, GIP 2012

The vehicle for the experiment was a reminder letter mailed out to individual sample members who had not registered with the GIP, within two weeks. This reminder letter was the first communication in the reminder process, thus we call it the first reminder letter. In total, 1,709 persons were eligible for the first reminder letter. However, 19 sample members had to be excluded from the experiment for survey practical reasons. This leaves 1,900 sample members that received this reminder letter and are part of the experiment. The reminder letter contained the full information on how to register for the GIP including login credentials and the URL of the GIP website.

6.2.2 EXPERIMENTAL DESIGN

About two-thirds of these sample members were randomly selected into the treatment group. To avoid confounding the results with respondents learning about different incentive treatments from other members of the same household we stratified the random selection by households and made sure all household members received the same treatment. 1,077 persons were selected into the treatment group, receiving reminder letters containing a € 5 note to thank respondents "for their efforts". 613 respondents were selected into the control group. The control group received a reminder letter that just reminded them to register without containing an incentive.

Two weeks after the mailing of the first reminder letter, respondents were eligible for additional reminders by telephone or mail in case they failed to register. Therefore, I am differentiating between sample members registering up until two weeks after they received the first reminder letter and those that registered after this time span. The pure effect of the reminder can only be examined in the first group as those respondents did not receive additional reminders. In contrast, sample members that registered at a later date could have received additional communication to remind them to register.

6.2.3 DATA

Data available for analyzing the sample composition are from the same data sources as described in chapter 5. I select variables available for all sample members that might potentially correlate with the participation decision based on the literature presented in subsection 2.2.4. A purchasing power indicator and the proportion of persons with a university degree in the area are available from block-level commercial area information. From public records of the INKAR database (Bundesinstitut für Bau- Stadt- und Rauminformation, 2015) I use an indicator for low, medium or high unemployment rate at the level of municipalities. A third set of information is interviewer observations collected by interviewers when conducting the face-to-face recruitment interview. From this interviewer observation data, information on building type and building maintenance is used. From the recruitment interview itself data on household size are used. Summary statistics

on these data are presented in table 6.2.1.

6.2.4 Methods

To answer the research questions, I first present recruitment rates for both experimental conditions. I check for registration within two weeks to exclude effects of other reminders. In addition, I examine whether there is an effect of the recruitment incentive for respondents that registered later than this two week period.

To asses cost efficiency, I will use information on data collection cost and registration rates.

6.3 Results

6.3.1 Main effect

Including a € 5 unconditional cash incentive in the reminder letter increased the recruitment rate substantively and significantly. Results are displayed in table 6.3.1.

In the first row, registrations within the first two weeks are tabulated by reminder incentive experiment condition. Among those sample members, the percentage difference in registrations is 17.3 percentage points in favor of the unconditional cash incentive treatment (control: 17.8% vs. treatment 35.1%; %d=17.3; t=-7.71). This contrast is highly significant.

In addition to this pure effect of the first reminder letter there also are spill over effects on those respondents that registered later than two weeks after the mailing of that letter. These registrations are tabulated in the second row on the results table 6.3.1 by experimental condition. Among those sample members we still observe an effect of increased registration numbers. The effect is weaker than the pure effect within two weeks, but still significant (control: 11.9% vs. treatment 14.5% ; %d=2.6; t=-3.46). Therefore, even though these sample members received additional reminders by mail and telephone I still observe a significant effect of the unconditional cash incentive.

Figure 6.3.1 illustrates the relative frequency of registered sample members over time.

Table 6.2.1: Summary statistics, recruitment paradata

	mean	min	max
Household size	2.94	1	8
Commercial area information			
Purchasing power			
Low	0.26	0	1
Medium	0.38	0	1
High	0.36	0	1
Public area information			
High Probability of:			
College graduates	0.21	0	1
Unemployment rate			
Low	0.37	0	1
Medium	0.34	0	1
High	0.25	0	1
Interviewer observations			
Building type			
Farm	0.01	0	1
Familiy home	0.31	0	1
Semi detached house	0.19	0	1
3-4 apartment	0.14	0	1
5-8 apartment	0.18	0	1
9+ apartment, no high rise	0.14	0	1
High rise	0.02	0	1
Building maintenance			
Good	0.21	0	1
Medium	0.75	0	1
Poor	0.04	0	1
Observations	1,690		

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

Note: standard deviation (household size) = 1.29

Table 6.3.1: Registration rate to the GIP online panel over time by reminder incentive experiment condition

	Control	*Incentive*	*Total*	%d	t-stat.
Registered within 2 weeks	109	378	487		
%	17.8	35.1	28.8	17.3	−7.71
Registered later	73	156	229		
%	11.9	14.5	13.5	2.6	−3.46
Did not register	432	542	974		
%	70.4	50.4	57.6		
Total	614	1,076	1,690		

Note: Reminder letter 1 recipients that registered within two weeks of reminder letter reception only; t-stat.: two-sample t-test

On the x-axis I plot days since sample members received the reminder letter. On the y-axis I graph the cumulative frequency of registered sample members. The curves show the cumulative percentages of registered respondents of all eligible sample members by days elapsed since receiving the first reminder letter. Three curves are plotted: A curve for the unconditional cash incentive group, one for the control group, and a curve for all sample members.

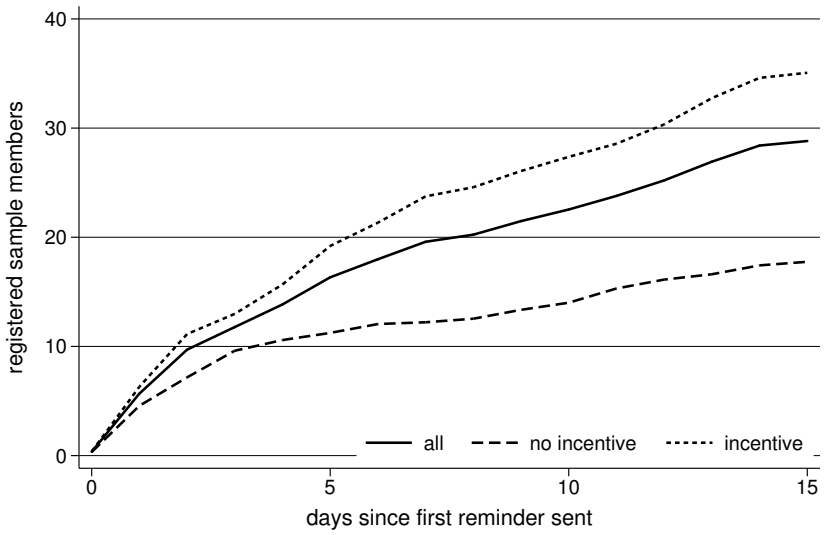


Figure 6.3.1: Sample members registered over time in percent

We see that the short dashed line depicting the registration rate of the uncondi-

tional cash incentive group has a steeper incline than the dashed line of the control group. The figure indicates that the reminder letter including the unconditional cash incentive is not only more efficient overall but also leads to sample members registering comparatively more quickly.

6.3.2 Sample composition

As the unconditional cash incentive greatly improves the registration rates and speed, I also examine whether these increased registration rates come at the cost of a bias in the sample composition. The mechanism for this bias could be that the incentive encourages different groups in the sample to participate selectively. This argument has already been outlined in section 5.2.

To compare the sample composition between the incentive and no incentive groups, I conduct t-tests for mean differences between incentives and control for each control variable. This analysis is restricted to those 487 respondents who registered in the time span between receiving the reminder letter and two weeks after receiving the letter, as here I can observe the pure effect of the reminder incentive experiment. Results of these tests are displayed in table 6.3.2.

The results show that there are no significant differences in the composition of the two samples with regards to the variables available. None of the t-tests indicate significant differences.

I conclude that using a € 5 cash incentive with the reminder letter does not introduce notable differential bias in the composition of the resulting GIP sample. There is a caveat however. These results are restricted to the variables available for analyzing the sample composition, which are limited to information available to both respondents and nonrespondents.

6.3.3 Response rates over time

As in all panel studies participation in the GIP decreases over time due to panel attrition. The reasons for this dropout are loss of contact to panelists, illness or death, a move abroad or just an unwillingness to participate further in the GIP. A figure of the number of participants in the GIP across the first 12 waves is displayed in figure 6.3.2.

Table 6.3.2: Sample composition at registration within two weeks by reminder incentive experiment groups, comparison of means, recruitment paradata

	€0	€5	t-stat.
Recruitment survey data			
Household size	2.83	2.74	0.66
Commercial area information			
Purchasing power			
Low	0.22	0.28	−1.15
Medium	0.41	0.32	1.86
High	0.37	0.41	−0.76
High probability of:			
College graduates	0.23	0.21	0.51
Public area information			
Unemployment rate			
Low	0.31	0.40	−1.61
Medium	0.33	0.33	0.04
High	0.31	0.24	1.44
Interviewer observations			
Building type			
Farm	0.00	0.02	−1.32
Familiy home	0.35	0.36	−0.21
Semi detached house	0.21	0.22	−0.31
3-4 apartment	0.10	0.13	−0.87
5-8 apartment	0.16	0.14	0.41
9+ apartment, no high rise	0.14	0.11	0.92
High rise	0.05	0.02	1.41
Building maintenance			
Good	0.25	0.30	−1.14
Medium	0.70	0.67	0.60
Poor	0.06	0.03	1.30
Observations	109	378	

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

Note: Reminder letter 1 recipients only; t-stat.: two-sample t-test

All GIP respondents are graphed using the solid line. 1,603 sample members registered for the GIP online panel and were invited to the first wave. The participation pattern is typical for panel studies: there is a larger drop in participation from wave one (1,483 respondents) to wave two (1,171 respondents) and a stable development thereafter. I added a line of all GIP panelists to serve as a reference to the participation pattern of the panelists in the experiment.

Comparing the overall participation numbers with those in the reminder experiment and registering within two weeks after receiving the first reminder letter (n=487 dash-dotted line) the pattern is very similar, with the initial drop being flatter than in the overall curve.

The short dashed line depicts the development in participation for those receiving a reminder incentive, the long dashed line the cases from the control group. There are no differences between the development of the treatment and the control group. All recruited panelists across the experimental groups behave similarly over time. In a next step I look at sample composition over time.

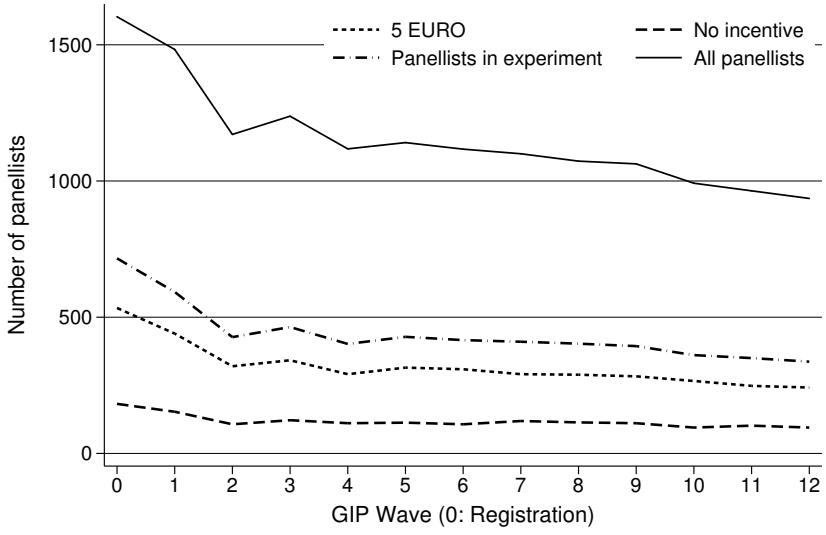


Figure 6.3.2: Participation in GIP over waves including reminder incentive experiment

6.3.4 Sample composition over time

The GIP is a panel study; respondents are repeatedly approached with survey requests. I therefore re-analyze the effects of reminder incentives after two years of panel surveys. Again, I compare those respondents that received an unconditional cash incentive with the first reminder letter to those respondents that did not receive this incentive.

Due to panel attrition, the number of observable cases are reduced to 135: Respondents in wave 12 of the GIP who registered within two weeks after receiving a first reminder letter. Again, I compare means between experimental groups and run a t-test. Results can be found in table 6.3.3.

The results are mixed. The comparison of means test yield two significant differences in building type variables: The proportion of respondents living in family homes is larger in the control group whereas the proportion of respondents living in semi detached houses is larger in the treatment group. Other contrasts in the building type variables and other variables are not significant.

I find 2 of 19 test of differences between experimental conditions to be significant. Overall, I do not find a consistent bias of using unconditional cash incentives. The evidence is not compelling enough to confirm long term effects of the reminder incentive on long term sample composition. The result may hint at some effects, but additional research is necessary on the true nature of long term sample composition effects of reminder incentives.

6.3.5 Costs

I have now established that introducing an unconditional cash incentive in the reminder mailing significantly increases registration rates without inducing notable bias in the sample composition. Thus, including an unconditional cash incentive has a scientific value in terms of a larger sample size. But is the measure cost-efficient as well?

It is evident that including unconditional cash incentives in the reminder letters is a considerable direct cost added to the budget. These direct costs might be offset by savings in other domains. I will now examine two possible cost saving mechanisms and their role in the GIP recruitment: cost savings by reduced reminder

Table 6.3.3: Sample composition at GIP wave 12 by reminder incentive experiment groups, Comparison of means, recruitment paradata

	€0	€5	t-stat.
Recruitment survey data			
Household size	3.03	2.65	1.60
Commercial area information			
Purchasing power			
Low	0.23	0.27	−0.48
Medium	0.37	0.34	0.33
High	0.40	0.39	0.10
High probability of:			
College graduates	0.29	0.24	0.50
Public area information			
Unemployment rate			
Low	0.34	0.47	−1.30
Medium	0.29	0.27	0.18
High	0.31	0.23	0.98
Interviewer observations			
Building type			
Farm	0.00	0.04	−1.20
Familiy home	0.54	0.32	2.37*
Semi detached house	0.09	0.29	−2.48*
3-4 apartment	0.09	0.10	−0.24
5-8 apartment	0.17	0.10	1.12
9+ apartment, no high rise	0.11	0.13	−0.24
High rise	0.00	0.02	−0.84
Building maintenance			
Good	0.34	0.37	0.29
Medium	0.63	0.60	0.30
Poor	0.03	0.03	−0.04
Observations	35	100	

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

Note: stat.: two-sample t-test

efforts during recruitment and savings by reducing total recruitment costs.

6.3.5.1 Reduced reminder effort

When speaking about total reminder effort I examine reminder cost in addition to the first reminder letter. Whenever the first reminder was effective by encouraging a sample member to register, no further reminders had to be fielded for this particular sample member. In contrast, those sample members that do not register timely after receiving the first reminder will be eligible for a telephone reminder and eventually a second reminder letter (see figure 6.2.1). This makes recruiting those cases more expensive.

By increasing registration rates as a result of the first reminder letter, unconditional cash incentives may be cost efficient by saving additional reminder effort. The average cost of additional reminders (telephone reminder, second reminder letter) is € 8.67 per piece for all 1,203 sample cases that did not register within two weeks. In this cost I include all variable costs of the reminder process like printing, shipping and handling of letters, and the telephone reminder. I did not include staff costs as these costs are more or less fixed.

When splitting those costs by incentive treatment and control group, I can see that, on average, reminder efforts for the treatment group amount to € 7.20 per case (698 cases) and € 10.70 for the control group (505 cases) (see table 6.3.4). As incentives are € 5.00 per piece, bringing the total cost per case in the treatment group to € 12.2, I can conclude that incentives were not cost efficient on reducing reminder effort when solely considering those cost elements.

Table 6.3.4: Costs of additional reminders after first reminder letter by reminder incentive experiment group

	Control	*Incentive*	*Total*
Number of cases	505	698	1,203
Costs of additional reminders in €	5,404	5,026	10,430
Average cost of add. reminders per case in €	10.7	7.2	8.7
Incentive cost per case in €	0	5	2.9
Total cost per case in €	10.7	12.2	11.6

6.3.5.2 Total recruitment cost

In addition, I examine whether the reminder incentive is cost effective by reducing the overall recruitment cost.

To illustrate the effect of the reminder letter incentive I compare the realized sample and a scenario, where the 2012 GIP would have been recruited without the reminder letter incentives in terms of online panel registration success and fieldwork cost. Results are displayed in table 6.3.5.

Table 6.3.5: Comparison of total recruitment cost with and without incentive treatment

	Scenario: no exp.	*GIP*2012
Reminder letters sent	1,690	1,690
Registration rate for reminded cases	29.6	42.4
Registered panellists from reminded cases	500	716
Overall registered panellists	1,387	1,603
Total fieldwork cost 2012 GIP recruitment in €	546,052	551,432
Fieldwork cost per recruited panelist in €	394	344

Note: "Scenario: no exp.": scenario 2012 GIP recruitment without reminder letter incentive experiment; "GIP 2012": data from GIP 2012 recruitment

In the third column I display the results from the GIP 2012 recruitment as it was conducted. 1,690 reminder letters were sent in the experiment. From these cases 42.4 percent or 716 persons registered in the GIP online panel. Those 716 cases are part of the 1,603 panellists who registered during the whole GIP 2012 recruitment. The total recruitment cost for the 2012 survey were € 551,432. If we set those costs in relationship to the 1,603 panellists the per case recruitment costs are € 344.

For the scenario without the reminder incentive experiment I assume the reminder registration rate was that of the control group in the experiment of 29.6 percent for all cases (see table 6.3.1). All other effects in the scenario are kept as in the real 2012 recruitment. This is a realistic assumption as the reminder incentive experiment was a randomized controlled experiment and all other factors were kept constant. The numbers for the scenario derived from this assumption are dis-

played in column 2. 1,690 reminder letters without any incentive would have resulted in 500 registered cases, 216 less than in the actual 2012 GIP recruitment.

The total number of registered panellists in the scenario is thus 1,387 216 cases less than in the real 2012 recruitment. Total fieldwork costs in the scenario are € 546,052, reduced by the direct cost of reminder letter incentives(€ 5,380) compared to the real GIP 2012 recruitment cost (the cost for increased reminder effort can be neglected). Fieldwork cost by recruited panellist in the scenario without reminder incentive experiment are € 394, a € 50 increase compared to the real per case cost in the 2012 GIP recruitment.

If we now want to replace those 216 cases we need to increase the sample size, conduct more face-to-face interviews, invite more respondents to the GIP. The total cost of recruiting 216 additional GIP panel members at € 394 a piece is about € 85,000. This is the sum required to recruit the same sample size without incentives than what has been achieved with the reminder incentive experiment.

Comparing the € 5,380 investment in unconditional cash incentives with the huge savings of € 85,000 I can conclude that for every Euro spent on incentives there is a return of savings of about € 15.8. The unconditional cash incentive is cost effective in reducing the total recruitment cost.

6.4 Discussion

In this chapter I have shown that including unconditional cash incentives of a modest amount in the reminder process greatly increases the registration rate in the GIP. Including a € 5 note in the reminder letter significantly increases registration rates in the GIP online panel. In addition, the sign up process is quicker than in the control group. Including an unconditional cash incentive does not influence sample composition of the resulting GIP online sample. Panel retention rates are not affected by the incentive treatment. Cost analysis show that including a € 5 cash incentive is cost efficient as the small incentive investment brings a large return in recruited panellists.

These findings are in line with findings of a positive effect of reminder incentives on participation rates as reported by McGonagle & Freedman (2017). In contrast to the study by McGonagle & Freedman (2017) I do not find the incentives to

influence sample composition when comparing subgroup effects.

The discussion of cost effectiveness is an important contribution to the literature. Building on work by Scherpenzeel & Toepoel (2012) I show that the € 5 incentive treatment is cost effective by reducing the number of face-to-face recruited sample members that do not sign up for the online panel. This is to a certain extent due to the design of the 2012 GIP recruitment, where a large part of the fieldwork cost, the face-to-face interview, are fixed. The situation in other studies might be different and needs to be examined closely. However, I am confident that the finding of cost effectiveness of unconditional reminder cash incentives can be transferred to other situations.

In summary, I highly recommend the use of unconditional cash incentives in reminder processes. For these reasons, in the GIP 2014 refreshment sample recruitment, the € 5 unconditional cash incentives scheme was adopted for all sample members with the login letter.

Some people say it's folly, but I'd rather have the lolly.

Eric Idle, The Money Song

7

The SHARE Germany wave 4 incentive experiment

7.1 Introduction and background

[1]

The study reported in this chapter concerns an incentive experiment conducted in the 2011 German refreshment sample Survey of Heath, Ageing and Retirement

---

[1] This paper uses data from SHARE Wave 4 (DOI: 10.6103/SHARE.w4.600,), see B(Börsch-Supan, Brandt, et al., 2013) for methodological details.
The SHARE data collection has been primarily funded by the European Commission through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812) and FP7 (SHARE-PREP: N°211909, SHARE-LEAP: N°227822, SHARE M4: N°261982). Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1AG455301, IAG_BSR0611, OGHA_04-064, HHSN271201300071C) and from various national funding sources is gratefully acknowledged (see www.share-project.org).

in Europe (SHARE, see chapter 4.2). The SHARE incentive study fits this dissertation nicely, as it examines the effect of offering no incentives versus unconditional cash incentives. In addition, the optimal amount of incentives is assessed. Testing unconditional cash incentives of € 40, € 20 and € 10 meant using incentives of high value for the German survey contex. The incentive experiment was part of a research project examining ways to increase response rates in face-to-face surveys by means of respondent incentives and interviewer training. In addition, a nonresponse follow-up study was conducted to measure nonresponse bias (see Börsch-Supan, Krieger, & Schröder, 2013). In this chapter, I will report only on the respondent incentive part of the experiment. The interviewer training part of the project had to be cancelled due to fieldwork issues and the nonresponse follow-up is beyond the scope of this dissertation. An earlier version of these results have been published in Börsch-Supan, Krieger, & Schröder (2013) and have been modified to fit this dissertation [2].

7.2 The use of large cash incentive sums in Europe

I have reported in chapter 2.2 that the research on the right amount of incentives is mixed. While some studies find that larger incentive values increase response rates, others find little to no effect. A unique feature of the SHARE incentive experiment is testing different values of large unconditional cash incentives in a face-to-face study in Europe. To date there have been only two other studies that experimented with similar amounts of unconditional cash incentives in Europe: The

[2] I gratefully acknowledge funding for the project "Respondent incentives, Interviewer Training and Survey Participation" by the German Research Foundation (DFG) through the "Priority Programme Survey Methodology 1292" (Coordinator: Uwe Engel). Principle Investigators of the Project were Axel Börsch-Supan and Matthis Fräßdorf (Schröder). Researchers working on this project were Annelies Blom and Ulrich Krieger. I am thankful for the help provided by our student research assistants Ute Hoffstätter, Anna Krüger, Elisa Leonhard, and Maximilian Weiß. At the Institut für angewandte Sozialforschung (infas) Anne Kersting and Birgit Jesske were working on the project and I thank them for their valuable input and oversight of the fieldwork implementation. I thank Frauke Kreuter, Josef Brüderl, Arie Kapteyn, Mick Couper, Peter Lynn, Peter Lugtig, Julie Korbmacher, and Stephanie Eckman for helpful comments. Parts of this work have been presented at the Allbus Methodenworkshop 2011, the International Panel Survey Methods Workshop 2012, the RC 33 research conference, and the International Conference of the DFG Priority Programme 1292 on Survey Methodology.

LISS panel in the Netherlands (Scherpenzeel & Toepoel, 2012) and PIAAC Germany field test (Martin et al., 2014). In the LISS panel incentive amounts of € 10, € 20 and € 50 were tested during the recruitment of the LISS probability-based online panel. Using € 20 and € 10 resulted in quite similar recruitment rates of 40.3% and 39.9% respectively. Using a € 50 cash incentive however, resulted in a lower recruitment rate of 37.3%. Overall, Scherpenzeel & Toepoel (2012) concluded that the € 10 was the most efficient incentive amount.

In contrast, following a field test PIAAC concluded that a € 50 unconditional cash incentive was the most cost-efficient resulting in a response rate of 41.7%. Response rates increased monotonously with increasing values of the offered cash incentive (Martin et al., 2014).

7.3 Data and methods

7.3.1 Sample

The SHARE recruitment incentive experiment was conducted during the 2011 SHARE Germany Wave 4 refreshment sample interviews. Fieldwork procedures are described in section 4.1.

7.3.2 Experimental design of the incentive experiment

When the project was started in September 2010, the threefold design of the experiments was finalized in cooperation with the team running the SHARE survey at the MEA and the agency contracted for fielding the survey, the Institut für angewandte Sozialforschung, Bonn (infas).

The first part of the study evaluates whether prepaid cash incentives increase cooperation rates and how these rates are influenced by different amounts of cash incentives. The implementation of this experiment is no trivial matter, as unconditional prepaid cash incentives are not commonly used to increase response rates in Germany. In contrast to the original plan of running the incentive experiment double blind (i.e. neither interviewers nor respondents would know about the experiment), we decided to run the experiment half blind, informing interviewers

about the treatment condition of respondents, while leaving the respondents uninformed.

The reasons for this decision were twofold. First, the agency had doubts about the practical implications of not informing the interviewers about such a considerable design feature. They deemed it important for the interviewers to be fully informed about all aspects of the survey to present the study correctly and positively when establishing contact. If respondents with questions about the incentives they received met an interviewer who was completely unaware of this particular study feature, it may have reflected badly on the interviewer, the agency, the survey sponsor, and the study as a whole.

Second, informing interviewers about the incentives creates a more realistic scenario. When using incentives in subsequent surveys (without experimental control), interviewers are well aware of the presence of incentives. They can refer to the incentives and use them as a reference when making contact. So in a real setting, incentives will always jointly affect interviewers and respondents. Running the experiment blindly on the respondents' side only thus increases the generalizability of our findings to future survey research projects.

The entire refreshment sample was divided into four different batches to be fielded sequentially. The batches were sized so that the target size of 4,000 cases would be realized if a response rate of 55 percent (5,247 addresses), 45 percent (1,050 additional addresses) or 40 percent (840 additional addresses) could be realized during fieldwork. 2,098 addresses were kept as an additional reserve. To allow sample members in the treatment group as much time as possible in the field, the experiment was run in the first batch of addresses that were sent out right at the beginning of the field time.

Individuals from PSUs in communities with fewer than 9,000 inhabitants were excluded from the experiment, because there was a considerable risk of respondents finding out about other respondents' incentives and thus of confounding the treatment conditions in these smaller towns and villages. Out of the total of 210 PSUs that the refreshment sample was drawn from, 54 PSUs were excluded for this reason.

After the addresses had been drawn from the registry, the survey agency delivered the sample to SHARE and the project team. Experimental conditions and the

control group were randomly allocated in all of the 156 larger PSUs in the survey. Therefore, all PSUs in the experiment contained sample members of the three experimental conditions and the control group. As a consequence, the experiment was run in all German regions and all interviewers had sample cases in all experimental conditions, as interviewers usually work in a certain PSU. Table 7.3.1 gives a summary of the experimental setup, where the size of the treatment groups was chosen according to considerations of statistical power for the significance tests.

Table 7.3.1: Summary of sample sizes

	PSUs	Sample persons
Full sample	210	9,235
First batch	210	5,247
Excluded (small communities)	54	1,347
Part of the experiment	156	3,900
Experimental conditions		
€ 40 prepaid	156	750
€ 20 prepaid	156	750
€ 10 prepaid	156	1,375
No prepaid incentive	156	1,025

7.3.3 Fieldwork issues - reduced sample size

Prior to proceeding to the results section, two developments during fieldwork have to be mentioned: problems during fieldwork and reactions of respondents to incentives.

Unfortunately, the survey operations conducted by infas did not proceed as planned. Due to high interviewer workloads across infas studies, too few interviewers worked actively on the SHARE survey. This led to massive delays in both the wave 4 refreshment sample and the SHARE Germany panel cases.

Confronted with the delays in fieldwork and continued capacity problems at the agency, the SHARE coordination decided to halt the refreshment sample fieldwork and concentrate on the panel cases. Nonetheless, fieldwork progress kept slow, panel cases took about 15 months to be completed. The refreshment sample fieldwork could not be resumed after having been halted.

This leads to the unfortunate situation that not all sample cases have been worked on completely and, thus, the incentive experiment is confounded.

7.3.4 Respondents' reaction to incentives

In addition to these issues related to the infas fieldwork capacity, there was another unforeseen development in the field. From the start of fieldwork in early June 2011, when the first advance letters reached the sampled households, the survey agency registered an unusually high volume of calls to the hotline. During the first week of fieldwork, a total of 85 calls were received. In a memo sent to the SHARE team, infas reported about calls of anxious respondents, who were unhappy with the cash provided in their advance letter.

Most of the people calling the hotline questioned the legitimacy of the survey, some even felt pressured into the survey – quite the opposite of the intended effect. Note, however, that the overwhelming majority of the respondents (almost 94 percent), did not voice concerns. In response to the complaints, the hotline staff spoke to respondents and also to police, lawyers, the Arbeitskreis deutscher Markt- und Meinungsforschungsinstitute (the business association representing the private-sector market and social research agencies), and representatives of municipalities that received complaints by sample members. One complaint to a member of the German Bundestag reached high levels at the Federal Ministry of Education and Research, the sponsors of SHARE in Germany.

Although the infas staff could reassure most callers of the sincere motives of the project, at the time infas informed the project team at SHARE, already 3,065 advance letters of the 5,421 respondents from the first batch had been mailed to the households, of which 1,723 contained a cash incentive. As a result of the infas memo, the advance letter was changed, putting even more emphasis on the voluntary nature of the survey and the unconditional gift character of the incentive.

In addition, a second advance letter was mailed to the 3,065 sample members who had already received the original version of the advance letter. There was little indication that this new letter helped to reduce skepticism among potential respondents, as hotline calls remained frequent during the mailing of the remainder of the advance letters and only dimmed off five weeks into the field period, when

all letters had reached sample members.

In summary, fieldwork was particularly difficult for SHARE wave 4 in Germany. Whereas the reactions to incentives turned out to be a smaller issue (see table 7.4.4) the reduction in sample size as a result of the halted fieldwork of the refreshment sample dampens the analytical potential of the incentive experiment. However, as shown in the following section, there are sill interesting lessons to be learned from the experiment.

7.3.5 Data

Data for this analysis has three sources: register data from the sampling frame, paradata from conducting the interviews and SHARE Germany wave 4 data.

7.3.5.1 Register data

The SHARE wave 4 refreshment sample was drawn from register data. Thus as a byproduct of sampling I have a set of information available on all sample members. More specifically, the register information on gender and being born 1957-1960 are available for all sample members. The information on younger age cohorts is used to generate the dummy variable "50-54 years old".

For sampling, the population figures of communities the PSUs were located in where used. (Lynn et al., 2013). Those population figures were used in the analysis. Communities with a population size smaller than 35,000 inhabitants are coded into the dummy variable "smaller sample points", communities with more than 200,000 inhabitants are coded into the dummy variable "larger sample points".

7.3.5.2 SHARE wave 4 paradata

From SHARE wave 4 interview paradata I use data to identify experimental groups. In addition, a dummy variable indicating the second version of advance letters with changed wording was used for the analysis.

Furthermore, the survey agency infas recorded hotline refusals and made them available for analysis. Contact is coded using SHARE sample management system data. All eligible cases that were contacted are coded 1 whereas all eligible non-contacts are coded 0. Cooperation is coded using SHARE sample management

system data outcome codes. All full interviews are coded as 1, all eligible sample members or those of unknown eligibility not providing an interview are coded as 0.

7.3.5.3 SHARE wave 4 data

For sample composition analysis I use variables from the SHARE Germany Wave 4 interview. From interview data I select variables that are of general interest to SHARE.

As a measure of poor health I constructed two dummy variables from self reported mediocre and poor health (PH003). Based on the question about the current occupation (EP005) I coded dummy variables for currently being in a job or self employment and for being in retirement. Number of children is directly taken from the SHARE question CH001. Information on respondents education is taken from variable DN010 on the highest school leaving certificate and Dummies for"Abitur" and "Hauptschulabschluss" are created. The question about the degrees of higher education or vocational training (DN012) was used to code the dummy variables for mentioning vocational training and university education.

Respondents saying they pray daily or multiple times a day (EX029) were coded 1 creating the dummy variable "praying often". Self placement on the left-right political ideology continuum is taken from the SHARE variable EX028. As proxy for cognitive ability I used results from a math problem SHARE respondents had to solve (CF108-CF112). Those who solved the problem are coded 1 into the variable "cognitive ability" and others are coded 0. Respondents who had taken part in at least one of the social activities presented in SHARE question AC035 (e.g. charity work or community-related organization) were coded 1 in the variable "social activities" and 0 otherwise.

7.4 Results

Due to the fieldwork problems, the analysis of incentive effects is restricted to addresses that either have a final disposition code (ineligible, refusal or interview) or could not be contacted during the field period while exhausting the minimum

number of eight personal visits to the housing address. These selected addresses have been worked on properly according to the field protocols.

7.4.1 Bivariate effect of incentive treatment on cooperation rates

Out of the 3,900 addresses originally in the experiment 2,241 households meet this condition. Cooperation with the survey request is measured at the household level. While SHARE is a survey of individuals and incentives were targeted to the persons drawn from the register, the survey also targets cohabiting household members. As the incentive may also have an influence on these partners of the sampled person, we look at cooperation of at least one household member. Table 7.4.1 shows the results of the three treatment conditions and the control group.

Table 7.4.1: Effect of incentives on cooperation

	Control		€10		€20		€40	
	(N)	(%)	(N)	(%)	(N)	(%)	(N)	(%)
Not cooperating	398	72.6	494	61.7	258	59.2	209	45.8
Cooperating	150	27.4	307	38.3	178	40.8	247	54.2
Total	548	100	801	100	436	100	456	100

The proportion of sample members cooperating with the survey request increases monotonously with the value of the incentives. The control group has the lowest cooperation propensity of the four groups (27.4%). Paying €10 cash unconditionally to sample members yields a cooperation rate of 38.3% and thus increases the likelihood of taking part in the survey by 10.9 percentage points compared to the control group that received no incentive upfront (38.3% versus 27.4%). The €20 incentive group has a larger cooperation rate, the percentage difference to the control group is 13.4 (40.8% versus 27.4%). The cooperation rate of the €40 group is the highest, 26.8 percentage points higher than that of the control group (54.2% versus 27.4%). All differences to the control group are statistically significant indicating that paying larger incentive amounts increases the participation to a larger extent than the smaller cash amounts.

Comparing between incentive groups, the difference between €10 and €20 unconditional cash incentive is not statistically significant (38.3% versus 40.8%,

%d=2.5). I observe a significant percentage point difference of over 15.9 between the € 20 and the € 40 incentive treatments (54.% versus 38.3%). This contradicts the work by Scherpenzeel & Toepoel (2012) who found a diminishing return of large incentive offerings in the LISS panel. Our findings are in line with a meta study conducted by Singer, Hoewyk, et al. (1999) that found a significant, monotonic effect for incentive sizes.

7.4.2 Multivariate analysis of incentive treatment on cooperation rate

To check whether these findings hold in multivariate analysis, a fixed-effects logistic regression on household cooperation was conducted. As households are nested in interviewers, we use interviewer fixed-effects. Explanatory variables are limited to frame variables available for all respondents and nonrespondents. In this regression gender, age group, having received the second version of the advance letter (see section 7.3.4), and the incentive treatment are controlled for. Population size of the municipalities of the sampling points is controlled by adding dummy variables for the lower third (9,000-35,0000 inhabitants) and the upper third of the population distribution (more than 200,000 inhabitants) respectively. Table 7.4.2 shows the results.

Table 7.4.2: Fixed-effects logistic regression on household cooperation

	odds ratio coefficients	standard error
€ 10 incentive	1.91***	0.27
€ 20 incentive	2.18***	0.38
€ 40 incentive	4.07***	0.72
Male target person	1.25*	0.14
50-54 years old	1.10	0.14
2nd Version of advance letter	0.89	0.20
Smaller sample point	0.49*	0.15
Large sample point	0.93	0.25

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

N: 2,751; McFadden's R^2: 0.05; 107 interviewers

In addition to the effects of the incentive treatment, two control variables yield significant results: gender and living in smaller municipalities. Households of male

target persons have 25 percent higher odds of cooperation than if the target person is female. The odds for cooperation of households in communities with less than 35,000 inhabitants are reduced by about 50 percent compared to households from towns that have between 35,000 and 200,000 inhabitants.

Regarding the main effect, we show the contrasts of the control group to the treatment groups. All those effects are positive and significant, meaning that the odds of responding are larger for all incentive groups other than the control group. For the € 10 and € 20, the odds are about twice as large as in the control group, for the € 40 group the odds of responding are four times as large as in the control group.

The positive effect of incentives on response propensity can thus also be shown in multivariate analysis. But given that fieldwork had to be terminated early, one possible explanation for these positive effects could be selective interviewer efforts.

7.4.3 Effect of incentives on contact attempts

Interviewers might choose a strategy where cases in the incentive condition are preferred by Interviewers or worked on harder by Interviewers. Such a strategy would potentially influence the survey outcome.

Given that interviewers are aware of the incentive treatments and they are primarily paid per completed case, interviewers have strong incentives to focus their efforts on those cases that have been offered an incentive due to the anticipated higher chances of a successful interview.

I therefore analyze the effect of the incentive treatment on the contact propensity of households. If interviewers prioritized contacting incentivized households, we would expect to find effects of incentives in this analysis. As above, we use gender, age, the advance letter version, and PSU population size dummies as control variables. Individual differences in interviewers' effectiveness are again controlled for by using interviewer fixed effects. Table 7.4.3 below shows the results of the fixed-effects logit regression on contact.

For this analysis all 3,900 cases from the experiment are included in the analysis, however, 778 addresses from 29 interviewers have to be dropped from the analysis

due to lack of variation on the dependent variable. These interviewers either made contact with all addresses or no addresses at all.

Table 7.4.3: Fixed-effects logistic regression on household contact

	odds ratio coefficients	standard error
€ 10 incentive	1.05	0.18
€ 20 incentive	0.96	0.18
€ 40 incentive	1.01	0.20
Male target person	0.87	0.10
50-54 years old	0.58^{***}	0.07
2^{nd} Version of advance letter	1.16	0.27
Smaller sample point	0.62	0.25
Large sample point	1.70	0.72

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

N: 2,824; McFadden's R^2: 0.02; 88 interviewers

The only significant effect in this analysis is the negative coefficient of being in the age cohort from 50 to 54 years. The odds of those sample persons being contacted are 38 percent smaller than those of the older sample persons. There is no significant effect of the incentive treatment groups, indicating that the contact propensity does not depend on whether a prepaid incentive has been sent or not. Hence we find no evidence that the interviewers used the information on the incentive treatment to make their contact decision.

This is plausible since we know from talking to the agency and the interviewers that driving distances between households is their foremost consideration when planning their work. While interviewers get some compensation for travel cost, their main income stems from finalized interviews. Thus, they try to optimize driving distances between addresses. Given that incentivized addresses are distributed randomly, incentives did not affect the contact strategies of interviewers.

7.4.4 Incentives leading to refusals

As mentioned in section 7.3.4, the unconditional incentive treatment also had negative consequences, since an unusually high number of sample persons were worried about the sincerity of the project and called the study hotline. Some of those

calls resulted in sample members refusing their participation right away. In those cases, no interviewer can be sent out to try to achieve their participation.

197 respondents out of 3,900 targeted sample members in the SHARE wave 4 refreshment experiment refused via the hotline, accounting for about 50 percent of the total call volume to the hotline. The distribution with regard to treatment groups is shown in Table 7.4.4.

Table 7.4.4: Hotline refusal by treatment groups

	Control	€ 10	€ 20	€ 40	Total
Number of refusals	16	86	45	50	197
Number of sampled households	1,050	1,350	750	750	3,900
% of hotline refusals per sampled hh	1.5	6.4	6.0	6.7	5.1

Note: hh: households

Table 7.4.4 clearly shows that refusals via hotline occurred more frequently among the treatment groups. The proportion of sample members refusing is about equal across treatment groups at about 6.5 percent and the differences is not statistically significant. Refusal in the control group is considerably lower at 1.5 percent of the sample members.

This result may be interpreted as an indication that the increased resentment towards the survey in this group is due to the sending of cash as such, not due to the amount of cash. The loss of sample members to hotline refusals has to be put into the perspective of the large gains in cooperation in the treatment conditions. Moreover, while one has to take the hotline refusals seriously, the great majority (almost 95 percent of sample members) did not voice concerns about the incentive treatment and cooperated in greater numbers with survey request than those not in the control group.

The lesson learned here is that unconditional cash incentives tend to come with a larger number of respondents voicing their concerns. A survey operation needs to be aware of this fact and have adequate response procedures in place to reassure respondents, the public, funders, and other stakeholders of the legitimacy of the survey and the use of incentives. As a direct consequence a report on incentive use in Germany has been commissioned showing the practice and success of incentive use in Germany (Pforr et al., 2015).

7.4.5 Cost efficiency of incentives

I have shown that using cash incentives comes with the benefit of an increased response rate and thus additional cases for analysis. But besides this scientific benefit, are incentives also cost efficient when fielding a face-to-face survey?

I analyze the cost structure by applying the cost calculations as has been suggested by Scherpenzeel & Toepoel (2012) when they analyze the cost effectiveness of a LISS panel experiment.

Using the cooperation rates for experimental conditions in the SHARE incentive experiment I determine the gross sample size needed to obtain a final sample of 500 panel participants. 500 participants is an arbitrary sample size used to make results comparable between groups and to the results obtained in Scherpenzeel & Toepoel (2012). I then determine the cost of such a sample by calculating the fieldwork cost and adding the incentive cost. Results of the cost calculations for the SHARE incentive experiment are displayed in table 7.4.5.

Table 7.4.5: Costs per cooperating household (hh) to obtain a final sample size of 500 households, calculated for different incentive conditions

Incentive	Cooperation rate[a] (%)	Gross sample size [b] (hh)	Total fieldwork cost[c] (€)	Total incentive cost (€)	Total costs per hh[d] (€)
No incentive	27.4	1,825	94,900	0	190
€ 10	38.3	1,305	67,860	13,050	162
€ 20	40.8	1,225	63,700	24,500	176
€ 40	54.2	923	47,996	36,920	170

[a] Rates from table 7.4.1

[b] 500 registered households/(predicted registration rate / 100)

[c] Fieldwork cost are on average € 52 per case

[d] (Total fieldwork cost + total incentive cost)/500 registered households

Using the observed cooperation rates from table 7.4.1 the required sample sizes to obtain 500 households in the final sample are calculated. Higher response rates require smaller gross samples. For the control group a sample size of 1,825 is needed to obtain 500 panel respondents given the cooperation rate of 27.4 percent, while in the € 40 group a sample of 923 cases is sufficient to reach the goal

of 500 respondents (cooperation rate: 54.2%). Total costs are then calculated using the € 52 per case fieldwork cost of the SHARE wave 4 Germany refreshment sample. For the control group, the costs are 1,825 cases x € 52 + € 0 incentive cost=€ 94900. For the € 40 condition the total costs are calculated as 923 cases x € 52 +36,920 incentive cost(€ 40 x 923 cases)=€ 84,916. The last column of table 7.4.5 displays the total cost per household (total cost divided by 500) for each experimental condition.

The table shows that the total cost per household is highest in the no incentive condition (€ 190) and lowest in the € 10 incentive condition (€ 162). Each of the incentive conditions results in lower total costs per household than the no incentive condition (€ 20 condition: € 176, € 40 condition: € 170). Offering an incentive is always more efficient than no incentive. The substantial direct cost of cash incentives are offset by the recruitment success when using incentives.

While the SHARE and the LISS survey have very few design characteristics in common, these findings are in line with the findings reported by Scherpenzeel & Toepoel (2012, table 4), who also found a € 10 cash incentive to be most cost effective with a per household cost of € 150 when using CAPI interviews without prior phone contact to respondents.

In contrast to my results, in the LISS experiment the high € 50 incentive condition compared more unfavorably to the € 10 condition with a cost of € 268 per household. In contrast in the SHARE results the high € 40 is almost as efficient as the € 10 condition. This is due to the high response rates in the € 40 condition which leads to a total household cost that is much lower (€ 170). The high incentive condition in the SHARE survey poses a more efficient alternative than the high incentive condition in the LISS experiment. I doubt that the € 10 difference in the incentive amount is the driving factor of this finding. There might be a different, unknown underlying explanation to this difference, possibly the effect is nonlinear but there is a threshold between "low" and "high" incentives.

This analysis shows that using prepaid cash incentives does not only increase response rates but can also be cost efficient by reducing the gross sample size needed and thereby reducing the total field work cost.

7.4.6 Sample composition

To check for effects of the different incentive treatments on sample composition I compare the sample persons characteristics in the achieved net sample. The net sample resulting from the gross sample used for the incentive experiment yielded 882 cases (see 7.4.1). Because of sample size restrictions, I group all incentive conditions together and compare them to the control group. For comparison I use basic demographic variables, variables indicating community involvement, and variables of interest to SHARE like retirement, cognitive ability, income, and health. Results are displayed in table 7.4.6.

Table 7.4.6: Comparison of means between groups (standard deviations in parentheses)

	Incentive		No Incentive	
	mean	sd	mean	sd
Male	.49	(.50)	.49	(.50)
Age 50-54	.17	(.38)	.24	(.43)
Health: fair	.30	(.46)	.24	(.43)
Health: bad	.13	(.33)	.12	(.33)
Working	.31	(.46)	.36	(.48)
Number of children	1.86	(1.35)	1.89	(1.39)
Low school education (Hauptschule)	.49	(.50)	.45	(.50)
High school (Abitur)	.17	(.37)	.18	(.39)
Vocational training	.54	(.50)	.58	(.49)
University education	.21	(.41)	.28	(.45)
Voluntary work	.20	(.40)	.19	(.39)
Social activities	.40	(.49)	.41	(.49)
Left right scale	4.96	(1.90)	4.88	(2.05)
Praying often	.25	(.43)	.22	(.41)
Cognitive ability	.73	(.44)	.73	(.44)
Retired	.51	(.50)	.43	(.50)
hh income	4,737	(7,635)	5,028	(7,088)
Observations	732		150	

Note: sd = standard deviation

When comparing SHARE respondents with and without incentives, three

mean comparisons are significant. The proportion of younger sample members (50 to 54 years) is significantly larger than among those respondents 55 and older (t=1.90, two-sample t-test). From frame data we know the true value in the gross sample for this age cohort is 18 percent. The incentivized sample is thus closer to the true value. In addition, two mean differences are significant. In the treatment groups, the proportion of retired respondents is 8 percentage points larger (t=-1.79, two-sample t-test) and the proportion of respondents reporting a university education is 7 percentage points lower (t=1.95, two-sample t-test) than in the control group. Other mean differences do not reach the level of statistical significance.

I provide some evidence for an effect of incentives on sample composition. For the youngest age cohort we have some evidence that incentives help to achieve a more balanced sample. For the effects incentives show when comparing university education and being retired we can not tell if incentives bring the findings in the realized sample closer to the true value or not as no frame data is available.

7.5 Summary

This chapter provides an overview of an experimental study implemented in the German sub-sample of SHARE. From the results we can draw five conclusions. First, prepaid cash incentives are effective in increasing response rates. Second, larger cash incentive values lead to higher response rates. Response rates are monotonously increasing with an increase in cash incentive value. Third, prepaid cash incentives are cost effective as they reduce the gross sample size to be contacted during fieldwork by increasing cooperation rates. The € 10 condition is the most cost efficient. Fourth, there is little evidence for influences of the incentive treatments on sample composition. Comparing samples across experimental groups, I observe more similarities than differences. Fifth, prepaid incentives (regardless of the amount) lead to higher initial refusals even before the interviewer attempts contact with the household. These refusals are outweighed by the gains in response rate.

Comparing our finding with prior studies with large incentives in Europe we find the SHARE results are similar to the results from the PIAAC Germany field

test (Martin et al., 2014) where response rates increased monotonously with incentive value. In contrast to the SHARE findings PIAAC found the € 50 incentive condition the most cost effective. In the SHARE experiment as in the LISS panel (Scherpenzeel & Toepoel, 2012), the € 10 condition is the most cost effective. However, in contrast to the SHARE findings, the LISS panel experiment found that the largest incentive condition of € 50 had lower recruitment rates than the € 10 and € 20 incentive conditions.

Comparing the three experiments in PIAAC, the LISS - Panel and SHARE, all studies generally find unconditional cash incentives to improve sample members cooperation. Their results differ when the results for different experimental conditions of different incentive values are compared.

There is a huge caveat with this analysis, as the fieldwork was aborted before completion. Therefore, I can´t tell if and how continued fieldwork would have affected the results. Maybe the cases coming in late in fieldwork would have been harder to reach and thus reduced overall response rates. If this would be the case the effect of large cash incentives reported in this study would be too high (compared to the response rates in the PIAAC incentive experiment by Martin et al. (2014)). However, I can also argue that the wave 4 SHARE Germany refreshment sample with its severe fieldwork problems is a very rigorous test for an incentive intervention. The true incentive effect, given regular fieldwork, might be even larger. Despite the caveat the SHARE study and the literature show how powerful large unconditional cash incentives are to improve survey participation.

As an outlook I see two further worthwhile avenues of research in the topic of large cash incentives, especially in Germany. One would be to further investigate the cost effectiveness of large cash incentives. In studies designed and conducted in cooperation with fieldwork companies the relationship between survey cost, fieldwork effort and respondent incentives could be examined in more detail. This would allow to better allocate survey budgets to incentives, fieldwork management or interviewers to maintain or even improve survey quality given budget restraints.

A second research question would be that of long-term effects of large cash incentive offerings. The economic literature on motivation crowding out effects of pro-social behavior (Frey & Jegen, 2001) describes the effect of people becoming unwilling to continue with a pro-social behavior once they get a financial reward

for it. If the crowding out effect can be applied to survey participation, offering incentives would, in the long run, have a negative effect on survey participation. While there is evidence for crowding out effects in other areas the effect on survey participation has not been tested.

With money you can make a splash!

Eric Idle, The Money Song

8 Conclusion

In this thesis I presented three experiments using unconditional cash incentives. In this concluding chapter I will recapitulate the findings and their implications for survey research in Germany.

In the first experiment, the GIP recruitment incentive experiment (chapter 5), the use of € 5 unconditional versus € 10 conditional incentives is tested. The incentive value is chosen to be cost neutral to the research budget as unconditional incentives are paid out to twice as many sample members as the conditional incentives given a response rate of about 50%.

The results show that unconditional cash incentives have a larger effect on response rates without introducing a bias into the resulting sample. Given these findings and that using unconditional cash incentives was cost neutral, unconditional incentives are clearly to be preferred to conditional incentives based on the results of this experiment.

The second experiment, the GIP reminder incentive experiment(chapter 6) is about testing the addition of a € 5 unconditional cash incentive to a reminder let-

ter against a control group without incentive treatment.

The results show that adding the incentive increases the registration rate, does not alter sample composition at registration, has a weak effect on sample composition after 12 panel waves, and is cost-efficient by reducing total recruitment cost. Based on the results of the reminder incentive experiment the use of unconditional reminder incentives is highly recommended.

The third experiment, the SHARE Germany wave four recruitment incentive experiment(chapter 7) tests the use of unconditional cash incentives of € 10, € 20, or € 40 against a no incentive control group. Results show a monotonously increasing response rate with increasing incentive value. There is litte evidence for sample composition influence of the incentive treatment. € 10 unconditional cash incentives is the most cost-efficient incentive treatment, closely followed by the € 40 treatment.

Table 8.0.1 shows an overview of the results of all experiments.

Table 8.0.1: Overview of results

Category	Experiments	Results
Main effect of incentives	SHARE, GIP rem	Incentives raise response rates
Prepaid versus promised incentives	GIP	Prepaid perform better
Optimal incentive value	SHARE	Higher values lead to higher response rates
Cost effectiveness	GIP rem, SHARE	Incentives are cost effective
Sample composition	GIP, GIP rem, SHARE	Little to no effect of incentives on sample composition.

Note: GIP=GIP recruitment incentive experiment; GIP rem=GIP reminder incentive experiment; SHARE= SHARE Germany wave 4 incentive experiment

SHARE and GIP reminder incentives tested the main effect of incentives finding that using incentives is more favorable than not using incentives. Prepaid versus promised incentives are tested in the GIP recruitment incentive experiment finding that unconditional incentives lead to higher response rates. The incentive value is examined in the SHARE experiment finding that higher incentives lead to

higher response rates. The GIP reminder incentive experiment and the SHARE experiment show that incentives are cost-efficient compared to not using incentives. All experiments provide little to no evidence for effects on sample composition.

The experiments in this dissertation confirm the majority of the literature findings that unconditional incentives have stronger effects on response rates than conditional incentives. Strong effects of large incentives on response rates have been shown in the literature before, however mainly in surveys outside of Europe. Widening the scope of incentive effects from response rates to sample bias is an important step in increasing the knowledge on incentive effects in surveys. The discussion of cost effectiveness of survey incentives in Germany is a new and important contribution to the literature on incentives. Showing that survey incentives are cost-efficient will help make the argument for the use of incentives when communicating with survey sponsors and the public.

While the experiments in this dissertation answer some questions, other topics of survey incentives remain to be researched in the future.

One aspect is the use of incentives in mixed mode surveys. Increasingly, surveys are conducted in multiple survey modes to mix the advantages of several survey modes or to reduce cost. When combining for example web, telephone and face-to-face modes new challenges for the use of incentives arise. Optimal incentive values or payment times or form of incentives may vary across survey modes. Thus, the incentive strategy needs to be optimized in conjunction with the mixed-mode design.

A second topic for future research is the cost effectiveness of incentives, especially to the survey environment in Germany. As mentioned in chapter 6, it would be helpful to examine the effects of incentives on the fieldwork, collaborating with the survey agency conducting the fieldwork. Only such a collaborative setting would allow to precisely measure how respondents and interviewers react to incentive treatments and how this influences the survey process.

A third aspect that should be examined in future research is the cost of incentives in comparison to other survey costs. When distributing a survey budget to get high quality survey data, investing in incentives is just one possibility to allocate the budget. Other aspects might include investing in interviewer training, interviewer incentives or improvements in questionnaire design. As budgets are lim-

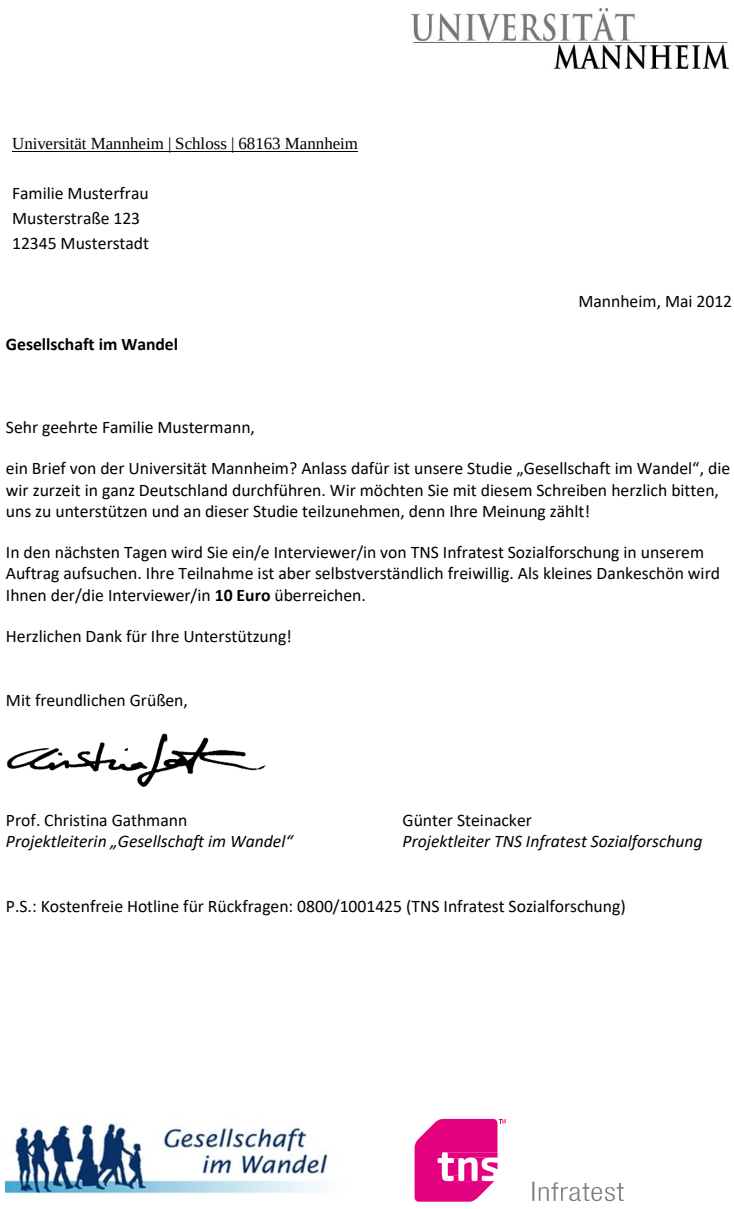
ited, spending money on one design feature of a survey prevents using the money to improve other aspects of a study. Studies that experimentally test these trade offs would help researchers make evidence based design decisions.

However, the basic question of this dissertation has been answered: unconditional cash incentives help raise response rates, are cost effective and do not influence sample composition. This is the same conclusion made by J. Scott Armstrong 43 years ago. He concludes his review paper stating:

"What have we learned after four decades of research on monetary incentives in mail surveys? The only type of monetary incentive that has an impact on the nonresponse rate is the prepaid monetary incentive ... . The reductions in nonresponse are substantial" (Armstrong, 1975, p. 116). 43 years later, I conclude that these findings are true for face-to-face surveys in this dissertation. In addition, I showed that there are no effects of incentives on sample composition and that incentives can be cost effective. While incentives are not the be-all and end-all solution to all problems in survey methodology, they are an important tool to keep respondents returning to our surveys.

9

Appendix



UNIVERSITÄT MANNHEIM

Universität Mannheim | Schloss | 68163 Mannheim

Familie Musterfrau
Musterstraße 123
12345 Musterstadt

Mannheim, Mai 2012

Gesellschaft im Wandel

Sehr geehrte Familie Mustermann,

ein Brief von der Universität Mannheim? Anlass dafür ist unsere Studie „Gesellschaft im Wandel", die wir zurzeit in ganz Deutschland durchführen. Wir möchten Sie mit diesem Schreiben herzlich bitten, uns zu unterstützen und an dieser Studie teilzunehmen, denn Ihre Meinung zählt!

In den nächsten Tagen wird Sie ein/e Interviewer/in von TNS Infratest Sozialforschung in unserem Auftrag aufsuchen. Ihre Teilnahme ist aber selbstverständlich freiwillig. Als kleines Dankeschön wird Ihnen der/die Interviewer/in **10 Euro** überreichen.

Herzlichen Dank für Ihre Unterstützung!

Mit freundlichen Grüßen,

Prof. Christina Gathmann
Projektleiterin „Gesellschaft im Wandel"

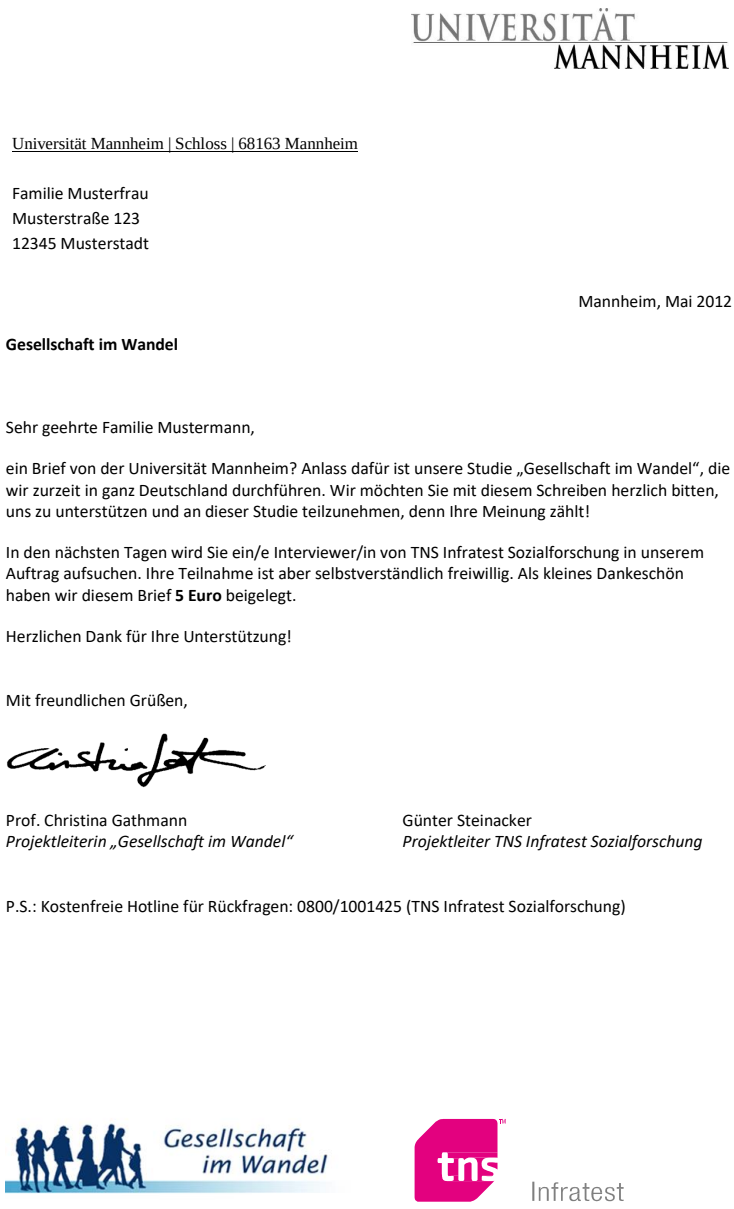
Günter Steinacker
Projektleiter TNS Infratest Sozialforschung

P.S.: Kostenfreie Hotline für Rückfragen: 0800/1001425 (TNS Infratest Sozialforschung)

Gesellschaft im Wandel

tns Infratest

Figure 9.0.1: Advance Letter GIP 2012 recruitment, conditional incentive



UNIVERSITÄT MANNHEIM

Universität Mannheim | Schloss | 68163 Mannheim

Familie Musterfrau
Musterstraße 123
12345 Musterstadt

Mannheim, Mai 2012

Gesellschaft im Wandel

Sehr geehrte Familie Mustermann,

ein Brief von der Universität Mannheim? Anlass dafür ist unsere Studie „Gesellschaft im Wandel", die wir zurzeit in ganz Deutschland durchführen. Wir möchten Sie mit diesem Schreiben herzlich bitten, uns zu unterstützen und an dieser Studie teilzunehmen, denn Ihre Meinung zählt!

In den nächsten Tagen wird Sie ein/e Interviewer/in von TNS Infratest Sozialforschung in unserem Auftrag aufsuchen. Ihre Teilnahme ist aber selbstverständlich freiwillig. Als kleines Dankeschön haben wir diesem Brief **5 Euro** beigelegt.

Herzlichen Dank für Ihre Unterstützung!

Mit freundlichen Grüßen,

Prof. Christina Gathmann
Projektleiterin „Gesellschaft im Wandel"

Günter Steinacker
Projektleiter TNS Infratest Sozialforschung

P.S.: Kostenfreie Hotline für Rückfragen: 0800/1001425 (TNS Infratest Sozialforschung)

Gesellschaft im Wandel

tns Infratest

Figure 9.0.2: Advance Letter GIP 2012 recruitment, unconditional incentive

Universität Mannheim | Schloss | 68131 Mannheim

Frau «VName» «NName»

«HHAdr_1» «HHAdr_2»
«HHAdr_3» «HHAdr_4»

Mannheim, 2. November 2017

Gesellschaft im Wandel

Sehr geehrte Frau «NName»,

vor ein paar Wochen haben wir Sie per Brief zur online durchgeführten Studie „Gesellschaft im Wandel“ eingeladen. Da Briefe manchmal verloren gehen, möchten wir Sie heute gerne noch einmal auf unsere Studie aufmerksam machen.

Für Ihre Teilnahme melden Sie sich möglichst bis zum (Datum + 2 Wochen) unter

www.gesellschaft-im-wandel.de

an, damit Sie dann schon an unserer nächsten Umfrage teilnehmen können. Die Anmeldung erfolgt mit Ihrer persönlichen Kennung und Ihrem Passwort.

Kennung:	«TNSID»
Passwort:	«password»

Ihre Zugangsdaten finden Sie auch im beigelegten Kärtchen. Bitte bewahren Sie diese für die Zukunft gut auf.

Vielen Dank für Ihre Teilnahme an unserem Projekt.

Mit freundlichen Grüßen,

Prof. Christina Gathmann
Projektleiterin „Gesellschaft im Wandel“

Bitte beachten Sie auch die Rückseite

Figure 9.0.3: GIP 2012, First reminder letter without incentive

Universität Mannheim | Schloss | 68131 Mannheim

Frau «VName» «NName»

«HHAdr_1» «HHAdr_2»
«HHAdr_3» «HHAdr_4»

Mannheim, 2. November 2017

Gesellschaft im Wandel

Sehr geehrte Frau «NName»,

vor ein paar Wochen haben wir Sie per Brief zur Fortsetzung der Studie „Gesellschaft im Wandel“ eingeladen. Da Briefe manchmal verloren gehen, möchten wir Sie heute gerne noch einmal auf unsere Studie aufmerksam machen.

Für Ihre Teilnahme melden Sie sich möglichst bis zum (Datum + 2 Wochen) unter

www.gesellschaft-im-wandel.de

an, damit Sie dann schon an unserer nächsten Umfrage teilnehmen können. Als kleines Dankeschön haben wir diesem Brief **5 Euro** beigelegt.

Die Anmeldung erfolgt mit Ihrer persönlichen Kennung und Ihrem Passwort.

Kennung:	«TNSID»
Passwort:	«password»

Ihre Zugangsdaten finden Sie auch im beigelegten Kärtchen. Bitte bewahren Sie diese für die Zukunft gut auf.

Vielen Dank für Ihre Teilnahme an unserem Projekt.

Mit freundlichen Grüßen,

Prof. Christina Gathmann
Projektleiterin „Gesellschaft im Wandel“

Bitte beachten Sie auch die Rückseite.

Figure 9.0.4: GIP 2012, First reminder letter with €5cash incentive

Gesellschaft im Wandel

Worum geht es?

In „Gesellschaft im Wandel“ geht es um Ihre Meinungen, Einstellungen und Erwartungen zu verschiedenen Themen wie Familie und Freunde, Arbeit und Freizeit, Wirtschaft und Politik. Die Befragungen werden alle 2 Monate über das Internet durch das LINK Institut für Markt- und Sozialforschung durchgeführt. Diese Befragungen sind Teil einer langfristigen wissenschaftlichen Studie der Universität Mannheim, die durch die Deutsche Forschungsgemeinschaft (DFG) gefördert wird. Die Studie dient keinerlei kommerziellen Zwecken.

Was haben Sie davon?

Mit Ihrer Teilnahme leisten Sie einen wertvollen Beitrag zur Wissenschaft und Sie helfen uns, das Zusammenleben in unserer Gesellschaft besser zu verstehen. Jeder Teilnehmer und jede Teilnehmerin ist für uns unersetzlich, und nur die von uns persönlich angeschriebenen Personen können an der Studie teilnehmen. Als Dankeschön bekommen Sie von uns für jede abgeschlossene Befragung 4 Euro und einen Bonus von bis zu 10 Euro, wenn Sie regelmäßig an den Befragungen teilgenommen haben.

Freiwilligkeit und Anonymität

Die Teilnahme an jeder einzelnen Online Umfrage, zu der wir Sie nach Ihrer Registrierung per Email einladen, ist freiwillig. Sie gehen mit der Registrierung keine dauerhafte Verpflichtung ein. Und natürlich erfolgen alle Auswertungen anonym gemäß den Vorgaben der Datenschutzgesetze.

Haben Sie Fragen oder Anregungen?

Internet: www.gesellschaft-im-wandel.de
E-Mail: info@gesellschaft-im-wandel.de
Telefon: 0800/5892604 (kostenlos aus dem dt. Fest- und Mobilnetz)

Figure 9.0.5: Reminder letter page 2, for both letters

Table 9.0.1: Fixed-effects logistic regression on GIP recruitment survey participation

Commercial area information		
Purchasing power: medium (ref.)		
Low	0.675	(−1.80)
High	0.638*	(−2.41)
Low x uncond inventive	1.115	(0.51)
High x uncond incentive	1.289	(1.21)
High probability of:		
University graduates	1.153	(0.86)
University graduates x uncond incentive	0.731	(−1.42)
Singles x uncond incentive	0.728	(−1.00)
Foreigners x uncond incentive	0.952	(−0.22)
Single households x uncond incentive	1.970*	(2.13)
Public area information		
Umployment quota: medium (ref.)		
Low x uncond incentive	0.742	(−1.44)
High x uncond incentive	1.049	(0.23)
Interviewer observations		
Intercom installed?	0.673**	(−3.14)
Intercom installed? x uncond incentive	1.244	(1.15)
Building type: farm (ref.)		
Familiy home	0.502	(−1.60)
Semi detatched house	0.554	(−1.34)
3-4 apts	0.564	(−1.28)
5-8 apts	0.547	(−1.35)
6:9+ apts	0.620	(−1.03)
High rise	1.429	(0.57)
Familiy home x uncond incentive	2.961	(1.65)
Semi detatched house x uncond incentive	2.809	(1.54)
3-4 apts x uncond incentive	2.784	(1.50)
5-8 apts x uncond incentive	2.822	(1.54)
6-9+ apts x uncond incentive	1.886	(0.92)
High rise x uncond incentive	0.987	(−0.02)
Building maintenance: good (ref.)		
Medium	0.647*	(−2.45)
Poor	0.548	(−1.76)
Medium x uncond incentive	0.954	(−0.21)
Poor x uncond incentive	0.658	(−1.00)
Perceived class: poor (ref.)		
Working	0.777	(−0.76)
Middle	1.038	(0.11)
Upper middle	1.461	(0.98)
Upper	1.623	(0.78)
Working x uncond incentive	0.944	(−0.12)
Middle x uncond incentive	1.093	(0.19)
Upper middle x uncond incentive	1.517	(0.73)
Upper x uncond incentive	0.657	(−0.43)
N	3,132	

Exponentiated coefficients; *t* statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

References

AAPOR. (2016). *Standard Definitions. Final Dispositions of Case Codes and Outcome Rates for Surveys* (Tech. Rep.). The American Association for Public Opinion Research.

Abreu, D. A., & Winters, F. (1999). Using Monetary Incentives to Reduce Attrition in the Survey of Income and Program Participation. In *Proceedings of the Survey Research Methods Section AMSTAT* (pp. 533–538).

Albaum, G. S., Evangelista, F., & Medina, N. (1998). Role of Response Behavior Theory in Survey Research: A Cross-National Study. *Journal of Business Research*, *42*, 115–125.

Allison, P. D. (2009). *Fixed effects regression models*. Thousand Oaks: Sage.

Al-Wazir, & Staatskanzlei Hessen. (1997). *Kleine Anfrage des Abg. Al-Wazir vom 26.3.97 und Antwort des Ministerpräsidenten -Staatskanzlei- Drucksache 14/2776.*

Armstrong, J. S. (1975). Monetary Incentives in Mail Surveys. *Public Opinion Quarterly*, *39*(1), 111–116.

Bell, A., & Jones, K. (2015). Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data. *Political Science Research and Methods*, *3*(1), 133–153.

Bethlehem, J. G. (2002). Weighting Nonresponse Adjustments Based on Auxiliary Information. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey Nonresponse* (pp. 275 – 288). Hoboken: Wiley.

Bethlehem, J. G., Cobben, F., & Schouten, B. (2011). *Handbook of Nonresponse in Household Surveys*. New York: Wiley.

Birnholtz, J. P., Horn, D. B., Finholt, T. A., & Bae, S. J. (2004). The Effects of Cash, Electronic, and Paper Gift Certificates as Respondent Incentives for a

Web-Based Survey of Technologically Sophisticated Respondents. *Social Science Computer Review*, *22*(3), 355–362.

Blohm, M., & Koch, A. (2013). Respondent Incentives in a National Face-to-Face Survey. *Methoden, Daten, Analysen (MDA)*, *7*(1), 89–122.

Blom, A. G., Gathmann, C., & Krieger, U. (2015). Setting Up an Online Panel Representative of the General Population: The German Internet Panel. *Field Methods*, *27*(4), 391–408.

Blom, A. G., Herzing, J. M. E., Cornesse, C., Sakshaug, J. W., Krieger, U., & Bossert, D. (2017). Does the Recruitment of Offline Households Increase the Sample Representativeness of Probability-Based Online Panels? Evidence From the German Internet Panel. *Social Science Computer Review*, *35*(4), 498–520.

Blom, A. G., Jäckle, A., & Lynn, P. (2010). The Use of Contact Data in Understanding Cross-National Differences in Unit Nonresponse. In J. A. Harkness et al. (Eds.), *Survey methods in multinational, multiregional, and multicultural contexts* (pp. 333–354). New York: Wiley.

Booker, C. L., Harding, S., & Benzeval, M. (2011). A systematic review of the effect of retention methods in population-based cohort studies. *BMC public health*, *11*(1), 249.

Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., … Zuber, S. (2013). Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). *International Journal of Epidemiology*, *42*(4), 992–1001.

Börsch-Supan, A., & Jürges, H. (2005). *The Survey of Health, Ageing and Retirement in Europe – Methodology*. Mannheim: Mannheim Research Institute for the Economics of Aging (MEA).

Börsch-Supan, A., Krieger, U., & Schröder, M. (2013). *Respondent incentives, interviewer training and survey participation.* München.

Boulianne, S. (2008). Incentives. In P. J. Lavrakas (Ed.), *Encyclopedia of Survey Research Methods* (pp. 328–331). Thousand Oaks: Sage.

Brüderl, J., & Ludwig, V. (2015). Fixed-Effects Panel Regression. In H. Best & C. Wolf (Eds.), *The SAGE Handbook of Regression Analysis and Causal Inference* (pp. 327–357). Thousand Oaks: Sage.

Bundesinstitut für Bau- Stadt- und Rauminformation. (2015). *INKAR: Indikatoren und Karten zur Raum- und Stadtentwicklung* (Tech. Rep.). Bonn.

Castiglioni, L., Pforr, K., & Krieger, U. (2008). The effect of incentives on response rates and panel attrition: Results of a controlled experiment. *Survey Research Methods*, *2*(3), 151–158.

Cernat, A., & Lynn, P. (2014). *The role of email addresses and email contact in encouraging web response in a mixed mode design.*

Church, A. H. (1993). Estimating the Effect of Incentives on Mail Survey Response Rates: A Meta-Analysis. *Public Opinion Quarterly*, *57*, 62–79.

Couper, M. (2013). Is the Sky Falling? New Technology, Changing Media, andthe Future of Surveys. *Survey Research Methods*, *7*(3), 145–156.

Cyffka, K., Dykema, J., Jaques, K., Ganci, R., Elver, K., & Stevenson, J. (2012). SHOW Me the Money?: Effects of Preincentives, Differential Incentives, and Envelope Messaging in an ABS Mail Survey. In *Aapor proceedings* (pp. 5407 – 5421).

De Luca, G., & Lipps, O. (2005). Fieldwork and survey management in SHARE. *The Survey of Health, Aging, and Retirement in Europe–Methodology*, 75–81.

Diekmann, A. (2006). *Methoden der Sozialforschung*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Diekmann, A., & Jann, B. (2001). Anreizformen und Ausschöpfungsquoten bei postalischen Befragungen : eine Prüfung der Reziprozitätshypothese. *ZUMA-Nachrichten*, *48*(25), 18–27.

Dillman, D. A. (1978). *Mail and Telephone Surveys: The Total Design Method*. Hoboken: Wiley.

Dillman, D. A. (2011). *Mail and Internet surveys: The Tailored Design Method – 2007 Update with New Internet, Visual, and Mixed-Mode Guide* (Second Edi ed.). Hoboken: Wiley.

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 and Community Health*, *59*(11), 987–999.

Frey, B. S., & Jegen, R. (2001). Motivation Crowding Theory. *Journal of Economic Surveys*, *15*(5), 589–611.

Gajic, A., Cameron, D., & Hurley, J. (2010). *The Impact of Alternative Incentives on Response and Retention in a Mixed-Mode Survey.*

Gelman, A., & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.

Gelman, A., Stevens, M., & Chan, V. (2003). Regression Modeling and Meta-Analysis for Decision Making. *Journal of Business & Economic Statistics*, *21*(2), 213–225.

Gendall, P., & Healey, B. (2008). Alternatives to Prepaid Monetary Incentives in Mail Surveys. *International Journal of Public Opinion Research*, *20*(4), 517–527.

Goldenberg, K. L., McGrath, D., & Tan, L. (2009). The effects of incentives on the consumer expenditure interview survey. In *Procedings Annual Conference, American Association for Public Opinion Research.*

Gouldner, A. W. (1960). The Norm of Reciprocity: A Preliminary Statement. *American Sociological Review*, *25*, 161.

Goyder, J. (2008). Nonresponse. In P. J. Lavrakas (Ed.), *Encyclopedia of Survey Research Methods.* Thousand Oaks: Sage.

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, *70*(5), 646–675.

Groves, R. M., & Couper, M. P. (2012). *Nonresponse in Household Interview Surveys*. Hoboken: Wiley.

Groves, R. M., Fowler, F. J. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2013). *Survey Methodology*. Hoboken: Wiley.

Groves, R. M., & Peytcheva, E. (2008). The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis. *Public Opinion Quarterly*, *72*(2), 167–189.

Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-Saliency Theory of Survey Participation: Description and an Illustration. *The Public Opinion Quarterly*, *64*(3), 299–308.

Hanly, M. J., Savva, G. M., Clifford, I., & Whelan, B. J. (2014). Variation in Incentive Effects across Neighbourhoods An Example from The Irish Longitudinal Study of Ageing. *Survey Research Methods*, *8*(1), 19–30.

Heckel, C., & Hofmann, O. (2014). Das ADM-Stichproben-System (F2F) ab 1997. In ADM Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V. (Ed.), *Stichproben-Verfahren in der Umfrageforschung. Eine Darstellung für die Praxis* (2nd ed., pp. 85–116). Wiesbaden: Springer VS.

Hogan, S. O. (2009). The Costs of Using Pre-Paid Incentives in a Physician Survey. *Survey Practice*, *2*(2), 1–5.

Homans, G. C. (1958). Social Behavior as Exchange. *American Journal of Sociology*, *63*(6), 597–606.

Hsu, J. W., Schmeiser, M. D., Haggerty, C., & Nelson, S. (2017). The Effect of Large Monetary Incentives on Survey Completion. *Public Opinion Quarterly*, *81*(3), 736–747.

Jäckle, A., & Lynn, P. (2008). Respondent incentives in a multi-mode panel survey: Cumulative effects on nonresponse and bias. *Survey Methodology*, *34*(1), 105–117.

Japec, L., Kreuter, F., Berg, M., Biemer, P., Decker, P., Lampe, C., … Usher, A. (2015). Big Data in Survey Research: AAPOR Task Force Report. *Public Opinion Quarterly*, *79*(4), 839–880.

Juster, F. T., & Suzman, R. (1995). An Overview of the Health and Retirement Study. *The Journal of Human Resources*, *30*, 7–56.

Kay, W. R., Boggess, S., Selvavel, K., & Mcmahon, M. F. (2001). The Use of Targeted Incentives to Reluctant Respondents on Response Rate and Data Quality 1. In *Proceedings of the Annual Meeting of the American Statistical Association.*

Keeter, S., Hatley, N., Kennedy, C., & Lau, A. (2017). *What Low Response Rates Mean for Telephone Surveys* (Tech. Rep.). Pew Reserch Center.

Kennedy, R., Wojcik, S., & Lazer, D. (2017). Improving election prediction internationally. *Science*, *355*(3), 515–520.

Klausch, T., Schouten, B., & Hox, J. J. (2015). Evaluating Bias of Sequential Mixed-mode Designs Against Benchmark Surveys. *Sociological Methods & Research*, 1–34.

Kretschmer, S., & Müller, G. (2017). The wave 6 NEPS adult study incentive experiment. *Methods, Data, Analyses*, *11*(1), 7–28.

Laurie, H., & Lynn, P. (2009). The Use of Respondent Incentives on Longitudinal Surveys. In P. Lynn (Ed.), *Methodology of Longitudinal Surveys* (pp. 205–234). Hoboken: Wiley.

Lin, I. F., & Schaeffer, N. C. (1995). Using Survey Participants to Estimate the Impact of Nonparticipation. *Public Opinion Quarterly*, *59*(2), 236.

Lynn, P., De Luca, G., Ganninger, M., & Häder, S. (2013). Sample Design in SHARE Wave Four. In F. Malter & A. Börsch-Supan (Eds.), *SHARE Wave 4: Innovations & Methodology* (pp. 74–123). Munich: MEA, Max Planck Institute for Social Law and Social Policy.

Mankiw, N. G. (2004). *Principles of economics*. Boston: Cengage Learning.

Martin, S., Helmschrott, S., & Rammstedt, B. (2014). The Use of Respondent Incentives in PIAAC: The Field Test Experiment in Germany. *Methods, Data, Analyses*, *8*(2), 223–242.

Mavletova, A., & Couper, M. P. (2016). Device use in web surveys: The effect of differential incentives. *International Journal of Market Research*, *58*(4), 523–544.

McGonagle, K. A., & Freedman, V. A. (2017). 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*(3), 221–237.

MEA - Max Planck Institute for Social Law and Social Policy. (2017). *SHARE Release Guide 6.0.0* (Tech. Rep.). Munich: MEA, Max Planck Institute for Social Law and Social Policy.

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*(4), 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*(1), 105–129.

Meuleman, B., Langer, A., & Blom, A. G. (2017). Can Incentive Effects be Generalized to Non-Western Countries ? Conditional and Unconditional Cash Incentives in a Web Survey of Ghanaian University Students. *Social Science Computer Review*, 1–20.

Meyer, B. D., Mok, W. K. C., & Sullivan, J. X. (2015). Household Surveys in Crisis. *Journal of Economic Perspectives*, *29*(4), 199–226.

Microm. (2011). *Microm Datenhandbuch. Arbeitsunterlagen für Microm MARKET & GEO* (Tech. Rep.). Microm Micromarketing-Systeme und Consult GmbH.

Millar, M. M., & Dillman, D. A. (2011). Improving Response to Web and Mixed-Mode Surveys. *Public Opinion Quarterly*, *75*(2), 249–269.

Olsen, R. J. (2005). The problem of respondent attrition: survey methodology is key. *Monthly Labor Review, 128*, 63 – 70.

Pforr, K., Blohm, M., Blom, A. G., Erdel, B., Felderer, B., Fräßdorf, 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*(3), 740–768.

Rat für Sozial- und WirtschaftsDaten. (2017). *Die sozial-, verhaltens- und wirtschaftswissenschaftliche Survey-Landschaft in Deutschland. Empfehlungen des RatSWD* (Tech. Rep.). Berlin: Rat für Sozial- und WirtschaftsDaten.

Rodgers, W. L. (2011). Effects of increasing the incentive size in a longitudinal study. *Journal of Official Statistics*, *27*(2), 279.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys.* Hoboken: Wiley.

Ryu, E., Couper, M. P., & Marans, R. W. (2005). Survey Incentives: Cash vs. In-Kind; Face-to-Face vs. Mail; Response Rate vs. Nonresponse Error. *International Journal of Public Opinion Research*, *18*(1), 89–106.

Scherpenzeel, A. (2014). Survey Participation in a Probability-Based Internet Panel in the Netherlands. In U. Engel, B. Jann, P. Lynn, A. Scherpenzeel, & P. Sturgis (Eds.), *Improving Survey Methods. Lessons from Recent Research.* (pp. 223–235). New York: Routledge.

Scherpenzeel, A., & Toepoel, V. (2012). Recruiting a Probability Sample for an Online Panel: Effects of Contact Mode, Incentives, and Information. *Public Opinion Quarterly*, *76*(3), 470–490.

Schnell, R. (1997). *Nonresponse in Bevölkerungsumfragen: Ausmaß, Entwicklung und Ursachen*. Opladen: Leske & Budrich.

Schnell, R., Hill, P. B., & Esser, E. (2013). *Methoden der empirischen Sozialforschung* (10., übera ed.). München: Oldenbourg.

Schoeni, R. F., Stafford, F., McGonagle, K. A., & Andreski, P. (2013). Response Rates in National Panel Surveys. *The ANNALS of the American Academy of Political and Social Science*, *645*(1), 60–87.

Schröder, M., Saßenroth, D., Körtner, J., Kroh, M., & Schupp, J. (2013). *Experimental Evidence of the Effect of Monetary Incentives on Cross-Sectional and Longitudinal Response: Experiences from the Socio-Economic Panel (SOEP).*

SHARE - Survey of Health, Ageing and Retirement in Europe. (2017). SHARE - Survey of Health, Ageing and Retirement in Europe. , Retrieved 2017-08-01, from http://www.share-project.org/.

Shuttleworth, F. K. (1931). A study of questionnaire technique. *Journal of Educational Psychology*, *22*(9), 652.

Singer, E. (2006). Nonresponse Bias in Household Surveys. *Public Opinion Quarterly*, *70*(5), 637–645.

Singer, E., & Bossarte, R. M. (2006). Incentives for Survey Participation. When Are They "Coercive"? *American Journal of Preventive Medicine*, *31*(5), 411–418.

Singer, E., Groves, R. M., & Corning, A. D. (1999). Differential Incentives: Beliefs About Practices, Perceptions of Equity, and Effects on Survey Participation. *Public Opinion Quarterly*, *63*(2), 251–260.

Singer, E., Hoewyk, J. V., Gebler, N., Raghunathan, T., & McGonagle, K. (1999). The Effect of Incentives on Response Rates in Interviewer-Mediated Surveys. *Journal of Official Statistics*, *15*(2), 217–230.

Singer, E., Hoewyk, J. V., & Maher, M. P. (1998). Does the Payment of Incentives Create Expectation Effects? *Public Opinion Quarterly*, *62*(2), 152–164.

Singer, E., Van Hoewyk, J., & Maher, M. P. (2000). Experiments with Incentives in Telephone Surveys. *Public Opinion Quarterly*, *64*(2), 171–188.

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*(1), 112–141.

Sinibaldi, J., Trappmann, M., & Kreuter, F. (2014). Which Is the Better Investment for Nonresponse Adjustment: Purchasing Commercial Auxiliary Data or Collecting Interviewer Observations? *Public Opinion Quarterly*, *78*(2), 440–473.

Skibba, R. (2016). The polling crisis: How to tell what people really think. *Nature*, *538*(7625), 304–306.

Sowell, T. (2007). *Basic economics: a common sense guide to the economy*. Cambridge: Basic.

Stoop, I. A. L. (2005). *The Hunt for the Last Respondent: Nonresponse in Sample Surveys*. The Hague: Sociaal en Cultureel Planbu.

Stoop, I. A. L. (2012). Unit Non - Response Due to Refusal. In L. Gideom (Ed.), *Handbook of Survey Methodology for the Social Sciences.* New York: Springer.

Student. (1908). The Probable Error of a Mean. *Biometrika*, *6*(1), 1–25.

Tourangeau, R. (2017). Mixing Modes. Tradeoffs Among Coverage, Nonresponse, and Measurement Error. In P. P. Biemer et al. (Eds.), *Total Survey Error in Practice* (pp. 115–132). Hoboken: Wiley.

Trappmann, M., Beste, J., Bethmann, A., & Müller, G. (2013). The PASS panel survey after six waves. *Journal for Labour Market Research*, *46*(4), 275–281.

Trussell, N., & Lavrakas, P. J. (2004). The Influence of Incremental Increases in Token Cash Incentives on Mail Survey Response. Is there an Optimal Amount? *Public Opinion Quarterly*, *68*(3), 349–367.

Warriner, K., Goyder, J., Gjertsen, H., Hohner, P., & Mcspurren, K. (1996). Charities, no; Lotteries, no; cash, yes: Main effects and interactions in a Canadian incentives experiment. *Public Opinion Quarterly*, *60*(4), 542–562.

Wijnant, A., Martens, M., Balster, E., & Das, M. (2013). Software Innovation in SHARE Wave Four. In F. Malter & A. H. Börsch-Supan (Eds.), *Share wave 4: Innovations & methodology* (pp. 62–73). Munich: MEA, Max Planck Institute for Social Law and Social Policy.

Yeager, D. S., Krosnick, J. A., Chang, L., Javitz, H. S., Levendusky, M. S., Simpser, A., & Wang, R. (2011). Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples. *Public Opinion Quarterly*, *75*(4), 709–747.

Yu, J., & Cooper, H. (1983). A Quantitative Review of Research Design Effects on Response Rates to Questionnaires. *Journal of Marketing Research*, *20*(1), 36–44.

Zagorsky, J. L., & Rhoton, P. (2008). The Effects of Promised Monetary Incentives on Attrition in a Long-Term Panel Survey. *Public Opinion Quarterly*, *72*(3), 502–513.