

Four Essays on Altruism and Compliance

INAUGURALDISSERTATION

zur Erlangung des akademischen Grades
eines Doktors der Wirtschafts- und Sozialwissenschaft
(Dr. rer. pol.)

der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Universität Potsdam

vorgelegt von

Florian Stolley, M.Sc.
geboren in Hamburg

September 2018

Erstgutachterin: Prof. Dr. Lisa V. Bruttel
Zweitgutachter: Prof. Dr. Marco Caliendo
Drittgutachter: Prof. Dr. Rainald Borck

Tag der Disputation: 13.03.2019

Published online at the
Institutional Repository of the University of Potsdam:
<https://doi.org/10.25932/publishup-42801>
<https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-428013>

Danksagung

Zuallererst möchte ich meiner Erstbetreuerin Lisa Bruttel danken, ohne die diese Doktorarbeit nicht möglich gewesen wäre. Neben der sehr guten Unterstützung, beispielsweise durch Hilfe und Rat zu verschiedenen wissenschaftlichen Themen und sehr schnelles Feedback zu Forschungsideen, möchte ich mich explizit für die interessante und spannende Zusammenarbeit bei den gemeinsamen Forschungsprojekten bedanken. Des Weiteren danke ich Marco Caliendo, der sofort bereit war, die Zweitbetreuung dieser Doktorarbeit zu übernehmen und durch wertvolle Kommentare und Vorschläge während meiner Seminarvorträge dazu beigetragen hat, die Qualität einzelner Kapitel zu verbessern.

Ein besonderes Dankeschön geht auch an meine Kollegin Sandra Bedenk und meinen Kollegen Andreas Orland für die gute Zusammenarbeit in den vergangenen vier Jahren und die ständige Bereitschaft, über die Herausforderungen in der Forschung zu sprechen und mit Rat und Tat bei der Bewältigung dieser zu helfen. Für die Unterstützung bei der Durchführung der Experimente, das Korrekturlesen und das Übersetzen von Anleitungen möchte ich mich bei den vielen studentischen Hilfskräften bedanken: Frederike Lannte, Bitja Terschüren, Sarah Skladny, Mia Teschner, Christoph Spyra, Janosch Hübner, Jas Holle, Luis Koch und Juri Nithammer.

Für die gute und erfolgreiche Zusammenarbeit danke ich meinen weiteren Co-Autoren Verena Utikal und Juri Nithammer. Außerdem danke ich allen, die mir bei Seminaren in Potsdam, bei Konferenzen in Chicago und Berlin oder in persönlichen Gesprächen Feedback zu meiner Forschung gegeben haben.

Ganz besonders möchte ich mich bei meiner Familie bedanken. Vielen Dank Euch für die immerwährende Unterstützung, Euer Interesse und natürlich auch für manches kurzfristige Korrekturlesen. Zum Schluss danke ich Dir, Maria, für Deine Unterstützung und Dein Verständnis in den vergangenen vier Jahren.

Contents

General Overview	1
1 Getting a Yes. An Experiment on the Power of Asking	5
1.1 Introduction	5
1.2 Related Literature	7
1.3 Experimental Design and Procedures	11
1.4 Results	14
1.4.1 Descriptive Statistics about the Giving Decision	14
1.4.2 Message Analysis	15
1.4.3 Regression Results	18
1.4.4 Different Behavior of Male and Female Dictators	21
1.5 Discussion	22
1.6 Conclusion	26
1.A Appendix	28
1.A.1 Supplementary Tables and Figures	28
1.A.2 Instructions and Control Questions	32
1.A.3 Screenshots	38
2 Gender Differences in the Response to Decision Power and Responsibility - Framing Effects in a Dictator Game	39
2.1 Introduction	39
2.2 Related Literature	41

2.3	Experimental Design and Procedures	43
2.4	Results	46
2.5	Discussion and Conclusion	49
2.A	Appendix	52
2.A.1	Supplementary Tables and Figures	52
2.A.2	Instructions	53
2.A.3	Pictures of the Material	57
3	“Thanks in Advance” - the Effect of a Polite but Empty Phrase on Compliance Behavior	59
3.1	Introduction	59
3.2	Related Literature	61
3.3	Experimental Design and Procedures	62
3.4	Results	64
3.5	Discussion and Conclusion	67
3.A	Appendix	69
3.A.1	Instructions	69
3.A.2	Screenshots	71
4	The Effects of (Imprecise) Defaults on Dishonesty - Tax Evasion in a Laboratory Experiment	73
4.1	Introduction	73
4.2	Related Literature	75
4.3	Experimental Design and Procedures	79
4.4	Behavioral Predictions	83
4.5	Results	86
4.5.1	Real-Effort Task	86
4.5.2	Relative Size of the Default	87
4.5.3	Tax Compliance	88

4.6	Discussion and Conclusion	93
4.A	Appendix	96
4.A.1	Supplementary Tables and Figures	96
4.A.2	Instructions and Control Questions	99
4.A.3	Screenshots	105
	Bibliography	106
	Abstract (English Version)	122
	Abstract (German Version)	124
	Declarations of Co-Authorship	127

List of Figures

1.1	Average marginal effects for the full sample	19
1.2	Average marginal effects for male and female dictators	21
1.A.1	A's decision screen in the first round (English translation)	38
1.A.2	A's decision screen from round 2 on (English translation)	38
2.1	Decision sheets (in the original, German language)	45
2.2	Mean giving separated by treatment and gender	47
2.A.1	Large envelope with a sticker with seat and id on it	57
2.A.2	Content of the envelope - instructions and two smaller envelopes	57
2.A.3	Content of the small envelope marked with a 1 - decision sheet	58
3.1	Mean time separated by treatment and gender	65
3.2	Mean length separated by treatment and gender	66
3.A.1	Screenshot of the 5x5-version of the bomb risk elicitation task	71
3.A.2	Screenshot of the request in the <i>ThanksInAdvance</i> treatment	72
3.A.3	Screenshot of the request in the <i>Control</i> treatment	72
4.1	Effect of the level of difficulty on the relative size of the default	88
4.2	Tax honesty separated by treatment and difficulty level	90
4.3	Tax honesty separated by treatment and difficulty level in round 1	92
4.A.1	Real-effort task (medium difficulty level; English translation)	105
4.A.2	Filing a tax return (English translation)	105

List of Tables

I	Overview of the four chapters	2
1.1	Matrix of endowments and payoffs	11
1.2	Objective characteristics of messages	15
1.3	Categories of messages resulting from the rating session	16
1.4	Logit regression results for four models	17
1.A.1	Message categories	28
1.A.2	Cross-correlation table of message characteristics	29
1.A.3	Logit regression results for subsamples	30
1.A.4	logit regression results for a subject fixed effects model	31
2.1	OLS regression results	48
2.A.1	Descriptive statistics of the distribution of participants	52
2.A.2	OLS regression results with gender interaction effects	52
4.1	Summary statistics for subjects in the experiment	82
4.2	Summary statistics for subjects' performance in the real-effort task	86
4.3	OLS regression results by difficulty level for two different models	91
4.A.1	Effect of learning on subjects' performance in RET	96
4.A.2	OLS regression results for the full sample	97
4.A.3	OLS regression results - differences between difficulty levels	98

General Overview

There are numerous situations in which people ask for something or make a request, e.g. asking a favor, asking for help or requesting compliance with specific norms. For this reason, how to ask for something in order to increase people's willingness to fulfill such requests is one of the most important question for many people working in various different fields of responsibility such as charitable giving, marketing, management or policy making.

This dissertation consists of four chapters that deal with the effects of small changes in the decision-making environment on altruistic decision-making and compliance behavior. Most notably, written communication as an influencing factor is the focus of the first three chapters. The starting point was the question how to devise a request in order to maximize its chance of success (Chapter 1). Here, the effect of communication on altruistic behavior is at the center of interest. The results of the first chapter originate the idea for the second chapter. How does communication by a neutral third-party, i.e. a text from the experimenters that either reminds potential benefactors of their responsibility or highlights their freedom of choice, affect altruistic decision-making. The third chapter builds on results of the first chapter as well. This chapter elaborates on the effect of thanking people in advance when asking them for help. More specifically, it is measured how much effort people spend on complying with a request when the request is accompanied by the phrase "thanks in advance." Thus, the first three chapters all deal with the effect of some kind of communication, either between the party that requests help and the one being asked for it (Chapter 1 and Chapter 3) or between a neutral third-party and the one being asked (Chapter 2).

The fourth chapter is not as closely related to the other chapters as the three first ones are. Nevertheless, this chapter deals as well with the question how compliance (here: compliance with norms and rules) is affected by subtle manipulations of the environment in which decisions are made. This chapter analyzes the effect of default settings in a tax return on tax compliance. Therefore, all chapters of

Table I: Overview of the four chapters

Chapter	Title	Co-Author(s)	Published version
1	Getting a Yes. An Experiment on the Power of Asking	Lisa Bruttel Verena Utikal	
2	Gender Differences in the Response to Decision Power and Responsibility - Framing Effects in a Dictator Game	Lisa Bruttel	<i>Games</i> , Vol. 9(2), 2018, 28.
3	“Thanks in Advance” - the Effect of a Polite but Empty Phrase on Compliance Behavior	Lisa Bruttel Juri Nithammer	
4	The Effects of (Imprecise) Defaults on Dishonesty - Tax Evasion in a Laboratory Experiment		

this dissertation - but particularly the fourth one - are related to the concept of nudging (Thaler and Sunstein, 2008).

Table I provides an overview of the four chapters, i.e. giving information about the title, the name of co-authors and the publication status. In order to study the research questions outlined above, controlled experiments were conducted. In three of the four chapters, these experiments were carried out in the laboratory, whereas Chapter 2 presents results from a pen and paper experiment, which was conducted in a classroom setting. In the following, the four chapters are introduced in more detail.

Chapter 1, which analyzes the effect of text messages on the decision to give something to another person while focusing on the content of messages, employs a mini-dictator game. Participants in the experiment randomly take on the role of either the dictator or the recipient. The recipient sends a free-form text message to the dictator before the latter makes a binary decision whether or not to give part of her or his endowment to the recipient. Giving additionally increases the sum of payoffs to both players. After the experiment, the text messages were categorized by independent raters.

We find that putting effort into the message by writing a long note without spelling mistakes increases dictators’ willingness to give. Moreover, writing in a humorous way and mentioning reasons why the money is needed pays off. Furthermore, men and women seem to react differently to some message categories. Only men react positively to efficiency arguments, while only women react to

messages that emphasize the dictator’s power and responsibility.

Building on this last result, Chapter 2 attempts to disentangle the effect of reminding potential benefactors of their responsibility for the potential beneficiary and the effect of highlighting their decision power and freedom of choice on altruistic decision-making. This is done by studying the effects of two different texts on giving in a dictator game. The texts are printed on top of the dictator’s decision sheet. One text stresses the dictator’s responsibility for the recipient’s payoff while the other text emphasizes the dictator’s decision power and freedom of choice. Three treatments are conducted in a between-subjects design, which includes a control treatment without a text. The fact that only two specific texts are used (and not free-form text messages) and that the texts are written by a neutral third-party, i.e. the experimenters, are the most prominent distinctions compared to the research design in Chapter 1. Due to the results of Chapter 1, we are especially interested in whether these texts trigger different reactions by female and male dictators.

Our results show a different reaction to these texts depending on the dictator’s gender. We find that only men react positively to a text that stresses their responsibility for the recipient by giving more to her or him, whereas only women seem to react positively to a text that emphasizes their decision power and freedom of choice.

While the first two chapters focus on altruistic decision-making, Chapter 3, which deals with the effect of the phrase “thanks in advance,” focuses on the compliance with a request. In the experiment, participants are asked to provide a detailed answer to an open question. The extent to which participants comply with the request is measured by the effort participants spend on answering the question. Complying with this request is associated with opportunity costs in terms of time. The treatment variable is whether or not they see the text “thanks in advance.” This phrase is often used in similar situations, i.e. in written requests for favors that can be fulfilled in a few minutes. Apart from not using a dictator game setting, the most noticeable difference to the design in the first two chapters is that participants do not help other participants, i.e. fellow students, but that they help the experimenters when complying with the request.

We find that participants react negatively by putting less effort into complying with the request in response to the phrase “thanks in advance.” This effect is predominantly driven by women. Male participants show no significant reaction to the treatment.

Chapter 4 studies the effect of prefilled tax returns with mostly inaccurate de-

fault values on tax compliance. In recent years, more and more tax returns are at least partially prefilled. While prefiling tax returns with correct values seems to have only advantages (Reeson and Dunstall, 2009; Goolsbee, 2006; Fonseca and Grimshaw, 2017; OECD, 2017; Bankman, 2008; Fochmann et al., 2018; Fonseca and Grimshaw, 2017), it is unclear whether prefiling a tax return with inaccurate values¹ has an effect on tax evasion. In the laboratory experiment conducted to study this research question, participants earn income by performing a real-effort task and must subsequently file a tax return for three consecutive rounds. In the main treatment, the tax return is prefilled with a default value, resulting from participants' own performance in previous rounds, which varies in its relative size. The results suggest that there is no lasting effect of a default value on tax honesty, neither for relatively low nor relatively high defaults. However, participants who face a default that is lower than their true income in the first round evade significantly and substantially more taxes in this round than participants in the control treatment without a default.

This dissertation adds to the literature in several ways. Primarily, the first three chapters contribute evidence on how the language of a request affects its success (Althoff et al., 2014; Mitra and Gilbert, 2014; Das et al., 2008) and, particularly in the first two chapters, how communication affects altruistic decision-making (Andreoni and Rao, 2011; Mohlin and Johannesson, 2008). Additionally, by employing dictator games, these two chapters show how context-dependent dictators' decisions are, i.e. how sensitive the results are to the specific framing of the dictator game (Brañas-Garza, 2007; Fehr and Schmidt, 2006; Gerlach and Jaeger, 2016). Moreover, the first three chapters contribute to the literature on gender differences in the field of social preferences in general (Croson and Gneezy, 2009) and in the area of altruism in particular (Andreoni and Vesterlund, 2001; DellaVigna et al., 2013; de Wit and Bekkers, 2016). Especially the last chapter, which primarily provides evidence on the topic of prefiling tax returns (Fochmann et al., 2018; Fonseca and Grimshaw, 2017; Bruner et al., 2015; Kotakorpi and Laamanen, 2016), adds to the literature on nudging as well (Thaler and Sunstein, 2008).

¹Most electronic tax return programs, for example, have the option to prefill a tax return with last year's values, which are often incorrect for the present year.

Chapter 1

Getting a Yes. An Experiment on the Power of Asking

1.1 Introduction

There are numerous situations in which one person asks another one a favor. A person on the street asking you for money, a roommate asking you to share your food with her, a charity organization asking you for a donation or a stranger asking you to share your place on a bench in a beer garden are just some examples of many. There are several aspects in which these situations differ, such as whether the person asking you is familiar or unfamiliar, whether she is more to your liking or less and whether she is the beneficiary of the favor or just an intermediary (e.g. a charity organization). However, there is also an aspect which all of these situations have in common and which seems to be at the heart of the decision whether to do the favor: communication between the potential beneficiary and the potential benefactor.

In this paper, we are interested in the effect of communication on the decision to give something to someone else. While existing studies have shown that communication as such has a positive effect on the giving decision (Andreoni and Rao, 2011; Mohlin and Johannesson, 2008; Langenbach, 2016), they provide only very limited insights into the effects of the content of written communication. This is the main focus of our study. We analyze which arguments of the recipient convince the person being asked the favor. Additionally, we study other aspects of written communication that systematically increase or decrease the chance of

a successful request. Finally, we are interested in potential gender differences in these effects since men and women have been shown to react differently to the social conditions of an experiment in general (Croson and Gneezy, 2009; Ellingsen et al., 2013; Kahn et al., 1971; Eagly, 1983) as well as to personal solicitation in particular (Meer and Rosen, 2011).

Closely related to our research question is a study by Althoff et al. (2014), in which the authors use data of requests for a free pizza in the online community Reddit. They analyze the content of the request by employing topic modeling and automatic detection to extract important aspects of the requests. Requests are analyzed regarding factors such as politeness, provision of evidence for the narrative, intention to reciprocate in the future, sentiment, length, gratitude, status of the one requesting a favor as well as whether the requests refer to narratives of money, job (loss), being a student, family or craving. Among other things, they find that being perceived as needy through the narrative (little money, job loss, family needs), signaling gratitude and the length of a request affect the success probability positively whereas politeness, for example, has no effect.

By using a laboratory experiment to study our research question, we can analyze the effects of differing content on the decision to do somebody a favor in a more controlled way. For example, participants do not self-select to this specific experiment or to a specific role and we provide participants with anonymity in order to prevent status effects. Furthermore, we collect data on the potential beneficiary as well as the potential benefactor which provides us with the possibility to study gender differences in the response to the content of a request.¹ Finally, we use a different technique to analyze messages: several human subjects categorize the messages independently, which has the advantage that we do not rely on lists of narrow keywords to extract differences in the content.

Apart from providing advice for communication in interpersonal interactions, our research may be useful for charitable organizations trying to convince potential donors to give. Our findings might furthermore contribute to research on negotiations by analyzing which aspects of communication can convince the other party to agree to a specific arrangement.

In order to study our research question, we conduct a laboratory experiment in which subjects participate in a modified dictator game. Dictators make a binary decision to either give part of their endowment to the receiver or to keep their endowment for themselves. Before making their decision, dictators receive a free-form message written by the receiver. The decision to give part of their

¹Althoff et al. (2014) only observe the identity of the benefactor in a few cases.

endowment to the receiver results in a decrease in the payoff to the dictator and an increase in the sum of payoffs to both players. This represents a welfare improvement which is inherent in most real-life situations of doing somebody a favor or giving something to someone since the party benefitting from the favor usually has a higher marginal utility for the requested good.

We find clear evidence that the content of a message as well as its spelling and length influence the dictator's decision to give. Messages showing that the author exerted effort by writing a long note without spelling mistakes, exhibit higher chances of success. The same holds for humorous messages. It seems that if the author of a message is "kind," the dictator will reciprocate this kindness by giving money. Additionally, informing potential givers as to the reasons why the money is specifically needed is an effective way of increasing generosity. Moreover, our results indicate that there are crucial differences in the responses of male and female dictators. Arguing that giving increases the sum of the payoffs improves the chances of a successful request only if the dictator is male. Women, on the other hand, react positively to statements acknowledging their power and responsibility, which may create subtle social pressure.

The rest of the paper is structured as follows. Next, the relevant literature is discussed. Following, we present the design of our experiment and introduce our message categories in Section 1.3. Section 1.4 presents the results of the experiment, which are subsequently discussed in Section 1.5. Section 1.6 concludes.

1.2 Related Literature

Situations in which one party can act altruistically² towards another party and in which there is no strategic interaction between these two parties have been studied extensively using the dictator game in the experimental economics literature (beginning with Kahneman et al., 1986; Forsythe et al., 1994). In the original game, one party (the dictator) can decide how to divide her endowment between herself and another party (the receiver). It has been found that a significant fraction of people behave altruistically to some extent (see Engel, 2011, for a review). Several explanations for this phenomenon have been mentioned in the literature over the previous years, including social preferences such as inequity aversion (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Charness and Rabin, 2002), feelings of empathy towards the other person (Batson et al.,

²In the following, a behavioral definition by which altruistic acts are "costly acts that confer economic benefits on other individuals" (Fehr and Fischbacher, 2003, p. 785) is used.

1.2. RELATED LITERATURE

1988; Batson, 2002; Andreoni and Rao, 2011; de Vignemont and Singer, 2006; Edele et al., 2013), following social norms regarding fairness (Bolton et al., 1998; Camerer and Thaler, 1995; Zhao et al., 2017) which, in turn, is closely related to guilt aversion (Charness and Dufwenberg, 2006; Battigalli and Dufwenberg, 2007)³ and self-image concerns (Dana et al., 2007; Tonin and Vlassopoulos, 2013; Ploner and Regner, 2013).

Even though communication plays an important role in a variety of social interactions, including the decision to do someone a favor, surprisingly few studies have implemented some sort of communication into their experimental design. Allowing communication in an experiment has the disadvantage that it reduces the level of control and that the analysis of the written data is less straightforward than it would be with numerical data. This is particularly true when free-form communication is used. Nevertheless, considering the great importance of communication, it cannot be disregarded when studying social interactions.

Closest to our experiment is the study by Andreoni and Rao (2011), in which either only the receiver, only the dictator, none of them, or both can write a free-form message. In the treatment most relevant to our research question, in which only the receiver can communicate, givings are significantly higher than without communication. The messages were categorized by independent raters into the categories of Fairness, Friendly Greeting, Flattery, Acknowledge Power and Need. While communication by itself (message + numerical request) has an effect on generosity, the content of the communication does not seem to have an effect. In their study, Andreoni and Rao hypothesize that the reason for this may be that the content is too obvious (nearly half of the messages appealed to fairness norms). However, also the number of observations is relatively small, since only 20 receivers were randomly matched to 20 dictators for two rounds. Furthermore, the additional numerical request has a significant effect on the dictator’s decision and, hence, may “crowd out” the content of the messages as is suggested by Andreoni and Rao (2011). In our study, we focus on one-sided, free-form communication from the receiver to the dictator without an additional numerical request and modify the design in such a way that it has more power to find systematic differences in dictator’s behavior resulting from differences in the content of the particular message. Primarily, this is done by letting dictators make multiple binary giving decisions in response to a sequence of messages sent by different receivers.

Mohlin and Johannesson (2008) compare the results of a dictator game in which

³However, Ellingsen et al. (2010) do not find evidence for guilt aversion in a dictator game.

1.2. RELATED LITERATURE

the receiver can either write a message to the dictator or not. They find that communication significantly increases giving. Additionally, they use a treatment in which a third party sends a message to the dictator to distinguish between a “content effect” and a “relationship effect.” Their results indicate that both effects are important and increase givings by a similar magnitude. Moreover, Mohlin and Johannesson try to analyze the content of the messages further by classifying the messages themselves into different categories. However, they do not find consistent differences in giving behavior with respect to their categories, which could be due to too few observations, as is mentioned by Mohlin and Johannesson, since only 55 dictators participated in their relevant treatment making one decision each. In contrast, in our study 126 dictators make multiple giving decisions leading to nearly 2,000 observations and messages are categorized by independent raters.

Similar to one of the treatments by Andreoni and Rao (2011), Langenbach (2016) studies the effect of a combination of a free-form message and a numerical request sent by the receiver in a dictator game. He also finds that giving is significantly higher with communication than without communication. Differences between the numerical request and the messages as well as differences in the content of the messages are not analyzed further in the paper.

Some studies have looked at the effect of communication in a dictator game without allowing free-form communication. Charness and Rabin (2005) let receivers express their preference about how they would like to be treated by either sending the message “Help me,” “Don’t help me” or, in some cases, no message at all to dictators before they decide about the allocation of funds in a series of binary dictator games. They find that more dictators chose the option that benefitted the receiver when the receiver asked for help. Rankin (2006) allows receivers to request a certain amount of money to be given to them by the dictator. When the request is for at most half of the total amount to be divided, the request has a positive effect on the generosity of the dictator compared to a treatment without a request-possibility. A request for more than half, however, has a negative effect on the amount transferred by the dictators to the receivers. This result has also been found by Andreoni and Rao (2011) and Yamamori et al. (2007).

Letting receivers only choose from a set of pre-written messages or letting them only send a numerical request, as it is done in these studies, increases the internal validity of the results because analyzing the data is straightforward. However, at the same time, it decreases the external validity, since such a reduced form of communication does not capture the richness of communication in the real world.

1.2. RELATED LITERATURE

Because we are interested in exactly these differences in the content of messages which may have an effect on the decision of dictators, free-form communication is used in our study.

Andreoni et al. (2017) designed a field experiment resembling the lab experiment in Andreoni and Rao (2011). In their study, solicitors for a fundraiser were positioned at one or both entrances of a supermarket. The solicitors either asked passersby verbally (“please give”) or merely stood there. Actively asking strongly increased the share of people donating as well as the average amount donated, but had the additional negative effect that people tried to avoid the solicitor by using a different entrance or exit. The result that actively asking people to donate increases givings has also been found by other authors using observational data on charitable donations (Yörük, 2009; Meer and Rosen, 2011). Related to this, Sanders and Smith (2016) conducted a field experiment in which lawyers asked clients during the process of drawing up a will if they wanted to leave a bequest for a charitable cause. The authors find that highlighting a social norm of charitable giving and mentioning emotional factors strongly increase the proportion of clients who decide to leave a bequest to charity. Furthermore, several field experiments in the area of charitable giving have shown that suggesting specific amounts to give has an effect on donation decisions (Edwards and List, 2014; Adena et al., 2014; Adena and Huck, 2016).

Apart from the effect of communication on altruistic behavior in a dictator game, communication has been studied in the experimental literature in other experimental settings that deal with social interactions. For instance, it has been shown that communication increases cooperation in the public good game (see, for example, Isaac and Walker, 1988, or Bochet et al., 2006).⁴

Besides communication, there are several factors that affect the decision to give in a dictator game which have been studied in the experimental literature so far. Relevant to our study are the following: decreasing the social distance between the dictator and the receiver leads to more generous giving decisions (Bohnet and Frey, 1999; Small and Loewenstein, 2003; Goeree et al., 2010; Hoffman et al., 1996).⁵ Subjects give more if the recipient is perceived as needy (Eckel et al., 2007). Increasing the efficiency of a kind act, i.e. multiplying the money transferred from the dictator to the receiver by a factor larger than one, has a positive

⁴Dal Bó and Dal Bó (2014) find that messages which appeal to moral norms and to the social benefit of a high contribution significantly increase contributions in a public good game. In their experiment, messages are not sent by other participants but by the experimenters themselves.

⁵However, decreasing the anonymity of the dictator can also have a negative effect on generosity (Dufwenberg and Muren, 2006).

Table 1.1: Matrix of endowments and payoffs

	A	B	Sum
Endowments	50	20	70
Payoffs if A decides not to give	50	20	70
Payoffs if A decides to give	40	40	80

effect on giving decisions (Engel, 2011). Also, the demographics of the dictators have an effect. For instance, women tend to give more than men, while students tend to give less than non-students (Engel, 2011).

1.3 Experimental Design and Procedures

In the experiment, we use a modified version of the dictator game. There are two types of players, player A (the dictator) and player B (the receiver). Prior to the start of the experiment, subjects are randomly assigned their role, which they keep for the duration of the experiment. At the beginning of each round, A receives an endowment of 50 points, whereas B receives an endowment of 20 points. Player A makes the binary decision to either keep the endowments unchanged or decrease her own points by 10 (to 40 points) while increasing player B's points by 20 (to 40 points). Thus, the decision to give points increases the sum of the points from 70 ($50 + 20$) to 80 ($40 + 40$). Table 1.1 illustrates the allocation of points to A and B depending on A's decision.

This simple game is played for multiple rounds. The number of rounds in a session is set in such a way that every A interacts with every B exactly once (absolute stranger matching). For example, if there are 32 subjects in a session, i.e. 16 A's and 16 B's, the game is played for 16 rounds. In the end, one of the rounds is randomly drawn. This round determines the payoff of the subjects in the experiment.

In the first round, there is no communication between A and B. Thus, A's decision in the first round can be seen as a baseline decision. From round 2 on, A receives a written message from the matched subject B before making her decision. B writes his message only once at the beginning of the experiment. B knows that this message will be shown to each subject A, except for the subject A with whom he is matched in the first round. Hence, subsequent to the first round, each message is shown to exactly one randomly determined dictator in every round

which ensures that the order of the messages does not affect the results. The reason for not letting B write a message every round is that we are interested in the performance of one message encountered by multiple A's, whereas we are not interested in learning effects of B. Subjects are told in the instructions that there are no restrictions regarding the content of the message except that they are not allowed to reveal their identity. The maximum length of a message is limited to 300 characters.

Eight sessions were conducted at the Laboratory for Experimental Research of the University of Erlangen-Nuremberg (LERN) in June 2016. In six of these sessions, 32 subjects took part, and in the remaining two sessions, 30 subjects took part, which led to 252 subjects and 1,986 giving decisions in total. Subjects were students, and nearly 80.0% of these studied management and economics. On average, subjects were 23.5 years old and 50.4% of them were male.

In each session, subjects were randomly assigned a seat. At the beginning, they read the instructions⁶ and had to complete a short quiz to ensure that everyone understood the experiment.⁷ While B's were writing their messages, A's made their decisions for the first round without communication and were then asked whether they could remember a situation in the real world that was similar to their task in the experiment.⁸ In the following rounds, A's decided whether or not to give while viewing the message of the matched B on-screen.⁹

Following the experiment, subjects answered a questionnaire¹⁰ and received their payoff in private. Points were converted to Euros at a rate of 5 Points = 1 Euro. Sessions lasted for approximately 45 minutes and subjects earned on average 11.45 Euros, including a show-up fee of 4 Euros. The experiment was computerized with z-Tree (Fischbacher, 2007) and the recruitment process was conducted using ORSEE (Greiner, 2015).

To determine which aspects of the messages are particularly important for the giving decision of the dictator, the free-form text messages were analyzed and sorted into eight categories after the main experiment took place. Some of the

⁶An English translation of the originally German instructions can be found in the Appendix.

⁷The control questions can be found in the Appendix.

⁸The reason for this was to make the experimental situation less abstract by thinking about a similar situation from the real world. Most of the situational descriptions were about beggars asking for money or, more generally, about solidarity within a society, a family or another group of people (e.g. friends, roommates).

⁹Screenshots of the decision screens for subjects A with and without communication can be found in the Appendix.

¹⁰Part of the questionnaire was the Interpersonal Reactivity Index of Davis (1983), which consists of four subscales. The subscale "Empathic Concern", which is later used in the regression analysis, consists of 7 items on a 5-point Likert scale.

categories are specific to the particular design of the experiment, others are more general. We followed the same procedure as Andreoni and Rao (2011) and let students from the same subject pool as in the main experiment decide whether a message belonged to one or multiple predetermined categories.¹¹ An English translation of the instructions for the rating session can be found in the Appendix. The eight categories are *Friendly Greeting*, *Sum of Points*, *Equity*, *Decency*, *Power / Responsibility*, *Humor*, *Need* and *Random Role*. Messages sorted into the category *Friendly Greeting* are characterized by a particularly friendly or polite greeting. The category *Sum of Points* contains messages in which the receiver argues that giving points to him will result in an increase in the sum of points, an efficiency gain or a welfare improvement. When the receiver tells the dictator that giving points will lead to an equal allocation of points, the message is part of the category *Equity*. When the receiver appeals to the dictator to behave fairly or decently, the message is sorted into the category *Decency*. Messages in which the receiver argues that the dictator has the sole decision power and / or the sole responsibility for both payments are sorted into the category *Power / Responsibility*. A particularly humorous message is sorted into the category *Humor*. Messages sorted into the category *Need* are those in which the receiver explains why he needs the money. Finally, when the receiver emphasizes in his message that the role allocation was random, and that the dictator could also have been a receiver or that the dictator should think about how she would like to be treated if she were a receiver, the message is sorted into the category *Random Role*.¹² Table 1.A.1 in the Appendix describes the eight categories in detail and gives one example message for each category.

¹¹27 subjects took part in the rating session. They were divided into three groups consisting of 9 raters each. Every rater read 42 (41 for the last group) messages and subsequently made her decision. The division into groups was done to reduce the number of messages each subject had to rate because it was important for us that raters made an effort and were concentrated when making their decision. Each message could be sorted into one, multiple or no category. Subjects were told that their help was important for our research and that they should take their time with their decisions. They earned 15 Euros for their participation. Thus, as in Andreoni and Rao (2011) we used a non-incentivized procedure in the rating session since we did not want to elicit subjects' expectations about the average opinion but rather their own opinion on whether a message belongs to a category or not.

¹²Regarding the categories used by Andreoni and Rao (2011), we adopted the categories *Friendly Greeting* and *Need*. However, in our study, a *friendly greeting* is defined as being more formal than it is in the study by Andreoni and Rao, in which it is defined as a "greeting usually used amongst friends." The category "Fairness" used in their study is divided into two categories, *Equity* and *Decency*. What Andreoni and Rao called "Acknowledge Power" is called *Power / Responsibility* in our study to emphasize that receivers did not only write about the power of the dictator but also about the dictator's specific responsibility for both payoffs. There were hardly any messages using "Flattery." Hence, this category was not used in our study.

1.4 Results

The following analysis uses 1,941 allocation decisions by subjects A - 1,815 with communication and 126 without communication (from round 1) - and 123 messages written by subjects B.¹³ First, some descriptive statistics about dictators' decisions are presented, then messages are analyzed, followed by the main part, in which the effects of the content of the messages on dictators' decisions are analyzed using logistic regressions.

1.4.1 Descriptive Statistics about the Giving Decision

Of all the 126 subjects A, 21 subjects (16.7%) always gave points to B, 24 subjects (19.0%) never gave points to B and 81 subjects (64.3%) varied their decisions to give. In round 1, which is the baseline decision without communication, subjects A decided to give points to B in 31.0% of the cases, while in later rounds they gave in 47.7% of the cases.¹⁴ Across all rounds, male A's gave slightly more often than female A's (50.6% compared to 42.7%).¹⁵ However, this difference between male and female dictators is not statistically significant when comparing the average giving decision of men and women per session (Wilcoxon Signed Rank Test $p = 0.1953$).

¹³The messages of three subjects B and the corresponding allocation decisions are excluded from the analysis. In the case of one subject B, there was a computer error which led to two messages being sent by this subject. We cannot be sure which message appeared on As' screen when deciding about the allocation of points. Therefore, the observations associated with this subject B had to be excluded. Furthermore, two subjects B wrote messages containing a deal offer in which they told the opposing party that they could meet after the experiment and exchange the welfare gains from giving. The observations associated with these messages are also excluded. The following results are robust to not excluding the observations belonging to these two deal messages.

¹⁴Since all subjects A in one session see the same messages, dictators' decisions after round 2 are not statistically independent observations. Hence, testing for differences between dictators' decisions without communication in round 1 and with communication after round 1 is merely possible at the session level, which leads to only 8 independent observations. Nevertheless, the difference between the average giving decision without and with communication is statistically significant (Wilcoxon Signed Rank Test $p = 0.0078$). Even though not the focus of this research, this finding confirms that communication as such has a positive effect on the decision to give. However, as this is only a within-subject comparison, the result could potentially also arise due to an experimenter demand effect (Zizzo, 2010) or due to an order effect (Harrison et al., 2005) since the round without communication is always the first one.

¹⁵In general, results from dictator games suggest that women behave more altruistically than men (Engel, 2011). In our setting, however, an altruistic act is relatively "cheap" or "efficient," since giving points to the receiver costs only 2 Euros while increasing the receiver's payoff by 4 Euros, which has been shown to increase the likelihood of altruistic acts more for male dictators than for female dictators (Andreoni and Vesterlund, 2001).

Table 1.2: Objective characteristics of messages

Variables	Full Sample	Gender of writer	
		Male	Female
Length (in characters)	219	216	222
Number of spelling mistakes	1.59	1.53	1.64
<i>Message contains (dummy variables):</i>			
Smileys	0.65	0.61	0.69
The word “Please”	0.19	0.25	0.12
The word “Thanks”	0.23	0.28	0.17
Number of messages	123	64	59

Note: Values in the table are averages for the full sample of messages, as well as for messages written by only male or female B’s.

1.4.2 Message Analysis

Table 1.2 shows characteristics of the 123 messages for the whole sample, as well as for male and female B’s separately.

On average, messages were 219 characters *long* and included 1.59 *spelling mistakes*. 65% contained at least one *smiley*, 19% contained the word “*please*”¹⁶ and 23% contained the word “*thanks*.”¹⁷ There are no significant gender differences in these objective message characteristics except for the use of the word “*please*,” which is used more frequently by male receivers; this difference is only significant at a 10% significance level (χ^2 Test $p = 0.062$).

As explained in detail in Section 1.3, the content of the messages was analyzed by letting 9 independent subjects decide for every message whether it belonged to none, one or multiple predetermined categories.¹⁸ We decided to use the average rating and no binary categorization in the following analysis since an average

¹⁶For “*please*,” different versions of the word were included (“*bitte*,” “*bitteschön*,” “*bitten*,” “*please*”).

¹⁷For “*thanks*,” different versions of the word were included (“*danke*,” “*dankeschön*,” “*thanks*,” “*thank you*,” “*merci*”).

¹⁸Each of the 123 messages was rated by 9 subjects leading to 1,107 rating decision. The decision that a message belonged to no category was made 46 times (4.2%), that it belonged to one category was made 208 times (18.8%), that it belonged to two categories was made 284 times (25.7%), that it belonged to three categories was made 253 times (22.9%), that it belonged to four categories was made 169 times (15.3%), that it belonged to five categories was made 107 times (9.7%) and the decision that it belonged to six, seven or eight categories was made 40 times combined (3.6%).

1.4. RESULTS

Table 1.3: Categories of messages resulting from the rating session

Variables	Full Sample	Gender of writer	
		Male	Female
Equity	0.54	0.56	0.52
Decency	0.50	0.47	0.53
Friendly Greeting	0.42	0.41	0.42
Power / Responsibility	0.35	0.34	0.35
Sum of Points	0.32	0.37	0.26
Humor	0.31	0.30	0.32
Need	0.14	0.11	0.17
Random Role	0.14	0.15	0.12
Number of messages	123	64	59

Note: Values in the table are averages for the full sample of messages, as well as for messages written by only male or female B's.

rating includes information about the level of agreement between raters.¹⁹

Table 1.3 summarizes the results of the rating session. An average message has a rating of 0.54 in the category *Equity*, which is the most frequent category. Following this category in order of frequency are *Decency*, *Friendly Greeting*, *Power and Responsibility*, *Sum of Points*, *Humor*, *Need* and *Random Role*, which is mentioned the least often.²⁰ Men appear to mention the increase in the *sum of points* more often, while women tend to appeal to *decent* behavior and address their particular *need* more often. However, these differences are not statistically significant.²¹ To sum up, there does not seem to be a considerable difference between the content written by male and female receivers.

¹⁹If, for example, 5 of the 9 raters decide that a message belongs in the category *Need*, this message will have a rating of 5/9 in the category *Need*. If we had used a binary rating instead, such a message would have had a rating of 1, regardless of whether 5 or 9 raters made this decision.

²⁰Table 1.A.2 in the Appendix shows the cross-correlation of all the message characteristics and categories. The highest (positive) correlation is between the three categories *Equity*, *Decency*, and *Power and Responsibility*. A factor analysis did not yield any additional insights.

²¹Wilcoxon Rank Sum Test for *Sum of Points* : $z = -1.433$, $p = 0.1517$; Wilcoxon Rank Sum Test for *Decency*: $z = 1.504$, $p = 0.1325$; Wilcoxon Rank Sum Test for *Need*: $z = 0.666$, $p = 0.5055$.

Table 1.4: Logit regression results for four models

VARIABLES	(1)	(2)	(3)	(4)
<i>Message Categories:</i>				
Equity	0.144 (0.144)		-0.042 (0.152)	-0.041 (0.203)
Decency	-0.013 (0.177)		0.079 (0.177)	0.096 (0.244)
Friendly Greeting	0.225** (0.093)		0.046 (0.101)	0.082 (0.134)
Power/Responsibility	0.534** (0.245)		0.416* (0.243)	0.666** (0.323)
Sum of Points	0.345** (0.141)		0.179 (0.135)	0.268 (0.177)
Humor	0.780*** (0.126)		0.491*** (0.126)	0.713*** (0.163)
Need	0.397** (0.156)		0.473*** (0.161)	0.630*** (0.210)
Random Role	0.369** (0.166)		0.239 (0.177)	0.205 (0.233)
<i>Message Characteristics:</i>				
Length		0.003*** (0.001)	0.002*** (0.001)	0.003*** (0.001)
# Spelling Mistakes		-0.071** (0.029)	-0.058* (0.030)	-0.082** (0.040)
Smiley		0.406*** (0.083)	0.339*** (0.083)	0.494*** (0.114)
Please		0.011 (0.091)	-0.091 (0.092)	-0.119 (0.124)
Thanks		0.179** (0.089)	0.159* (0.093)	0.214* (0.123)
<i>Characteristics Dictator A:</i>				
Baseline Give				2.441*** (0.367)
Male				0.343 (0.277)
Age				-0.016 (0.047)
Management/Economics				0.537 (0.409)
Empathy				0.857*** (0.233)
Round				0.031*** (0.009)
Constant	-1.265*** (0.410)	-1.268*** (0.405)	-1.510*** (0.429)	-5.697*** (1.305)
Observations	1,815	1,815	1,815	1,815

Note: The dependent variable in all four regressions is *give*, which takes a value of 1 if A decided to give points to B. Robust standard errors (clustered on the level of subjects A) in parentheses. All regressions include session dummies.
* significant at 10%; ** significant at 5%; *** significant at 1%.

1.4.3 Regression Results

Table 1.4 shows the results of a logistic regression²² in which the dependent variable is *give*, a variable that takes the value of 1 if the dictator decided to give points to B and 0 otherwise. The first model in column (1) includes only the category ratings of a message as explanatory variables. The second model in column (2) includes only the objective characteristics of a message as explanatory variables. The third model in column (3) includes the category ratings as well as the objective characteristics of a message as explanatory variables. These models show that the content as well as spelling and length of a message influence the decision to give.²³

Including the objective characteristics of a message as explanatory variables in the third model changes the coefficients for the category variables compared to the first model. Especially controlling for the length of a message has an impact, since messages scoring high in the categories *Sum of Points*, *Equity*, *Random Role*, *Power / Responsibility* and also - to a lesser extent - *Decency* and *Friendly Greeting* are on average relatively long messages, as can be seen from the cross-correlation Table 1.A.2 in the Appendix. The length of a message, in turn, has a positive impact on the probability of a successful request. Controlling for the use of *smileys* in a message, which has a positive impact on the likelihood of giving, increases the coefficient for *Need* since messages scoring high in this category include on average fewer *smileys*. On the other hand, messages in the category *Humor* use a lot of *smileys* and, hence, controlling for *smileys* decreases the estimated effect for such a message.

In the fourth model in column (4) of Table 1.4, characteristics of dictators A are included as explanatory variables. The predicted effects of the explanatory variables regarding the content as well as spelling and length of the messages change only slightly compared to the third model. Unless otherwise stated, the fourth model is used in the following discussion of the results. Additionally, Figure 1.1 displays the average marginal effects of all eight categories on the probability of success to provide an idea of the magnitude of the effects.

²²Using a Probit model or an OLS model does not yield essentially different results. Table 1.A.4 in the Appendix shows regression results for a fixed effects model in which subjects A define the panel. The results are mostly robust to the model used here.

²³A Wald Test rejects the joint hypothesis that the category variables for the first model (Wald-statistic with 8 restrictions: $\chi^2 = 56.84$, $p < 0.0001$) as well as the subjective category variables and objective characteristics variables for the third model (Wald-statistic with 13 restrictions: $\chi^2 = 68.21$, $p < 0.0001$) are simultaneously equal to 0, i.e. the explanatory variables in the first and the third model do have predictive power.

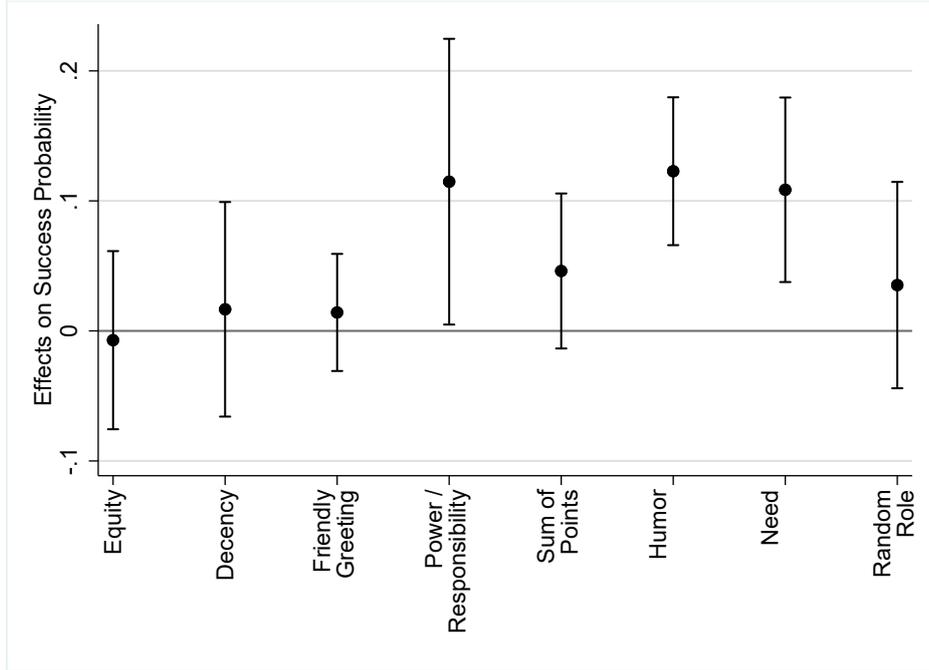


Figure 1.1: Average marginal effects for the full sample

Note: The figure shows point estimates as well as 95% confidence intervals.

The regression results show no statistically significant effect of the categories *Equity*, *Decency* and *Friendly Greeting*. Messages that fall in the category *Power and Responsibility*, though, affect the decision to give points positively and statistically significant. The model predicts that, on average, a message with a rating of 1²⁴ in the category *Power and Responsibility* has a success probability that is around 11.5 percentage points higher than a message with a rating of 0 in this category as can be seen in Figure 1.1. We find no statistically significant effect for the categories *Sum of Points* and *Random Role* on the likelihood that the dictator will give points to the receiver. The estimated effects are positive but only statistically significant at a 5% significance level when not controlling for the observable characteristics of a message - in particular, the length of a message. The category *Humor* is estimated to have the largest positive impact on the success probability. The model predicts that, on average, a message with a rating of 1 in this category has a success probability that is roughly 12 percentage points higher than a message with a rating of 0 in this category. This effect is highly significant. Regarding the category *Need*, the regression results show that mentioning the specific need of the receiver results in a higher chance of the dictator giving points to the particular receiver. The model predicts a statistically

²⁴A rating of 1 means that all 9 raters decided that a message belongs in the respective category.

1.4. RESULTS

significant increase in the success likelihood by nearly 11 percentage points for a message with a rating of 1 in the category *Need*, compared to a message with a 0 in this category.

As for the objective message characteristics, Table 1.4 shows that longer messages affect the giving decision positively.²⁵ The effect is also highly significant. Additionally, *spelling mistakes* in a message are punished by the dictator by reducing her willingness to give (significant at the 5% level). A *smiley* has a positive and highly significant effect on the probability that a dictator gives points to the receiver. This effect is estimated to be around 8.5 percentage points. Although every child is taught that the word “*please*” is mandatory or at least expected when asking someone a favor in a polite way, surprisingly, the model predicts that using “*please*” in a message does not improve the dictator’s willingness to act kindly. Eventually, messages containing the word “*thanks*” seem to have a higher success chance. This effect, however, is only statistically significant at the 10% level ($p = 0.083$).

Considering the characteristics of the dictator, our data confirms that dictators who gave points to the receiver in the baseline round without communication are clearly more inclined to give points in the later rounds when communication is present. Table 1.A.3 in the Appendix additionally displays regression results divided according to dictators who did or did not give points to the receiver in the baseline round (columns (3) and (4)). It could be that dictators who are less inclined to give and, hence, have to be actively convinced to give, and dictators who are already inclined to give and, hence, have to be actively dissuaded from giving, react differently to certain aspects of a message. However, the results show no substantial differences in the behavior of dictators who either gave or did not give points in the baseline round.²⁶

Male dictators and dictators who study management and economics seem to be more generous; however, this is not statistically significant for the gender of the dictator and only significant at a 10% level for the dictator’s field of study. Subjects scoring high on the subscale “Empathic Concern” in the post-experimental questionnaire are more willing to give.²⁷ Finally, the willingness to give increases in later rounds.²⁸

²⁵The coefficient is relatively small because the length is measured in characters.

²⁶The results indicate some differences in the reaction to an argument concerning the *sum of points*, which seems to be only successful when trying to convince a dictator who did not give points without communication.

²⁷The subscale has a scale reliability coefficient (Cronbach’s α) of 0.8126.

²⁸Since every message is shown to exactly one A in each round in random order, the positive time trend is of no concern when analyzing the effect of these messages.

1.4. RESULTS

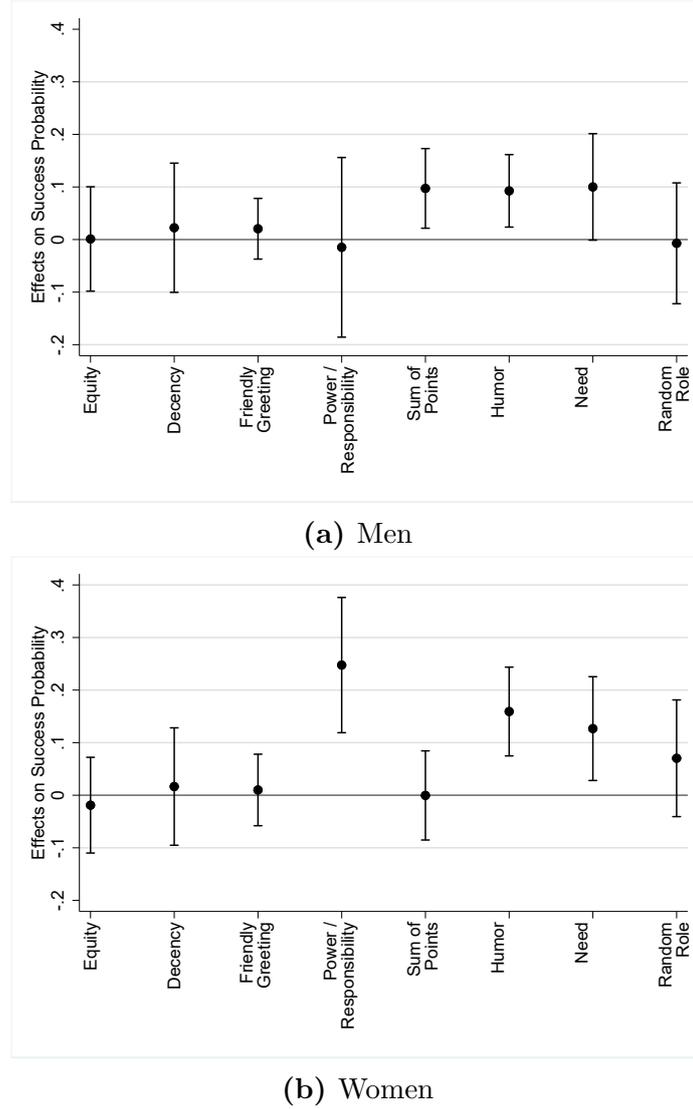


Figure 1.2: Average marginal effects for male and female dictators

Note: The figures show point estimates as well as 95% confidence intervals.

1.4.4 Different Behavior of Male and Female Dictators

Regarding gender differences in the reaction to particular aspects of the message, Figure 1.2 displays the average marginal effects for the eight categories for male as well as female dictators resulting from a logistic regression, which is shown in Table 1.A.3 in the Appendix. This regression uses the same model as before (fourth model in Table 1.4) and shows results once only for male dictators in column (1) and once only for female dictators in column (2).

First of all, men and women seem to react differently to messages which emphasize the *power and responsibility* relationship between the dictator and the receiver.

The effect of such a message is estimated to be very close to 0 for male dictators ($p = 0.865$) while being positive, very large and highly significant for female dictators ($p < 0.001$); for female dictators an increase in the success probability of nearly 25 percentage points is predicted by the model if a message has a rating of 1 in this category, compared to a rating of 0.

Furthermore, arguments concerning the efficiency of giving points to the receiver seem to yield different reactions by men and women. For male dictators, the effect of the category *Sum of Points* on the success probability is positive, statistically significant ($p = 0.016$) and relatively large; an increase of 9.7 percentage points is predicted by the model for a message that has a rating of 1 in this category, compared to having a rating of 0 in that category. For female dictators, though, the effect is close to 0.

Finally, men show no reaction to using the word “*thanks*” in a message ($p = 0.811$) whereas it increases the success probability significantly for female dictators ($p = 0.018$). Hence, the statistically significant effects of the category *Power and Responsibility* and of the word “*thanks*” on the willingness to give in the full sample are solely driven by female dictators.

1.5 Discussion

In the following we will discuss our results about the effect of message content on giving decisions. We will briefly repeat our main results and link them with previous results from the literature as well as answers in the post-experimental questionnaire.

First of all, the content of the receiver’s request influences the dictator’s decision to give. This result is additionally supported by statements from dictators in the post-experimental questionnaire, in which they were asked how they came to their decision. Ten out of 126 dictators mentioned that they had planned to either never give points to the receiver or always give points, but changed their minds after reading particularly “good” or “bad” messages.

Result 1: Messages including *Humor* affect the dictator’s decision to give positively. The *length* of a message and using a *smiley* affect the success probability positively, while it is negatively affected by *spelling mistakes*.

It seems that dictators are more willing to give if the receiver puts effort into writing the message (by writing a longer message without spelling mistakes) and

if the message has the potential to contribute to the dictator’s well-being (e.g. a message that is humorous or contains a smiley). This finding can be interpreted in terms of reciprocity (Rabin, 1993): since the receiver has done something for the dictator (putting effort into writing the message or contributing to the dictator’s well-being), the dictator reciprocates and gives points to B. When asked in the questionnaire about what influenced their decision, dictators most commonly gave answers related to this line of argumentation. They said they were more inclined to give when a message was nice and friendly so that the writer seemed more likeable (41 mentions), when the message was creative (23 mentions), funny and humorous (21 mentions), thoughtful and showed effort (12 mentions) or contained few mistakes (6 mentions).

Result 2: We do not find an effect of a *friendly greeting* or of using the word “*please*” on the willingness to give.

In the light of the above argumentation, it may be surprising that a friendly greeting has no effect on the decision of a dictator. One likely reason is that norms concerning politeness in electronic chats are different to standard norms in personal settings. Likewise, Althoff et al. (2014) also finds no effect of the politeness of a request on the willingness to do somebody a favor in an online community. One dictator wrote in the questionnaire that she does not like a greeting that is too formal or polite. This could explain why a friendly greeting (and also the use of the word “please”) has no effect.

Result 3: Messages including *Need* affect the dictator’s decision to give positively.

A message of “need” may serve as an “empathetic stimulus” (Andreoni et al., 2017) since it makes a need salient and also heightens the attachment to the other person by letting the dictator adopt the perspective of the receiver. It has been argued that both the “magnitude of the perceived need” as well as the “strength of the attachment” to the other person increase the empathetic concern for this person (Batson, 1991). The empathy-altruism hypothesis (Batson, 1991; Andreoni and Rao, 2011; Andreoni et al., 2017) states that feelings of empathy towards the person in need may be the main driving force for altruistic actions. Hence, it seems that by increasing the empathic concern, such a message raises the probability of an altruistic act. Additionally, heightening the attachment to the receiver can also be thought of as reducing the social distance between the receiver and the dictator (Mohlin and Johannesson, 2008) which would increase

the generosity of the dictator (Bohnet and Frey, 1999; Small and Loewenstein, 2003; Goeree et al., 2010; Hoffman et al., 1996). However, while explaining the particular need has a positive effect on generosity, the specific wording seems to be important: a message inducing pity or showing obvious signs of begging was mentioned most often in the questionnaire as a reason for not giving points to a receiver (14 mentions).

Result 4: We do not find an effect of messages including *Equity* on the success probability.

Arguing that giving points to the receiver also equalizes payoffs of the two players has no significant effect on the decision to give, neither for male nor for female dictators. This result may be surprising since highlighting the aspect that giving points leads to equal payoffs could make inequity averse (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000) dictators more inclined to give. However, in our setting this aspect of the design is very obvious and, thus, it may not be possible to make it more salient by mentioning it in a message.

Result 5: Messages including *Power and Responsibility* as well as using the word “*thanks*” affect the willingness to give positively for female dictators.

The positive effect of emphasizing the power and responsibility relationship in a message as well as using the word “thanks” is driven solely by female dictators. We believe that both aspects of a message could lead to a subtle increase in the social pressure to behave altruistically without being strongly demanding. Highlighting the responsibility of the dictator for the receiver is a rather clear way to remind the dictator that she is responsible for her (potentially) selfish action and to think about the consequences for the other party, which has been shown to increase giving in a dictator game (Brañas-Garza, 2007). The word “thanks” is mostly used in the end of a message (“thanks in advance”) and, hence, signals gratitude but also an expectation that the dictator behaves kindly, which could also increase the social pressure to give in a subtle way.

It has been shown that social pressure has a strong impact on generosity in the area of charitable donations both in field experiments (DellaVigna et al., 2012; Andreoni et al., 2017) as well as in field data (Bekkers and Wiepking, 2011; de Wit and Bekkers, 2016). Men and women may initially feel a different level of social pressure to help the receiver. Psychological research (Eagly, 2009; Babcock and Laschever, 2003) has shown that gender-specific stereotypes exist and that these stereotypes work not only by describing how men and women usually behave

(descriptive), but also by claiming how men and women should behave (prescriptive). Hence, such a gender-specific role or stereotype serves as a social norm and men as well as women are expected to behave accordingly, thereby creating social pressure to do so (Eagly, 2009; Babcock and Laschever, 2003; de Wit and Bekkers, 2016). Since women are assumed to be more caring, friendly and “other-oriented” (Deaux and Major, 1987; Babcock and Laschever, 2003; Eagly, 2009), women experience a higher social pressure to act in such a way, i.e. they are expected to be more generous and help the person in need. In a field experiment, DellaVigna et al. (2013) show that women increase donations to a charitable cause significantly more than men when put under pressure.²⁹ Hence, a different reaction to a subtle increase in the social pressure to act altruistically may be an explanation for why only women react to arguments concerning the power and responsibility relationship as well as to using the word “thanks” in a message.

Result 6: We do not find an effect of messages including *Decency* on the success probability.

Social pressure to give can also be exerted by messages that explicitly make fairness norms more salient. Such rather demanding messages have no significant effect on dictators’ decision and there are no gender difference in this case. In line with this result, several dictators mentioned in the questionnaire that they did not like messages which were too demanding or which told them explicitly how they should behave or what they should do (9 mentions). In this laboratory setting, it is relatively easy to avoid such an explicit form of social pressure since clicking twice is enough to let a message disappear. Hence, it is likely that the effect of explicit and direct social pressure is different in settings in which avoidance is more difficult, as it is the case in most real-life interactions (DellaVigna et al., 2013; Andreoni et al., 2017; DellaVigna et al., 2012).

Result 7: Messages including *Sum of Points* affect the dictator’s decision to give positively for male dictators.

Our results indicate a gender difference when the dictator is faced with an argument mentioning the increase in total points when giving. Men react positively to such an argument, while women show no reaction. This finding corresponds to results by Andreoni and Vesterlund (2001), in which the authors vary the budget and the relative price of altruism in a dictator game and find that men react more strongly to the efficiency of a donation, i.e. they behave more altruistically than

²⁹However, for field data from the Netherlands, de Wit and Bekkers (2016) find no evidence that social pressure can explain the different behavior of men and woman in donations.

women when altruism is relatively “cheap” and less altruistically than women when altruism is relatively “expensive.” This gender difference with respect to the relative price of a donation has also been found in field data on charitable donations (Andreoni et al., 2003). A message arguing that giving points to the receiver increases the total sum of points may make this aspect more salient and, thereby, triggers a positive reaction by male dictators.

Due to the fact that our subject pool mainly consists of management and economics students, it could very well be that the effect of the efficiency argument is stronger in this case than it would be in a sample which is more representative of the general population (Engelmann and Strobel, 2006; Fehr and Schmidt, 2006).³⁰ However, since the proportion of management and economics students is similarly high for male and female A’s (78 % of the female A’s and 82 % of the male A’s study management and economics), the different reactions to the efficiency argument cannot be explained by our particular subject pool.

1.6 Conclusion

In this study, we implemented free-form written communication in a modified dictator game and analyzed its effect on the dictators’ decision to give. The content of a message has a clear effect on the dictator’s decision. Humorous and friendly messages, as well as messages showing that the author exerted effort, lead to reciprocal behavior by the dictator, i.e. if people feel like they get something in return for their generosity - and it may be just a smile(y) - they are willing to help. Mentioning the receiver’s specific need also increases generosity, which emphasizes the importance of empathetic feelings towards the person in need for altruistic behavior. Only men react positively to an efficiency argument, whereas emphasizing the power and responsibility relationship increases giving only if the dictator is female.

Our research supports the results of Andreoni and Rao (2011), Mohlin and Johannesson (2008) and Langenbach (2016) that communication has a positive effect on dictators’ willingness to act kindly. Furthermore and most importantly, we fill the gap in analyzing systematic reactions of dictators in response to certain aspects of written communication, which, to the best of our knowledge, has not

³⁰Vice versa, an appeal for equity may work better if the subject pool consisted of fewer management and economics students (Engelmann and Strobel, 2006; Fehr and Schmidt, 2006).

1.6. CONCLUSION

been done before in this level of detail. Additionally, our research contributes to the current research on gender differences in the area of altruism (Andreoni and Vesterlund, 2001; DellaVigna et al., 2013; Croson and Gneezy, 2009; de Wit and Bekkers, 2016).

Primarily, our results are relevant for numerous everyday life situations in which one party asks another one a favor. Additionally, such diverse fields as charity work, negotiations and marketing might benefit from our findings, even though, especially the latter two differ considerably from our setting. Nevertheless, since negotiating is about convincing the other party with specific arguments to agree to something and marketing is likewise to a large extent about convincing somebody (i.e. a customer) to do something (i.e. to buy the product), some of our findings may be transferable to these fields.

Especially for charity work, it is important to distinguish between content- and relationship-effects when using our results, since the person who asks a favor is not the person who benefits from the favor in that case. It is very likely that relationship-specific effects are mitigated if an intermediary tries to convince a potential benefactor and not the potential beneficiary of the favor herself. Future research may study the effect of communication in a setting in which an intermediary asks a favor and analyze any differences to our findings.

1.A Appendix

1.A.1 Supplementary Tables and Figures

Table 1.A.1: Message categories

Category	Description	Example Message (translated from German)
Equity	Changing the score of points leads to an equal distribution of points.	“...and accomplish identical payments for both of us. Please reduce your endowment by 10 points, so that 20 points are added to my endowment and that there is equity in payments. ...”
Decency	Appealing to A to behave fair and decent	“Please be fair and distribute justly,... . Please do not be selfish and inconsiderate, since you would harm me that way and leave with a bad conscience in the end.”
Friendly Greeting	Friendly or polite address	“Hello dear participant A, ...”
Power/ Responsibility	A has the sole decision power but also the sole responsibility for both payments.	“I cannot overrule your decision and maybe not even influence it. ... THE CHOICE IS YOURS :)”
Sum of Points	B argues with an efficiency gain, i.e. an increase in the sum of points.	“... Please share your points with me! That way, the overall welfare will increase. You will receive a little less, in return I get a little more. In total, this would put the two of us in a better position. ...”
Humor	The message is particularly humorous	“Two economists are on a quest for personal happiness (i.e. the maximum utility). Let us do it: Let’s increase the overall welfare! In favor of shifting the aggregated budget line, towards higher indifference curves, until we reach the heaven of the best possible Nash equilibrium.”
Need	B explains why he needs the points / money	“Hi. On the weekend, I would like to visit my girlfriend. In order to do this, I need some money to buy a bus ticket. So please be a bit fair and share your points with me. Do it for the sake of Love. :)”
Random Role	B mentions the random allocation of roles and / or how A would feel being a B.	“...Perhaps you keep in mind that you could find yourself in my position and you would surely be glad about a friendly A.”

Table 1.A.2: Cross-correlation table of message characteristics

Variables	Eq	D	FG	P/R	SoP	H	N	RR	L	SM	S	PI	T
Equity (Eq)	1.000												
Decency (D)	0.604	1.000											
Friendly Greeting (FG)	-0.006	0.090	1.000										
Power / Responsibility (P/R)	0.580	0.572	0.105	1.000									
Sum of Points (SoP)	0.366	0.225	-0.076	0.408	1.000								
Humor (H)	-0.322	-0.318	-0.013	-0.360	-0.235	1.000							
Need (N)	-0.069	-0.024	0.056	-0.058	-0.151	0.149	1.000						
Random Role (RR)	0.244	0.275	0.195	0.316	0.103	-0.048	-0.090	1.000					
Length (L)	0.371	0.259	0.186	0.338	0.422	0.090	0.017	0.340	1.000				
# Spelling Mistakes (SM)	0.063	-0.043	-0.032	-0.105	-0.029	0.031	0.037	0.179	0.121	1.000			
Smileys (S)	0.014	-0.120	0.077	-0.097	-0.117	0.224	-0.083	0.039	-0.068	0.332	1.000		
Please (PI)	0.153	0.282	0.044	0.160	0.045	-0.114	0.189	0.008	-0.057	-0.054	-0.173	1.000	
Thanks (T)	0.038	-0.040	0.179	-0.005	0.066	-0.007	0.062	0.034	0.033	0.139	-0.009	0.187	1.000

1.A. APPENDIX

Table 1.A.3: Logit regression results for subsamples

VARIABLES	Gender		Decision Round 1	
	Men (1)	Women (2)	Baseline=1 (3)	Baseline=0 (4)
<i>Message Categories:</i>				
Equity	0.007 (0.309)	-0.113 (0.278)	-0.201 (0.315)	0.105 (0.258)
Decency	0.137 (0.381)	0.098 (0.339)	-0.036 (0.390)	-0.091 (0.297)
Friendly Greeting	0.125 (0.184)	0.060 (0.207)	0.071 (0.408)	0.060 (0.146)
Power / Responsibility	-0.090 (0.534)	1.475*** (0.395)	1.057 (0.752)	0.494 (0.372)
Sum of Points	0.594** (0.246)	-0.002 (0.258)	-0.500 (0.313)	0.379* (0.204)
Humor	0.565*** (0.211)	0.949*** (0.243)	0.791** (0.327)	0.693*** (0.196)
Need	0.611* (0.325)	0.755** (0.298)	0.770* (0.410)	0.675*** (0.247)
Random Role	-0.044 (0.359)	0.419 (0.335)	0.520 (0.545)	0.203 (0.267)
<i>Message Characteristics:</i>				
Length	0.003*** (0.001)	0.003* (0.001)	0.006*** (0.001)	0.002** (0.001)
# Spelling Mistakes	-0.103* (0.056)	-0.072 (0.060)	-0.192** (0.075)	-0.065 (0.043)
Smileys	0.547*** (0.144)	0.487*** (0.177)	0.892*** (0.237)	0.385*** (0.123)
Please	0.002 (0.177)	-0.243 (0.192)	0.117 (0.344)	-0.144 (0.141)
Thanks	0.049 (0.204)	0.378** (0.160)	0.596 (0.465)	0.154 (0.137)
Round	0.039*** (0.015)	0.027** (0.013)	0.041** (0.020)	0.032*** (0.011)
Constant	-4.437** (2.183)	-7.208*** (1.749)	-4.027* (2.096)	-6.206*** (1.860)
Observations	893	922	501	1,254

Note: The dependent variable in all regressions is *give*, which takes a value of 1 if A decided to give points to the receiver. Robust standard errors (clustered on the level of subjects A) in parentheses. All regressions include session dummies and control for individual characteristics of dictator A (see Table 1.4).

* significant at 10%; ** significant at 5%; *** significant at 1%.

1.A. APPENDIX

Table 1.A.4: logit regression results for a subject fixed effects model

VARIABLES	Full Sample (1)	Gender	
		Men (2)	Women (3)
<i>Message Categories:</i>			
Equity	0.161 (0.309)	0.389 (0.461)	0.022 (0.422)
Decency	-0.170 (0.448)	-0.315 (0.645)	-0.032 (0.631)
Friendly Greeting	0.194 (0.228)	0.232 (0.341)	0.191 (0.314)
Power / Responsibility	0.890 (0.571)	-0.254 (0.842)	1.834** (0.793)
Sum of Points	0.396 (0.293)	1.034** (0.443)	-0.138 (0.402)
Humor	1.165*** (0.293)	1.135*** (0.438)	1.238*** (0.403)
Need	1.195*** (0.317)	1.231*** (0.477)	1.166*** (0.429)
Random Role	0.331 (0.368)	0.161 (0.547)	0.455 (0.508)
<i>Message Characteristics:</i>			
Length	0.005*** (0.001)	0.005** (0.002)	0.004** (0.002)
# Spelling Mistakes	-0.133** (0.067)	-0.158 (0.102)	-0.127 (0.091)
Smileys	0.726*** (0.193)	0.722** (0.287)	0.720*** (0.268)
Please	-0.200 (0.218)	-0.073 (0.320)	-0.345 (0.302)
Thanks	0.356* (0.196)	0.074 (0.294)	0.578** (0.268)
Round	0.050*** (0.017)	0.071*** (0.026)	0.036 (0.024)
Observations	1,124	488	636
Number of Subjects	78	34	44

Note: The dependent variable in all regressions is *give* which takes a value of 1 if A decided to give. The panel is defined by subjects A.

* significant at 10%; ** significant at 5%; *** significant at 1%.

1.A.2 Instructions and Control Questions

Instructions for Participants of the Main Experiment

The following section provides the English translation of the original German instructions that the participants received in the main experiment:

General Instructions

Today, you are participating in a decision-making experiment. If you read the following instructions carefully, you can earn money. The amount of money that you will receive depends on your decisions and the decisions of other participants. For the entire duration of the experiment, it is not allowed to communicate with other participants. Thus, we ask you not to talk to each other. We also ask you to turn off your mobile phone. If you violate these rules, you will be excluded from the experiment and the payment.

In case there is something you do not understand, please take another look at the instructions or raise your hand. We will then come to your seat and answer your question personally.

During the experiment, we are not using Euros but rather points. The points you earn in the experiment will be converted into Euros as follows:

5 points = 1 Euro

At the end of the experiment, you will receive the points you have earned, converted into Euros, in **cash**. Additionally, you will receive 4 Euros for showing up on time.

On the following pages, we will explain to you the experiment in detail. Before the start of the experiment, we will ask you a few control questions on the screen, which are supposed to help you understand the procedure. The experiment only starts when all participants are completely familiar with the procedure of the experiment and have answered the control questions correctly.

The Experiment

In the experiment, you and the other participants each take on a role. There are two different role types. Half of the participants is participant A, the other half is participant B. Your role will be assigned to you randomly at the beginning of the

experiment. You only decide for your role and you keep this role for the whole experiment. In the following, a participant who takes on role A will be called participant A.

The experiment consists of several rounds in which you interact with alternating participants. In each round, the computer program selects groups of two at random of all the participants present. Each group consists of exactly one participant A and one participant B. Every round, new groups are randomly selected in such a way that you will always be matched with a participant with whom you have not interacted yet. Neither you nor the other participants get to know something about the identities of the participants in the groups; neither before nor after the experiment.

The experiment ends as soon as every participant A has interacted with every participant B once. If there are, for example, 20 participants in this experiment, i.e. 10 participants A and 10 participants B, the experiment will end after 10 rounds. Thus, in every round you are matched with a participant with whom you have not interacted yet.

As soon as the experiment starts, you will be informed about your role and the number of rounds on screen.

Round 1

At the beginning of each round, all participants of this experiment receive an endowment of points. Each participant A receives 50 points, each participant B receives 20 points. Now participant A can decide whether she wants to change the score of points as follows:

- Participant A can reduce her own points from 50 to 40 points which leads to an increase of participant B's points from 20 to 40.
- Participant A can keep her own 50 points; then, participant B also keeps his 20 points.

From round 2 on

As in round 1, participant A can decide about changing the score of points. Before participant A decides about a possible change of the score of points, participant B sends a text message to participant A. The maximum length of the

message is 300 characters. Being a participant B, you may write whatever you like with the only exception that you are not allowed to give any information about your identity. Participant A reads this message before she decides about a possible change of points.

The message that participant B sends to participant A is written only once, namely in round 1. In round 1, this message is not shown but from round 2 on, the message is shown to the matched participant A in the respective round.

Thus, participants B only take an active part in the experiment once (in the first round). In the other rounds, they do not write messages anymore. However, at the end of every round, they are informed about their payoff in points in this round.

Payment

At the end of the experiment, the computer program will choose one of the previously carried out rounds at random. Only the group composition of this selected round and only the decision of participant A for this round are relevant for your payment. For the other rounds, which are not selected, you will not receive any payment. The first round, in which participant A makes a decision without having received a message from participant B before, is equally likely to be selected as every other round. Thus, as a participant A, you should decide very carefully in every round because only after the experiment, you will learn which of your decisions is relevant for the payment. As a participant B, you should think carefully about your message in the first round because, from round 2 on, every participant A you interact with reads this message.

After you have made all decisions, you will learn which round has been selected by the computer program for the payment. Participants A get to know whether they decided for a change of points in this particular round and which payment they receive. Participants B get to know whether their points have been changed by the respective participant A and which payment they receive.

Following the experiment, we are going to ask you to fill in a questionnaire. Then, you will receive your payment in cash. Points are converted to Euro at an exchange rate of 5 points for 1 Euro.

Control Questions in the Main Experiment

- If participant A decides not to change the score of points, participant A will receive 50 points and participant B will receive 20 points. (true)
- If participant A decides to change the score of points, participant A will receive 50 points and participant B will receive 40 points. (false)
- In the end of each round, participant B learns which decision participant A made in this round. (true)
- In the beginning of each round, participant B can send a message to participant A. (false)
- Every round is paid out. (false)

Instructions for Participants of the Rating Session

The following section provides the English translation of the original German instructions that the participants received in the rating session:

General Instructions

We welcome you at LERN. Thank you very much for being here today.

Please notice: During the whole stay in the laboratory, it is not allowed to communicate with other participants. Thus, we ask you not to talk to each other. Please turn off your mobile phone as well. If you have any questions, please give us a hand signal. We will come to your seat and answer your question personally. Your task today is the following:

A while ago, we conducted an experiment in which one part of the participants could write messages to the other participants. We would like you to read those messages and sort them into one or several categories.

At the end of the task, you will receive a payment of 15 Euros in cash.

On the next page, you find the instructions of the conducted experiment. Please read these instructions carefully to make sure that the context of the messages becomes clear.

The independent categorization is an important part of our project. We appreciate your collaboration very much.

- Instructions of the main experiment -

Your task today

You are going to read messages on screen which have been written by a participant B of the experiment explained above. You should sort these messages into the following categories:

- **Friendly greeting.**
Participant B addresses participant A in a very friendly or polite way.
- **B argues with an efficiency gain due to a change of the score of points.**
In this context, efficiency gain means that the sum of points of participant A and B increases due to a change of points, i.e. the “pie” that is divided becomes larger.
- **B uses equity as an argument.**
By changing the score of points the points are distributed evenly. Participant A and B will only receive the same number of points if A decides to change the points.
- **B asks A to act decently.**
Participant B appeals to participant A to act friendly or fair, for instance by appealing to A’s conscience.
- **B emphasizes that A bears the responsibility for both payments.**
Participant B notices that participant A has the exclusive power to make decisions. He may explicitly point out to A that because of her special responsibility she should make her decision very responsibly.
- **The message is particularly humorous.**
This category includes, for example, messages in which participant B tells a joke. However, you can also sort messages into this category that you believe to be particularly humorous for other reasons.
- **B explains why he needs the points / the money urgently.**
Participant B mentions his special need and his dependency on the income

from the experiment. In this context he might also explain specifically, what he would spend the additional points on.

- **B mentions the random allocation of roles and / or how A would feel being a B.**

Messages belonging to this category are about the random allocation of role A and role B in the beginning of the experiment. Participant A could have become a participant B with the same probability.

Important: You can sort each message into **none**, **one** or **several** categories!

1.A.3 Screenshots

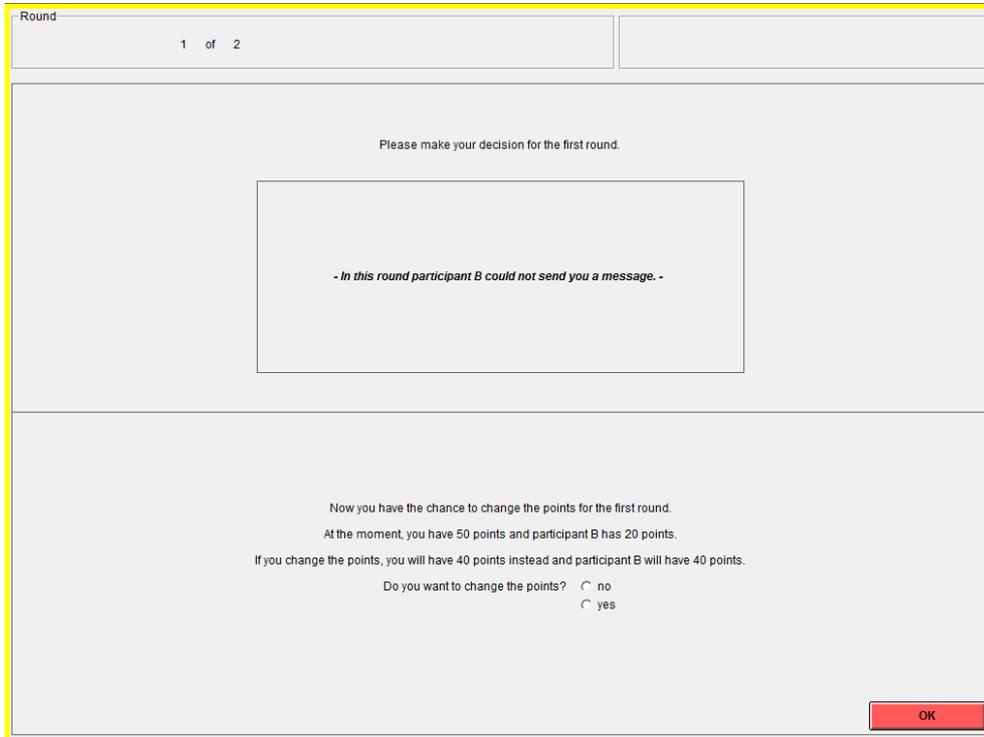


Figure 1.A.1: A's decision screen in the first round (English translation)

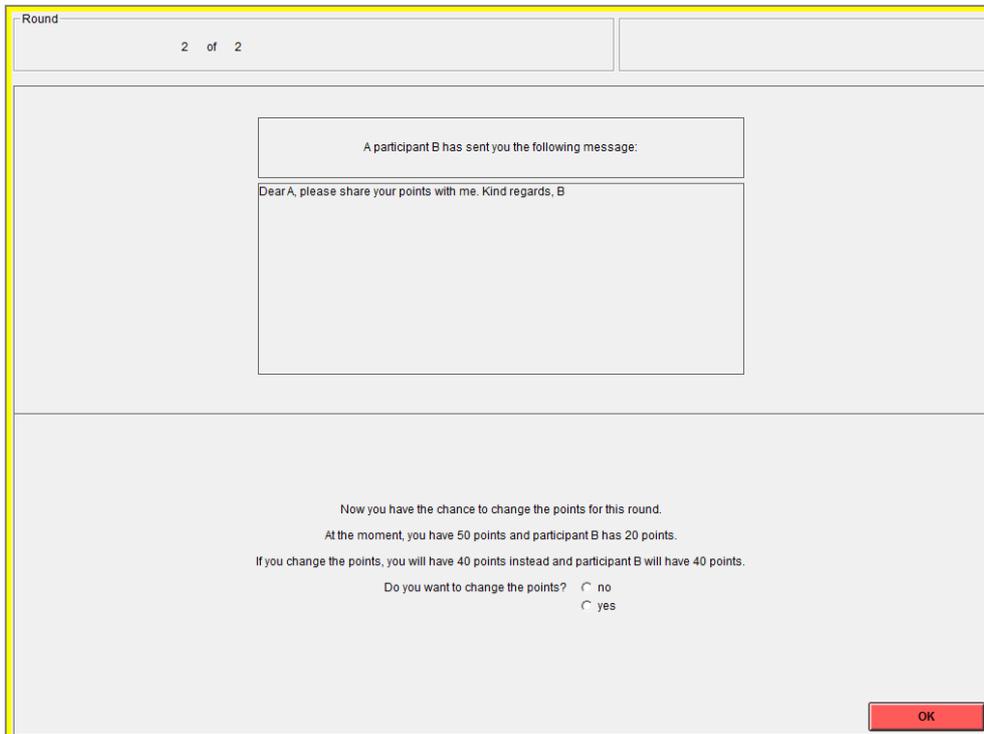


Figure 1.A.2: A's decision screen from round 2 on (English translation)

Chapter 2

Gender Differences in the Response to Decision Power and Responsibility - Framing Effects in a Dictator Game

2.1 Introduction

The dictator game (Kahneman et al., 1986; Forsythe et al., 1994) has been used extensively in experimental economics and related fields in order to study social preferences such as fairness and altruism. In the basic game, a dictator has an endowment that she can allocate between herself and a receiver. While most results suggest that dictators do not always behave selfishly (see Engel (2011) for a review and meta-analysis), it has been argued that the dictator game is rather sensitive to the context and the frame in which the game is presented (Fehr and Schmidt, 2006; Krupka and Weber, 2013; List, 2007; Levitt and List, 2007; Zizzo, 2010).

In this paper, we study the effect of two different context frames (Gerlach and Jaeger, 2016) on dictators' decisions, specifically the effect of two different texts, which participants read before making their decision. In one treatment, this text affirms the dictators' freedom of decision and decision power¹ while in the other

¹The exact wording was: "You alone decide about the payoffs, and you are completely free in your decision."

treatment the text reminds dictators of their responsibility for the other player.² Both texts have in common that they refer to the dictator’s power, but they differ in the implied recommended use of this power: while the first one explicitly refrains from implying either a selfish or an altruistic decision and instead emphasizes using the power at one’s discretion, the second one highlights that power comes with responsibility for others, which may increase giving to the recipient. Reminding people that they are responsible for the other person (treatment *Responsibility*), but also that they are the sole decision maker (treatment *DecisionPower*) may make people perceive stronger accountability for their own decision. Brañas-Garza et al. (2009) show that feelings of accountability for the decision (or responsibility as it is called in their paper), which the authors define as having control over the decision, affect dictator’s altruism. Following this definition, the treatment *Responsibility* may directly increase feelings of accountability while the treatment *DecisionPower* may indirectly increase feelings of accountability by reminding dictators that they are in control over the allocation decision. If people feel more accountable for their (selfish) action, it will be more difficult to maintain a positive self-image when acting in a selfish way. Thus, people may respond to both treatment conditions by giving more to the receiver (Dana et al., 2007; Etang et al., 2016).³ Additionally, increasing the salience of being responsible for someone else (treatment *Responsibility*) makes it more likely that a dictator thinks about the consequences for the other person when making her decision. Besides increasing accountability, this may increase feelings of shame if a dictator would give little or nothing to the recipient. Feelings of shame have been shown to increase altruistic behavior (Ellingsen and Johannesson, 2008; Xiao and Houser, 2009; Dana et al., 2007; Petit et al., 2012).

We are especially interested in whether these texts trigger different reactions by female and male dictators. From previous research, we know that females tend to react stronger than males to the context of an experiment in general (Croson and Gneezy, 2009; Ellingsen et al., 2013; Kahn et al., 1971; Eagly, 1983). Moreover, Chapter 1 has shown that male and female dictators react differently to recipients’ messages emphasizing the dictator’s responsibility and decision power. Furthermore, psychological research has shown that women are expected to be more caring, altruistic and other-oriented (Deaux and Major, 1987; Babcock and Laschever, 2003; Eagly, 2009; Heilman and Chen, 2005). In a dictator game,

²The exact wording was: “You bear the responsibility for both payoffs, your own and the one of participant B.”

³See the theory of self-concept maintenance by Mazar et al. (2008).

women (Aguiar et al., 2009) or men and women likewise (Brañas-Garza et al., 2018) believe that female dictators are more generous. Higher expectations to behave altruistically lead to higher social pressure to act accordingly (Eagly, 2009; Babcock and Laschever, 2003; de Wit and Bekkers, 2016). Hence, women may feel a higher pressure to help a person in need. This form of social pressure to act altruistically is, therefore, mainly exerted by society but also by participants themselves if they want to maintain their self-image of being someone who follows social norms or fulfills expectations. We believe that the two frames in this experiment affect the social pressure to give differently. While the text highlighting dictator’s freedom of choice attempts to decrease the pressure (“...you are completely free in your decision.”), the text emphasizing dictator’s responsibility for the recipient may increase social pressure (“You bear the responsibility for both payoffs, ...”). Since women and men may initially feel a different level of social pressure, these texts could potentially trigger a different reaction depending on whether a male or female dictator reads them. Thus, we expect that both genders may respond differently to our treatment conditions.

In order to study this research question, we conduct a pen and paper dictator game experiment, comprising three treatments in a between-subjects design. In two of the three treatments, either the text emphasizing the dictator’s responsibility or the text highlighting the dictator’s decision power is printed on top of the decision sheet. The third treatment serves as a control treatment without a text. In total, 207 participants in the role of the dictator participated in the experiment.

We find that highlighting the dictator’s responsibility for the payoffs of both players has a positive effect on the amount given. This effect is driven by the behavior of men, who react strongly positive, whereas women show no response to this text. Emphasizing the dictator’s decision power and freedom of choice has no significant effect; if at all, it tends to increase giving by female dictators.

2.2 Related Literature

Closest to our study is the one by Brañas-Garza (2007), in which the text “Note that your recipient relies on you” is added to the bottom of the instructions of a dictator game. The author finds that adding this text increases dictators’ giving significantly. This text is similar to our first text about the responsibility of the dictator. Besides replicating their findings in this respect, our paper adds the

DecisionPower frame and a focus on gender differences with respect to the two different frames explained above. A recent study by Capraro et al. (2018) asks dictators what they think would be the right thing to do or what they think society would consider the right thing to do before making their decision. They find that such a question increases dictator’s giving significantly. Remotely related are two studies by Haley and Fessler (2005) and Rigdon et al. (2009), which also manipulate the decision environment in a dictator game by adding figures of “watching eyes” on the decision screen of the dictator and which also find that these manipulations increase dictators’ giving. Furthermore, it has been shown that communication between the dictator and the receiver affects giving positively and that the content of the communication matters (Andreoni and Rao, 2011; Charness and Rabin, 2005; Mohlin and Johannesson, 2008). Dal Bó and Dal Bó (2014) use a similar approach to ours in a different context, i.e., a repeated public good game. After a first sequence of rounds, they show participants one of five different messages. They find that moral appeals increase contributions significantly but transitory.⁴

The present study is closely related to the study in Chapter 1, in which the effect of free-form text messages written by the receiver and read by the dictator prior to a binary dictator decision is analyzed. These text messages are categorized into eight categories, one of which contains statements that “the dictator has the sole decision power and / or the sole responsibility for both payments.” Such messages increase dictators’ willingness to give if the dictator is a women. However, this study cannot separate the effects of messages referring to the dictator’s responsibility from those mentioning the dictator’s decision power. Furthermore, the present paper analyzes the effect of the two different statements in a more controlled setting by not using free-form messages of the recipient but different texts written by the experimenters.

Regarding the context-dependence of dictator giving, it has been shown, for example, that dictators’ giving decreases if anonymity of decisions is increased (Hoffman et al., 1994, 1996; Cherry et al., 2002; Franzen and Pointner, 2012)⁵, that the social distance between dictator and receiver affects dictators’ decision (Hoffman et al., 1996; Bohnet and Frey, 1999; Small and Loewenstein, 2003; Goeree et al., 2010; Charness and Gneezy, 2008), that giving increases if the receiver is

⁴These messages either reminded participants “that moral actions are those that treat others as you would like to be treated” or “that actions are moral to the extent that they contribute to maximizing collective payoffs.”

⁵However, also paying the dictator in public can lead to lower generosity (Dufwenberg and Muren, 2006).

perceived as more deserving of help (Eckel and Grossman, 1996; Fong, 2007), that giving decreases if the endowment is earned instead of given by the experimenters (Hoffman et al., 1994; Cherry et al., 2002; List, 2007), that giving decreases if the dictator has the additional option of taking money from the receiver (List, 2007; Bardsley, 2008), that giving decreases if the dictator can disguise her decision (Dana et al., 2007) and that giving increases if the receiver is referred to as a partner and that the joint endowment should be allocated between oneself and the partner (Banerjee and Chakravarty, 2012). There are mixed results regarding the question whether framing dictator’s decision as taking something from the receiver in contrast to the normally used frame of giving something to the receiver has an effect on dictator’s decision. While Dreber et al. (2013), Suvoy (2003) and Grossman and Eckel (2015) find no effect, Cox et al. (2016) and Korenok et al. (2014) as well as Korenok et al. (2017), who additionally let the respective party earn the endowment, find that receivers’ earnings are higher if dictators have to take funds from them.

Finally, a meta-analysis on gender differences in dictator games has shown that women give significantly more to receivers than men (Engel, 2011). However, this result seems to depend heavily on the specific design and context of the experiment (Croson and Gneezy, 2009). For example, Andreoni and Vesterlund (2001) find that women are more concerned with equality of payoffs, whereas men are more concerned with maximizing the sum of payoffs. Furthermore, female dictators are found to be more sensitive to receivers’ gender, i.e. women give less to women than to men (Ben-Ner et al., 2004). In a review of gender differences in the experimental economics literature, Croson and Gneezy (2009) state that they believe women react stronger to the social conditions of an experiment on social preferences such as the dictator game.

2.3 Experimental Design and Procedures

In the experiment, participants play either the role of a dictator (named participant A) or the one of a receiver (participant B). Each dictator has an endowment of 30 Euros⁶ and can decide how much of this endowment she wants to give to the receiver.⁷ Every integer between 0 and 30 is a valid decision. This dictator

⁶30 Euros \approx 35 Dollars at the time of the experiment.

⁷Only 10% of the participants were paid after the experiment took place as is explained in detail in the end of this Section. The instructions, however, stress that dictators should

game is played only once (one-shot).

The experiment consists of three different treatments in a between-subjects design. In two of these treatments (*Responsibility* and *DecisionPower*), a specific text is shown to the dictators before they make their decision while the third treatment serves as a control treatment without such a text (*Control*). The text in the treatment *Responsibility* emphasizes the dictator’s responsibility for the payoff of the receiver by saying “You bear the responsibility for both payoffs, your own and the one of participant B.”⁸ The text in the treatment *DecisionPower* emphasizes the dictator’s decision power and freedom of choice by saying “You alone decide about the payoffs, and you are completely free in your decision.”⁹ At all other instances, the instructions were neutrally formulated and carefully avoided mentioning the hierarchical relationship between the dictator and the recipient.

The experiment was conducted in December 2017 in two lecture rooms at the University of Potsdam using pen and paper. In one of these rooms, 207 students arrived and were all assigned the role of the dictator.¹⁰ The receivers, who were also students, were in another lecture room at the same time. The dictators were informed about this in the instructions.¹¹

The experiment took place before a first-year bachelor course in microeconomics. Participants for the role of the dictators were recruited via mailing invitations to students of this course as well as an announcement in this course one week prior to the experiment. Participants in the role of recipients were invited to participate immediately before the start of their lecture at the day of the experiment. They were told that they would obtain the money other participants gave them and that the money would be paid out to them immediately after the end of that lecture. No one in that room refused to participate.

Of the 207 dictators, 105 (51%) were women and 101 (49%) were men. One dic-

think carefully about their decision as it has real consequences if they are drawn for payment. Paying only a fraction of participants is not uncommon in experimental economics, especially in classroom settings (Erat and Gneezy, 2011) or in survey experiments (Exadaktylos et al., 2013).

⁸The original text in German was: “Sie tragen die Verantwortung für beide Auszahlungen, Ihre eigene und die von Teilnehmer/-in B”.

⁹The original text in German was: “Sie allein entscheiden über die Auszahlungen, und Sie sind völlig frei in Ihrer Entscheidung”.

¹⁰Prior to the experiment, we did a power calculation in which we estimated effect sizes by using the results from Brañas-Garza (2007). We concluded that we would need around 25 observations for each gender for each treatment to find significant effects. Thus, we intended to conduct the experiment with a minimum number of 150 dictators.

¹¹While it may potentially affect dictators’ decisions that receivers are not in the same room, we do not see a reason why such an effect should differ between the treatments.

Entscheidung über die Aufteilung der 30 Euro

Ich entscheide mich für die folgende Aufteilung:
(nur **ganzzahlige** Beträge zwischen 0 und 30 sind erlaubt)

Teilnehmer/-in B bekommt von mir: _____ Euro

Demnach behalte ich: _____ Euro

(Die Summe der beiden Beträge **muss** 30 Euro betragen.)

(a) *Control*

<div style="border: 1px solid black; padding: 5px;"> <p style="text-align: center;">Entscheidung über die Aufteilung der 30 Euro</p> <p style="border: 1px dashed red; padding: 2px;"><u>Bitte beachten Sie:</u> Sie tragen die Verantwortung für beide Auszahlungen, Ihre eigene und die von Teilnehmer/-in B.</p> <p>Ich entscheide mich für die folgende Aufteilung: (nur ganzzahlige Beträge zwischen 0 und 30 sind erlaubt)</p> <p style="text-align: center;">Teilnehmer/-in B bekommt von mir: _____ Euro</p> <p style="text-align: center;">Demnach behalte ich: _____ Euro</p> <p style="text-align: center;">(Die Summe der beiden Beträge muss 30 Euro betragen.)</p> </div>	<div style="border: 1px solid black; padding: 5px;"> <p style="text-align: center;">Entscheidung über die Aufteilung der 30 Euro</p> <p style="border: 1px dashed red; padding: 2px;"><u>Bitte beachten Sie:</u> Sie allein entscheiden über die Auszahlungen, und Sie sind völlig frei in Ihrer Entscheidung.</p> <p>Ich entscheide mich für die folgende Aufteilung: (nur ganzzahlige Beträge zwischen 0 und 30 sind erlaubt)</p> <p style="text-align: center;">Teilnehmer/-in B bekommt von mir: _____ Euro</p> <p style="text-align: center;">Demnach behalte ich: _____ Euro</p> <p style="text-align: center;">(Die Summe der beiden Beträge muss 30 Euro betragen.)</p> </div>
--	---

(b) *Responsibility*

(c) *DecisionPower*

Figure 2.1: Decision sheets (in the original, German language)

tator did not provide information on their gender. On average, participants were around 21 years old. The most frequent field of study was business administration (35% of the dictators), followed by economics (31%) and political sciences and economics (28%).

Before entering the room for the experiment, dictators received a randomly drawn card on which they found their assigned seat and their ID number. At their seat, participants found a large envelope with a sticker on it, which repeated the information from their card. Hence, they were randomly assigned to one of the three treatments. 56 dictators participated in the treatment *Control*, 77 in the treatment *Responsibility* and 74 in the treatment *DecisionPower*. Table 2.A.1 in the Appendix provides information about the distribution of participants' characteristics (gender and age) across treatments. The envelope at the participants'

seats contained general and specific instructions, one small envelope marked with the number 1, which contained the decision sheet, and one small envelope marked with the number 2, which contained the questionnaire.¹² Figure 2.1 shows the decision sheets for all three treatments. The additional texts on the decision sheets in the treatments *Responsibility* and *DecisionPower* are marked by a red dotted rectangle.

In the beginning of the experiment, a short introduction was read aloud to all participants by an experimenter, who was not the lecturer of the microeconomics course. Then, participants read the general and specific instructions by themselves.¹³ Afterwards, they opened the first envelope, made their decision and closed the envelope again. Next, they opened the second envelope, completed the questionnaire and closed the envelope again. Finally, they put everything back in the large envelope and sealed it. Subsequently, the envelopes were collected. Participants were informed in verbal and written form that only 1 in 10 participants would be randomly selected for payment after all envelopes were collected. This was done by letting one of the participants publicly roll a 10-sided die. Each participant whose last digit of their ID number was the number diced was randomly matched with one receiver in the second room and received their payment subsequently. The participants in the second room randomly drew a card with a number on it to determine whether they would receive a payment. Each dictator chosen by the roll of the die was matched randomly to one of those receivers. Receivers who were not matched to one of the dictators did not receive a payment.¹⁴ In total, 21 dictators were randomly matched to 21 receivers. Thus, 42 participants received a payment after the experiment.

2.4 Results

Participants in the role of a dictator on average give 40.2% of their endowment of 30 Euros to the receiver, i.e. 12.05 Euros. Dictators in the treatment *Control* give 36.8%, dictators in the treatment *Responsibility* give 43.0% and dictators in the treatment *DecisionPower* give 39.7% of their endowment. Averaged over all

¹²Pictures of the material used for the experiment can be found in the Appendix.

¹³An English translation of the originally German verbal instructions, general instructions and specific instructions can be found in the Appendix.

¹⁴There were fewer participants in the second room than in the first room. Participants in the role of the dictator who sat in the first room received no information about the number of participants present in the second room.

2.4. RESULTS

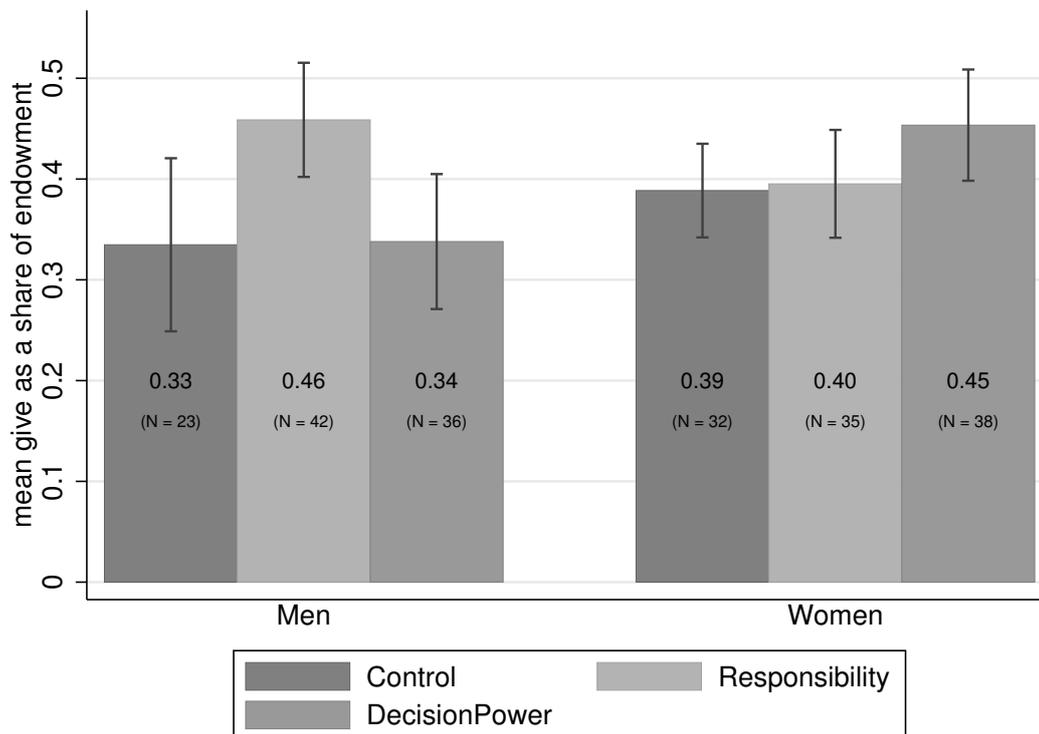


Figure 2.2: Mean giving separated by treatment and gender

Note: The figure shows means as well as 95% confidence intervals.

treatments, women are slightly more generous by giving 41.4% compared to men giving 38.7% of their endowment. This difference, however, is not statistically significant (Wilcoxon Rank Sum test: $z = 0.210$, $p = 0.8340$).

Figure 2.2 shows average dictators' decisions as a share of endowment separated by treatment and gender. On average, men give 33.5% in the *Control* treatment, 45.9% in the *Responsibility* treatment and 33.8% of their endowment in the *DecisionPower* treatment. Female dictators seem to react completely different than men to the two frames. On average, they give 38.9% in the *Control* treatment, 39.5% in the *Responsibility* treatment and 45.4% of their endowment in the *DecisionPower* treatment.

Since participants' characteristics are not perfectly balanced between treatments (see Table 2.A.1 in the Appendix),¹⁵ we use OLS regressions to test for treatment and gender differences.¹⁶ Table 2.1 presents the regression results. The first

¹⁵For example, there are more men in the treatment *Responsibility* than in the treatment *Control*, who, on average, give slightly less than women. Hence, simply comparing the means may understate a potential effect.

¹⁶A Tobit model may be more appropriate for the analysis of the dictators' decision if left-censoring in the dependent variable is a concern. (See Brañas-Garza et al. (2017) or Engel

2.4. RESULTS

Table 2.1: OLS regression results

VARIABLES	Full Sample (1)	Men (2)	Women (3)
<i>Responsibility</i>	0.063 ** (0.031)	0.125 ** (0.050)	−0.002 (0.037)
<i>DecisionPower</i>	0.030 (0.031)	−0.003 (0.052)	0.068 * (0.037)
male	−0.028 (0.025)		
age	−0.003 (0.004)	−0.003 (0.008)	−0.001 (0.005)
Constant	0.429 *** (0.099)	0.418 ** (0.189)	0.402 *** (0.106)
Observations	205	100	105
R-squared	0.100	0.171	0.136

Note: The dependent variable in all regressions is the share of the endowment given by the dictator. The regression depicted in column (2) only includes observations by male dictators, the one in column (3) only observations by female dictators. All regressions additionally control for participants' field of study.

* significant at 10%; ** significant at 5%; *** significant at 1%.

column uses the full sample, the second and third column present the results for male and female participants separately.¹⁷

The regression results for the full sample in column (1) show that the text emphasizing the dictator's responsibility has a positive and statistically significant effect on dictators' giving compared to the *Control* treatment. Adding a text that highlights the dictator's freedom of choice and decision power is also estimated to increase giving. However, this effect is not statistically significant ($p = 0.336$). In line with the descriptive analysis in Figure 2.2, columns (2) and (3) in Table 2.1

(2011) for good discussions of the appropriateness of different models to analyze dictator game decisions.) However, only 7% of the dictators decided to give 0 Euro in our study, which is a very low fraction compared to other studies (A meta-analysis found 36% of dictators to give nothing to the recipient (Engel, 2011), while in a survey-experiment, which also paid only 1 in 10 participants, around 17% of dictators decided to give 0 Euro (Exadakylos et al., 2013)). Therefore, we do not believe that censoring from below is an issue here and decided to use an OLS model due to the slightly easier interpretation of coefficients. Nevertheless, all of the results in Tables 2.1 and 2.A.2 are robust to using a Tobit model.

¹⁷Table 2.A.2 in the Appendix additionally shows the results of an OLS regression including interaction effects of gender with the treatment variables.

show that the positive effect of the *Responsibility* treatment on giving is entirely driven by male dictators ($p = 0.015$). For female dictators, the effect of the *Responsibility* treatment is close to zero ($p = 0.955$). On the contrary, the effect of the *DecisionPower* treatment is estimated to be close to zero for men ($p = 0.946$), while the results suggest a positive effect for women, which is significant at the 10% level ($p = 0.068$).

2.5 Discussion and Conclusion

In this study, we analyzed the effects of two different short texts, which were printed on the decision sheet, on giving in a dictator game. The *Responsibility* treatment reminded dictators of their responsibility for the recipient's payoff, the *DecisionPower* treatment highlighted the dictator's freedom of decision. Furthermore, we analyzed differences in behavior depending on dictator's gender.

We expected the *Responsibility* treatment to increase giving (because of the increased salience of responsibility), and more so for females than for males (because women are presumably under stronger social pressure to behave pro-socially (Deaux and Major, 1987; Babcock and Laschever, 2003; Eagly, 2009; Heilman and Chen, 2005; de Wit and Bekkers, 2016)). In our data, however, the *Responsibility* treatment increases giving by male dictators only. One possible explanation for the lack of an effect when reminding women that they are responsible for the payoff (and, thus, for the well-being) of the receiver is that this text does not provide them with new information; they already feel responsible for the receiver due to the general expectation that females care for others who depend on them (Deaux and Major, 1987; Babcock and Laschever, 2003; Eagly, 2009; Heilman and Chen, 2005). Hence, our treatment text would not increase the feelings of responsibility and the pressure to behave altruistically anymore. For men, however, who probably feel a much lower pressure to help the receiver and a lower level of responsibility for the well-being of the receiver initially, this text may be able to increase the feelings of responsibility and, thus, the pressure to help. This could then lead to the observed effect that men increase their giving after seeing the responsibility text.

With respect to the *DecisionPower* treatment, our prediction was less clear. If it reminds participants of their accountability for their action by highlighting their control over the decision, it may have increased giving compared to the *Control* treatment. On the contrary, if participants perceived the text in this treatment

as a justification for selfish behavior, it may even have reduced giving compared to the *Control* treatment. Indeed, the insignificant overall effect we find may reflect that these behavioral forces counteract each other. Future research should try to disentangle these mechanisms. The weakly significant, positive effect for female dictators may be explained by the explicit reduction of social pressure to which female participants seem to respond positively. Presumably, they appreciate the reduction of the general pressure they usually face to behave altruistically. If future research should indeed confirm that women are more willing to do somebody a favor or to comply with a request when they are reminded of their freedom of choice than when they are told that they are responsible for the potential beneficiary of a request, this could have important implications. Besides direct applications such as everyday life situations of asking somebody a favor or charity organizations asking somebody for money, these insights may also be applicable to leadership strategies in the workplace: it may be that female employees' compliance with a request could be increased by formulating such a request in a way as to highlight their freedom of choice and, thereby, presumably decrease the pressure to comply.

Our study contributes to the literature in several ways. The first contribution can be seen as a series of robustness checks or extensions: it replicates the findings of Brañas-Garza (2007) that reminding the dictator that the other party depends on her has a significantly positive effect on dictators' giving, using a different version of the text for the frame.¹⁸ Furthermore, this paper provides additional evidence of framing effects in dictator games (Fehr and Schmidt, 2006; Gerlach and Jaeger, 2016). In particular, we contribute evidence of how these frames are sensitive to the dictator's gender. More generally speaking, we thereby contribute to the literature on gender differences in the field of social preferences (Croson and Gneezy, 2009).

As another, probably more important contribution, the present paper provides novel insights into what has driven dictators' reaction to pre-play messages sent by the recipient in Chapter 1. The results of Chapter 1 show that only female dictators react positively to messages emphasizing the dictator's power position. The dataset did not allow differentiating between messages emphasizing freedom of choice and those stressing the dictator's responsibility so that only a joint positive effect of these types of messages could be reported. If our result regarding the weakly positive effect of highlighting the freedom of choice for female dictators is confirmed by future research, it will suggest that the aspect of the freedom

¹⁸Brañas-Garza (2007) does not present results depending on gender.

2.5. DISCUSSION AND CONCLUSION

of choice rather than the responsibility argument is responsible for this gender difference in Chapter 1. The relative weakness of this effect in the present study furthermore indicates that text messages sent by the recipient to the dictator have a stronger impact on giving decisions than a framing text written by the experimenters. However, particularly this last finding needs further substantiation by future research.

2.A Appendix

2.A.1 Supplementary Tables and Figures

Table 2.A.1: Descriptive statistics of the distribution of participants

	Treatments		
	<i>Control</i>	<i>Responsibility</i>	<i>DecisionPower</i>
male	0.42 (a)	0.55	0.49
age	21.79	21.55	20.88
N	56	77	74

(a) One participant in the treatment *Control* did not provide information on their gender.

Table 2.A.2: OLS regression results with gender interaction effects

VARIABLES	Give
<i>Responsibility</i>	-0.004 (0.041)
<i>DecisionPower</i>	0.066 (0.041)
<i>Responsibility</i> x male	0.128** (0.061)
<i>DecisionPower</i> x male	-0.066 (0.061)
male	-0.050 (0.047)
age	-0.002 (0.004)
Constant	0.419*** (0.098)
Observations	205
R-squared	0.155

Note: The dependent variable in the regression is the share of the endowment given by the dictator. The regression controls for participants' field of study.

* significant at 10%; ** significant at 5%; *** significant at 1%.

2.A.2 Instructions

The following section provides the English translation of the original German instructions, which participants playing the role of the dictator received in the experiment:

Verbal Instructions / Welcome Address

Hello, my name is Florian Stolley and I am a research assistant at the University of Potsdam. Today, you are participating in an experiment in which you make a simple decision. Please think carefully about this decision as it has real consequences.

By participating in this experiment, you support our research. Additionally, you can earn money. Following the experiment, a ten-sided die will be thrown. If the last digit of your participant ID - the so-called payoff number - is thrown, you will receive a payment immediately after the lecture. The amount of the payment depends on your decision in this experiment. You will find more details about this in the instructions.

In front of you, you find a large envelope. This envelope contains:

1. the instructions.
2. a small envelope with the number 1 on it.
3. a small envelope with the number 2 on it.

Please start by reading the instructions carefully. Please open envelope 1 and make a decision only **after** reading the instructions. Afterwards, please open envelope 2 and fill in a questionnaire.

We ask you to be quiet and to not talk to each other during the experiment.

You can start now by opening the large envelope and reading the instructions carefully.

Welcome to this experiment!

Thank you very much for participating in this experiment!

Thereby, you support our research and can earn money in addition.

- Please do not open the two small envelopes yet.
- Please read the instructions carefully and follow the instructions.
- Please stay quietly in your seat for the entire duration of the experiment and do not talk to other participants.
- Please refrain from using your cell phone during the experiment.
- Please do not look at other participants' sheets and do not let other participants look at yours.
- Please keep your place card until the end of the lecture. Only with this place card, you are able to receive a payment.

Use of data:

Following today's experiment, the data will be analyzed anonymously. If you are chosen for payment by the throw of the die, you will have to acknowledge the amount received with your name and signature. These data are only used for the financial accounting of this experiment. Your name will not be linked to your decision or to your answers on the questionnaire. You can abort your participation in this experiment at any time; however, this would then result in you being ineligible for payment.

Instructions

In the experiment, you will make a decision. After all participants in this room are finished with the experiment, it is going to be decided by chance (a public throw of a die) who of you receives a payment. The probability to receive a payment is 1 to 10 or 10%.

The amount of money you receive, if you are chosen by the throw of the die, depends on your decision in the experiment. Thus, your decision has real consequences. You should think carefully about your decision.

In the experiment, you and the other participants each assume a role. There are two different roles which will be called A and B from now on. You and the other participants in this room will all assume the role of “A.” The participants, who will assume the role of “B,” are in a parallel lecture, which takes place right now in another room on the Campus Griebnitzsee.

Should you be chosen by the throw of the die, you will be matched randomly with a participant B. You will not learn the identity of this person, neither before nor after the experiment. The participant B you were matched to will not learn your identity as well.

Decision:

You, as “A,” receive an endowment of 30 Euros. The “B”s do not receive any endowment. You can decide on how much of your endowment you want to give to the participant B you are matched to. You can choose any integer amount between 0 and 30 Euros.

Throw of the die:

The last digit of your participant ID, which you can find, for example, on your place card or on your large envelope, is the payoff-number. The payoff-number is a number between 0 and 9. Following the experiment, a ten-sided die will be thrown here in the lecture room. The number thrown decides who of you receives a payment.

Payment:

After this lecture, the participants who have been chosen by the throw of the die for payment will receive their payment. The amount paid out depends on the individual decisions by the participants.

Please keep your place card until the end of the lecture because **only with the place card, you are able to receive a payment.**

Summary:

- You are a participant A. The participants B are in another room in this building at this very moment.
- You decide on how much of your endowment of 30 Euros you want to give to the participant B you are matched to (if your payoff-number should be chosen by the throw of the die).
- You will receive your payment at the end of the lecture for handing in your place card (if your payoff-number should be chosen by the throw of the die).

Single steps:

1. Open the envelope with the number 1 and note your decision on the inlying decision sheet.
2. Fold the decision sheet and put it back into the envelope number 1.
3. Open the envelope with the number 2 and fill in the inlying questionnaire.
4. Fold the questionnaire and put it back into the envelope number 2.
5. Put envelope number 1, envelope number 2 and the instructions into the large envelope and seal it.
6. Please wait quietly until all other participants are finished and all envelopes have been collected.

2.A.3 Pictures of the Material

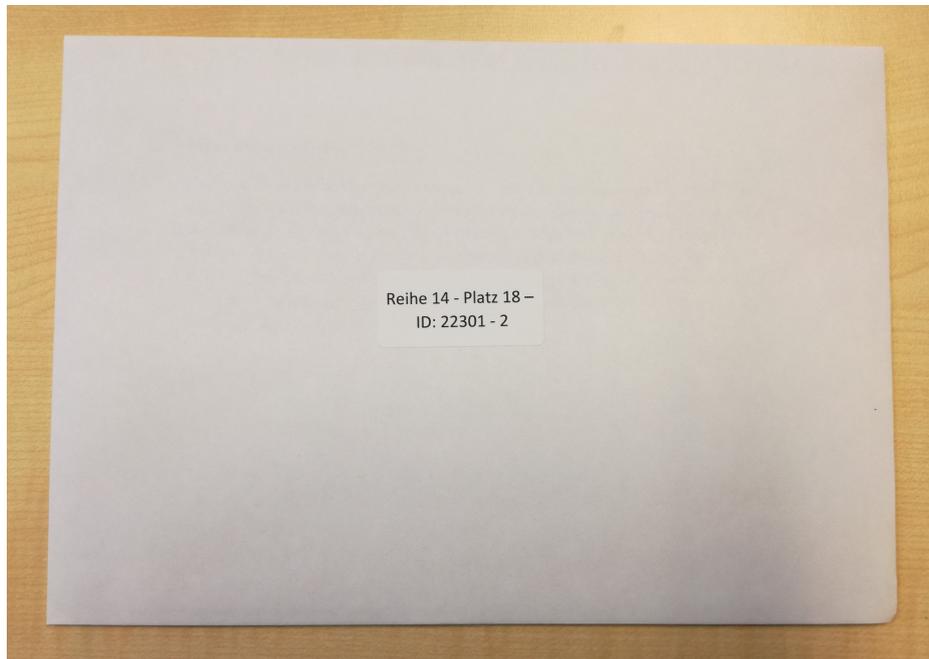


Figure 2.A.1: Large envelope with a sticker with seat and id on it

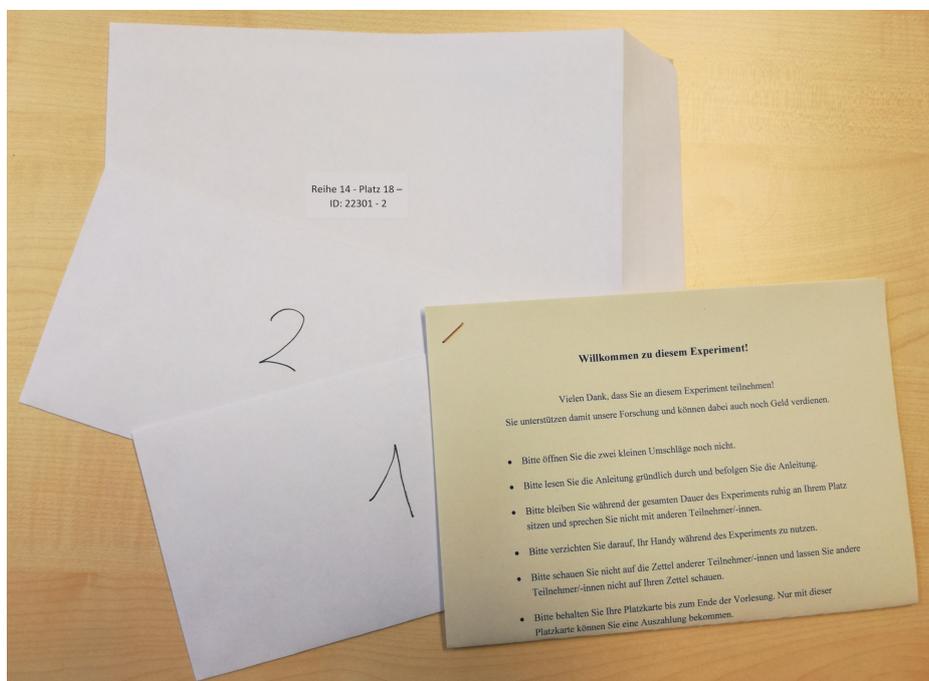


Figure 2.A.2: Content of the envelope - instructions and two smaller envelopes

2.A. APPENDIX

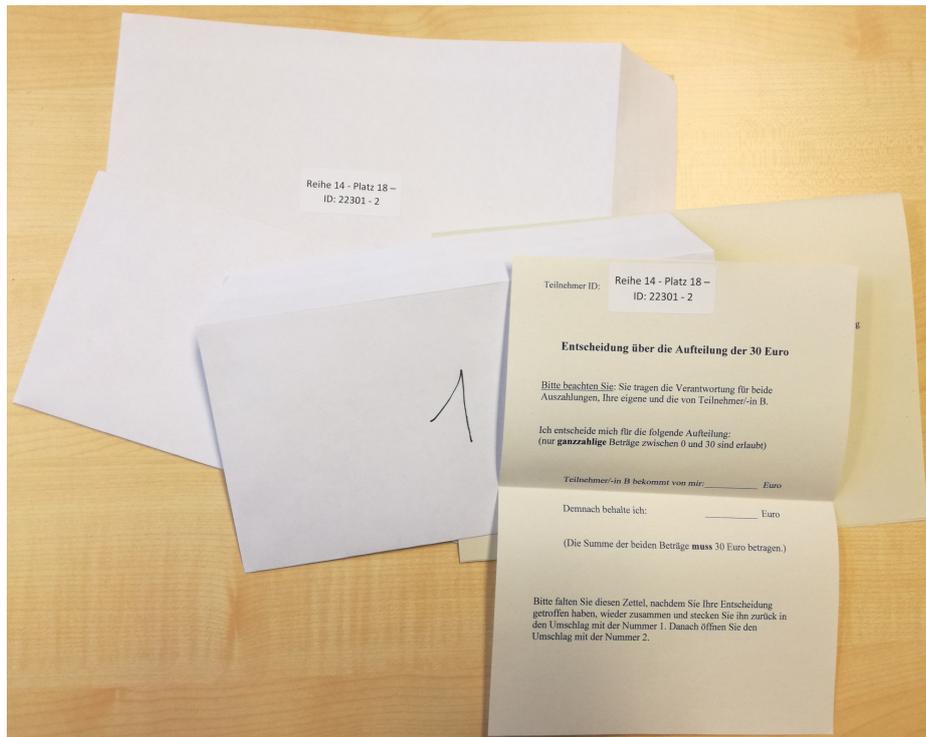


Figure 2.A.3: Content of the small envelope marked with a 1 - decision sheet

Chapter 3

“Thanks in Advance” - the Effect of a Polite but Empty Phrase on Compliance Behavior

3.1 Introduction

In social or market interactions there are many situations in which one party asks another one a favor that the latter is free to refuse. Therefore, the asking party is in need of means to achieve compliance with its request. Such challenges can be found in marketing applications,¹ they can be important to policy makers or government,² overcoming them is an essential part of the work of charitable organizations,³ and they are faced by people making altruistic requests.⁴ Closely related are principal-agent problems in the workplace.⁵

In this paper, we study the effect of a widely used phrase - “thanks in advance” - on the compliance with a request. This phrase is often used as a closing in written communication such as letters and emails when asking the recipient for her or his help. We focus on situations in which the phrase is frequently used:

¹E.g. newspapers which want to increase their subscription numbers (Scott, 1976).

²E.g. for social marketing purposes (Brennan and Binney, 2010; Supphellen and Nelson, 2001) such as the advancement of energy savings (Katzev and Johnson, 1984).

³Bell et al. (1994), Das et al. (2008), Andreoni et al. (2017), Sanders and Smith (2016), Supphellen and Nelson (2001).

⁴Andreoni and Rao (2011), Mohlin and Johannesson (2008), Althoff et al. (2014).

⁵Although employees are typically not in the position to refuse a request for carrying out a task, they often have substantial leeway how much effort to spend on carrying out the task (Eisenhardt, 1989).

asking for favors which can be done in just a few minutes and which are asked in a rather impersonal manner. Examples include companies sending form letters asking their clients to fill in questionnaires about their products, political activists asking people online or via mail to sign a petition, and charity organizations sending solicitation letters by mail or email to potential donors. In this study, we will use the phrase in a context in which we, as researchers, ask participants to fill in a questionnaire in detail.

How do recipients react to such a phrase? On the one hand, it may be that recipients increase compliance, because they like that gratitude is already promised and, thus, not uncertain (Norrick, 1978). Moreover, the phrase may cause a feeling of guilt on the side of the one being asked, if she or he does not grant the favor (Norrick, 1978). Hence, guilt averse people may be more inclined to grant the favor (Charness and Dufwenberg, 2006; Battigalli and Dufwenberg, 2007).

On the other hand, the recipient of a request in which the phrase is used, may feel like the asker is already certain that the favor will be granted (Mizutani and Mizutani, 1989; Ohashi, 2006). This may be perceived as a reduction of the freedom to decide whether or not to grant the favor, which may negatively affect the way in which the recipient complies with the request (Falk and Kosfeld, 2006). If so, asking someone a favor using “thanks in advance” may even be perceived as a command rather than a request. Furthermore, it may also decrease compliance if the recipient sees it as an empty phrase that does not convey “real” gratitude but rather a cheap imitate of gratitude.

Moreover, it may be that men and women react differently to the phrase “thanks in advance”. Results of Chapter 1 suggest that male and female dictators react differently to the use of the word “thanks” in a message in terms of giving. Additionally, women have been shown to react more strongly to the context of an experiment in general (Croson and Gneezy, 2009; Ellingsen et al., 2013; Kahn et al., 1971; Eagly, 1983). Hence, we focus explicitly on analyzing potential gender differences in the response to the phrase.

In order to study the effect of the phrase “thanks in advance” on compliance, we conduct a laboratory experiment in which we ask participants a favor, i.e. telling us in as much detail as possible about their strategy in a separate task of the experiment performed beforehand. Two treatments are conducted in a between-subjects design. In the main treatment, the phrase “thanks in advance” is displayed as a closing of the request while no phrase is used in the *Control* treatment. Participants’ compliance with the request is measured by the time participants spend answering the request as well as the length of the answer.

Participants are allowed to leave the experiment as soon as they are finished. Hence, answering in more detail is associated with opportunity costs in terms of time.

We find that participants spend significantly less effort on answering the request when being thanked in advance. This effect is mainly driven by women who not only spend less time answering, they also tend to write shorter answers when seeing the phrase. Interestingly, this also holds true for women who indicate that “thanks in advance” is a polite closing in an email. We do not find a statistically significant reaction of men, but the qualitative treatment differences are similar to those of women.

3.2 Related Literature

Several studies analyzed the effects of a request’s language on compliance. In Chapter 1 the effect of free-form text messages on giving in a binary dictator game is analyzed. The messages are written by the receiver and read by the dictator prior to making her or his decision. Messages containing the word “thanks” in any variant, which is mostly used as a closing in these messages, affect the probability for a successful request differently depending on the dictator’s gender. Female dictators react positively to these messages whereas male dictators show no significant reaction. Althoff et al. (2014) study factors that influence the compliance with a written request in an online community. They find that gratitude expressed in advance increases the probability of the request being successful. Mitra and Gilbert (2014) find that an expressed intention of reciprocal behavior by the receiver of a favor increases compliance. However, the magnitude of the announced reciprocal behavior matters: Too small offers can have no or a negative effect on compliance. Hence, even if “thanks in advance” is perceived as an act of politeness or as a sincere expression of gratitude, which participants usually would react to reciprocally, it could be too small to influence behavior.

Goei et al. (2003) find that an initial favor by a person A towards a person B has a positive effect on B’s liking of A, which, in turn, has a positive effect on compliance with a request by A towards B, that is made after the initial favor. It seems plausible that this effect on compliance can be transferred to other actions than an initial favor by A, if these have an effect on liking. Hence, a positive or negative effect of “thanks in advance” on compliance could be explained by a positive or negative effect of the phrase on participants’ liking of the person

asking the favor. Thus, this argument as well is based on reciprocal behavior on the part of the one being asked the favor.

Closely related to the present paper are linguistic studies analyzing cultural differences in the perception of "thanks in advance." Lee et al. (2012) find differences in the use and perception of "thanks in advance" between the United States and Japan. This phrase is commonly used in the United States, although more as a neutral phrase than as an actual expression of gratitude. In Japan, it is very uncommon to use the phrase when asking someone a favor. Thanking someone in advance is viewed as extremely impolite in Japan, because the asking party assumes that the request will be fulfilled (Mizutani and Mizutani, 1989; Ohashi, 2006). Economidou-Kogetsidis (2015) asks English native speaking faculty members to rate the politeness of a pool of emails that were written by Greek students in English. An email containing "Thanking you in advance" was rated to be more or equally polite than all other emails. However, this effect cannot be causally attributed to the use of "Thanking you in advance," as the emails also differ considerably in other dimensions, e.g. the greeting. In our paper, we provide direct and causal evidence for a potential effect of the phrase "thanks in advance" on compliance behavior, as we are able to observe a partial effect of that phrase through the experimental nature of our study. Additionally, we explicitly ask the participants to rate the politeness of "thanks in advance." If participants react to politeness reciprocally and find "thanks in advance" to be a polite expression, one would assume that participants increase their compliance and vice versa if they find it to be rude. We add to the literature analyzing cultural differences by using a pool of German students.

3.3 Experimental Design and Procedures

The experiment is very short and consists of two parts. In the first part, participants perform a 5x5-version of the bomb risk elicitation task (Crosetto and Filippin, 2013). In the second part, participants answer a short questionnaire, which includes an open question where they are asked to explain their behavior in the bomb risk elicitation task.

In the bomb risk elicitation task in the first part, participants collect packages. Each collected package is worth 0.50 Euro. There are 25 packages which are numbered consecutively from 1 to 25. In one of the packages, a bomb is hidden. If participants collect this package, all packages collected will be destroyed and

participants will earn 0 Euro for this task. On their screen, participants see a rectangle with 25 buttons labeled “package 1” to “package 25”. They can decide which individual package to collect and how many packages they wish to collect.⁶ Only after finishing the questionnaire, each participant picks a note with a number between 1 and 25 on it. The picked number determines the package in which the bomb is hidden for this particular participant. This task is used, because it is short and it leaves participants with sufficient substance to write more than a few words about their decision making process. For example, they can explain whether they chose packages according to a special pattern or lucky numbers, or they may elaborate on their risk preferences.

In the questionnaire (after finishing the collection of packages and before picking the number of the “bomb” package), participants are asked to explain their strategy in the bomb risk elicitation task in as much detail as possible.⁷ The experiment consists of two treatments in a between-subjects design. In the main treatment, the phrase “thanks in advance” is displayed directly below the request.⁸ In the *Control* treatment, no such additional text is displayed.⁹ Participants are told explicitly that they can leave the experiment as soon as they have completed the questionnaire and that they will individually receive their payment immediately afterwards. Hence, complying with the request to answer in as much detail as possible is associated with noticeable opportunity costs.¹⁰

12 sessions were conducted at the Potsdam Laboratory for Economic Experiments (PLEx) in May 2018, divided into 6 sessions for each treatment. The number of participants per session varied between 9 and 20. In total, 179 students participated in the experiment, 92 in the main treatment and 87 in the *Control* treatment. In the main treatment, 57 participants were female (62%) while in the *Control* treatment, 59 participants were female (68%).

In each session, participants were randomly assigned a seat. At the beginning,

⁶Participants are allowed to collect between 1 and 24 packages. A screenshot of the bomb risk elicitation task can be found in the Appendix.

⁷Translated to English, the request is as follows: “Please explain in as much detail as possible how you decided in the previous part which packages to collect.” The exact wording in the original, German language is: “Erläutern Sie bitte möglichst ausführlich, wie Sie im vorherigen Teil entschieden haben, welche Päckchen Sie einsammeln.”

⁸The exact wording in the original, German language is: “Vielen Dank im Voraus.”

⁹Screenshots of the two questionnaire screens, one for the main and one for the *Control* treatment can be found in the Appendix.

¹⁰Participants spend on average around 20 percent (around 3 minutes) of their time in the laboratory on answering this request. Participants are allowed to provide no answer. To make sure that participants do not skip the request by accident, they are asked whether they are done answering when leaving the request screen.

they read the instructions for the bomb risk elicitation task on screen¹¹ and, subsequently, carried out the task. Then, participants answered the questionnaire. After answering the strategy request (with or without the “thanks in advance” text), participants were asked to rate the politeness of the closing “thanks in advance” and two other closings,¹² which were displayed in random order, on a 7-point Likert scale. After finishing the questionnaire, participants were told to go to the experimenters. They drew a number which determined the number of the package containing the “bomb” (and thus their individual payoff) and received their payment in cash. Sessions lasted for approximately 15 minutes and participants earned on average 7.60 Euros,¹³ including a show-up fee of 5 Euros. The experiment was computerized with z-Tree (Fischbacher, 2007) and the recruitment process was conducted using ORSEE (Greiner, 2015).

3.4 Results

We use two proxy variables to estimate the extent to which participants comply with our request: the time spent on answering the request in seconds¹⁴ as well as the length of their explanation in characters.¹⁵

On average, participants in the *Control* treatment type an answer which is 330 characters long and spend 185 seconds on doing that.¹⁶ Participants in the *ThanksInAdvance* treatment type on average 295 characters and spend 152 seconds on doing that. The difference in time spent is statistically significant at the 5% level (two-sided t-test: $p = 0.0196$). Hence, participants who are thanked

¹¹An English translation of the originally German instructions can be found in the Appendix.

¹²These other closings, translated into English, are: “I would be delighted to receive a response.” and “I expect your answer.” We asked participants about these other closings to avoid making it too obvious, what we were interested in.

¹³7.60 Euros \approx 8.85 Dollars at the time of the experiment.

¹⁴To be precise, we measure the time from when the screen with the strategy request is initially displayed to exiting this screen.

¹⁵Three participants chose not to provide an answer, all of whom were women, two of whom were in the *ThanksInAdvance* treatment.

¹⁶In their answer, participants often discussed their level of risk aversion or the probability of hitting the bomb. Many wrote that they chose numbers which had special meaning to them or, on the contrary, numbers that were completely random. Some wanted to achieve or avoid a certain pattern, and a few had a certain payoff goal. Typical examples are “I was thinking about collecting 50% of the packages to keep the risk of hitting the bomb about the same as the chance of not hitting the bomb. I chose the packages at random.” or “I decided to collect 10 packages, because that would lead to 5 Euros and the probability of success is still more than 50 percent. When selecting the packages, I started with my favorite numbers and then selected random numbers.” Thus, it seems that participants took the request seriously.

3.4. RESULTS

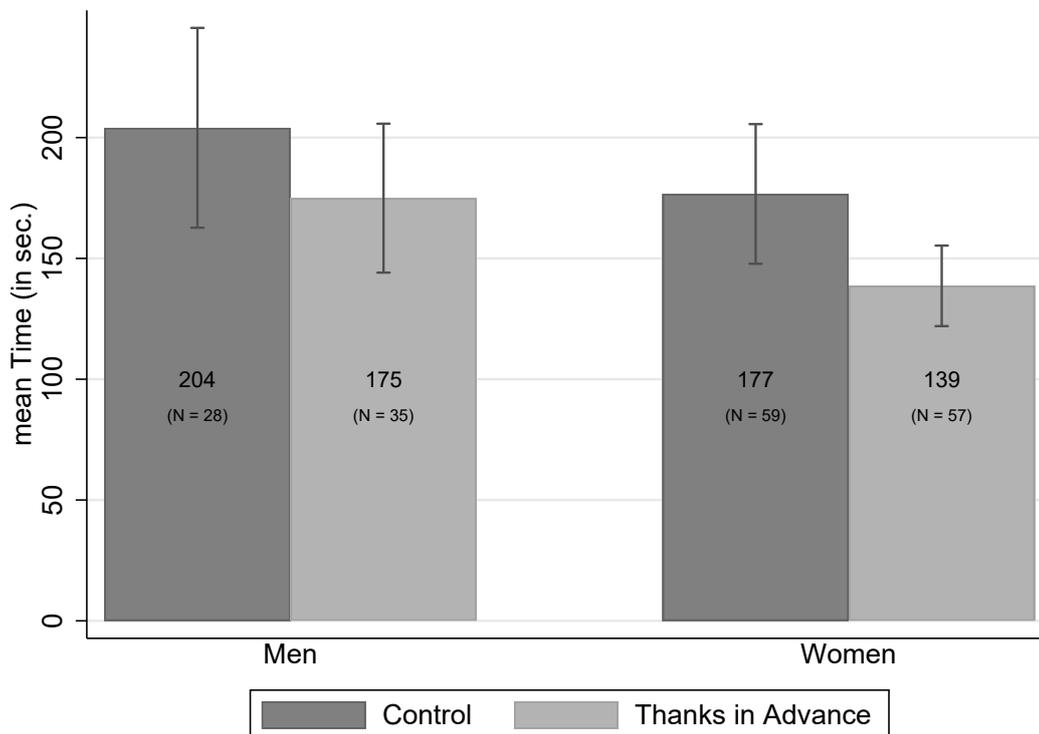


Figure 3.1: Mean time separated by treatment and gender

Note: The figure shows means as well as 95% confidence intervals.

in advance spend less time answering our request. The difference in the length of the answers between the two treatments is not statistically different from 0 (two-sided t-test: $p = 0.1659$).

Figure 3.1 shows the average time participants spend on answering our request, separated by treatment and gender. Women in the *Control* treatment spend an average of 177 seconds on answering while women in the *ThanksInAdvance* treatment spend only 139 seconds. This difference is statistically significant (two-sided t-test: $p = 0.0256$). The reaction of men to the treatment seems to go in the same direction. They spend on average 204 seconds on answering in the *Control* treatment and only 175 seconds in the *ThanksInAdvance* treatment. However, this difference is not statistically significant (two-sided t-test: $p = 0.2437$).

Figure 3.2 shows the average length of participants' answers to our request in characters, separated by treatment and gender. In the *ThanksInAdvance* treatment, women use an average of 275 characters while they use an average of 329 characters in the *Control* treatment. This difference is, however, only statistically significant at the 10% level (two-sided t-test: $p = 0.0923$). There is barely a difference in the average length of men's answers between the two treatments.

3.4. RESULTS

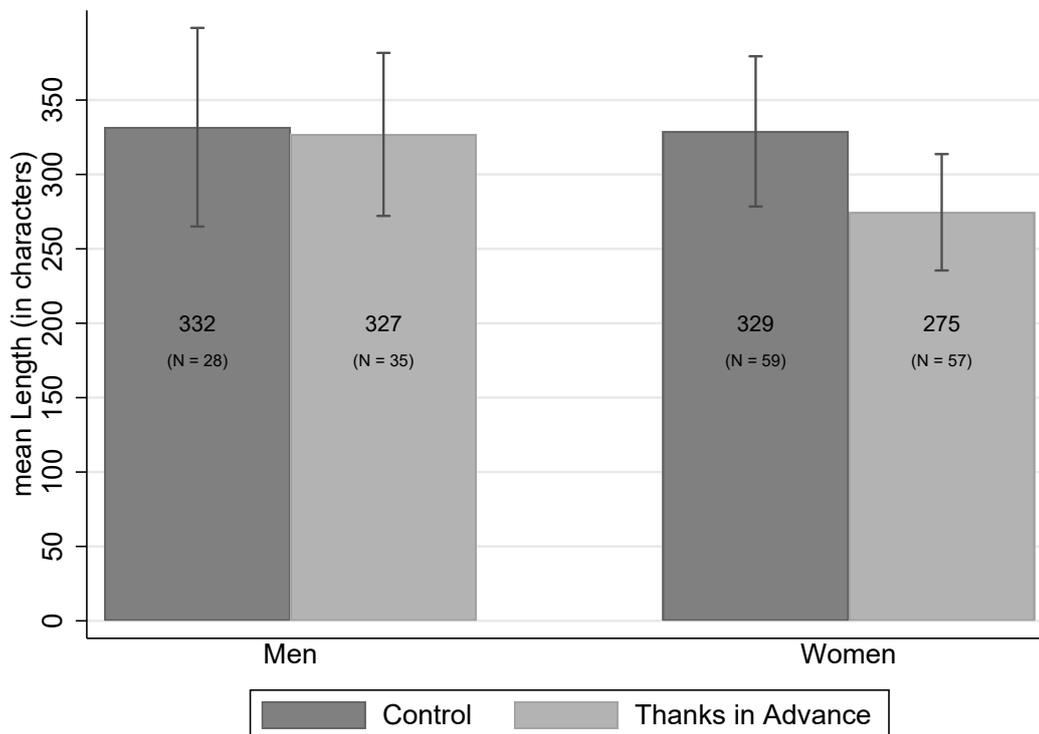


Figure 3.2: Mean length separated by treatment and gender

Note: The figure shows means as well as 95% confidence intervals.

On average, men’s answers are 332 characters long in the *Control* treatment and 327 characters long in the *ThanksInAdvance* treatment. This difference is not statistically significant (two-sided t-test: $p = 0.9085$).

To test whether the perceived politeness of the phrase “thanks in advance” explains the results, we additionally split the sample by the median politeness rating of “thanks in advance” in the questionnaire and conduct the tests only for those, who rate “thanks in advance” at or above the median, i.e. the ones who believe “thanks in advance” to be polite.¹⁷ Women, who perceive the phrase to be relatively polite, tend to spend less time answering (two-sided t-test: $p = 0.0516$) and tend to write shorter answers (two-sided t-test: $p = 0.0697$) when being exposed to it, even though the number of observations decreases due to the median split. Men, who rate “thanks in advance” at or above the median, do not react statistically significantly to the treatment (two-sided t-tests: time: $p = 0.5591$, length: $p = 0.4754$). Women and men, who rate “thanks in advance” below the median,

¹⁷The median politeness rating on a Likert scale from one to seven is 6, the mean is 5.73. Thus, “thanks in advance” is generally viewed as a very polite phrase by our participants. The test results are robust to only including those participants assigning “thanks in advance” an above-median rating, i.e. a rating of 7.

show similar reactions than those who perceive the phrase to be relatively polite. However, none of the effects are statistically significant.

3.5 Discussion and Conclusion

In this study, we analyzed the effect of the text “thanks in advance” on compliance with a request in a sample of German students. In a controlled laboratory experiment, we conducted two treatments in a between-subjects design, in which we asked participants a favor they were free to refuse¹⁸ and that was associated with opportunity costs. The text “thanks in advance” was added to the request in the main treatment, whereas this text was not displayed in the *Control* treatment. As proxies for the effort participants exert in order to comply with our request, we measured the time participants spent on answering the request and the length of their answers.

We find that the text “thanks in advance” has a negative effect on compliance with our request: Participants, who are thanked in advance, spend significantly less time answering our request than participants, who are in the *Control* treatment. This effect is mainly driven by women. Women also tend to write shorter answers in response to the treatment.

Surprisingly, even those women, who rated “thanks in advance” as (very) polite, seem to react negatively to the phrase. Although they see “thanks in advance” as polite, they might also see it as insincere and therefore unfriendly, which they react to reciprocally (Goei et al., 2003), i.e. negatively by putting less effort into compliance. This possible differentiation between the effects and causes of politeness vs. the effects and causes of friendliness leaves space for future research. Furthermore, it may be that participants perceive the closing “thanks in advance” as a reduction of their autonomy to decide whether to comply with the request (Norrick, 1978; Mizutani and Mizutani, 1989). In Chapter 2, the effect of two different texts by the experimenters on dictators’ giving is analyzed. The results suggest that female dictators increase their giving when reminded of their freedom of choice. If it is indeed the case that participants perceive the closing “thanks in advance” as a reduction of their autonomy to decide whether to comply with

¹⁸Participants were asked to explain to us in detail the strategy they used in a previous part of the experiment. Thus, the request is comparable in extent to e.g. answering a small questionnaire or looking up a small information. Participants took the request seriously as can be seen from their answers. See footnote 16 for examples.

a request, the results of both chapters concerning gender differences complement each other well: It seems that women are more likely to do something for someone (giving in a dictator game) when their freedom of choice is highlighted while being less likely to do something for someone (complying with a request) when they feel restricted in their freedom of choice. Hence, the perceived reduction of the autonomy to decide may explain the stronger negative reaction by female participants to the phrase “thanks in advance.” However, this interpretation of the results needs further substantiation.

The results of this chapter seem to contradict those of Chapter 1, as the results of the first chapter suggest that women are more willing to give to a recipient in response to reading a message which contains the word “thanks.” However, the experimental setting in the two chapters differs considerably. Most importantly, the effect of personal free-form text messages written by fellow students is studied in Chapter 1, whereas the effect of only one impersonal wording, written by the experimenters, is studied in the present chapter. Thus, in contrast to the first chapter, there is no personal and direct communication between human participants. Moreover, the first chapter uses a dictator game setting while this chapter analyzes the effort participants spend on complying with our request. Therefore, the comparison of the results of the two chapters primarily imply that the context in which communication is used is crucial for the resulting effect and, furthermore, that women may respond more strongly to differences in the context (Croson and Gneezy, 2009; Ellingsen et al., 2013).

Our study contributes to the literature in several ways: We provide additional evidence on the effects of the language of a request on perceived politeness (Economidou-Kogetsidis, 2015) as well as on compliance with that request (Althoff et al., 2014; Mitra and Gilbert, 2014; Das et al., 2008) and we add to studies analyzing cultural differences in the use and effects of “thanks in advance” (Lee et al., 2012; Mizutani and Mizutani, 1989; Ohashi, 2006). Moreover, we find indications of reciprocal behavior that could be explained by the findings in Goei et al. (2003) on the effect of liking on compliance.

Our findings imply that the commonly used phrase “thanks in advance” should be avoided or used carefully in requests as it tends to decrease compliance, at least for female, German students. It may be advisable to skip expressions of gratitude *ex ante* and, instead, express gratitude only after the favor has been granted. Alternatively, one should at least try to use a more personal and sincere expression of future gratitude in requests.

3.A Appendix

3.A.1 Instructions

The following section provides the English translation of the original German instructions, which the participants received in the experiment on screen:

Screen 1

Welcome to PLEx

Please abide by the following rules:

1. Communication is forbidden as long as you are in the laboratory.
2. If you have any questions, please ask the experimenters.
3. You may only use those functions of your computer that are intended for the experiment.

Thank you for supporting our research. You will find the instructions for the experiment on the next pages.

Screen 2

General Instructions

You will receive 5 Euros for showing up on time and your participation in the experiment.

Furthermore, you can earn additional money in the main part of the experiment. The amount of money you earn depends on your decision in the experiment.

At the end of today's experiment, you will be paid in cash.

The experiment consists of a main part, which we will explain to you in detail on the next page, and a subsequent questionnaire. Please read the instructions for the main part of the experiment carefully.

If there is anything you do not understand, please raise your hand. We will then come to your seat and answer your question personally.

Screen 3

Instructions for the main part of the experiment

In the main part of the experiment you will see a rectangle on the left side of the screen, which consists of 25 numbered packages.

Your task is to collect packages. This is done by a left-click on the desired package. You earn 0.50 Euro for each collected package.

Once you have collected a package, it disappears from the rectangle. On the right side of the screen you will see current information about the number of packages you have collected so far and how your earnings may look like accordingly.

But beware: Your earnings are merely hypothetical at first, because behind one of the 25 packages there is a bomb! If you have collected the package behind which the bomb is hidden, all packages you have collected will be destroyed and you will receive 0 Euro for this part of the experiment.

You do not know behind which one of the packages the bomb is. The bomb can be hidden anywhere in the rectangle with equal probability.

Only when leaving the laboratory, you are going to pick a number from a bag of 25 numbered notes. The number on your note determines behind which package the bomb is in your case. Therefore, you will learn only after the experiment whether or not you have collected the package behind which the bomb is located. If you have collected the package behind which the bomb is located, you will receive 0 Euro. However, if you have not collected the package behind which the bomb is located, you will receive 0.50 Euro for each package you have collected. In any case, you may keep the 5 Euro for your punctual appearance.

If you have collected a package by accident, you can always return all packages by clicking “reset” and start collecting the packages again.

There is no time limit for this task. Once you are done, you will start with the questionnaire.

3.A.2 Screenshots

The following section provides screenshots of the bomb risk elicitation task (Figure 3.A.1), of the open question in the *ThanksInAdvance* treatment (Figure 3.A.2) as well as in the *Control* treatment (Figure 3.A.3) in the original, German language.

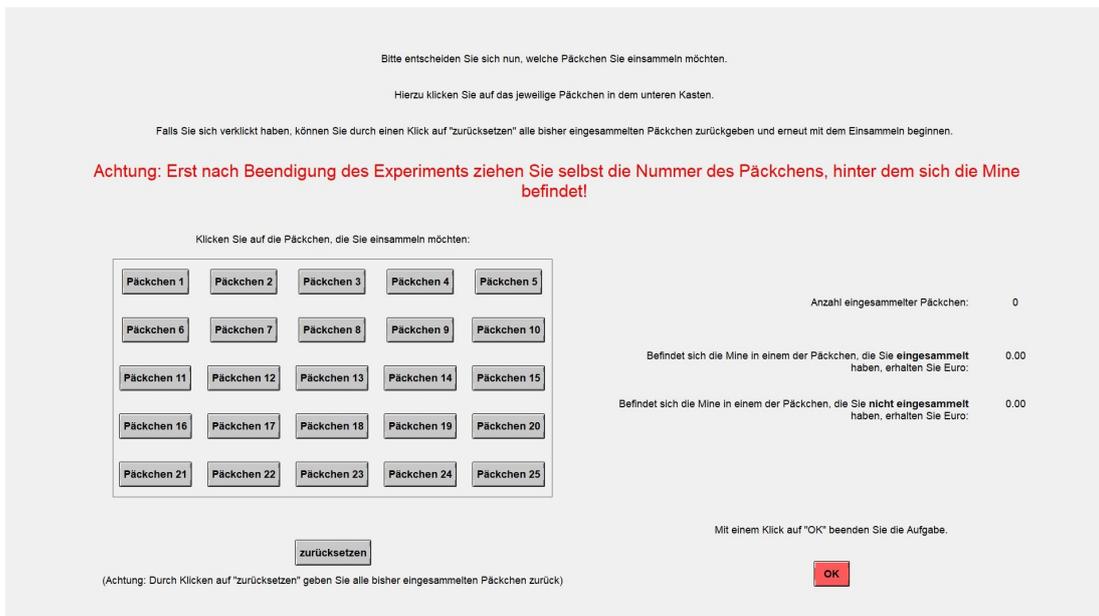


Figure 3.A.1: Screenshot of the 5x5-version of the bomb risk elicitation task

3.A. APPENDIX



Figure 3.A.2: Screenshot of the request in the *ThanksInAdvance* treatment

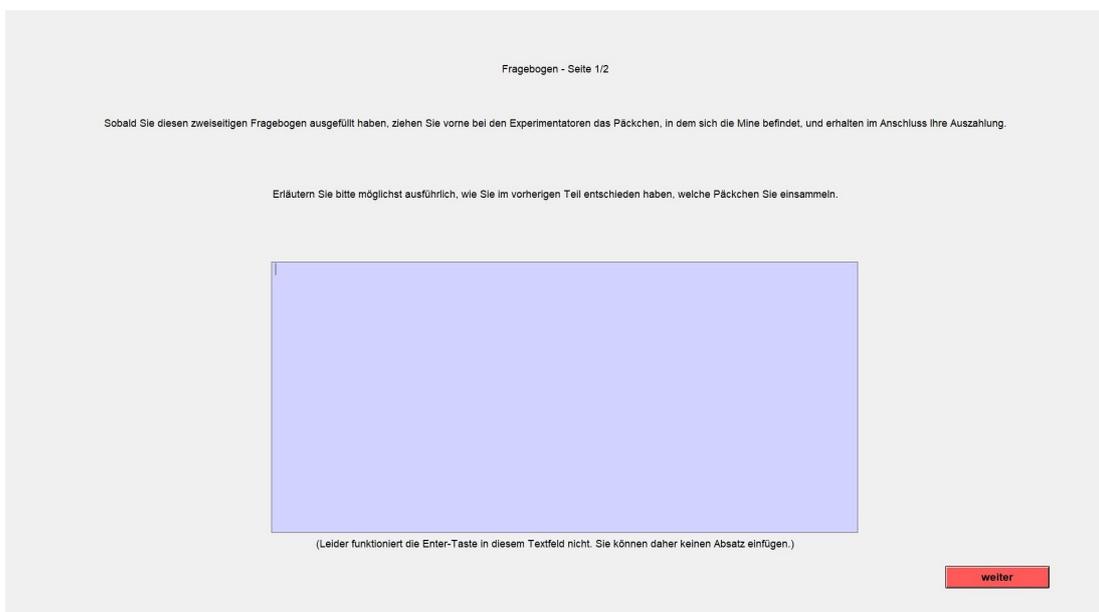


Figure 3.A.3: Screenshot of the request in the *Control* treatment

Chapter 4

The Effects of (Imprecise) Defaults on Dishonesty - Tax Evasion in a Laboratory Experiment

4.1 Introduction

For many individuals, the process of filing a tax return has changed considerably in past years, as it is more and more common for some fields of a tax return to be prefiled (OECD, 2017). The main source for prefilling is third-party reported data, which is collected by tax authorities and provided to individuals before filing their tax return. Mainly, this is data on salaries and wages, dividends and interest, as well as pension payments. Prefilling a complex form such as a tax return with accurate data has the main advantage that it reduces the high costs of filing the tax return (Reeson and Dunstall, 2009; Goolsbee, 2006; Fonseca and Grimshaw, 2017; OECD, 2017; Bankman, 2008). Furthermore, recent studies suggest that prepopulating a tax return with correct values increases tax compliance (Fochmann et al., 2018; Fonseca and Grimshaw, 2017) or even prevents tax evasion almost completely if these default values are based on third-party reported data, since the probability of detection is higher when modifying such a default (Bankman et al., 2015; Kleven et al., 2011; Slemrod, 2007; OECD, 2017). Hence, prepopulating a tax return with correct values has only advantages.

However, is it also beneficial to prepopulate a tax return with incorrect values? Most electronic tax return programs have the option to prepopulate fields with last year's values. In most cases, these values are incorrect for the present year. While these defaults still help taxpayers accomplish the complex task of filing a tax return by giving guidance and orientation as to what kind of income or deduction has to be inserted into which field, the effect on taxpayers' compliance behavior is not clear.

A default value in a tax return may affect tax compliance via three channels. First of all, there may be a status quo bias (Samuelson and Zeckhauser, 1988), i.e. taxpayers may be reluctant to change the status quo, which in this case is the default value in the tax return. Secondly, an anchoring effect (Tversky and Kahneman, 1974; Epley and Gilovich, 2001) may lead subjects to report an income which is closer to the default value than they would report without the default, because subjects may assimilate their final decision towards this starting value. Lastly, subjects may be more prone to evading taxes if the default value already implies cheating, i.e. if the default value is lower than the true income. In such a case, evading taxes would only be a passive act of "omission" and not an active act of "commission," which has been shown to increase cheating (Mazar and Hawkins, 2015).

This paper tries to answer the question of whether defaults in a tax return affect taxpayers' compliance behavior while focusing on default values that are incorrect, i.e. not equal to the true value. Hence, we attempt to analyze whether defaults higher and lower than the true value have an effect on tax compliance and whether these effects are different. Thus, this paper does not attempt to give answers as to whether or not prepopulating tax returns with accurate third-party reported data is beneficial, but rather whether or not there are effects on tax compliance of prepopulating fields for which there is no third-party reported data, such as cash income or deductibles, with last year's data or estimates of the true value. Furthermore, it is of interest whether the potential effects are transitory or lasting.

In order to study this research question, a controlled laboratory experiment is conducted in which subjects earn money by completing a real-effort task and must subsequently file a tax return for three consecutive rounds.¹ Taxes are de-

¹In the field of tax evasion research, a laboratory experiment has the main advantage that compliance behavior can be observed completely, which is rarely possible with field data (Torgler, 2002). For a discussion about using laboratory experiments, such as tax evasion experiments, in the area of public economics, see Alm (2010). The external validity of findings from laboratory experiments on tax compliance is discussed in Alm et al. (2015).

ducted on reported income. Underreporting may lead to an additional payment if a tax return is audited, which happens with a fixed probability. The experiment consists of two treatments in a between-subjects design. In the first treatment, the tax return is prepopulated with a constant default value, while the second treatment serves as a control treatment without a default. The default value in the first treatment is the hypothetical income which subjects would earn if they showed the same performance as they did in previous rounds of the real-effort task. Hence, the default value results from individual decisions in the past, similar to an electronic tax return program prepopulating a tax return with the previous year's data. The real-effort task has three levels of difficulty: easy, medium and hard. If the level of difficulty is easy (hard), subjects will be likely to earn more (less) money and, hence, their true income will probably be higher (lower) than the constant default value, which is based on the average performance in all three difficulty levels. Thus, the design provides a natural and transparent way of coming up with a default which varies in its relative magnitude and for which subjects have no reason to believe that changing the default would affect the probability of an audit. Hence, potential effects are driven solely by a direct default effect (status quo bias, anchoring effect, omission vs. commission) and not by an indirect effect on subjects' beliefs about the audit probability.

The results suggest that there is no lasting effect of an incorrect or inaccurate default value on tax compliance. Compared to no default, neither a relatively low nor a relatively high default have a significant effect on tax honesty over the course of the three rounds. However, when only looking at the first encounter with a default, a default value lower than the true income decreases tax honesty significantly and substantially, whereas there is again no significant effect for defaults higher than the true income.

The rest of the paper is structured as follows. First, the relevant literature is discussed. Following that, the design of the experiment is presented in Section 4.3. Then, behavioral predictions are derived from theory in Section 4.4. Section 4.5 presents the results. Finally, Section 4.6 discusses these results and concludes.

4.2 Related Literature

Recently, two studies have attempted to answer research questions similar to that of this paper. Closest to this paper is the laboratory experiment of Fochmann et al. (2018), in which the authors let participants earn income by completing a

4.2. RELATED LITERATURE

real-effort task for six consecutive rounds and then have them file a tax return in which, depending on the treatment, fields may be prepopulated with a default value. Four treatments are conducted, one in which there is no default value, one in which the default value is the correct income for all 6 fields, and two in which 3 fields are correctly prepopulated and 3 fields are incorrectly prepopulated - either with a random number smaller than the actual income or with a random number larger than the actual income.² They find that tax compliance is higher with a correct default than without a default, and that tax compliance is lower with a lower default than with a correct default. They do not find a difference in tax compliance between a correct default and a higher default, as well as between no default and a lower default.

In contrast to their paper, this study focuses on the comparison between no default and an incorrect default because this is, in my opinion, the more interesting question when thinking about whether or not defaults in a tax return may have negative effects on tax compliance. Prepopulating a tax return with correct values should always be preferred because it decreases taxpayers' cost of filing a tax return (Reeson and Dunstall, 2009; Goolsbee, 2006; Fonseca and Grimshaw, 2017); furthermore, it might increase tax compliance (Fochmann et al., 2018; Fonseca and Grimshaw, 2017), due either to a potential default effect (status quo bias, anchoring effect, omission vs. commission) or to a change in taxpayers' belief about the probability of an audit, which is particularly likely to occur if a correct default value is based on data from third-party reporting. Moreover, electronic tax return programs often do not even have the objective of providing correct default values, since they prepopulate a lot of fields with the previous year's data, which is, at most, an unreliable guess for the true value of the current year. Nevertheless, these defaults may help taxpayers when completing the complex task of filing a tax return. By providing a natural way of coming up with default values, which result from subjects' actions in previous periods, this paper tries to answer the question as to whether or not providing an incorrect default has an effect on tax compliance compared to the other available option of providing no default at all.

A second paper which looks at the effects of prepopulating a tax return in an experimental setting is the one by Fonseca and Grimshaw (2017). In their study,

²By prepopulating 3 fields with a correct value and 3 fields with an incorrect value, an experimenter demand effect (Zizzo, 2010) may increase potential differences in tax compliance between correct and incorrect defaults if participants are under the impression that they should only alter the three incorrect fields and not the three correct ones. Hence, this may be a problem when comparing the effects of correct and incorrect defaults.

participants in an online experiment filed a tax return in which, depending on the treatment, some fields were prepopulated with either correct or incorrect default values. Contrary to this paper, participants did not have to earn their income; they received information about their income and deductible costs in the beginning. The authors find that a correct default value in one field of a tax return increases tax compliance for this field significantly compared to no default value. A default value in one field which is lower than the actual income decreases tax compliance and tax revenue compared to both a correct default and no default. They conclude that defaults should only be used in a tax return if the underlying data is very reliable.

However, one has to be careful with this conclusion, since the authors clearly focus on prepopulated tax returns by the government resulting from third-party reporting. Participants in the experiment are told in the instructions that the “pre-filled values are estimates by the experimental tax authority of the values in your profile.” Furthermore, the audit probability is not fixed and participants are told that the likelihood of being audited varies with the values they enter in the tax return.³ Since the default value gives participants information about the tax authority’s knowledge and since the audit probability varies, it is likely that participants assume the audit probability to be different when changing a correct default value to underreport income than when sticking to an incorrectly low default value to underreport income (Kotakorpi and Laamanen, 2016). Hence, while being more realistic for third-party reported data used by the government, the study may confound two effects: an effect on taxpayers’ beliefs about the audit probability and the sole effect of a default value. Therefore, the negative effect of an incorrectly low default and the positive effect of a correct default could be overstated when thinking about default values which clearly do not affect the audit probability of a tax return by the tax authority, such as default values in an electronic tax return program stemming from the previous year’s data.

While not focusing only on the effects of prepopulated tax returns, Bruner et al. (2015) is a third study which uses prepopulated fields in a tax return in a laboratory experiment. Their experimental design features 2 exogenous income streams (one that is known by the tax authority ex ante and one that is not known) and deductions. They find that prepopulating the fields of “known” income and deductions leads to more tax evasion for the non-prepopulated field of “not known” income, especially when deductions are prepopulated with a value higher than the true deductible amount. However, as in Fonseca and Grimshaw (2017), the

³The audit probability varies according to a specific rule unknown to the participants.

4.2. RELATED LITERATURE

effect may not only be the result of a default, but also due to changing beliefs about the knowledge of the tax authority and, thus, about the probability of an audit. Participants may think that if two fields are prepopulated while one is not, that the probability of detection is lower for this particular field than it would be when all three fields were not prepopulated. Furthermore, the design is relatively complex, including, for example, two different audit probabilities for different items in the tax return, one of which is changed three times during the course of the experiment, which could potentially make the varying beliefs about the audit probability more pronounced depending on which fields are prepopulated.

One study that looks at the effects of a partially prepopulated tax return in the field is the study by Kotakorpi and Laamanen (2016). The authors analyze data from Finland between 1995 and 2004, in which a proportion of taxpayers received a partially prepopulated tax return while others did not. The default values in the prepopulated fields were all based on third-party reported data. Since the authors do not have data on actual tax compliance behavior or income data, they compare reporting behavior between taxpayers who received a partially prepopulated tax return and those who did not. They find that taxpayers who received a partially prepopulated tax return claim fewer deductions in fields that are not prepopulated, claim more deductions in fields that are prepopulated, and declare slightly less income in fields that were not prepopulated compared to those who did not receive a partially prepopulated tax return. For prepopulated income, such as labor income, they do not find a difference. The effects of partially prepopulated tax returns on reporting behavior are larger for small amounts of deductions and income. On the aggregate level, however, a partially prepopulated tax return has no effect on the total taxable income or taxes paid. For data from Denmark, Gillitzer and Skov (2018) find that prepopulating a deduction field - charitable donations - through the use of third-party reported data increased claimed deductions considerably, particularly of small amounts.

Apart from prepopulated tax returns, a large body of literature exists on the topic of tax evasion. The theoretical tax evasion literature started with the model by Allingham and Sandmo (1972), which is based on the economics of crime approach by Becker (1968). This theory states mainly that tax evasion is a rational choice and, hence, is negatively influenced by the audit probability and the fine, and positively influenced by the tax rate. Since then, a large body of literature has expanded the study of tax evasion in several directions. Excellent reviews of the literature, focusing on laboratory experiments, can be found in Andreoni et al. (1998), Torgler (2002), Hofmann et al. (2008), Alm (2012) and Slemrod

(2016), as well as in two reviews focusing on field experiments by Hallsworth (2014) and Mascagni (2018).

Besides tax evasion, Mazar and Hawkins (2015) study the effect of defaults on cheating in a laboratory experiment. They find that subjects cheat more when cheating requires only accepting an incorrect but financially superior default (omission), and that subjects cheat less when cheating requires actively overriding a correct and financially inferior default (commission). They argue that the active act of cheating (commission) is seen as more morally reprehensible than the passive act of cheating (omission).⁴

Several studies have analyzed the effects of defaults in other settings. Well known are the positive effects on retirement savings from an automatic enrollment when signing an employment contract (Madrian and Shea, 2001), or the large differences in the proportion of people consenting to be organ donors between opt-in and opt-out systems (Johnson and Goldstein, 2003). Modifying defaults in order to change people’s behavior is a key aspect of nudging, a concept defined by Richard Thaler and Cass Sunstein as altering the decision environment in order to achieve a specific change in behavior without imposing restrictions or modifying economic incentives (Thaler and Sunstein, 2008). The goal of nudging is to make affected people “better off, as judged by themselves,” (Thaler and Sunstein, 2008, p. 5) a criterion that has been discussed controversially (Sugden, 2017; Sunstein, 2018).

4.3 Experimental Design and Procedures

The experiment consists of two parts. In both parts, subjects earn points by carrying out a real-effort task for three rounds. Only in the second part do subjects declare their income to a hypothetical tax authority by filing a tax return after each round of the real-effort task. Hence, each subject files three tax returns in the experiment. Every tax return is audited by the tax authority with a fixed probability. An audit leads to an additional payment if subjects declared less income than they actually earned. The purpose of the first part of the experiment is to determine each subject’s average performance on the real-effort task. The experiment consists of two treatments in a between-subjects design. In the treatment *Default*, the income field in the tax return in all three rounds is

⁴A recent review of the literature on lying and cheating can be found in Jacobsen et al. (2018).

prepopulated with a value equal to the income which would have been generated by the subject's average performance in the real-effort task in the three rounds of the first part. Hence, the default value can be seen as a previous period's value, similar to prepopulating a tax return with the previous year's data. The treatment *NoDefault* serves as a control treatment, in which the income field in the tax return is not prepopulated.

In the real-effort task, subjects encode words into numbers. The task is based on Benndorf et al. (2018),⁵ which is a modification of the task by Erkal et al. (2011). Depending on the level of difficulty (easy, medium or hard), a word consists of either two, three or four random letters. In a table at the bottom of the page, a three-digit number is assigned to each of the 26 letters from A to Z in random order. The task is to find the correct three-digit number for each letter of the word.⁶ After correctly encoding one word, a new random word appears, a new three-digit number is randomly assigned to each letter of the alphabet, and a new order of letters in the table randomly emerges. In each round, subjects have 120 seconds to encode as many words as possible. In each of the two parts, they carry out the real-effort task three times, once for every level of difficulty. The order of the task's level of difficulty is random in both parts. In the first part, subjects receive 3 points for every correctly encoded word, and the payoff is equal to the sum of the points earned over the three rounds.

The reason for choosing this real-effort task is that, according to Benndorf et al. (2018), subjects' performance does not increase considerably due to learning. This is important, because in the treatment *Default*, subjects' average performance in the first part is used as a benchmark value for their performance in the second part. Furthermore, this task provides a setting in which the level of difficulty can be easily modified. In the treatment *Default*, the default value in the tax return stays constant over all three rounds (the average performance in the first part). However, the level of difficulty changes from round to round. Therefore, when the level of difficulty is easy, subjects should earn more points and the default value in the tax return should be lower than the true income. In the opposite scenario, when the level of difficulty is hard, the default value should be higher than the true income. A medium level of difficulty should result in a default value similar to the true income. Hence, this design provides a natural way of coming up with a default value which presumably is at least one time lower and one time higher than the true income and which results from subjects' own behavior in the past.

⁵The author gratefully acknowledges using the z-Tree code provided by the authors.

⁶A screenshot of the real-effort task can be found in the Appendix.

Furthermore, since the origin of the default value is clear and transparent, there is no reason for subjects to believe that altering such a default value would affect the audit probability.

In the second part, subjects have to file a tax return after each round of the real-effort task.⁷ Subjects have to pay taxes on declared income⁸ with a tax rate of $t = 0.3$. They are informed that the tax authority may audit their tax return with a fixed probability of $p = 0.15$. In the treatment *Default*, instructions additionally emphasize that changing the default value will not affect the audit probability. If a tax return is audited and subjects have declared less than their actual income, they will have to pay back the evaded taxes plus a fine of the same amount. The expected value of a subject's net income depending on subject's gross income (GI) and declared income (DI) if $DI \leq GI$ is given by:

$$\begin{aligned} E[GI, DI] &= (1 - 0.15) \cdot (GI - 0.3DI) + 0.15 \cdot (GI - 0.3DI - 0.6(GI - DI)) \\ &= 0.9GI - 0.21DI \end{aligned}$$

The expected value is maximized by declaring an income of 0 and, hence, risk-neutral or risk-loving subjects should declare an income of 0. Subjects receive feedback on whether or not they were audited in a particular round only after all three rounds are completed. Subjects receive 30 points per correctly solved word in the real-effort task in the second part, which is more than in the first part. However, they only receive payment for one randomly drawn round to prevent hedging.

In total, 201 subjects took part in the experiment. Sixteen sessions were conducted at the Potsdam Laboratory for Economic Experiments (PLEx) between June and November 2017. A total of 104 subjects participated in the treatment *Default*, 97 subjects in the treatment *NoDefault*. Table 4.1 shows summary statistics for the subjects in the experiment according to treatment.⁹ All subjects were students. The most common field of study was business and economics (30%), followed in order of frequency by social sciences (20%), law (14%), humanities (13%), sciences (9%) and others (13%).

In each session, subjects were randomly assigned a seat in visually separated cu-

⁷A screenshot of the tax return can be found in the Appendix.

⁸The minimum income subjects can declare is 0. The maximum income subjects can declare is the gross income in the treatment *NoDefault* and either the gross income or the default value in the treatment *Default*, whichever is greater.

⁹There are no significant differences between treatments.

Table 4.1: Summary statistics for subjects in the experiment

	Treatment	
	<i>NoDefault</i>	<i>Default</i>
male	0.44	0.35
age	23.23	22.68
semester	4.18	4.48
experience tax declaration	0.32	0.23
risk aversion	3.41	3.16
encoded words 2nd part	10.59	10.79
N	97	104

Note: Risk aversion is measured using a question in the questionnaire (“In general, how risk-seeking would you say you are on a scale from 1 to 6?”), 1 being very risk averse and 6 being very risk-loving. Dohmen et al. (2011) show that responses to such a simple question can predict actual risky behavior reliably.

bicles. At the beginning, they read the instructions for the first part¹⁰ and had to answer control questions correctly in order to ensure that they had understood the experiment. Subsequently, subjects became acquainted with the real-effort task in a trial round, in which every subject had to encode 5 words for each of the three levels of difficulty. After every subject had finished the trial round, the first part started and subjects had to encode as many words as possible for three rounds. Following this, subjects received the instructions for the second part of the experiment. After reading the instructions, subjects had to answer control questions again¹¹ before the second part of the experiment started, in which subjects earned points and then had to file a tax return for three rounds.

Following the experiment, subjects answered a questionnaire and received their payoff in private. Points were converted to Euros at a rate of 35 Points = 1 Euro. Sessions lasted for approximately 55 minutes and subjects earned on average 12.68 Euros, including a show-up fee of 3 Euros.¹² The experiment was computerized with z-Tree (Fischbacher, 2007) and the recruitment process was conducted using ORSEE (Greiner, 2015).

¹⁰Subjects were informed that the experiment consisted of two parts and received general information about the second part.

¹¹An English translation of the instructions and the control questions for the first and second parts of the experiment can be found in the Appendix.

¹²12.68 Euros \approx 14.50 Dollars at the time of the experiment.

4.4 Behavioral Predictions

For a fully rational subject, such as the one described in the tax evasion model by Allingham and Sandmo (1972), prefilling a tax return with a default has no influence on tax compliance. However, for human beings, whose decision-making process is often better described by the concept of bounded rationality (Simon, 1957; Kahneman, 2011), a default option may influence their decision making. Since it is not clear whether a default is equal to, greater or less than the true income when the level of difficulty is medium, behavioral predictions are only derived for the difficulty levels easy and hard, in which the default value is very likely to be either lower or higher than the income.

First of all, a status quo bias may affect subjects' behavior, i.e. subjects may prefer the current status and be reluctant to change it. The most frequent explanations for this bias include loss aversion, inertia, and a reluctance toward making one's own decisions (Samuelson and Zeckhauser, 1988; Kahneman et al., 1991; Ritov and Baron, 1992). The default option in the tax return may serve as the status quo, and changing it may be perceived as an additional loss. In the control treatment, no status quo exists and, thus, no additional loss is incurred. Hence, a status quo bias leads to relatively fewer changes in the default value than one would expect for fully rational subjects. For a default value higher than the true income, this bias has a positive effect on tax honesty compared to no default. For a default value lower than the true income, the effect is not clear. For subjects who are relatively tax-honest, i.e. subjects who would declare an income greater than the default value in a hypothetical situation without the default, the default leads to more tax evasion compared to no default. For subjects who are relatively tax-dishonest, i.e. subjects who would declare an income less than the default value in a hypothetical situation without the default, the default has a positive effect on tax honesty. Hence, the overall effect depends on the share of tax-honest to tax-dishonest subjects and their hypothetical income declaration decision without a default.

Secondly, a default may influence subjects' decisions due to the anchoring effect (Tversky and Kahneman, 1974; Epley and Gilovich, 2001).¹³ The anchoring effect works by providing a starting point from which subjects begin their decision-making process. If subjects assimilate their final decision insufficiently, their declared income will be biased towards this starting value. The default in the present study is by definition an irrelevant anchor, since it does not provide

¹³See Furnham and Boo (2011) for a literature review.

any new information about the decision subjects are facing. However, even completely irrelevant anchors have been shown to influence behavior (Tversky and Kahneman, 1974; Ariely et al., 2003). If we assume that there is no anchor serving as a starting value in the control treatment, the same predictions as in the case of the status quo bias are derived: For a default value higher than the true income, tax honesty increases compared to no default. For relatively tax-honest (dishonest) subjects, a default lower than the income decreases (increases) tax honesty.

Predictions status quo bias and anchoring effect: (assuming no anchor in *NoDefault*)

- Hard level of difficulty: The default has a positive effect on tax honesty.
- Easy level of difficulty: The effect of the default on tax honesty is not clear.

However, if we assume that in the absence of a default the gross income serves as an anchor, the prediction regarding a default lower than the true income will change slightly for the anchoring effect. In this case, a relatively low default will have an unambiguous negative effect on tax honesty compared to the control treatment, since the starting point from which subjects begin assimilating their decision is lower than the gross income.

Predictions anchoring effect: (assuming gross income is the anchor in *NoDefault*)

- Hard level of difficulty: The default has a positive effect on tax honesty.
- Easy level of difficulty: The default has a negative effect on tax honesty.

Lastly, the default may influence the psychological costs of lying. It has been shown that people are more inclined to cheat if cheating requires only a passive act of “omission” and are less inclined to cheat if cheating requires an active act of “commission” (Mazar and Hawkins, 2015). This finding has been explained by the action principle, which states that harmful acts are perceived as less morally reprehensible if the act does not require an active action and more morally reprehensible if the act does require an active action (Cushman et al., 2006; Spranca et al., 1991). The more morally reprehensible an act is perceived to be and, hence, the stronger the “signal of malicious intent” is (Mazar and Hawkins, 2015), the

harder it is to maintain a self-concept of being a decent and moral person, which increases the psychological costs of this act (Mazar et al., 2008). Hence, an active act of cheating, which signals malicious intent, is associated with higher psychological costs of cheating than passively accepting to cheat.

According to Mazar and Hawkins (2015), actively overriding a correct default in order to cheat is the strongest signal of malicious intent, while passively accepting a default which implies cheating is the weakest signal of malicious intent. An active act of cheating without a default lies in-between these two. It seems plausible that the fourth relevant option of overriding an incorrect default in order to cheat lies in-between the former two as well, and is a comparable signal of malicious intent to cheating without a default.

Following this line of argumentation, overriding a default which is higher than the true income in order to cheat is a comparable signal of malicious intent to cheating without a default and, thus, the psychological costs of cheating are similar. Hence, no effect of such a default on tax honesty is expected. If the default is lower than the income, the effect of the default depends on subjects' hypothetical behavior without a default. For relatively tax-honest subjects (those who would declare an income greater than a default in the hypothetical situation without the default), cheating in the treatment *Default* by passively accepting the default is associated with lower psychological costs than it is in the treatment *NoDefault*, in which subjects have to make an active choice in order to cheat. For subjects who are relatively tax-dishonest (those who would declare an income less than a default in the hypothetical situation without the default), cheating in the treatment *Default* requires overriding an incorrect default, which is associated with similar psychological costs as the active decision in the treatment *NoDefault*. Hence, the overall effect of a default lower than the true income is expected to be negative, since relatively tax-honest subjects cheat more while tax-dishonest subjects do not change their behavior.

Predictions “omission” vs. “commission:”

- Hard level of difficulty: The default has no effect on tax honesty.
- Easy level of difficulty: The default has a negative effect on tax honesty.

Thus, all three described biases do not yield clearly opposing predictions. For a default lower than the income, either a negative effect on tax honesty is predicted or no clear prediction is derived. For a default greater than the income, either

4.5. RESULTS

Table 4.2: Summary statistics for subjects' performance in the real-effort task

	Level of Difficulty			Part		Sum
	Easy	Medium	Hard	Part 1	Part 2	
encoded words (mean)	14.3	9.8	7.4	10.3	10.7	10.5
N	402	402	402	603	603	1,206

a positive effect on tax compliance or no effect is predicted. Hence, it is likely that the design will not be able to distinguish clearly between the three biases as explanations for potential effects of a default, which is, however, not the main purpose of the design.

4.5 Results

There were 201 participants in the experiment, leading to a total of 603 tax returns and 1,206 attempts to encode as many words as possible. The following section presents the results. First, subjects' performance in the real-effort task is shown, focusing on the effect of the difficulty level as well as on subjects' learning behavior. Subsequently, the effect of the difficulty level on the size of the default relative to its true value is described. Then, results on the default value's effect on tax compliance are presented.

4.5.1 Real-Effort Task

Each subject carried out the real-effort task three times in each of the two parts. Table 4.2 shows descriptive statistics for the real-effort task. On average, subjects encoded 10.5 words per round. The level of difficulty did affect subjects' performance: Subjects encoded roughly 4.5 more words in easy than in medium levels, and roughly 2.5 more words in medium than in hard levels.¹⁴ The differences in

¹⁴The difference in performance between easy and medium levels is larger than the difference between medium and hard. Therefore, the average performance overall is higher than the average performance in the medium level. This may be beneficial, because it could mitigate potential learning effects in the sense that the default value in the second part (which is the average performance from the first part) will be closer to the medium performance in the second part if subjects' performance increases in the second part.

performance between the levels of difficulty are highly significant.¹⁵ Furthermore, on average, subjects encoded 0.4 more words in the second part than in the first part. The difference in performance between the two parts is highly significant.¹⁶ This result suggests that subjects' performance may increase due to learning. However, the difference in the performance between the first and second parts could also be a result of the differences in incentives between part 1 and part 2. An OLS regression for only the first part, in which the number of correctly encoded words is the dependent variable and which controls for the level of difficulty, confirms that subjects' performance increases significantly, most likely due to learning.¹⁷ The model estimates that performance increases by 0.15 words per round ($p = 0.013$; standard errors clustered on subject's level). Using the same model for the second part, the effect of the round on performance decreases (but is still positive) and is no longer statistically significant ($p = 0.113$). Hence, there seems to be learning, but the learning effect is unsurprisingly stronger in the first part than in the second.

4.5.2 Relative Size of the Default

The size of the default relative to the true income (default / gross income) depends on the level of difficulty in the respective round, since the income is affected by the level of difficulty while the default value is constant over the course of the three rounds. Figure 4.1 shows in detail how the level of difficulty affects the relative size of the default. The default is always lower than the actual income when the level of difficulty is easy, since subjects earn more points in this condition. The relative size of the default in this condition varies between 45% and 90% (average of 70%, median of 69%) of the gross income. When the level of difficulty is hard, subjects earn fewer points and, hence, see a default value which is greater than their actual income. With one exception, this is true for all subjects.¹⁸ The relative size of

¹⁵Wilcoxon Signed Rank test used:

easy vs. medium: $z = 24.867$; $p < 0.0001$

medium vs. hard: $z = 25.040$; $p < 0.0001$

For each subject, the average of his or her performance in the two rounds with the respective level of difficulty is used.

¹⁶Wilcoxon Signed Rank test used: $z = 7.541$; $p < 0.0001$. For each subject, the average of his or her performance in the two parts is compared.

¹⁷Table 4.A.1 in the Appendix shows the regression results.

¹⁸One subject performed very poorly in the first part. Hence, his or her default was rather small. Even in the hard condition in part 2, this subject earned as many points as he or she did on average in the first part and, therefore, did not experience a default greater than his or her income. Excluding this subject from the following analysis does not change any of the results.

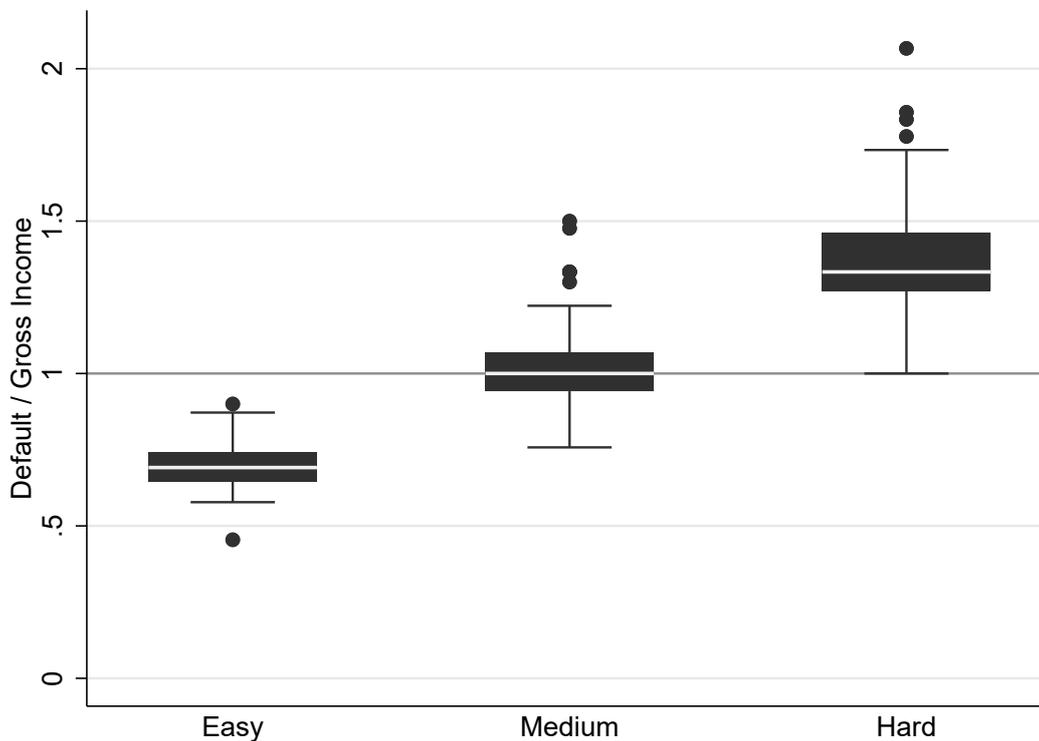


Figure 4.1: Effect of the level of difficulty on the relative size of the default
Note: The figure shows all observations belonging to the 104 subjects in treatment Default.
 The whiskers include all observations within a 1.5 interquartile range.

the default in this condition varies between 100% and 207% (average of 138%, median of 133%) of the gross income. When the level of difficulty is medium, subjects earn an income that is similar to the default value in the tax return. However, only for 11 subjects (11%) was the default exactly equivalent to their income earned in this condition. The relative size of the default varies between 76% and 150% (average of 102%, median of 100%) of the gross income.

4.5.3 Tax Compliance

This section presents the main results, i.e. the effects of a default on tax honesty. We begin by briefly looking at the effect of a default value irrespective of its relative size. Then, this effect is separated depending on the size of the default relative to the income.

When pooling data from both treatments, subjects declared on average 67% of their gross income on their tax return. Twenty-eight subjects (14% of all sub-

4.5. RESULTS

jects) always declared an income which was at least as high as their gross income¹⁹ and 16 subjects (8% of all subjects) always declared an income of 0 on their tax return. There are no significant differences in the frequency of subjects cheating fully or being completely honest between treatments.²⁰

The level of tax honesty of a tax return is defined by the declared income relative to the gross income (DI/GI). In the treatment *NoDefault*, subjects declared on average 65% of their income, whereas subjects in the treatment *Default* declared 69% of their income. Since subjects' characteristics are not perfectly balanced between treatments, as shown in Table 4.1, an OLS regression in which DI/GI is the dependent variable is used to test for a statistically significant effect of a default on tax honesty.²¹ The results of this OLS regression are shown in Table 4.A.2 in the Appendix. A default has no significant effect on tax honesty ($p = 0.631$) when not taking into account the relative size of the default. However, this result does not have to imply that a default has no effect on tax honesty. It may very well be that the effect of a relatively high default and a relatively low default cancel each other out.²²

Our design allows us to analyze the effect of a default which naturally varies in relative size in a between-subjects comparison, i.e. we compare subjects' behavior with and without a default for all three levels of difficulty separately. Figure 4.2 shows the average share of declared income relative to gross income for the three levels of difficulty and the two treatments separately. Regardless of whether or not subjects see a default, there is hardly any difference in the average tax honesty for the easy level, in which the default is lower than the gross income. In case of the medium level (default similar to gross income) and the hard level (default almost always higher than gross income), subjects declared on average 5 to 6 percentage points more income relative to their gross income when the tax return was prepopulated.²³

To control for differences in subjects' characteristics between treatments, OLS regressions are used to test for statistically significant effects of a default on tax

¹⁹The maximal declarable income in the treatment *Default* is either the gross income or the default, whichever is greater. Hence, when the default is greater than the gross income, subjects can declare an income higher than their true income. Two subjects each did this once by leaving the default value unchanged, which led to a declared income that was one time 14% higher and one time 4% higher than their gross income.

²⁰ χ^2 test used: $p = 0.538$ for completely honest and $p = 0.235$ for cheating fully.

²¹The regression controls for subjects' characteristics, income, level of difficulty and the respective round. Standard errors are clustered on subject's level, since every subject files three tax returns.

²²In addition, Table 4.A.2 shows that men and older subjects are less tax-honest. In later rounds, subjects declare relatively less income and, not surprisingly, risk averse subjects declare relatively more income. There is no significant effect of the level of difficulty, gross income or

4.5. RESULTS

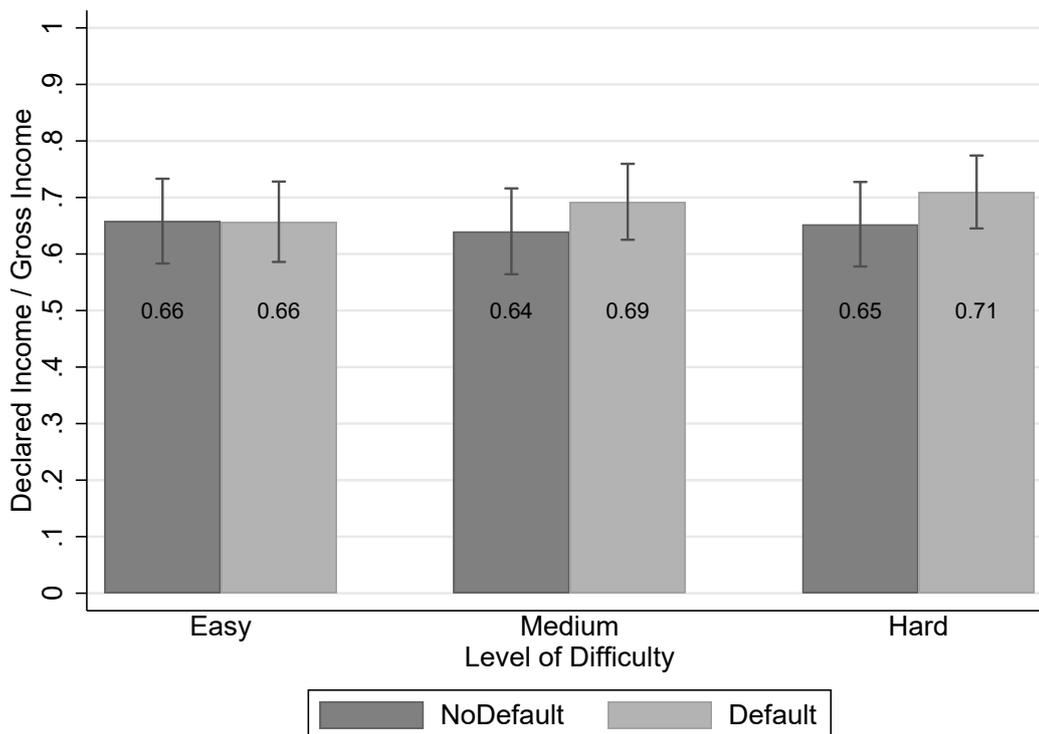


Figure 4.2: Tax honesty separated by treatment and difficulty level

Note: The figure shows means and 95% confidence intervals.

honesty depending on the relative size of the default. Table 4.3 shows two different models for each level of difficulty, i.e. 6 regressions in total. Columns (1), (3) and (5) show the results for the basic model without interaction effects. The effect of a default is very close to 0 for the medium as well as the hard level of difficulty in columns (3) and (5) ($p = 0.903$ for medium; $p = 0.847$ for hard). When the level of difficulty is easy (default < gross income), a default is estimated to decrease tax honesty by nearly 6 percentage points, as can be seen in column (1). However, this effect is also not statistically different from 0 ($p = 0.250$). Furthermore, the effects of a default do not differ significantly between the three levels of difficulty.²⁴

experience with tax declarations on tax honesty.

²³None of these differences are statistically significant when comparing means between treatments by using the Wilcoxon Rank Sum test:

Easy: $z = -0.315$; $p = 0.7530$

Medium: $z = -1.094$; $p = 0.2740$

Hard: $z = -0.981$; $p = 0.3266$.

²⁴Column (1) of Table 4.A.3 in the Appendix provides the results of an OLS regression which include interaction effects of the treatment variable and the level of difficulty. These interaction effects are not statistically different from 0.

4.5. RESULTS

Table 4.3: OLS regression results by difficulty level for two different models

VARIABLES	Easy		Medium		Hard	
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Default</i>	-0.056 (0.048)	-0.189** (0.083)	-0.005 (0.043)	0.036 (0.076)	0.009 (0.045)	0.063 (0.077)
<i>Default</i> x laterrounds		0.200** (0.101)		-0.062 (0.093)		-0.083 (0.094)
Constant	1.420*** (0.256)	1.477*** (0.255)	1.594*** (0.223)	1.563*** (0.229)	1.683*** (0.232)	1.648*** (0.236)
Control Variables	YES	YES	YES	YES	YES	YES
Observations	201	201	201	201	201	201
R-squared	0.235	0.251	0.366	0.368	0.291	0.294

Note: The dependent variable in all regressions is declared income over gross income. The variable *Default* is a dummy variable which is 1 if subjects participated in the treatment *Default*. The variable *laterrounds* is a dummy variable which is 1 if it is either the second or the third round and 0 if it is the first round. All regressions control for gross income, subjects' characteristics (gender, age, field of study, experience tax declaration, risk aversion) and the variable *laterrounds*.

* significant at 10%; ** significant at 5%; *** significant at 1%.

It may be that the effect of a default on the income declaration decision decreases when encountering the same default for a second or third time. Figure 4.3 shows the share of declared income for the three levels of difficulty and the two treatments separately only for income declarations in round 1. On the one hand, it seems that a default has a negative effect on tax honesty in the first round when the level of difficulty is easy, i.e. when the default is lower than the true income. On the other hand, a default seems to increase tax honesty in the first round for the medium and hard levels, i.e. when the default is of similar size or greater than the actual income.

To test for differences in the reaction depending on whether or not the fixed default is encountered for the first time, columns (2), (4) and (6) of Table 4.3 show the results of regressions in which an interaction effect of the variable *laterrounds* and the treatment variable is added to the basic model. *Laterrounds* is a dummy variable that takes the value 1 if it is the second or third round and 0 if it is the first round. The regression results in column (2) show that in the first round a default has a significantly negative effect on tax honesty when the level of difficulty is easy: Subjects are estimated to declare nearly 19 percentage points less

4.5. RESULTS

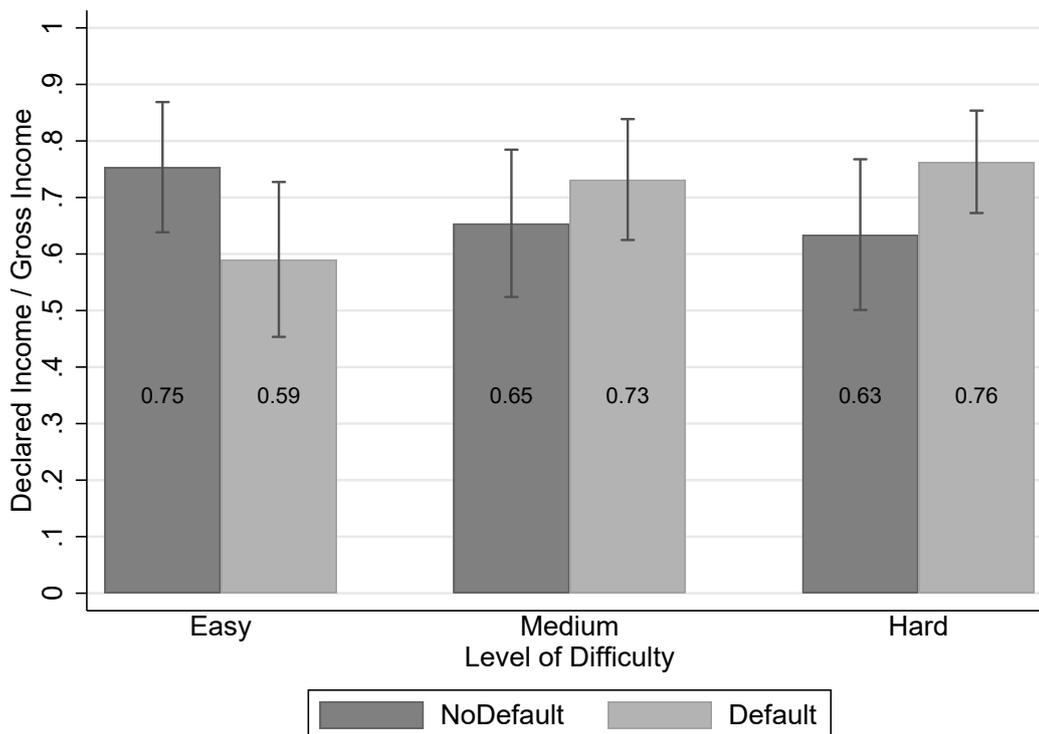


Figure 4.3: Tax honesty separated by treatment and difficulty level in round 1

Note: The figure shows means and 95% confidence intervals.

income relative to their actual income in response to the default. Furthermore, the positive and significant interaction effect for this level of difficulty shows us, first of all, that there is a significant difference in the response to such a default when subjects have already seen the default in previous rounds compared to seeing it for the first time and, secondly, that the effect of such a default is very close to zero in later rounds. When the level of difficulty is medium or hard, the point estimates suggest that a default has a small positive effect on tax honesty in the first round. However, neither of these effects are statistically significant ($p = 0.636$ for medium; $p = 0.410$ for hard). Additionally, the differences in the effect of a default between the first and later rounds are also not statistically significant in these cases. Finally, subjects' response to a default in the first round is significantly different between the easy level on the one hand (negative and statistically significant) and the medium and hard levels on the other hand (positive but not significantly different from 0), as can be seen from the positive and significant interaction effect of the treatment variable with the medium and hard levels of difficulty in column (2) of Table 4.A.3 in the Appendix.

4.6 Discussion and Conclusion

In this paper, a laboratory experiment was conducted in order to study the effect on tax compliance of prefilling tax returns with mostly inaccurate data. Subjects first earned money by performing a real-effort task and then had to declare their income to a hypothetical tax authority. In the main treatment, the tax return was prefilled with a default value resulting from subjects' average performance in previous rounds of the real-effort task, similar to electronic tax return programs that use the previous year's data to prefill tax returns. Since the level of difficulty of the real-effort task varied, subjects' true income varied as well and was at least one time lower²⁵ and at least one time higher than the default. In the control treatment, the tax return was not prefilled.

A lasting effect of a default on tax honesty was not found, neither of a default lower or higher than the income, nor of a default value similar to the income. Additionally, transitory effects were analyzed by looking only at the first encounter with a default. Compared to the control treatment without a default, a default lower than the actual income decreased tax compliance significantly and substantially in this case. No transitory effects were found for a default higher than or similar to the actual income.

The finding that a default higher than the income does not seem to influence tax compliance can be explained by similar psychological costs of cheating when overriding an incorrect default or when having to make an active decision in the absence of a default, since both actions may signal similar levels of malicious intent (Mazar and Hawkins, 2015). Both an anchoring effect and a status quo bias predict a positive effect on tax compliance of a default higher than the income and, thus, are not able to explain the lack of an effect. However, it seems plausible that at least the status quo bias is not very pronounced in this setting, since it should be very obvious to participants that a default higher than the income is unambiguously financially inferior and, hence, there is a strong incentive to change the status quo, i.e. to change the default.

The negative (but transitory) effect of a default lower than the true income on tax honesty can be explained by a status quo bias, by an anchoring effect and by various psychological costs of cheating. If the status quo bias or anchoring effect (assuming that no anchor exists without a default) were responsible for this finding, it would suggest that the negative effect for relatively tax-honest people

²⁵One subject did not experience a default higher than the income, as explained in Section 4.5.

4.6. DISCUSSION AND CONCLUSION

outweighs the positive effect for relatively tax-dishonest people. Assuming that the gross income serves as an anchor in the case of no default, the anchoring effect and various psychological costs of cheating do predict the negative effect found. Hence, it seems that various psychological costs, which depend on the relative size of the default, can best explain both results, at least in the short run.

The question remains as to why the negative effect of a relatively low default does not persist. One explanation may be that subjects do not take the default seriously in later rounds because it is always the same fixed number. It may be that the negative effect would persist if subjects saw a different default value than the one they saw in previous rounds. The persistence of default effects in general, as well as in tax returns in particular, is a fruitful and important area for future research.

This study contributes to the literature in several ways. First of all, it presents further evidence on the effect of prefilling tax returns on tax compliance (Fochmann et al., 2018; Fonseca and Grimshaw, 2017; Bruner et al., 2015; Kotakorpi and Laamanen, 2016). In contrast to the above-listed studies, this paper explicitly focuses on inaccurate defaults, resembling electronic tax return programs that prefill tax returns with last year's data. By using subjects' average performance in the past to determine the default, this study provides a natural and transparent way of coming up with a default which results from subjects' own individual past behavior and which ensures that potential default effects are not confounded with other effects such as a change in subjects' beliefs about the knowledge of the tax authority. Secondly, this study is, to the best of my knowledge, the first to analyze the transitory nature of the effect of default values on tax compliance. Finally, this study contributes evidence on the effect of defaults on the decision to lie or cheat (Mazar and Hawkins, 2015) and thereby contributes to the literature on nudging in general (Thaler and Sunstein, 2008).

There are some limitations of this study. By focusing explicitly on the effect of inaccurate defaults, this paper cannot make statements about the effect on tax honesty of prefilling tax returns with correct defaults. However, previous evidence suggests that using accurate third-party reported data to prefill a tax return has only advantages by decreasing taxpayers' costs of filing the tax return, as well as increasing tax compliance (Reeson and Dunstall, 2009; Goolsbee, 2006; Fonseca and Grimshaw, 2017; OECD, 2017; Bankman, 2008; Fochmann et al., 2018; Fonseca and Grimshaw, 2017). Moreover, in this simple experiment, subjects have full knowledge about the true value of their income. In reality, though, there may be some level of uncertainty due to the complexity of a tax return, which could

influence the effect of a default.

The results presented in this study suggest that there is no lasting negative or positive effect of a default on tax honesty. Since prefilling may reduce taxpayers' costs of filing a tax return even with inaccurate values by providing hints as to which kind of income stream or deduction has to be inserted in which field, prefilling seems to be beneficial. However, one should be careful with defaults that are lower than the true income amount. Filing a tax return is usually performed only once a year and it seems reasonable to believe that seeing a default one year later is similar to encountering it for the first time. Hence, the negative effect of a default that is less than the true income may be persistent in real life.

In particular, there may be negative effects on tax compliance following a recession. Last year's income values, which are the default values for the present year, are likely to be lower than the true values in this year when the economy is expanding again. Hence, taxpayers might be inclined to evade more taxes due to the relatively low default values in this situation.

4.A Appendix

4.A.1 Supplementary Tables and Figures

Table 4.A.1: Effect of learning on subjects' performance in RET

VARIABLES	Part 1	Part 2
round	0.150** (0.060)	0.101 (0.063)
# letters / word	-3.351*** (0.052)	-3.606*** (0.059)
Constant	20.043*** (0.276)	21.311*** (0.305)
Observations	603	603
R-squared	0.654	0.693

Note: The dependent variable in these OLS regressions is the number of correctly encoded words. # letters / word is a variable that determines the level of difficulty in the respective round. Standard errors are clustered on subject's level.

* significant at 10%; ** significant at 5%; *** significant at 1%.

Table 4.A.2: OLS regression results for the full sample

VARIABLES	DI/GI
<i>Default</i>	-0.018 (0.038)
gross income	-0.000 (0.000)
# letters / word	0.009 (0.035)
male	-0.191*** (0.045)
age	-0.014** (0.006)
experience tax declaration	0.000 (0.048)
risk aversion	-0.092*** (0.016)
round	-0.022** (0.011)
Constant	1.562*** (0.245)
field of study	YES
Observations	603
R-squared	0.276

Note: The dependent variable is declared income over gross income. The variable *Default* is a dummy variable which is 1 if subjects participated in the treatment *Default*. # letters / word is a variable that determines the level of difficulty in the respective round. Standard errors are clustered on subject's level.

* significant at 10%; ** significant at 5%; *** significant at 1%.

Table 4.A.3: OLS regression results - differences between difficulty levels

VARIABLES	All Rounds (1)	Round 1 only (2)
<i>Default</i>	-0.055 (0.047)	-0.190** (0.077)
<i>Default</i> x Medium	0.053 (0.036)	0.223** (0.105)
<i>Default</i> x Hard	0.058 (0.048)	0.218** (0.104)
Constant	1.577*** (0.201)	1.639*** (0.229)
Observations	603	201
R-squared	0.276	0.330

Note: The dependent variable is declared income over gross income. The variable *Default* is a dummy variable which is 1 if subjects participated in the treatment *Default*. The base level is Easy. All regressions control for gross income, subjects' characteristics (gender, age, field of study, experience tax declaration, risk aversion) and, in the case of the first model, the variable laterrounds. Standard errors are clustered on subject's level.

* significant at 10%; ** significant at 5%; *** significant at 1%.

4.A.2 Instructions and Control Questions

Instructions for Participants of the Experiment - Part 1

The following section provides the English translation of the original German instructions that the participants received at the beginning of the experiment:

General Instructions

Today, you are participating in a decision-making experiment. If you read the following instructions carefully, you can earn money. The amount of money that you will receive depends on the decisions you make and on chance, but it does not depend on other participants' decisions.

For the entire duration of the experiment, it is not allowed to communicate with other participants. Thus, we ask you not to talk to each other. We also ask you to turn off your mobile phone. If you violate these rules, you will be excluded from the experiment and the payment.

In case there is something you do not understand, please take another look at the instructions or raise your hand. We will then come to your seat and answer your question personally.

During the experiment, we are not using Euros but rather points. The points you earn in the experiment will be converted into Euros as follows:

35 points = 1 Euro

At the end of the experiment, you will receive the points you have earned, converted into Euros, in **cash**. Additionally, you will receive 3 Euros for showing up on time.

The experiment consists of **two** parts. In the first part, you can earn points by solving tasks. In the beginning of the second part, you can again earn money by solving the same tasks from the first part once more. Based on the points you earned in the second part, you will then file a tax return. You will receive the instructions for the second part after the first part is completed.

On the following pages, we will explain to you the exact sequence of events in the first part of the experiment. Before the start of the experiment, we will ask you a few control questions on the screen, which are supposed to help you understand the procedure. These control questions do not affect your payment. The experiment only starts when all participants are completely familiar with the procedure of the experiment and have answered the control questions correctly.

Detailed description of the first part of the experiment

In this part of the experiment, you can earn points by solving tasks. The tasks ask you to encode words by translating these words into numbers. Each word consists of random letters, and each letter must be translated into a **three-digit** number. You can find the correct number in a table at the bottom of the screen. In this table, each letter has the correct three-digit number assigned to it. The screen for the encoding task looks like this:

— *Screenshot of the encoding task* —

There are **three** different levels of difficulty in the encoding task. In the **easy level**, the word consists of **two** letters. In the **medium level**, the word consists of **three** letters. And in the **hard level**, the word consists of **four** letters. You will work on every level of difficulty once in random order.

In each round, you will have 120 seconds to encode as many words as possible. You will receive 3 points for each word you encode correctly. Your payoff in this part of the experiment will be the sum of the points that you earned in all three rounds.

<i>Only in treatment Default:</i> The average of your performance in the first part of the experiment will be shown to you in the second part of the experiment as an orientation value. However, it will not affect your payoff in the second part.
--

In the beginning of this part of the experiment, there will be a trial round in which you will become familiar with the encoding task and be able to test it. In the trial round, you will have to encode 5 words for each level of difficulty without any time restriction. The actual experiment will not start until all participants have completed the trial round successfully.

Following the entire experiment, we will ask you to fill out a questionnaire. You will then receive your payment in cash. The exchange rate is 35 points = 1 Euro.

Control Questions for Participants of the Main Experiment - Part 1

- In the first part of the experiment, you will file a tax return. (false)
- In each round, you have 2 minutes to encode as many words as possible. (true)
- In order to earn points, you must translate numbers into letters. (false)
- Your payoff in this part of the experiment will be the average of the number of points you have earned in the three rounds. (false)

Instructions for Participants of the Experiment - Part 2

The following section provides the English translation of the original German instructions that the participants received after completing the first part of the experiment:

Detailed description of the second part of the experiment

1. In this part of the experiment, you can again earn points by first solving the same tasks you were given in the previous part of the experiment. You will again work on all three different levels of difficulty in random order. The only difference from the task in the first part is that you will now earn **30** points for every correctly encoded word.
2. Subsequently to **each** round of encoding, you will declare your income in a tax return. This declared income will be subject to an income tax at a rate of 30%. For example, imagine you have earned an income in the encoding task of 90 points but chose only to declare an income of 50 points; you will then pay an amount of $0.3 \times 50 = 15$ points in income tax.

gross income	declaration	paid taxes
90	50	15

Table 1: Example with gross income = 90 and declaration = 50

The screen of the income declaration looks like this:

— *Screenshot of the income declaration* —

You enter the income that you want to declare to the tax authority into the highlighted box.

Only in treatment Default: By default, this box is prefilled with the income you would have earned with your average performance from the first part of the experiment. If, for example, you encoded an average of five words in the first part, you will see a default value of $5 \times 30 = 150$. You can change this value as you wish without any consequences. In particular, changing this value will not affect the likelihood of your tax declaration being audited by the tax authority (see 3).

Your declared income can be lower than your actual income. The maximum possible value for your declared income is either your actual income or the above-mentioned default value - whichever is greater.

Only in treatment NoDefault: Your declared income can be lower but not higher than your actual income.

By clicking “Calculate,” you will see how many points you would pay in income tax and how large your remaining net income would be. Besides providing this information, clicking on “Calculate” does not have any consequences. Only clicking “Declare Income” finalizes your input and declaration.

3. Tax audit

At the end of each round, your tax return will have a probability of 15% of being audited. If you are audited and found to have declared less income than your actual income, you will have to pay the unpaid taxes plus a fine of the same amount. Hence, if you are audited you will pay double the amount of taxes that you initially did not pay.

Only in treatment Default: If your declared income is more than your actual income, you will not receive a refund from the tax authority.

4.A. APPENDIX

In the following two tables, the above-mentioned example is shown completed, once in the case of a tax audit and once in the case of no tax audit.

gross income	declaration	paid taxes	unpaid taxes	deduction of points	net income
90	50	15	12	24	51

Table 2a: Example case with actual income > declared income **with** a tax audit

gross income	declaration	paid taxes	unpaid taxes	deduction of points	net income
90	50	15	12	0	75

Table 2b: Example case with actual income > declared income **without** a tax audit

For example, imagine having earned an income of 90 points and declaring an income of 50 points; you will receive a net income of 51 points if your tax return is audited (with a 15% likelihood) and a net income of 75 points if your tax return is not audited (with a 85% likelihood).

The second part of the experiment consists of **three** rounds as well, so you will again work on every level of difficulty in random order. You will only learn whether your tax return in each round has been audited after all three rounds have been completed. Hence, you will not get this information at the end of each round. At the end of every round, you will only receive information about your earned income, how many points are being deducted as income tax based on your declared income, and your net income in the case of a tax audit as well as in the case of no tax audit.

At the end of this part of the experiment, the computer will randomly choose one of the three previously conducted rounds. Only your decisions in this round will be relevant for your payoff. You will not receive any payoff for the other two rounds that were not chosen by the computer. Therefore, you should think carefully about your decisions in each round.

Before starting this part of the experiment, we will again ask you a few questions on your screen. As in the first part, these questions do not affect your payment and this part of the experiment will only start when all participants are completely familiar with the procedure of the experiment and have answered the

questions correctly. This part of the experiment is followed by a short questionnaire. You will then receive your payment in cash. The exchange rate is 35 points = 1 Euro.

Control Questions for Participants of the Main Experiment - Part 2

- If you earn an income of 300 points and declare an income of 150 points, you will pay 90 points in taxes. (false)
- If you declare an income that is lower than your actual income and you are audited, you will have to pay double the amount of unpaid taxes (additional tax payment + fine). (true)
- The probability that your tax return will be audited is 50%. (false)
- You will only learn at the end of the experiment whether or not your tax returns have been audited by the tax authority. (true)
- Your payoff from this part of the experiment is equal to your net income (income after taxes and possible fines) earned from one randomly selected round. (true)

4.A.3 Screenshots

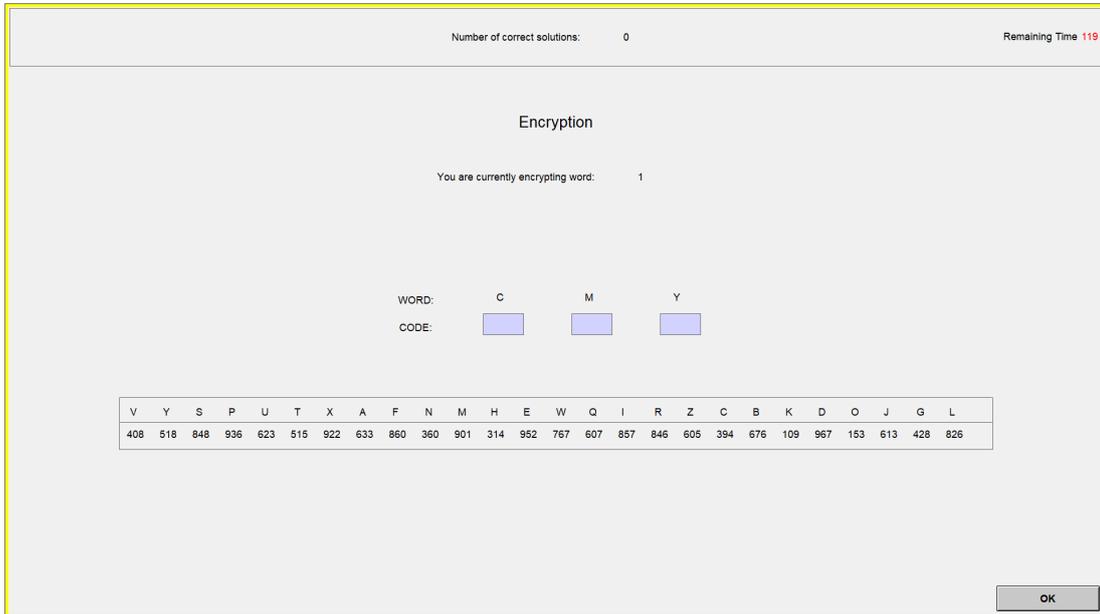


Figure 4.A.1: Real-effort task (medium difficulty level; English translation)

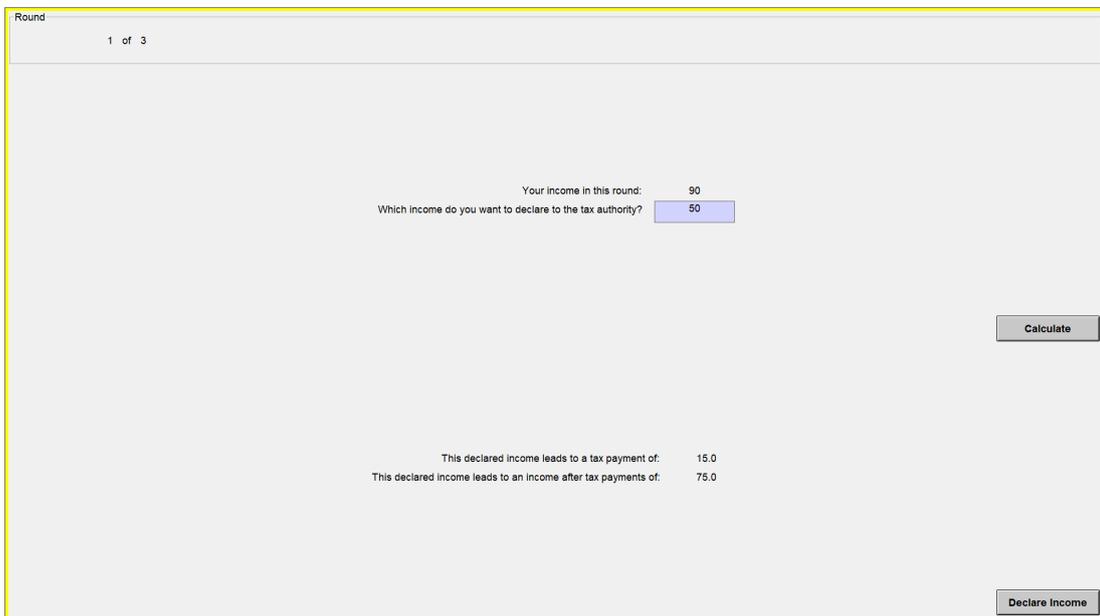


Figure 4.A.2: Filing a tax return (English translation)

Bibliography

- Adena, Maja and Steffen Huck (2016) “A field experiment on crowdfunding for a club good,” Discussion Paper (SP II 2016-308), Wissenschaftszentrum Berlin für Sozialforschung, URL: <https://bibliothek.wzb.eu/pdf/2016/ii16-308.pdf>, accessed on September 18, 2018.
- Adena, Maja, Steffen Huck, and Imran Rasul (2014) “Charitable Giving and Non-binding Contribution-Level Suggestions - Evidence from a Field Experiment,” *Review of Behavioral Economics*, 1 (3), pp. 275–293.
- Aguiar, Fernando, Pablo Brañas-Garza, Ramón Cobo-Reyes, Natalia Jimenez, and Luis M. Miller (2009) “Are women expected to be more generous?” *Experimental Economics*, 12 (1), pp. 93–98.
- Allingham, Michael G. and Agnar Sandmo (1972) “Income tax evasion: A theoretical analysis,” *Journal of Public Economics*, 1 (3), pp. 323–338.
- Alm, James (2010) “Testing Behavioral Public Economics Theories in the Laboratory,” *National Tax Journal*, 63 (4), pp. 635–658.
- (2012) “Measuring, explaining, and controlling tax evasion: Lessons from theory, experiments, and field studies,” *International Tax and Public Finance*, 19 (1), pp. 54–77.
- Alm, James, Kim M. Bloomquist, and Michael McKee (2015) “On the External Validity of Laboratory Tax Compliance Experiments,” *Economic Inquiry*, 53 (2), pp. 1170–1186.
- Althoff, Tim, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky (2014) “How to Ask for a Favor: A Case Study on the Success of Altruistic Requests,” Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media, Association for the Advancement of Artificial Intelli-

BIBLIOGRAPHY

gence, URL: <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/download/8106/8101>, accessed on September 18, 2018.

Andreoni, James, Eleanor Brown, and Isaac Rischall (2003) “Charitable Giving by Married Couples: Who Decides and Why Does it Matter?” *Journal of Human Resources*, 38 (1), pp. 111–133.

Andreoni, James, Brian Erard, and Jonathan Feinstein (1998) “Tax Compliance,” *Journal of Economic Literature*, 36 (2), pp. 818–860.

Andreoni, James and Justin M. Rao (2011) “The power of asking: How communication affects selfishness, empathy, and altruism,” *Journal of Public Economics*, 95 (7–8), pp. 513–520.

Andreoni, James, Justin M. Rao, and Hannah Trachtman (2017) “Avoiding The Ask: A Field Experiment on Altruism, Empathy, and Charitable Giving,” *Journal of Political Economy*, 125 (3), pp. 625–653.

Andreoni, James and Lise Vesterlund (2001) “Which is the Fair Sex? Gender Differences in Altruism,” *Quarterly Journal of Economics*, 116 (1), pp. 293–312.

Ariely, Dan, George Loewenstein, and Drazen Prelec (2003) “‘Coherent Arbitrariness’: Stable Demand Curves Without Stable Preferences,” *Quarterly Journal of Economics*, 118 (1), pp. 73–106.

Babcock, Linda and Sara Laschever (2003) *Women don’t ask: Negotiation and the gender divide*: Princeton University Press, Princeton, NJ.

Banerjee, Priyodorshi and Sujoy Chakravarty (2012) “The effect of minimal group framing in a dictator game experiment,” Working Paper, URL: <http://dx.doi.org/10.2139/ssrn.2071006>, accessed on September 18, 2018.

Bankman, Joseph (2008) “Using Technology to Simplify Individual Tax Filing,” *National Tax Journal*, 61 (4), pp. 773–789.

Bankman, Joseph, Clifford Nass, and Joel Slemrod (2015) “Using the ‘Smart Return’ to Reduce Evasion and Simplify Tax Filing,” *Tax Law Review*, 69, p. 459.

Bardsley, Nicholas (2008) “Dictator game giving: Altruism or artefact?” *Experimental Economics*, 11 (2), pp. 122–133.

BIBLIOGRAPHY

- Batson, C. Daniel (1991) *The Altruism Question: Toward a Social-Psychological Answer*: Lawrence Erlbaum Associates, Hillsdale, NJ.
- (2002) “Addressing the altruism question experimentally,” in *Altruism & altruistic love: Science, philosophy, & religion in dialogue* eds. by Stephen G. Post, Lynn G. Underwood, Jeffrey P. Schloss, and William B. Hurlbut: Oxford University Press, New York, NY, pp. 89–105.
- Batson, C. Daniel, Janine L. Dyck, J. Randall Brandt, Judy G. Batson, Anne L. Powell, M. Rosalie McMaster, and Cari Griffitt (1988) “Five studies testing two new egoistic alternatives to the empathy-altruism hypothesis,” *Journal of Personality and Social Psychology*, 55 (1), pp. 52–77.
- Battigalli, Pierpaolo and Martin Dufwenberg (2007) “Guilt in Games,” *The American Economic Review*, 97 (2), pp. 170–176.
- Becker, Gary (1968) “Crime and Punishment: An Economic Approach,” *Journal of Political Economy*, 76 (2).
- Bekkers, René and Pamala Wiepking (2011) “A Literature Review of Empirical Studies of Philanthropy: Eight Mechanisms That Drive Charitable Giving,” *Nonprofit and Voluntary Sector Quarterly*, 40 (5), pp. 924–973.
- Bell, Robert Alan, Matthew Cholerton, Kevin E. Fraczek, Guy S. Rohlf, and Brian A. Smith (1994) “Encouraging donations to charity: A field study of competing and complementary factors in tactic sequencing,” *Western Journal of Communication*, 58 (2), pp. 98–115.
- Ben-Ner, Avner, Fanmin Kong, and Louis Putterman (2004) “Share and share alike? Gender-pairing, personality, and cognitive ability as determinants of giving,” *Journal of Economic Psychology*, 25 (5), pp. 581–589.
- Benndorf, Volker, Holger A. Rau, and Christian Sölch (2018) “Minimizing Learning Behavior in Repeated Real-Effort Tasks,” Working Paper, URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2503029, accessed on September 18, 2018.
- Bochet, Olivier, Talbot Page, and Louis Putterman (2006) “Communication and punishment in voluntary contribution experiments,” *Journal of Economic Behavior & Organization*, 60 (1), pp. 11–26.

BIBLIOGRAPHY

- Bohnet, Iris and Bruno S. Frey (1999) “The sound of silence in prisoner’s dilemma and dictator games,” *Journal of Economic Behavior & Organization*, 38 (1), pp. 43–57.
- Bolton, E. Gary, Elena Katok, and Rami Zwick (1998) “Dictator game giving: Rules of fairness versus acts of kindness,” *International Journal of Game Theory*, 27 (2), pp. 269–299.
- Bolton, Gary E. and Axel Ockenfels (2000) “ERC: A Theory of Equity, Reciprocity, and Competition,” *The American Economic Review*, 90 (1), pp. 166–193.
- Brañas-Garza, Pablo (2007) “Promoting helping behavior with framing in dictator games,” *Journal of Economic Psychology*, 28 (4), pp. 477–486.
- Brañas-Garza, Pablo, Valerio Capraro, and Ericka Rascon (2018) “Gender Differences in Altruism on Mechanical Turk: Expectations and Actual Behaviour,” Working Paper, URL: <https://dx.doi.org/10.2139/ssrn.2796221>, accessed on September 18, 2018.
- Brañas-Garza, Pablo, Miguel A. Durán, and Maria Paz Espinosa (2009) “The Role of Personal Involvement and Responsibility in Unfair Outcomes: A Classroom Investigation,” *Rationality and Society*, 21 (2), pp. 225–248.
- Brañas-Garza, Pablo, Ismael Rodríguez-Lara, and Angel Sánchez (2017) “Humans expect generosity,” *Scientific Reports*, 7 (42446).
- Brennan, Linda and Wayne Binney (2010) “Fear, guilt, and shame appeals in social marketing,” *Journal of Business Research*, 63 (2), pp. 140–146.
- Bruner, David, Michael Jones, Michael McKee, and Christian Vossler (2015) “Tax Reporting Behavior: Underreporting Opportunities and Prepopulated Tax Returns,” Department of Economics Working Paper (15-11), Appalachian State University, URL: <http://econ.appstate.edu/RePEc/pdf/wp1511.pdf>, accessed on September 18, 2018.
- Camerer, Colin and Richard H. Thaler (1995) “Anomalies: Ultimatums, Dictators and Manners,” *Journal of Economic Perspectives*, 9 (2), pp. 209–219.
- Capraro, Valerio, Glorianna Jagfeld, Rana Klein, Mathijs Mul, and Iris van de Pol (2018) “What’s the Right Thing to Do? Increasing Pro-Sociality with

BIBLIOGRAPHY

- Simple Moral Nudges,” Working Paper, URL: <http://dx.doi.org/10.2139/ssrn.3064004>, accessed on September 18, 2018.
- Charness, Gary and Martin Dufwenberg (2006) “Promises and Partnership,” *Econometrica*, 74 (6), pp. 1579–1601.
- Charness, Gary and Uri Gneezy (2008) “What’s in a name? Anonymity and social distance in dictator and ultimatum games,” *Journal of Economic Behavior & Organization*, 68 (1), pp. 29–35.
- Charness, Gary and Matthew Rabin (2002) “Understanding Social Preferences with Simple Tests,” *Quarterly Journal of Economics*, 117 (3), pp. 817–869.
- (2005) “Expressed preferences and behavior in experimental games,” *Games and Economic Behavior*, 53 (2), pp. 151–169.
- Cherry, Todd L., Peter Frykblom, and Jason F. Shogren (2002) “Hardnose the Dictator,” *The American Economic Review*, 92 (4), pp. 1218–1221.
- Cox, James C., John A. List, Michael Price, Vjollca Sadiraj, and Anya Samek (2016) “Moral Costs and Rational Choice: Theory and Experimental Evidence,” NBER Working Paper (No. 22234), National Bureau of Economic Research, URL: <http://www.nber.org/papers/w22234>, accessed on September 18, 2018.
- Crosetto, Paolo and Antonio Filippin (2013) “The ‘bomb’ risk elicitation task,” *Journal of Risk and Uncertainty*, 47 (1), pp. 31–65.
- Croson, Rachel and Uri Gneezy (2009) “Gender Differences in Preferences,” *Journal of Economic Literature*, 47 (2), pp. 448–474.
- Cushman, Fiery, Liane Young, and Marc Hauser (2006) “The role of conscious reasoning and intuition in moral judgment: testing three principles of harm,” *Psychological Science*, 17 (12), pp. 1082–1089.
- Dal Bó, Ernesto and Pedro Dal Bó (2014) “‘Do the right thing:’ The effects of moral suasion on cooperation,” *Journal of Public Economics*, 117, pp. 28–38.
- Dana, Jason, Roberto A. Weber, and Jason Xi Kuang (2007) “Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness,” *Economic Theory*, 33 (1), pp. 67–80.

BIBLIOGRAPHY

- Das, Enny, Peter Kerkhof, and Joyce Kuiper (2008) “Improving the Effectiveness of Fundraising Messages: The Impact of Charity Goal Attainment, Message Framing, and Evidence on Persuasion,” *Journal of Applied Communication Research*, 36 (2), pp. 161–175.
- Davis, Mark H. (1983) “Measuring individual differences in empathy: Evidence for a multidimensional approach,” *Journal of Personality and Social Psychology*, 44 (1), pp. 113–126.
- Deaux, Kay and Brenda Major (1987) “Putting gender into context: An interactive model of gender-related behavior,” *Psychological Review*, 94 (3), pp. 369–389.
- DellaVigna, Stefano, John A. List, and Ulrike Malmendier (2012) “Testing for Altruism and Social Pressure in Charitable Giving,” *Quarterly Journal of Economics*, 127 (1), pp. 1–56.
- DellaVigna, Stefano, John A List, Ulrike Malmendier, and Gautam Rao (2013) “The importance of being marginal: Gender differences in generosity,” *The American Economic Review: Papers and Proceedings*, 103 (3), pp. 586–590.
- Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner (2011) “Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences,” *Journal of the European Economic Association*, 9 (3), pp. 522–550.
- Dreber, Anna, Tore Ellingsen, Magnus Johannesson, and David G. Rand (2013) “Do people care about social context? Framing effects in dictator games,” *Experimental Economics*, 16 (3), pp. 349–371.
- Dufwenberg, Martin and Astri Muren (2006) “Generosity, anonymity, gender,” *Journal of Economic Behavior & Organization*, 61 (1), pp. 42–49.
- Eagly, Alice H. (1983) “Gender and social influence: A social psychological analysis,” *American Psychologist*, 38 (9), pp. 971–981.
- (2009) “The His and Hers of Prosocial Behavior: An Examination of the Social Psychology of Gender,” *American Psychologist*, 64 (8), pp. 644–658.
- Eckel, Catherine C. and Philip J. Grossman (1996) “Altruism in Anonymous Dictator Games,” *Games and Economic Behavior*, 16 (2), pp. 181–191.

BIBLIOGRAPHY

- Eckel, Catherine, Philip J. Grossman, and Angela Milano (2007) “Is More Information Always Better? An Experimental Study of Charitable Giving and Hurricane Katrina,” *Southern Economic Journal*, 74 (2), pp. 388–411.
- Economidou-Kogetsidis, Maria (2015) “Teaching email politeness in the EFL/ESL classroom,” *ELT Journal*, 69 (4), pp. 415–424.
- Edele, Aileen, Isabel Dziobek, and Monika Keller (2013) “Explaining altruistic sharing in the dictator game: The role of affective empathy, cognitive empathy, and justice sensitivity,” *Learning and Individual Differences*, 24, pp. 96–102.
- Edwards, James T. and John A. List (2014) “Toward an understanding of why suggestions work in charitable fundraising: Theory and evidence from a natural field experiment,” *Journal of Public Economics*, 114, pp. 1–13.
- Eisenhardt, Kathleen M. (1989) “Agency Theory: An Assessment and Review,” *Academy of Management Review*, 14 (1), pp. 57–74.
- Ellingsen, Tore and Magnus Johannesson (2008) “Anticipated verbal feedback induces altruistic behavior,” *Evolution and Human Behavior*, 29 (2), pp. 100–105.
- Ellingsen, Tore, Magnus Johannesson, Johanna Mollerstrom, and Sara Munkhammar (2013) “Gender differences in social framing effects,” *Economics Letters*, 118 (3), pp. 470–472.
- Ellingsen, Tore, Magnus Johannesson, Sigve Tjøtta, and Gaute Torsvik (2010) “Testing guilt aversion,” *Games and Economic Behavior*, 68 (1), pp. 95–107.
- Engel, Christoph (2011) “Dictator games: a meta study,” *Experimental Economics*, 14 (4), pp. 583–610.
- Engelmann, Dirk and Martin Strobel (2006) “Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments: Reply,” *The American Economic Review*, 96 (5), pp. 1918–1923.
- Epley, Nicholas and Thomas Gilovich (2001) “Putting Adjustment Back in the Anchoring and Adjustment Heuristic: Differential Processing of Self-Generated and Experimenter-Provided Anchors,” *Psychological Science*, 12 (5), pp. 391–396.
- Erat, Sanjiv and Uri Gneezy (2011) “White Lies,” *Management Science*, 58 (4), pp. 723–733.

BIBLIOGRAPHY

- Erkal, Nisvan, Lata Gangadharan, and Nikos Nikiforakis (2011) “Relative Earnings and Giving in a Real-Effort Experiment,” *The American Economic Review*, 101 (7), pp. 3330–3348.
- Etang, Alvin, David Fielding, and Stephen Knowles (2016) “Who votes expressively, and why? Experimental evidence,” *Bulletin of Economic Research*, 68 (2), pp. 105–116.
- Exadaktylos, Filippos, Antonio M. Espín, and Pablo Brañas-Garza (2013) “Experimental subjects are not different,” *Scientific Reports*, 3 (1213).
- Falk, Armin and Michael Kosfeld (2006) “The Hidden Costs of Control,” *The American Economic Review*, 96 (5), pp. 1611–1630.
- Fehr, Ernst and Urs Fischbacher (2003) “The nature of human altruism,” *Nature*, 425, pp. 785–791.
- Fehr, Ernst and Klaus M. Schmidt (1999) “A Theory of Fairness, Competition, and Cooperation,” *Quarterly Journal of Economics*, 114 (3), pp. 817–868.
- (2006) “Chapter 8: The Economics of Fairness, Reciprocity and Altruism – Experimental Evidence and New Theories,” in *Handbook of the Economics of Giving, Altruism and Reciprocity : Foundations* eds. by Serge-Christophe Kolm and Jean Mercier Ythier, 1: Elsevier, Amsterdam, pp. 615–691.
- Fischbacher, Urs (2007) “z-Tree: Zurich toolbox for ready-made economic experiments,” *Experimental Economics*, 10 (2), pp. 171–178.
- Fochmann, Martin, Nadja Müller, and Michael Overesch (2018) “Less Cheating? The Effects of Prefilled Forms on Compliance Behavior,” Working Paper, URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3206410, accessed on September 18, 2018.
- Fong, Christina M. (2007) “Evidence from an Experiment on Charity to Welfare Recipients: Reciprocity, Altruism and the Empathic Responsiveness Hypothesis,” *Economic Journal*, 117 (522), pp. 1008–1024.
- Fonseca, Miguel A. and Shaun B. Grimshaw (2017) “Do Behavioral Nudges in Prepopulated Tax Forms Affect Compliance? Experimental Evidence with Real Taxpayers,” *Journal of Public Policy & Marketing*, 36 (2), pp. 213–226.

BIBLIOGRAPHY

- Forsythe, Robert, Joel L. Horowitz, N. E. Savin, and Martin Sefton (1994) “Fairness in Simple Bargaining Experiments,” *Games and Economic Behavior*, 6 (3), pp. 347–369.
- Franzen, Axel and Sonja Pointner (2012) “Anonymity in the dictator game revisited,” *Journal of Economic Behavior & Organization*, 81 (1), pp. 74–81.
- Furnham, Adrian and Hua Chu Boo (2011) “A literature review of the anchoring effect,” *The Journal of Socio-Economics*, 40 (1), pp. 35–42.
- Gerlach, Philipp and Bastian Jaeger (2016) “Another frame, another game? Explaining framing effects in economic games,” Working Paper, URL: <https://dx.doi.org/10.17605/OSF.IO/YF36N>, accessed on September 18, 2018.
- Gillitzer, Christian and Peer Ebbesen Skov (2018) “The use of third-party information reporting for tax deductions: Evidence and implications from charitable deductions in Denmark,” *Oxford Economic Papers*, 70 (3), pp. 892–916.
- Goei, Ryan, Lisa L. Massi Lindsey, Franklin J. Boster, Paul D. Skalski, and Jonathan M. Bowman (2003) “The Mediating Roles of Liking and Obligation on the Relationship between Favors and Compliance,” *Communication Research*, 30 (2), pp. 178–197.
- Goeree, Jacob K., Margaret A. McConnell, Tiffany Mitchell, Tracey Tromp, and Leeat Yariv (2010) “The 1/d Law of Giving,” *American Economic Journal: Microeconomics*, 2 (1), pp. 183–203.
- Goolsbee, Austan (2006) “The Simple Return: Reducing America’s Tax Burden Through Return-Free Filing,” Discussion Paper (2006-04), The Brookings Institution, URL: http://www.hamiltonproject.org/papers/the_simple_return_reducing_americas_tax_burden_through_return-free_fil/, accessed on September 18, 2018.
- Greiner, Ben (2015) “Subject pool recruitment procedures: organizing experiments with ORSEE,” *Journal of the Economic Science Association*, 1 (1), pp. 114–125.
- Grossman, Philip J. and Catherine C. Eckel (2015) “Giving versus taking for a cause,” *Economics Letters*, 132, pp. 28–30.

BIBLIOGRAPHY

- Haley, Kevin J. and Daniel M.T. Fessler (2005) “Nobody’s watching? Subtle cues affect generosity in an anonymous economic game,” *Evolution and Human Behavior*, 26 (3), pp. 245–256.
- Hallsworth, Michael (2014) “The use of field experiments to increase tax compliance,” *Oxford Review of Economic Policy*, 30 (4), pp. 658–679.
- Harrison, Glenn W., Eric Johnson, Melayne M. McInnes, and E. Elisabet Rutström (2005) “Risk Aversion and Incentive Effects: Comment,” *The American Economic Review*, 95 (3), pp. 897–901.
- Heilman, Madeline E. and Julie J. Chen (2005) “Same behavior, different consequences: Reactions to men’s and women’s altruistic citizenship behavior,” *Journal of Applied Psychology*, 90 (3), pp. 431–441.
- Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, and Vernon Smith (1994) “Preferences, Property Rights, and Anonymity in Bargaining Games,” *Games and Economic Behavior*, 7 (3), pp. 346–380.
- Hoffman, Elizabeth, Kevin McCabe, and Vernon L. Smith (1996) “Social Distance and Other-Regarding Behavior in Dictator Games,” *The American Economic Review*, 86 (3), pp. 653–660.
- Hofmann, Eva, Erik Hoelzl, and Erich Kirchler (2008) “Preconditions of Voluntary Tax Compliance,” *Journal of Psychology*, 216 (4), pp. 209–217.
- Isaac, R. Mark and James M. Walker (1988) “Communication and Free-Riding Behavior: The Voluntary Contribution Mechanism,” *Economic Inquiry*, 26 (4), pp. 585–608.
- Jacobsen, Catrine, Toke Reinholt Fosgaard, and David Pascual-Ezama (2018) “Why do we lie? A practical guide to the dishonesty literature,” *Journal of Economic Surveys*, 32 (2), pp. 357–387.
- Johnson, Eric J. and Daniel Goldstein (2003) “Do Defaults Save Lives?” *Science*, 302 (5649), pp. 1338–1339.
- Kahn, Arnold, Joe Hottes, and William L. Davis (1971) “Cooperation and optimal responding in the Prisoner’s Dilemma game: Effects of sex and physical attractiveness,” *Journal of Personality and Social Psychology*, 17 (3), pp. 267–279.

BIBLIOGRAPHY

- Kahneman, Daniel (2011) *Thinking, fast and slow*: Farrar Straus and Giroux, New York.
- Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler (1986) “Fairness and the Assumptions of Economics,” *The Journal of Business*, 59 (4), pp. S285–S300.
- (1991) “Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias,” *Journal of Economic Perspectives*, 5 (1), pp. 193–206.
- Katzev, Richard D. and Theodore R. Johnson (1984) “Comparing the Effects of Monetary Incentives and Foot-in-the-Door Strategies in Promoting Residential Electricity Conservation,” *Journal of Applied Social Psychology*, 14 (1), pp. 12–27.
- Kleven, Henrik Jacobsen, Martin B. Knudsen, Claus Thustrup Kreiner, Søren Pedersen, and Emmanuel Saez (2011) “Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark,” *Econometrica*, 79 (3), pp. 651–692.
- Korenok, Oleg, Edward L. Millner, and Laura Razzolini (2014) “Taking, giving, and impure altruism in dictator games,” *Experimental Economics*, 17 (3), pp. 488–500.
- Korenok, Oleg, Edward Millner, and Laura Razzolini (2017) “Feelings of ownership in dictator games,” *Journal of Economic Psychology*, 61 (C), pp. 145–151.
- Kotakorpi, Kaisa and Jani-Petri Laamanen (2016) “Prefilled income tax returns and tax compliance: Evidence from a natural experiment,” Tampere Economic Working Papers (104), University of Tampere, URL: <http://tampub.uta.fi/bitstream/handle/10024/98986/978-952-03-0150-7.pdf>, accessed on September 18, 2018.
- Krupka, Erin L. and Roberto A. Weber (2013) “Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?” *Journal of the European Economic Association*, 11 (3), pp. 495–524.
- Langenbach, Pascal (2016) “The values of ex-ante and ex-post communication in dictator games,” Preprints (2014/7), Max Planck Institute for Research on Collective Goods, URL: <http://hdl.handle.net/11858/00-001M-0000-002A-3379-7>, accessed on September 18, 2018.

BIBLIOGRAPHY

- Lee, Hye Eun, Hee Sun Park, Tatsuya Imai, and Daniel Dolan (2012) “Cultural Differences Between Japan and the United States in Uses of ‘Apology’ and ‘Thank You’ in Favor Asking Messages,” *Journal of Language and Social Psychology*, 31 (3), pp. 263–289.
- Levitt, Steven D. and John A. List (2007) “What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?” *Journal of Economic Perspectives*, 21 (2), pp. 153–174.
- List, John A. (2007) “On the Interpretation of Giving in Dictator Games,” *Journal of Political Economy*, 115 (3), pp. 482–493.
- Madrian, Brigitte C. and Dennis F. Shea (2001) “The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior,” *Quarterly Journal of Economics*, 116 (4), pp. 1149–1187.
- Mascagni, Giulia (2018) “From the lab to the field: A review of tax experiments,” *Journal of Economic Surveys*, 32 (2), pp. 273–301.
- Mazar, Nina, On Amir, and Dan Ariely (2008) “The Dishonesty of Honest People: A Theory of Self-Concept Maintenance,” *Journal of Marketing Research*, 45 (6), pp. 633–644.
- Mazar, Nina and Scott A. Hawkins (2015) “Choice architecture in conflicts of interest: Defaults as physical and psychological barriers to (dis)honesty,” *Journal of Experimental Social Psychology*, 59, pp. 113–117.
- Meer, Jonathan and Harvey S. Rosen (2011) “The ABCs of charitable solicitation,” *Journal of Public Economics*, 95 (5), pp. 363–371.
- Mitra, Tanushree and Eric Gilbert (2014) “The language that gets people to give,” in *CSCW ’14* eds. by Susan Fussell, Wayne Lutters, Meredith Ringel Morris, and Madhu Reddy, pp. 49–61: ACM, New York.
- Mizutani, Osamu and Nobuko Mizutani (1989) *Nihongo Notes 9: Situational Japanese 4*: The Japan Times, Tokyo.
- Mohlin, Erik and Magnus Johannesson (2008) “Communication: Content or relationship?” *Journal of Economic Behavior & Organization*, 65 (3–4), pp. 409–419.
- Norrick, Neal R. (1978) “Expressive illocutionary acts,” *Journal of Pragmatics*, 2 (3), pp. 277–291.

BIBLIOGRAPHY

- OECD (2017) *Tax Administration 2017: Comparative Information on OECD and Other Advanced and Emerging Economies*: OECD Publishing, Paris, URL: http://dx.doi.org/10.1787/tax_admin-2017-en, accessed on September 18, 2018.
- Ohashi, Jun (2006) “Japanese culture specific face and politeness orientation: A pragmatic investigation of yoroshiku onegaishimasu,” *Multilingua*, 22 (3), pp. 257–274.
- Petit, Emmanuel, Anna Tcherkassof, and Xavier Gassmann (2012) “Sincere Giving and Shame in a Dictator Game,” *Cahiers du GREThA* (2012-25), Groupe de Recherche en Economie Théorique et Appliquée, URL: <https://ideas.repec.org/p/grt/wpegrt/2012-25.html>, accessed on September 18, 2018.
- Ploner, Matteo and Tobias Regner (2013) “Self-image and moral balancing: An experimental analysis,” *Journal of Economic Behavior & Organization*, 93, pp. 374–383.
- Rabin, Matthew (1993) “Incorporating Fairness into Game Theory and Economics,” *The American Economic Review*, 83 (5), pp. 1281–1302.
- Rankin, Frederick W. (2006) “Requests and social distance in dictator games,” *Journal of Economic Behavior & Organization*, 60 (1), pp. 27–36.
- Reeson, Andrew and Simon Dunstall (2009) “Behavioural Economics and Complex Decision-Making - Implications for the Australian Tax and Transfer System,” CMIS Report (09/110), CSIRO Australia, URL: https://mosaicprojects.com.au/PDF/Complex_Decision_Making.pdf, accessed on September 18, 2018.
- Rigdon, Mary, Keiko Ishii, Motoki Watabe, and Shinobu Kitayama (2009) “Minimal social cues in the dictator game,” *Journal of Economic Psychology*, 30 (3), pp. 358–367.
- Ritov, Ilana and Jonathan Baron (1992) “Status-quo and omission biases,” *Journal of Risk and Uncertainty*, 5 (1), pp. 49–61.
- Samuelson, William and Richard Zeckhauser (1988) “Status quo bias in decision making,” *Journal of Risk and Uncertainty*, 1 (1), pp. 7–59.

BIBLIOGRAPHY

- Sanders, Michael and Sarah Smith (2016) “Can simple prompts increase bequest giving? Field evidence from a legal call centre,” *Journal of Economic Behavior & Organization*, 125, pp. 179–191.
- Scott, Carol A. (1976) “The Effects of Trial and Incentives on Repeat Purchase Behavior,” *Journal of Marketing Research*, 13 (3), pp. 263–269.
- Simon, Herbert A. (1957) *Models of man: social and rational.*: Wiley, New York.
- Slemrod, Joel (2007) “Cheating Ourselves: The Economics of Tax Evasion,” *Journal of Economic Perspectives*, 21 (1), pp. 25–48.
- Slemrod, Joel B. (2016) “Tax Compliance and Enforcement: New Research and Its Policy Implications,” Ross School of Business Working Paper (1302), University of Michigan, URL: <https://dx.doi.org/10.2139/ssrn.2726077>, accessed on September 18, 2018.
- Small, Deborah A. and George Loewenstein (2003) “Helping a Victim or Helping the Victim: Altruism and Identifiability,” *Journal of Risk and Uncertainty*, 26 (1), pp. 5–16.
- Spranca, Mark, Elisa Minsk, and Jonathan Baron (1991) “Omission and commission in judgment and choice,” *Journal of Experimental Social Psychology*, 27 (1), pp. 76–105.
- Sugden, Robert (2017) “Do people really want to be nudged towards healthy lifestyles?” *International Review of Economics*, 64 (2), pp. 113–123.
- Sunstein, Cass R. (2018) “‘Better off, as judged by themselves’: a comment on evaluating nudges,” *International Review of Economics*, 65 (1), pp. 1–8.
- Supphellen, Magne and Michelle R. Nelson (2001) “Developing, exploring, and validating a typology of private philanthropic decision making,” *Journal of Economic Psychology*, 22 (5), pp. 573–603.
- Suvoy, Ryan (2003) “The Effects of Give and Take Framing in a Dictator Game,” Honors Thesis, URL: <http://hdl.handle.net/1794/896>, accessed on September 18, 2018.
- Thaler, Richard H. and Cass R. Sunstein (2008) *Nudge: Improving Decisions about Health, Wealth, and Happiness*: Yale University Press, New Haven.

- Tonin, Mirco and Michael Vlassopoulos (2013) “Experimental evidence of self-image concerns as motivation for giving,” *Journal of Economic Behavior & Organization*, 90, pp. 19–27.
- Torgler, Benno (2002) “Speaking to Theorists and Searching for Facts: Tax Morale and Tax Compliance in Experiments,” *Journal of Economic Surveys*, 16 (5), pp. 657–683.
- Tversky, Amos and Daniel Kahneman (1974) “Judgment under Uncertainty: Heuristics and Biases,” *Science*, 185 (4157), pp. 1124–1131.
- de Vignemont, Frederique and Tania Singer (2006) “The empathic brain: how, when and why?” *Trends in Cognitive Sciences*, 10 (10), pp. 435–441.
- de Wit, Arjen and René Bekkers (2016) “Exploring Gender Differences in Charitable Giving: The Dutch Case,” *Nonprofit and Voluntary Sector Quarterly*, 45 (4), pp. 741–761.
- Xiao, Erte and Daniel Houser (2009) “Avoiding the sharp tongue: Anticipated written messages promote fair economic exchange,” *Journal of Economic Psychology*, 30 (3), pp. 393–404.
- Yamamori, Tetsuo, Kazuhiko Kato, Toshiji Kawagoe, and Akihiko Matsui (2007) “Voice matters in a dictator game,” *Experimental Economics*, 11 (4), pp. 336–343.
- Yörük, Barış K. (2009) “How responsive are charitable donors to requests to give?” *Journal of Public Economics*, 93 (9–10), pp. 1111–1117.
- Zhao, Kun, Eamonn Ferguson, and Luke D. Smillie (2017) “Individual Differences in Good Manners Rather Than Compassion Predict Fair Allocations of Wealth in the Dictator Game,” *Journal of Personality*, 85 (2), pp. 244–256.
- Zizzo, Daniel John (2010) “Experimenter demand effects in economic experiments,” *Experimental Economics*, 13 (1), pp. 75–98.

Abstract (English Version)

There are numerous situations in which people ask for something or make a request, e.g. asking a favor, asking for help or requesting compliance with specific norms. For this reason, how to ask for something in order to increase people's willingness to fulfill such requests is one of the most important question for many people working in various different fields of responsibility such as charitable giving, marketing, management or policy making.

This dissertation consists of four chapters that deal with the effects of small changes in the decision-making environment on altruistic decision-making and compliance behavior. Most notably, written communication as an influencing factor is the focus of the first three chapters. The starting point was the question how to devise a request in order to maximize its chance of success (Chapter 1). The results of the first chapter originate the ideas for the second and third chapter. Chapter 2 analyzes how communication by a neutral third-party, i.e. a text from the experimenters that either reminds potential benefactors of their responsibility or highlights their freedom of choice, affects altruistic decision-making. Chapter 3 elaborates on the effect of thanking people in advance when asking them for help. While being not as closely related to the other chapters as the three first ones are, the fourth chapter deals as well with the question how compliance (here: compliance with norms and rules) is affected by subtle manipulations of the environment in which decisions are made. This chapter analyzes the effect of default settings in a tax return on tax compliance.

In order to study the research questions outlined above, controlled experiments were conducted. Chapter 1, which analyzes the effect of text messages on the decision to give something to another person, employs a mini-dictator game. The recipient sends a free-form text message to the dictator before the latter makes a binary decision whether or not to give part of her or his endowment to the recipient. We find that putting effort into the message by writing a long note without spelling mistakes increases dictators' willingness to give. Moreover, writ-

ing in a humorous way and mentioning reasons why the money is needed pays off. Furthermore, men and women seem to react differently to some message categories. Only men react positively to efficiency arguments, while only women react to messages that emphasize the dictator's power and responsibility.

Building on this last result, Chapter 2 attempts to disentangle the effect of reminding potential benefactors of their responsibility for the potential beneficiary and the effect of highlighting their decision power and freedom of choice on altruistic decision-making by studying the effects of two different texts on giving in a dictator game. We find that only men react positively to a text that stresses their responsibility for the recipient by giving more to her or him, whereas only women seem to react positively to a text that emphasizes their decision power and freedom of choice.

Chapter 3 focuses on the compliance with a request. In the experiment, participants are asked to provide a detailed answer to an open question. Compliance is measured by the effort participants spend on answering the question. The treatment variable is whether or not they see the text "thanks in advance." We find that participants react negatively by putting less effort into complying with the request in response to the phrase "thanks in advance."

Chapter 4 studies the effect of prefilled tax returns with mostly inaccurate default values on tax compliance. In a laboratory experiment, participants earn income by performing a real-effort task and must subsequently file a tax return for three consecutive rounds. In the main treatment, the tax return is prefilled with a default value, resulting from participants' own performance in previous rounds, which varies in its relative size. The results suggest that there is no lasting effect of a default value on tax honesty, neither for relatively low nor relatively high defaults. However, participants who face a default that is lower than their true income in the first round evade significantly and substantially more taxes in this round than participants in the control treatment without a default.

Abstract (German Version)

Es fallen einem zahlreiche Situationen ein, in denen Menschen um etwas bitten, sei es um Unterstützung oder um das Befolgen bestimmter Regeln oder Normen. Die Frage, wie man eine Bitte formulieren muss, um die Bereitschaft der Menschen zu erhöhen, diese Bitte zu erfüllen, ist daher eine äußerst wichtige für viele Menschen, die in ganz unterschiedlichen Bereichen arbeiten, beispielsweise in Hilfsorganisationen, im Marketing, im Management oder auch in der Politikberatung.

Diese Dissertation besteht aus vier Kapiteln, die sich mit den Auswirkungen kleiner Veränderungen in der Entscheidungsumgebung auf altruistische Entscheidungen sowie unterstützendes und regelkonformes Verhalten befassen. Schriftliche Kommunikation als Einflussfaktor steht hierbei im Mittelpunkt der ersten drei Kapitel. Ausgangspunkt war die Frage, wie man eine Bitte formulieren muss, um ihre Erfolgsaussichten zu erhöhen (Kapitel 1). Aus den Ergebnissen des ersten Kapitels resultieren die Ideen für das zweite und dritte Kapitel. In Kapitel 2 wird untersucht, wie sich Kommunikation durch einen neutralen Dritten - ein Text der Experimentatoren, der potenzielle Wohltäter entweder an ihre Verantwortung oder ihre Entscheidungsfreiheit erinnert - auf eine altruistische Entscheidung auswirkt. In Kapitel 3 wird analysiert, was für einen Effekt es auf die Hilfsbereitschaft von Teilnehmern hat, wenn man ihnen bereits im Voraus für ihre Unterstützung dankt. Obwohl das vierte Kapitel weniger eng mit den anderen Kapiteln verbunden ist, als es die ersten drei sind, geht es auch hier um die Frage, wie sich die Bereitschaft der Teilnehmer, Normen zu befolgen, durch kleine Änderungen der Entscheidungsumgebung erhöhen lässt. In diesem Kapitel wird untersucht, welchen Einfluss vorausgefüllte Steuererklärungen auf die Steuerhinterziehung haben.

Zur Beantwortung der oben genannten Forschungsfragen wurden Experimente durchgeführt. In Kapitel 1 wird ein Mini-Diktatorspiel verwendet, um herauszufinden, welchen Effekt Textnachrichten auf die Entscheidung haben, einer anderen

Person etwas abzugeben. Die Person in der Rolle des Diktators liest eine vom Empfänger geschriebene Nachricht, bevor sie oder er die binäre Entscheidung trifft, einen Teil der Anfangsausstattung an den Empfänger abzugeben. Unsere Ergebnisse zeigen, dass Diktatoren eher bereit sind, etwas abzugeben, wenn sich die Empfänger beim Schreiben der Nachricht bemüht haben, indem sie längere Nachrichten mit weniger Rechtschreibfehlern schreiben. Außerdem erhöhen humorvolle Nachrichten und welche, die Gründe nennen, wofür die Empfänger das Geld brauchen, die Bereitschaft der Diktatoren, Geld abzugeben. Darüber hinaus scheinen Männer und Frauen auf einige Nachrichteninhalte unterschiedlich zu reagieren. Nur Männer reagieren positiv auf Effizienzargumente, während nur Frauen auf Nachrichten reagieren, die die Macht und Verantwortung des Diktators unterstreichen.

Aufbauend auf diesem letzten Ergebnis versucht das zweite Kapitel, die Effekte der Betonung der Entscheidungsmacht und der Betonung der Verantwortung auf altruistisches Handeln zu trennen. Hierzu wird der Effekt von zwei Texten auf Entscheidungen im Diktatorspiel untersucht. Die Ergebnisse zeigen, dass nur Männer positiv auf die Betonung ihrer Verantwortung reagieren, während Frauen tendenziell mehr abgeben, wenn ihre Entscheidungsmacht und Entscheidungsfreiheit hervorgehoben wird.

In Kapitel 3 geht es um das Befolgen einer Bitte. Die Bitte besteht darin, im Rahmen eines Experiments eine Frage möglichst detailliert zu beantworten. Das Befolgen der Bitte wird daran gemessen, wie viel Mühe sich die Teilnehmer beim Beantworten der Frage geben. Die Teilnehmer sehen dabei entweder den Text „Vielen Dank im Voraus“ oder nicht. Es zeigt sich, dass die Teilnehmer auf den Text negativ reagieren, indem sie sich weniger Mühe geben.

In Kapitel 4 werden die Auswirkungen von (zumeist ungenau) vorausgefüllten Steuererklärungen auf die Steuerhinterziehung untersucht. In dem Experiment verdienen die Teilnehmer zunächst Einkommen durch das Ausführen einer Aufgabe, bevor sie eine Steuererklärung ausfüllen. Dies wiederholt sich für insgesamt drei Runden. Im Haupttreatment ist die Steuererklärung mit einem Wert vorausgefüllt, der sich aus der Leistung der Teilnehmer in vorherigen Runden ergibt und dessen relative Größe von Runde zu Runde variiert. Die Ergebnisse deuten darauf hin, dass es keinen nachhaltigen Effekt solch einer Standardvorgabe gibt, weder wenn dieser Wert relativ niedrig noch wenn dieser relativ hoch ist. Teilnehmer, bei denen die Steuererklärung in der ersten Runde mit einem relativ niedrigen Wert vorausgefüllt ist, hinterziehen jedoch signifikant mehr Steuern, als wenn die Steuererklärung nicht vorausgefüllt ist.

Declarations of Co-Authorship

Aus datenschutzrechtlichen Gründen wird in der publizierten Version auf die drei Seiten "Declarations of Co-Authorship" verzichtet.

Aus datenschutzrechtlichen Gründen wird in der publizierten Version auf die drei Seiten "Declarations of Co-Authorship" verzichtet.

Aus datenschutzrechtlichen Gründen wird in der publizierten Version auf die drei Seiten "Declarations of Co-Authorship" verzichtet.