# Economic Inquiries into

# Pro-Social and Anti-Social Behavior

Dissertation

Zur Erlangung der Würde des Doktors der

Wirtschafts- und Sozialwissenschaften (Doctor rerum politicarum)

des Fachbereichs Volkswirtschaftslehre

der Universität Hamburg

vorgelegt von

Menusch Khadjavi

geboren am 14. Januar 1985 in Braunschweig

Hamburg, den 30. September 2013

| | |
|---|---|
| Vorsitzende: | Prof. Dr. Anke Gerber |
| Erstgutachter: | Prof. Dr. Andreas Lange |
| Zweitgutachter: | Prof. Dr. Gerd Mühlheußer |
| Datum der Disputation: | 20. Dezember 2013 |

## Acknowledgements

I am indebted to Andreas Lange for his dedicated supervision, continuous encouragement and co-authorship of some of the work presented in this dissertation. Working at Andreas' chair and doing research with him has been inspiring and a great privilege – making me feel like a child in a candy store. It is seldom that one may enjoy great freedom and caring support simultaneously.

I am grateful to Anke Gerber and Gerd Mühlheußer for their commitment to my dissertation. Discussions on academic and private topics with Anke, Gerd and other colleagues at the University of Hamburg sweetened my everyday life. Amongst others, I would like to thank Leonie Baumann, Berno Büchel, Angelika Höllendörfer, Ute Lampalzer, Martin Leroch, Jannis Liedtke, Jakob Neitzel, Andreas Nicklisch and Philipp Schliffke for many good times.

For their generous invitations and dedicated hosting I would like to thank Uri Gneezy at the University of California San Diego and Max Auffhammer at the University of California Berkeley. At conferences, courses and stays abroad, I have received valuable comments and encouragement for my research from and spent fun times with Anastasia Danilov, Susann Fiedler, Orsola Garofalo, Diogo Geraldes, Martin Kesternich, Marco Kleine, Johannes Müller-Trede, Silvia Saccardo, Paul Smeets, Jan Stoop, Janna Ter Meer, Daan van Soest, Roel van Veldhuizen and Lilia Zhurakhovska.

The laboratory team led by Olaf Bock, including Sarah Mörtenhuber and Jan Papmeier, has been elementary for conducting experiments in the laboratory in Hamburg and in the field in Vechta. Thank you all for your help. Financial support by the Department of Economics and by Faculty of Economic and Social Sciences is gratefully acknowledged.

On a personal note, I thank my fantastic friends in and around Braunschweig, friends from my times in Bayreuth and Delaware, my friends in Amsterdam, and my friends in the great city of Hamburg for their support and love during the last years. Finally, I am indebted to my caring family. You are the fundament of my motivation, courage and happiness. Thank you all, especially my parents, my grandmother and Kacana, for making my life beautiful.

Hamburg, September 2013

*"The maxims are, first, that the individual is not accountable to society for his actions, in so far as these concern the interests of no person but himself. Advice, instruction, persuasion, and avoidance by other people, if thought necessary by them for their own good, are the only measures by which society can justifiably express its dislike or disapprobation of his conduct. Secondly, that for such actions as are prejudicial to the interests of others, the individual is accountable, and may be subjected either to social or to legal punishments, if society is of opinion that the one or the other is requisite for its protection."*

John Stuart Mill (1859, On Liberty, p. 104)

## Contents

# 1 Introduction

Inspired by social philosophers like Adam Smith and John Stuart Mill (and his quote above), modern-day economics rests on the account that it is regarded as highly desirable for individuals to make their own choices on how to live their lives without external restrictions for they are the bearers of their personal tastes and preferences. Such liberty is exhausted when individuals' actions interfere with the liberty of others. In this case, it is often regarded as just for society to exercise social or legal punishment in order to protect the liberty of others (Mill 1859).

Today's discipline of normative public economics investigates such cases, i.e. cases in which it is desirable for society, represented by the government in a democracy, to intervene with actions of agents in markets and other affairs in the public domain. Classic examples span from the passing of criminal laws by the legislative, the enforcement of such laws via crime detection and prosecution (Becker 1968), the introduction of incentives in the form of taxes and subsidies to steer behavior (Pigou 1920) to the definition of property rights (Coase 1960).

This dissertation provides new inquiries into the behavior of individuals in cases in which their actions interfere with the liberty of others. More specifically, the aim of this dissertation is to provide new behavioral economic insights into individual and group decision making of pro-social, voluntary giving (and sharing) of resources to others on the one hand and especially anti-social acts of taking and stealing from others on the other hand. In this spirit, my dissertation serves two major purposes. First, it provides new insights for behavioral economic research, for instance on the motivation of individuals to steal from a group and other individuals. Second, it answers questions of practical importance, for instance, concerning the consequences of institutions like transparency, peer punishment and criminal prosecution.

As an advancement of the homo economicus paradigm (i.e. narrowly selfish and rational preferences), the last decades have witnessed substantial progress in the detection of other-regarding, social preferences in economic research (e.g. Güth et al. 1982, Henrich et al. 2001). Consequently, a number of social preference theories have emerged (Meier 2007). Seminal works include outcome-based models like inequity aversion and altruism (e.g. Fehr and Schmidt 1999, Bolton and Ockenfels 2000, Levine 1998), action-based reciprocity models (e.g. Rabin 1993, Fischbacher et al. 2001, Dufwenberg and Kirchsteiger 2004, Falk and Fischbacher 2006), the model of impure altruism (warm-glow) of Andreoni (1990), and models of utility from a positive self-image (e.g. Brekke et al. 2003, Bénabou and Tirole 2006). On the contrary, works like Herrmann et al. (2008) and Abbink and Sadrieh (2009) provide evidence for the existence of preferences for anti-social behavior such as punishing contributors to a public good, and destroying endowments of others.

The above literature on other-regarding and social preferences draws most of its evidence from experimental economic research. For this reason it seems natural to tackle the related questions I aim to answer with experimental methods. Roth (1993) speculates that Bernoulli (1738) may be the first attempt of an economic experiment and Bardsley et al. (2010) name Hume (1739-40) among the first experimental approaches to questions in economics. Especially in the last fifty years the body of economic research that employs experimental methods has exploded (Bardsley et al. 2010). In appreciation of such research, the Nobel Prize in Economics has been awarded to experiment-affine laureates like Daniel Kahneman, Elinor Ostrom, Vernon Smith and Alvin Roth in recent years.

Based on the mature experimental economic literature, I can rely on established methods in experimental economics to answer my research questions. In every chapter I present in this dissertation I refine an established decision problem or game-theoretic problem and the respective experimental environment to answer my research questions.

The first experimental environment that I employ in this dissertation is the linear public good game. Surveys by Ledyard (1995) and Chaudhuri (2011) document the vast literature that has emerged on this game in recent years. For me, the attractiveness to study the private provision of public goods stems from both its extensive applicability and the inherent social dilemma between the group optimum and the Nash equilibrium (for standard homo economicus preferences). Chapters 2, 3 and 4 aim at fostering the understanding of norms and social preferences that motivate contributions to and especially appropriation from public goods in a number of different settings.

chapter 2, *Doing Good or Doing Harm – Experimental Evidence on Giving and Taking in Public Good Games*, explores the voluntary provision of public goods, when individuals can *give and take*, i.e. contribute to a public good or appropriate an existing public good to their private advantage. Using different payoff-equivalent versions of a linear public good game, I find that the resulting provision level of the public good when agents can contribute or appropriate is similar to the provision in the pure giving frame. This contrasts the replication of Andreoni (1995)'s finding that agents behave significantly more selfishly when taking is the only option. Similarly, a pure extension of the action space to the taking domain while keeping the giving domain leads to a significant reduction of provision levels. Chapter 2's results do not support the standard impure altruism (warm glow) model of Andreoni (1990). Rather, the data suggests a combination of the standard impure altruism model with a component which is referential with respect to the action set at hand. Consequently, an individual may regard her action of modest appropriation from a public good as pro-social if she had the opportunity to appropriate a much greater amount. Chapter 2 provides a novel extension of Andreoni's model.

Besides the insights it yields on giving and taking actions and possible underlying motivation, chapter 2 effectively provides a framework for chapters 3 and 4 titled *Transparency and Accountability* and *Setting the Bar* respectively. In the following, I will argue that the giving-and-taking framework of decision making is appropriate and insightful for both subsequent chapters.

Chapter 3, *Transparency and Accountability*, employs an *asymmetric* version of the linear giving-and-taking public good game framework outlined above. Here I place three standard players with a giving action set and one special player with a giving-and-taking action set into a social dilemma. The analogy here is an environment in which one agent holds more power than others. More specifically, she holds the power to give and take from a public good, for instance when employees of a firm can exert effort to foster the returns to the firm, while the manager can exert effort but is also able to take from the firm's account. Another example is the asymmetry between households and industrial firms in providing environmental quality. While households, at least on a larger scale, can typically only help to protect the environment, industrial firms also hold significant power to pollute it.

In order to solve issues of corruption and of pollution, policy suggestions typically include greater transparency and measures to hold actors accountable by punishment for offenses. The purpose of chapter 3 is to analyze the separate and joint effects of such policies on public good provision and the underlying motivation of agents. The results underline the importance of combining the two institutions of transparency and accountability in order to achieve high levels of welfare. If it is not possible to hold a special player accountable, then a high level of transparency of actions motivates her to exploit the maximal amount. With a low level of transparency of actions, the special player appears to try to hide her exploitative actions by keeping exploitation within moderate limits. Hence, chapter 3 reports adverse effects of transparency when it is not possible to hold exploiters accountable. On the contrary, transparency in combination with accountability achieves high levels of welfare. The practical advice is that both institutions should be available jointly.

Chapter 4, titled *Setting the Bar – An Experimental Investigation of Immigration Requirements,* also employs the giving-and-taking public goods framework, yet with another twist. In the face of poverty-motivated migration streams from developing countries as well as demand of and competition for highly skilled talents, migration is one of the pressing issues worldwide. It however remains unclear what perceptions and procedures shape efficient migration policy. In order to study migration policy settings, chapter 4 employs the framework with *endogenous groups* to investigate how perceived out-group potential (and associated bargaining power) and debate among in-group members shape immigration requirements. Importantly, in-group members hold the power to voluntarily contribute to a public good and to appropriate part of the public good to their own advantage. The analogies are those of paying taxes (contributing) and receiving social transfers (appropriating). Chapter 4 employs the minimal group paradigm (Tajfel et al. 1971) and immigration requirements are set by in-group voting.

The results show that debate appears to foster fair and efficient immigration requirements if perceived migrant potential is high. Conversely, debate in the perceived low-potential case leads to the tightest requirements. Consequently, fewer out-group subjects immigrate. Individuals in the experiment in chapter 4 appear to rate the fairness of an immigration requirement with regard to its divergence from in-group members actions. This finding suggests that relative fairness (how much do immigrants have to contribute compared to in-group members?) is more important than absolute fairness (how much do immigrants have to contribute as an absolute amount?). I will discuss the findings in light of their importance for migration policy implications.

Taken together, chapters 2, 3 and 4 demonstrate the importance and usefulness of analyzing pro-social and anti-social behavior in a common framework. For a better understanding of means to curb anti-social behavior and to promote pro-social behavior, I turn to individual decision making. Chapters 5 and 6 provide insights into the effectiveness of the state's intervention against stealing actions via deterrence incentives.

Chapter 5 analyzes deterrence schemes and their impact on stealing by a series of experiments with two different subject pools, students and prison inmates. For this purpose I organized a lab-in-the-field experiment (also referred to as artefactual field experiment, Harrison and List 2004) in the *JVA für Frauen*, a penitentiary for women, in Vechta, Germany. Subjects either face potential punishment when stealing, or they can steal without deterrence. My data confirms Becker's (1968) deterrence hypothesis for both students and prison inmates, i.e. deterrence incentives decrease stealing. This first result supports the hypothesis that criminals are not systematically different from other groups in the population with respect to their rationality of decision making (a point previously questioned by Anderson 2002). This first finding is good news for the extensive literature of Becker-inspired theoretical work on criminal law and economics. It can rest assured that their fundamental hypothesis is indeed true – a question which had not been answered by experimental methods before. Second, both groups also show crowding out of subsequent pro-social behavior due to deterrence incentives. That is, if deterrence incentives exist, subsequent behavior is more selfish. This second finding constitutes a challenge for policy making. In the spirit of Gneezy and Rustichini (2000b)'s title "Pay enough or don't pay at all" for crowding out of intrinsic motivation for volunteering, one may conclude 'punish enough or don't punish at all' for deterrence incentives.

Chapter 6 provides evidence that (part of) this crowding out takes place via the change of emotions. I collect data on stealing behavior and associated emotions with and without deterrence incentives in place. Without deterrence incentives – a variant of the dictator game – players with pronounced pro-social emotions steal less. That is, when predicting the amount stolen, a cluster of pro-social emotions influences this amount significantly. Conversely, when players face expected punishment pro-social emotions are deactivated and do not decrease stealing; in this case self-centered emotions get activated and motivate greater stealing. I use the principal component analysis to cluster emotions, a method inspired by Reuben and van Winden (2010). This chapter thus provides support for theory on emotions in behavioral criminal law and economics (van Winden and Ash 2012). I hope that it serves as an example and motivates more behavioral economic work on the interaction of emotions and decision making. I regard research on this interaction as highly useful to better understand pro-social and anti-social motivation and as a largely under-researched area.

As I outline above, my dissertation yields new behavioral economic insights into the motivation and consequences of pro-social and anti-social behavior. With regard to the motivation of actions, relative concepts are prevalent. Doing good or doing harm in chapter 2 appear to depend on the available action set and not on the absolute action, in chapter 3 cooperation of standard players appears to be mainly conditioned on other standard players (but not or to a lesser extent on special players), and the fairness of an minimum contribution immigration requirement appears to be relative to what in-group members contribute to the public good in chapter 4. Chapter 5 demonstrates that criminals make decisions when taking and giving alike students. This finding supports the idea that some concepts of decision making may be ubiquitous. Finally, emotional motivation of decision making are measured and established in chapter 6. Results also indicate difficulties and solutions to introduce effective and efficient policy, may it be to curb corruption (chapter 3), to achieve migration policy which enhances welfare of in-group and out-group members (chapter 4), or to understand and use laws to curtail crime (chapters 5 and 6).

# 2 Doing Good or Doing Harm – Experimental Evidence on Giving and Taking in Public Good Games[1]

*"It is absurd to divide people into good and bad. People are either charming or tedious."*
Oscar Wilde

## 2.1 Introduction

A functioning society relies on a sufficient provision of public goods. Overcoming free-riding in their production through taxes or voluntary individual contributions is thereby as important as deterring individuals from exploiting existing public goods for their own personal advantage.

While *doing good* has been explored in numerous laboratory and field experiments on the economics of charities, the alternative where individuals are *doing harm* to the public for their own private benefit has received much less attention in the literature.[2] A notable exception is

---

1 This Chapter is co-authored by Andreas Lange.

2 For an overview of earlier studies of public good games, see Ledyard (1995). Social dilemma and public good games often find that behavior of individuals differs from the standard game theoretic prediction. While the prediction in standard linear public good games is that participants give nothing to the public good, studies present robust evidence that group contributions to the public good are significantly greater than zero (often around 50 to 60 percent of total possible contributions) in the first period, and – even though with a declining trend – mostly remain significantly greater than zero in subsequent periods (see e.g. Isaac and Walker 1988, Isaac et al. 1994, Gächter et al. 2008).

Andreoni (1995) who examines the settings where agents can either only take or only give and finds that the resulting public good provision levels are significantly lower in the taking than in the giving frame. Dufwenberg et al. (2011), Park (2000), and Sonnemans et al. (1998) also identify similar differences between contribution decisions to public goods vs. public bads. [3] Cox et al. (2013), however, do not establish these differences in their provision vs. appropriation games.

While these articles examine either giving or taking behavior, many natural examples exist where individuals may *both* contribute to a public good and reduce its provision level. They range from environmental pollution where agents can emit or try to reduce overall pollution by investing in carbon offsets (Kotchen 2009), legal vs. illegal forms of employment, illegal claims of social services, and tax evasion, to managers whose actions may enhance the performance of a firm or exploit the work contribution of others.

In this chapter, we investigate how individuals behave when their action space allows for *giving and taking*, i.e. contributing to a public good or exploiting it. Using a standard linear public game setting, we perform four experimental treatments that differ with respect to the initial allocation given to the private and the public accounts and the actions available to agents. We first study the one-sided cases of (i) a voluntary contribution mechanism with no initial resources in the public account where agents can solely give and (ii) an inverse public good treatment where all resources are in the public account and subjects can only take. We examine these cases to replicate Andreoni (1995)'s finding based on neutral language instructions. We then consider two treatments that start with a positive allocation in both the private and the public account. In the first treatment, the action space allows for both giving and taking decisions. We find that the resulting provision level of the public good is similar to the provision in the pure giving frame. Our results thereby indicate that Andreoni (1995)'s pessimistic finding on lesser provision in a taking frame may be driven by the complete elimination of giving options: with an intermediate initial endowment in the public account, despite the taking option, subjects reach similar provision levels as when having a zero initial endowment in the public account.

---

Positive giving decisions in public good games, dictator and ultimatum, and other games are in conflict with the Nash prediction of payoff-maximization and have led to a series of theoretical explanations based on other-regarding, social preferences. See Meier (2007) for a survey.

[3] Cubitt et al. (2011) follow a similar approach in a one-shot setting with second-stage punishment option and ex-post elicitation of emotions, but largely find insignificant results. For a broader comparison of framing effects in public good experiments, see Cookson (2000).

In the final treatment, we again start with an intermediate level of the public account, but limit actions to the giving domain. Comparing this treatment with the intermediate give and take treatment, we find that provided with the opportunity to *take*, fewer individuals give positive amounts even though we keep the giving domain unchanged. With this result, we extend findings on the impact of a taking domain in dictator games (List, 2007; Bardsley, 2008) to the linear public good setting: provided with the opportunity to *take*, fewer individuals give positive amounts.

Overall, our results demonstrate that the option to 'take' significantly changes individuals' contribution decisions. However, the negative impact of taking on the final provision level of the public good is much less pronounced than found in Andreoni (1995) and others, when individuals' giving options to the public good are preserved. The remainder of this chapter is organized as follows. Section 2 presents the experimental design of the study. Results are presented and discussed in section 3. Section 4 summarizes insights for theory and policy while section 5 provides a concluding discussion.

## 2.2 Experimental Design

Our experimental design consists of four treatments. We use the structure of a linear public good game:

$$\pi_i = w^t - a_i + h\left(E^t + \sum_{j=1}^{n} a_j\right)$$

where $h$ denotes the per capita return to the public good with $0 < h < 1 < hn$, $w^t$ represents the initial endowment of *i* in treatment *t* (and is the same for all *n* group members), $a_i \in A^t$ denotes *i*'s transfer to the public good account, $A^t$ is the available action set and $E^t$ is the initial allocation to the public good account. In the experiment, we chose $n = 4$ and $h = 0.4$.

The treatments differ in the initial allocation to the public good, $E^t$, the initial endowment, $w^t$, and in the action set that is available to agents.

We first replicate the standard voluntary contribution mechanism (*GIVE*), where agents can only give ($E^{GIVE} = 0$, $w^{GIVE} = 20$, $A^{GIVE} = [0,20]$). Second, we look at the inverse setting (*TAKE*), where the whole endowment is in the public account and agents can only take

($E^{TAKE} = 80$, $w^{TAKE} = 0$, $A^{TAKE} = [-20{,}0]$). These two treatments replicate Andreoni (1995) give and take treatments using a neutral framing, i.e. we avoided using words like "give" or "take" in the instructions.[4] We complement these extreme settings with an intermediate treatment (*GITA*=*GI*ve and *TA*ke), where agents can either give or take, i.e. add or subtract from an existing public account take ($E^{GITA} = 32$, $w^{GITA} = 12$, $A^{GITA} = [-8{,}12]$). Here, 40 percent of the budget is initially allocated to the public good while agents receive 60 percent of the budget in their private accounts. They are informed that they may either contribute to or take from the public good. Agents are able to take up to 8 units out of the public good (by choosing $a_i = -8$) or to give up to 12 units to the public good ($a_i = 12$).[5] These three treatments are calibrated such that the range of final allocations to the public and private accounts is identical, i.e. they are payoff-equivalent. We only vary the starting point: either all wealth is in the private account (*GIVE*), all is in the public account (*TAKE*), or the endowment is intermediate (*GITA*). These starting points can be summarized as follows:

$$w^t = 20 - \Delta^t \qquad E^t = n\Delta^t \qquad a_i \in A^t = [-\Delta^t, 20 - \Delta^t]$$

with $\Delta^{GIVE} = 0$, $\Delta^{TAKE} = 20$, and $\Delta^{GITA} = 8$. By comparing behavior in *GITA* with *GIVE* and *TAKE* we will explore how simultaneous giving and taking options compare to the one-sided settings where only taking or only giving is allowed.

A final treatment *GIVE** also uses the same intermediate initial endowment ($E^{GIVE^*} = E^{GITA} = 32$, $w^{GIVE^*} = w^{GITA} = 12$, but limits the action space to the giving domain ($A^{GIVE^*} = [0{,}12]$), such that we get a second VCM. Comparing *GIVE** with *GITA* allows us to explore how a simple extension of the action space to the taking domain affects behavior in a public good setting. Relative to *GIVE*, *GIVE** resembles a situation where agents' income is reduced (for example by taxes) in order to provide the public good. In an abstract way, this comparison allows to study potential crowding out of voluntary contributions through public provision of public goods (e.g. Bergstrom et al. 1986).

The standard game theoretic prediction for selfish agents clearly is that agents will contribute no units of their endowments to the public good and – if taking is possible – transfer the

4 Subjects were asked to transfer tokens between their private and the group account.

5 We deliberately chose asymmetric bounds in the negative and positive domain in order to avoid a potential focal point at the mid-point of the action space to coincide with '0' contribution decisions.

maximal allowable amount to their private account. We therefore would predict $a^{GIVE} = 0$, $a^{TAKE} = -20$, $a^{GITA} = -8$, and $a^{GIVE^*} = 0$.

All experimental sessions were conducted in the computer laboratory of the Faculty of Economic and Social Sciences, University of Hamburg, Germany in January and April 2011. Each session lasted approximately one hour. We used z-Tree (Fischbacher 2007) to program and ORSEE (Greiner 2004) for recruiting. In total, 160 subjects participated in the experiment. All were students with different academic backgrounds, including economics.

**Table 1.** Summary of Experimental Sessions.

| Session | Number of groups | Number of participants | Treatment |
|---|---|---|---|
| 1 | 5 | 20 | TAKE |
| 2 | 6 | 24 | GITA |
| 3 | 5 | 20 | GIVE |
| 4 | 6 | 24 | TAKE |
| 5 | 4 | 16 | GITA |
| 6 | 4 | 16 | GIVE |
| 7 | 5 | 20 | GIVE* |
| 8 | 5 | 20 | GIVE* |

Note: Numbers of groups across treatments are not equal due to some registered subjects not showing up.

Each of our 8 sessions consisted of 10 periods. Once the participants were seated and logged into the terminals, a set of instructions was handed out and read out loud by the experimenter.[6] In order to ensure that subjects understood the respective game, experimental instructions included several numerical examples and participants had to answer nontrivial control questions via their computer terminals.[7] At the beginning of the experiment subjects were randomly assigned to groups of four. The subjects were not aware of whom they were

6 We mainly followed the instructions of Fehr and Gächter (2000), but slightly changed the wording. For instance, instead of 'contributions to a project', instructions asked participants to divide tokens between a private and a group account. Instructions can be found in appendix A.

7 In case a participant did not answer the questions correctly, she was given a help screen that explained the correct sample answers in detail. We believe this might further reduce experimenter demand effects compared to individual talks with subjects. See Zizzo (2010) for more information on experimenter demand effects.

grouped with, but they did know that they remained within the same group of subjects for all periods.

At the end of each period, participants received information about their earnings, the cumulative group contribution to or extraction from of the group account and the final amount of units in the group account. Subjects were never able to identify individual behavior of group members. At the end of the experiment, one of the periods was randomly selected as the period that determined earnings with an exchange rate between Euro and token of 3 EUR = 10 tokens. Including a show-up fee of 4 EUR, the average payment over all treatments was 11.70 EUR. Table 1 summarizes the information for all 8 sessions.

## 2.3 Results

We craft the results summary by both pooling the data across all periods and reporting treatment differences for the first period.

To allow a comparison between treatments with different action spaces, Table 2 reports the decisions along with the corresponding public good contribution level per agent ($a_i + E^t/4$). At the group level, this normalized contribution coincides with the provision level of the public good.

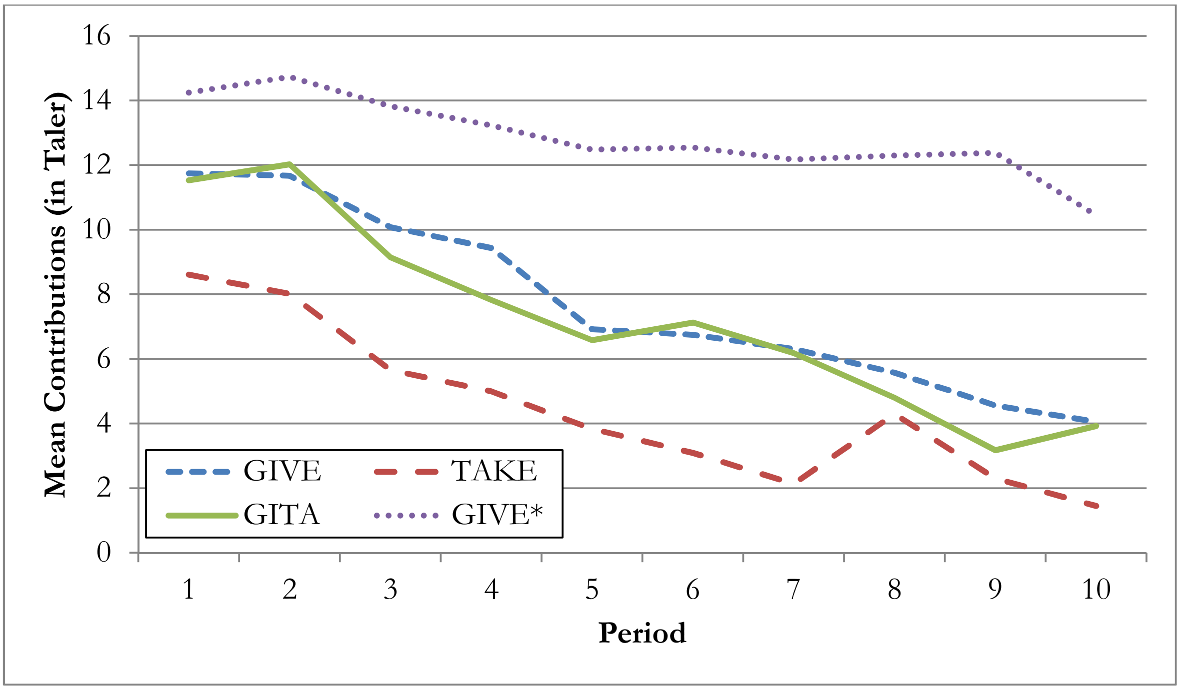
**Table 2.** Summary Statistics of GIVE, GIVE*, TAKE and GITA.

| | First period | | | | All 10 periods (means) | | | |
|---|---|---|---|---|---|---|---|---|
| Statistic | GIVE | GIVE* | TAKE | GITA | GIVE | GIVE* | TAKE | GITA |
| Mean decision | 11.75<br>(7.52) | 6.25<br>(4.18) | -11.39<br>(7.80) | 3.53<br>(7.55) | 7.71<br>(7.69) | 4.84<br>(4.42) | -15.56<br>(6.19) | -0.77<br>(7.87) |
| Mean contribution | 11.75<br>(7.52) | 14.25<br>(4.18) | 8.61<br>(7.80) | 11.53<br>(7.55) | 7.71<br>(7.69) | 12.84<br>(4.42) | 4.44<br>(6.19) | 7.23<br>(7.87) |
| % of decisions MaxSelf | 11.11 | 12.50 | 36.36 | 20.00 | 33.89 | 25.25 | 52.73 | 45.25 |
| % of decisions >0 | 88.89 | 87.50 | - | 65.00 | 66.11 | 74.75 | - | 41.50 |

Note: Standard deviations for individual level data in parentheses.

Across all periods, in *GIVE*, each agent contributed 7.71 tokens on average, resulting in a public good provision level of 4*7.71=30.84 tokens. Contributions are substantially smaller in *TAKE* with 4.44 tokens, resembling the Andreoni (1995) result. Importantly, the average contribution in *GITA* was 7.23 tokens, and thereby almost indistinguishable from the one in *GIVE*. This is particularly surprising as taking was possible, and the range of final contribution levels was identical to the one in *GIVE* and *TAKE*. Dropping the taking option in *GIVE** leads to the highest average contribution of 12.84 tokens.[8] Figure 1 depicts contribution levels by period.

**Figure 1.** Average Contributions over all Periods, by Treatment.




The differences between the treatments are confirmed by a series of non-parametric tests (see Table 3) as well as by a series of linear regression models that we report in Table 4. The regressions predict the contribution to the public good (in tokens) as a result of the different treatments. Table 4 also separates the effects between the first five and last five periods. It should be noted that our regressions indicate a declining trend in contributions over the periods that also reduces the differences between the treatments.

8 Note that the minimum contribution in the *GIVE** was 8 tokens, because the public account contained 8 tokens per person already and 'taking' was no option.

Taking the average of all periods, the *TAKE* treatment leads to less contribution than *GIVE* (3.3 tokens, statistically significant at the 10% level comparing the coefficients using a t-test). While contributions in *GITA* are significantly larger than in *TAKE* (5% level of significance in Table 4, confirmed by Mann-Whitney test), they are almost identical as in *GIVE* (see Table 3 and 4).

While confirming the significant difference between a giving and a taking environments (see Andreoni 1995), we obtain the following result when starting from an intermediate level of the public account

**Result 1.** *There is no significant difference in the provision of the public good between a pure giving frame, GIVE, and a treatment where taking is possible when starting with an intermediate endowment of the public account (GITA). Starting with an even higher allocation in the public account such that only taking is possible (TAKE) leads to significantly smaller final provision levels of the public good.*

**Table 3.** Results of Test Statistics for Comparison of Group Contributions, all 10 Periods.

| (row vs. column comparison) | | Treatment: GIVE | Treatment: GIVE* | Treatment: TAKE |
|---|---|---|---|---|
| Treatment | GIVE* | > (p = 0.0412) | | |
| Treatment | TAKE | < (p = 0.0557) | < (p = 0.0001) | |
| Treatment | GITA | = (p = 0.8703) | < (p = 0.0012) | > (p = 0.0411) |

Note: All test statistics are (nonparametric) Mann Whitney tests except for TAKE vs. GIVE which is done using a Welch t test because of unequal variances. The table is to be read row vs. column. For instance, group contributions are significantly greater in the GIVE* compared to the GIVE.

Result 1 provides an important caveat to findings by Andreoni (1995) and Park (2000): reversing the public good giving game to a taking game in *TAKE* may reduce the contribution levels only because of the one-sided nature of the action space that does not allow for giving. In our *GITA* treatment where the public good and private account are endowed with intermediate levels and agents can give and take, the final provision of the public good is almost identical to a standard voluntary contribution mechanism (*GIVE*). In fact, Figure 1 also displays almost identical contribution rates. Note, however, that *GIVE* vs. *GITA* allows for the same range of possible contribution levels, i.e. the agent always has the option to effectively contribute between 0 and 20 to the public good, and – correspondingly – to secure himself an allocation in his private account between 0 and 20.

**Table 4.** Linear Regression of Contribution Levels of all Four Treatments.

| | Dependent Variable: Contribution Level | | | | |
|---|---|---|---|---|---|
| Independent Variables | (I) All 10 periods | (II) All 10 periods | (III) All 10 periods | (IV) Periods 1 to 5 | (V) Periods 6 to 10 |
| GIVE | 0.481 (1.964) | 0.481 (1.964) | 0.552 (2.082) | 0.552 (2.082) | 0.41 (2.022) |
| GIVE* | 5.605*** (1.311) | 5.605*** (1.311) | 4.28*** (1.544) | 4.28*** (1.544) | 6.93*** (1.297) |
| TAKE | -2.789** (1.236) | -2.789** (1.236) | -3.193** (1.513) | -3.193** (1.513) | -2.385** (1.145) |
| Period 6-10 | | -3.528*** (0.394) | -4.38*** (0.920) | | |
| Period 6-10_GIVE | | | -0.142 (1.181) | | |
| Period 6-10_GIVE* | | | 2.65** (1.122) | | |
| Period 6-10_TAKE | | | 0.807 (1.039) | | |
| Constant | 7.231*** (1.024) | 8.994*** (1.084) | 9.42*** (1.039) | 9.42*** (1.272) | 5.04*** (0.952) |
| Observations | 1600 | 1600 | 1600 | 800 | 800 |
| Individuals | 160 | 160 | 160 | 160 | 160 |
| Groups | 40 | 40 | 40 | 40 | 40 |

Note: Random effects estimation clustered at group level; GITA is the baseline. Standard errors in parentheses, significance: *p < 0.10, **p < 0.05, ***p < 0.01.

The comparison between *GITA* and *GIVE** allows us to study how a pure extension of the action space into the taking domain affects behavior. We observe significantly less contributions in *GITA* than in *GIVE** (1 percent level of significance, see Mann-Whitney test in Table 3 and regression results in Table 4): averaged across 10 periods, 7.23 tokens are contributed in *GITA* and 12.84 tokens in *GIVE**.

We can therefore formulate the following result:

**Result 2.** *Extending the action space to taking domain leads to a reduction of the average provision of the public good (GIVE* vs. GITA).*

While Result 2 may not look particularly surprising as some agents will start taking when this option is present, it is worthwhile to further explore individual behavior. When considering *GITA* and *GIVE**, the action space and the resulting effects on private and public account are identical when concentrating on the giving domain. We can therefore compare the percentage of individuals who give a strictly positive amount (see Table 2). In order to identify how taking options change the behavior of agents due to changed intentions rather than due to reactions to behavioral changes of others, we concentrate on period 1 decisions. While 87.5 percent give in *GIVE** and 88.9 percent in *GIVE*, only 65 percent give a positive amount in the first period in *GITA*. The differences between *GIVE** and *GITA* as well as *GIVE* and *GITA* are both significant at the 5 percent level based on Fisher exact tests. The differences are even more pronounced when averaging across all 10 periods (41.5 percent in GITA vs. 74.8 percent in *GIVE**).

**Result 3.** *Fewer agents give a positive amount if the action space allows for giving and taking (GITA vs. GIVE* and GITA vs. GIVE).*

A random effects Probit regression for the probability of positive giving in Table 5 confirms this finding. The estimated coefficients for *GIVE* and *GIVE** are positive and statistically significant (baseline *GITA*). Result 3 thereby extends the findings by List (2007) and Bardsley (2008) from dictator games to public good games: not all subjects who take when taking is allowed pool at zero contributions when the action set allows for giving only.

Complementing this result, we finally consider the percentage of agents who choose the most selfish option (see Table 2). In *TAKE*, 36.4 percent of agents take out the maximum amount in period 1, while fewer agents choose the most selfish action in *GITA* (20.0 percent, difference significant at the 10 percent level based on Fisher exact test), *GIVE** (12.5 percent, at 5 percent level) and in *GIVE* (11.1 percent, at 1 percent level). Introducing a taking option and reducing giving options therefore appears to make subjects less hesitant to behave in the most selfish manner.

**Table 5.** Probit Regression of Giving a Positive Amount in GIVE, GIVE* and GITA.

| | Dependent Variable: Binary variable on whether a positive amount was given (yes = 1) | |
|---|---|---|
| Independent Variables | (VI) All 10 periods | (VII) All 10 periods |
| GIVE | 1.005*** (0.305) | 1.141*** (0.349) |
| GIVE* | 1.353*** (0.301) | 1.537*** (0.345) |
| Period 6-10 | | -1.043*** (0.106) |
| Constant | -0.319 (0.204) | 0.162 (0.238) |
| Observations | 1160 | 1160 |
| Individuals | 116 | 116 |
| Groups | 29 | 29 |

Note: Random effects estimation clustered at group level; GITA is the baseline. Standard errors in parentheses, significance: *p < 0.10, **p < 0.05, ***p < 0.01.

## 2.4 Implications for Theory and Policy

In this section we draw conclusions from our results that we believe are worthwhile considering. Taking jointly, our results allow us further insights into the motives of giving. Giving in public good games is often interpreted as a sense for efficiency, conditional cooperation or agents gaining a warm glow from giving (Andreoni 1990). Cox et al. (2013) already show for a whole class of social preference theories (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Charness and Rabin, 2002; Cox and Sadiraj, 2007) would predict identical allocation to players in both the *GIVE* and the *TAKE* treatments. Different from their results on symmetric contribution vs. appropriation games, but consistent with Andreoni (1995), we identify significant differences in the resulting provision levels of the public good which cannot be explained by this class of preference theories. We further find that a pure extension of the action space to the taking domain in *GITA* vs. *GIVE** decreases the percentage of players who contribute positive amounts to the public good.[9] This finding is not consistent with a strict version of warm glow that agents receive from contributing a positive amount. Our results are, however, consistent with a modified version in which an agent's utility depends on the chosen action relative to the available set of actions. In order to capture this idea, we posit that warm-glow may be the larger, the larger the difference between actual contributions and the minimal contribution is. We capture this through a warm-glow component of utility

$$f(a_i - \theta \min a)$$

where $\theta \in [0,1]$ is a parameter determining the extent to the minimal contribution level enters and $f$ is assumed to be increasing and concave. Form $\theta = 0$, this model coincides with the standard version of warm-glow. For $\theta = 1$, agents compare their action only relative to the worst action, e.g., taking $18$ in *TAKE* and thereby leaving $2$ in the public account would again be equivalent to giving $2$ in *GIVE*. For $\theta > 0$, this evaluation of action is similar to the kindness definition in Rabin (1993).

Note that the simultaneous existence of different $\theta$-types explains our data: the optimal choice would be given by

$$\max_{a_i} w^t - a_i + h\left(E^t + \sum_{j=1}^{n} a_j\right) + f(a_i - \theta \min a)$$

which leads to the following first-order condition for an interior solution

$$0 = -1 + h + f'(a_i - \theta \min a)$$

such that $a_i$ is increasing in $\min a$. Hence, such modified version of the traditional warm-glow model explains the larger positive giving rates in *GIVE** vs. *GITA* as well as the observed difference between *GIVE* and *TAKE*: the introduction of taking options reduces contributions from some agents. A test of this model modification remains the subject of further research.

We designed our experiment with a primal focus on exploring the voluntary provision of public goods when giving and taking is possible. However, our results shed new light on the interaction between voluntary, private contributions to a public good and government

---

[9] Note that this comparison could not have been made based on *GIVE* and *TAKE* and therefore goes beyond the existing literature.

contributions which are financed through taxes (e.g., Andreoni 1993, Chan et al. 2002, Bergstrom et al. 1986). In an abstract way, the payoff structure of *GIVE** relative to *GIVE* can be reinterpreted as having private income of agents reduced (taxed) while simultaneously providing a public good at the corresponding level (tax-based financing of public good). In the terms of this public finance literature, we find evidence for incomplete crowding out even in our linear public good setting where the reduced allocation to the private account was not framed as ex ante taxation. However, Result 1 indicates that this finding is driven by the assumption of non-negative contributions. When extending the action set to the taking domain, i.e. when agents have options to diminish the public good to their own private advantage, we find complete crowding out (*GITA* vs. *GIVE*).

## 2.5 Conclusion

The last decades have seen an enormous interest of economists in providing insights into why people give to public goods. By mainly focusing on the giving decision, however, the public good game literature has largely ignored a simple and obvious twist to how individual actions may affect the provision of public goods: agents may not only engage in giving, but may also choose actions that diminish the public good. Environmental amenities serve as a prominent, yet not exclusive example where both, doing good and doing harm, are feasible options.

In this chapter, we report findings from linear public good games that allow a direct comparison of the impact of taking options on the provision of public goods. First, if the action set only allows for taking from an initially provided public good, the resulting provision level of the public good is smaller than in any situation where agents can (also) contribute positive amounts. Starting from an intermediate level of the public account, taking options lead to similar final provision levels than starting from zero public good provision within a giving frame. The difference between a pure giving and a pure taking frame (Andreoni 1995) may therefore be driven by of the respective one-sided action spaces.

Second, extending the action space to the taking domain leads to a larger share of agents to choose the most selfish action, while fewer agents give positive amounts to the public good. We therefore identified important channels through which extending the action space to taking from the public good impacts the provision level.

Our findings allow further insights into the motives of giving. They are neither consistent with a strict version of warm glow nor other social preference theories. Rather, subjects appear to see their actions relative to the available action set. Similarly to List (2007) and Bardsley (2008) who observe that 'giving' in dictator games is not the same as 'doing good', 'taking' (from a public good or directly from other subjects) may not readily be interpreted as 'doing harm'. As such, when subjects compare their actions to all feasible actions in situations that allow for taking, 'doing good' may simply mean 'not doing (too much) harm'.

Naturally, this chapter can only provide initial insights into how and why individuals contribute to or diminish the provision of public goods. It provides an interesting avenue for future research. For a better understanding how to overcome social dilemmas, it is necessary to both explore which institutions induce agents to provide public goods and which ones discipline agents to refrain from exploiting them.

## 2.6 Appendix A: English Translation of the Experimental Instructions

General Instructions for Participants

**Welcome to the Experiment Laboratory!**

You are now taking part in an economic experiment. You will be able to earn a considerable amount of money, depending on your decisions and the decisions of others. It is therefore important that you read these instructions carefully.

The instructions which we have distributed to you are solely for your private information. **It is prohibited to communicate with other participants during the experiment.** Should you have any questions please raise your hand and an experimenter will come to answer them. If you violate this rule, we will have to exclude you from the experiment and from all payments.

During the experiment you will make decisions **anonymously**. Only the experimenter knows your identity while your personal information is confidential and your decisions will not be traceable to your identity.

In any case you will earn 4 Euros for participation in this experiment. The additional earnings depend on your decisions. During the experiment your earnings will be calculated in tokens. At the end of the experiment your earned tokens will be converted into Euros at the following exchange rate:

**1 Token = 0,30 €,**

and they will be paid to you in **cash**.

The experiment consists of **10 periods** in which you always play the same game. The participants are divided into groups of 4. Hence, you will interact with 3 other participants. The composition of the groups will remain the same for all 10 periods. Please mind that you and all other participants decide anonymously. Therefore group members will not be identifiable over the periods.

At the end of the experiments you will receive your earning from **one out of the ten periods** converted in Euros (according to the exchange rate above) in addition to the 4 Euros for your participation. The payout period will be determined **randomly**. You should therefore take the decision in **each** period seriously, as it may be determined as the payout period.

The following pages describe the course of the experiment in detail.

Rules of the Game

Each player faces the same assignment. Your task (as well as the task of all others) is to allocate tokens between your private account and a group account.

At the beginning of each period each participant receives **20 tokens** in a private account. You have to decide how many of these 20 tokens you transfer to a **group account**, and how many you keep in your **private account**. Your transfer can be between 0 and 20 tokens (only whole numbers).

*[TAKE: At the beginning of each period there are **80 tokens** in the **group account** and **no tokens** in your **private account**. You have to decide how many of the 80 tokens you leave in the group account and how many tokens you transfer to your private account. Your transfer can be between 0 and 20 tokens (only whole numbers).]*

*[GITA: At the beginning of each period each participant receives **12 tokens** in a **private account**. There are **32 tokens** in a **group account**. You have to decide how many of these 32 tokens you leave in the group account and how many of the 12 tokens you transfer from your private account to the group account respectively. Your transfer input is related to the group account, so that a negative input means a transfer from the group account to your private account and positive inputs mean transfers from your private account to the group account. Your transfer can be between -8 and 12 tokens (only whole numbers).]*

*[GIVE*: At the beginning of each period each participant receives **12 tokens** in a **private account**. There are **32 tokens** in a **group account**. You have to decide how many tokens you transfer to the group account. Your transfer can be between 0 and 12 tokens (only whole numbers).]*

Your total income consists of two parts:

(1) the tokens which you have kept in your private account,

(2) the **income from the group account**. This income is calculated as follows:

*[TAKE: (1) the tokens which you have transferred to your private account]*

Your income from the group account =

0,4 *times* the total amount of tokens in the group account

Your income in tokens in a period hence amounts to

**(20 - your transfer) + 0.4 *(total amount of tokens in the group account).**

*[TAKE:* ***(transfer to the private account) + 0.4*(total amount of tokens in the group account)****]*

*[GITA, GIVE*:* ***(12 – your transfer) + 0.4*(total amount of tokens in the group account)****]*

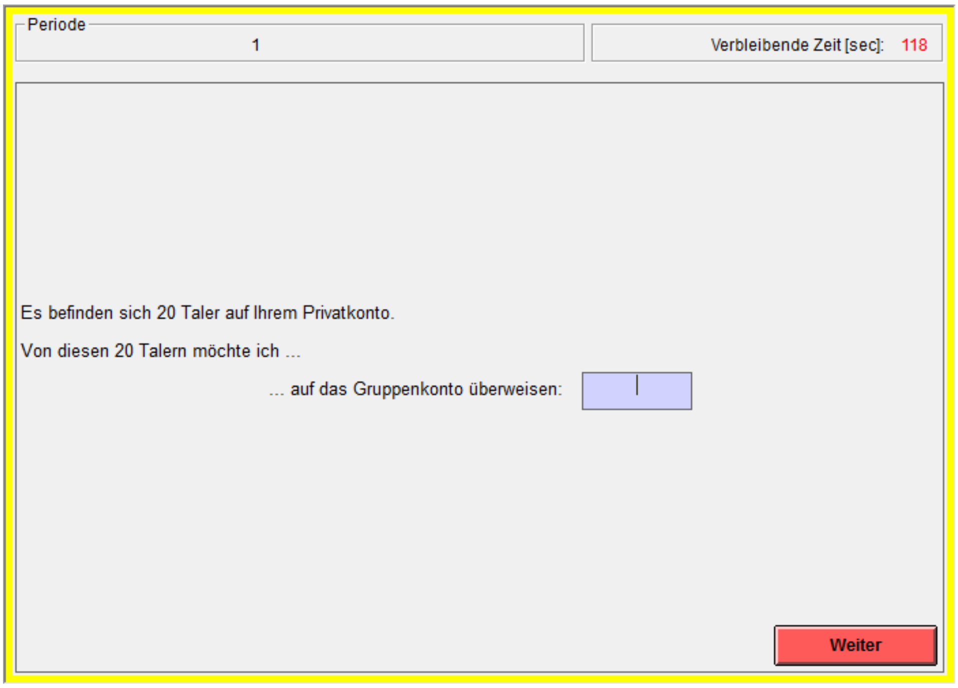
The income of each group member from the group account is calculated in the same way, this means that each group member receives the same income from the group account. Suppose the sum of transfers to the group account of all group members is 60 tokens. In this case each member of the group receives an income from the group account of 0.4*60 = 24 tokens. If you and your group members transfer a total amount of 9 tokens to the group account, then you and all other group members receive an income of 0.4*9 = 3.6 tokens from the group account. Every token that you keep in your private account yields 1 token of income to you.

*[TAKE, GITA, GIVE*: similar or same examples.]*

Information on the Course of the Experiment

At the beginning of each period the following input screen is displayed:

**The Input Screen:**




The **period number** is displayed on the top left. The top right shows the **time** in seconds. This is how much time is left to make a decision.

At the beginning of **each period** your **endowment** contains **20 tokens** (as described above). You decide about your transfer to the group account by typing a whole number between 0 and 20 into the input window. You can click on it by using the mouse.

*[TAKE: At the beginning of* ***each period*** *the group account contains* ***80 tokens****. You decide about your transfer to your private account by typing a whole number between 0 and 20 into the input window. You can click on it by using the mouse.]*

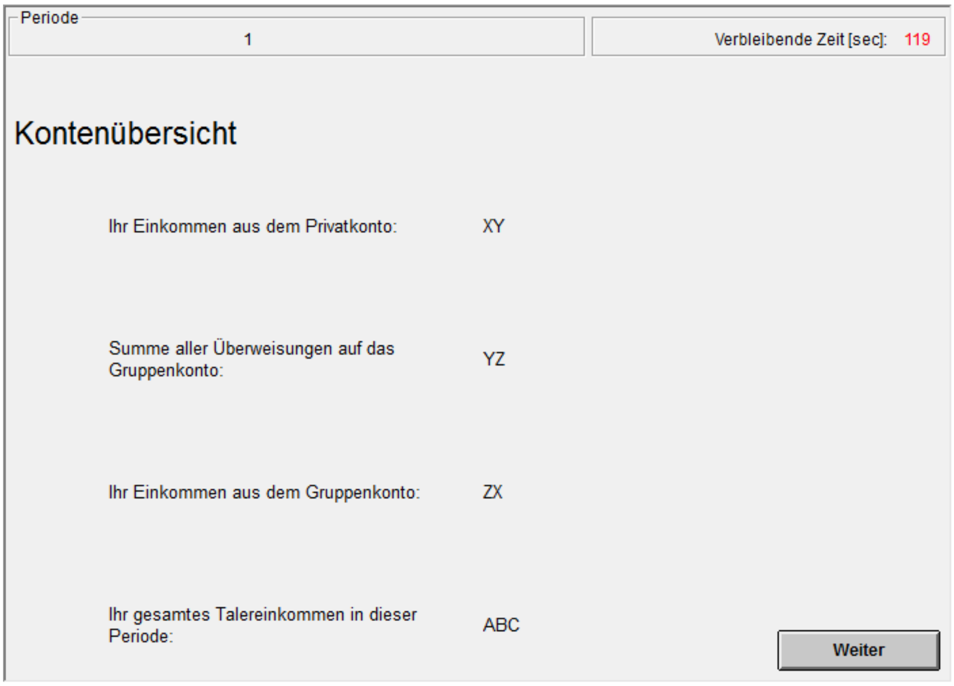
*[GITA: At the beginning of* ***each period*** *the group account contains* ***32 tokens****. You decide about your transfer to your private account or your transfer to the group account by typing a whole number between -8 and 12 into the input window. You can click on it by using the mouse.]*

*[GIVE*: At the beginning of* ***each period*** *the group account contains* ***32 tokens****. You decide about your transfer to the group account by typing a whole number between 0 and 12 into the input window. You can click on it by using the mouse.]*

When you have decided about your transfer to the group account, you have also chosen how many tokens you keep to yourself, that is (**20 - your transfer**) tokens *[differs by treatment]*. When you have typed in your decision, you need to press the **Enter Button** (by use of the mouse). By pressing the Enter Button your decision for the period is final and you cannot go back.

After all group members have made their decisions, your income from the period will be displayed on the following income screen. You will see the sum of transfers to the group account and your income from your private account. You will also see your total income in that period.

**The Income Screen:**




As described above, your **income** is

**(20 – your transfer) + 0,4*(total amount of tokens in the group account).**

*[TAKE: **(transfer to the private account) + 0.4*(total amount of tokens in the group account)**]*

*[GITA, GIVE*: **(12 – your transfer) + 0.4*(total amount of tokens in the group account)**]*

Good luck in the experiment!

# 3 Transparency and Accountability – Evidence from an Asymmetric Public Good Game[10]

*"Good governance comprises […] transparency and accountability in the management of public affairs…"*

The United Nations Development Agenda

## 3.1 Introduction

The question of how to deal with individuals who gain from some kind of (mis-)behavior for which the public bears the costs is a fundamental challenge for every socio-economic system. The economic literature describes this case as individuals' free-riding in social-dilemmas. There are numerous examples for this type of problem like the extensive depletion of ecosystems, or (in the context of labor economics) free-riding on the effort of co-workers in team production and the exploitation of administrative power by corruption.[11] Key features for the avoidance of free-riding are transparency and accountability: making individual actions transparent allows the identification of free-riders, while bringing free-riders to account for their behavior eliminates its attractiveness. Therefore, economists have been examining the effect of transparency in diverse settings ranging from principal-agent and moral hazard problems (Holmström, 1979, Prat, 2005) to central bank policies (Geraats, 2002, Walsh, 2007).

---

[10] This Chapter is co-authored by Andreas Lange and Andreas Nicklisch.

[11] Especially for the latter case, the substantial economic severity of this misbehavior is well documented (e.g., Mauro, 1995, Mo, 2001).

Although the importance of transparency and accountability has been acknowledged widely, the specific interplay between transparency and accountability is largely underexplored.

Our chapter attempts to resolve this shortcoming providing new insights on the interaction of transparency and accountability using a controlled, laboratory environment. More specifically, we analyze the isolated effect of a variation in the degree of transparency without any accountability for the incentives to free-ride. We ask whether a higher degree of transparency alone sufficiently eliminates the incentives for misbehavior. On the other hand, we want to explore the effect of a variation in the degree of transparency on the level of cooperation once we allow for peer-punishment. That is, we explore which degree of transparency creates sufficient accountability that sustains the cooperation within groups. Thereby, we focus on situations with heterogeneous agents. Agents differ with respect to their options to contribute to or hurt the wellbeing of the group. Such asymmetries apply to almost all instances where transparency of and accountability for actions are discussed. One prominent example is corruption: There is a substantial difference between corrupt citizens and corrupt officials. The harm that is caused potentially by misbehavior of the latter type of agents is substantially higher than by misbehavior of the former type of agents. Likewise, households and manufacturers differ with respect to their ability to protect the ecosystem by a reduction of greenhouse gas emission or fresh water pollution. Facing the decision of whether or not to engage in ecosystem protection, many households endeavor to protect ecosystems today by making their lives greener, for example by buying products with lower carbon footprints, remodeling homes to save heating, and recycling waste. However, while environmental awareness of manufacturers has also been increasing over the past decades, they typically hold considerable abilities to exploit ecosystems. That is, a manufacturer's action space with respect to environmental damages is substantially larger than the one for a household.

Guided by such examples on asymmetries in the action sets of agents, we set up a repeated standard (linear) public good game which allows one (*special*) agent to extract money unilaterally from the public good – thereby reducing the payoff to other (*standard*) agents. Meanwhile, all (special and standard) agents may or may not cooperate by voluntary contributions to the public good. We vary the experimental set-up along two dimensions: (i) the transparency of individual agents' identity and actions (ranging from the case where agents neither know the individual cooperation rates nor standard agents can identify the special agent – *non-transparency* – to the case where individual cooperation rates are unknown, but special agents are identifiable for standard agents – *low transparency* – to the case where both individual cooperation rates and the identity of the special agent are common knowledge –

*high-transparency*;[12] (ii) the opportunity to sanction agents for their (potential) actions (here, we consider games with/without implemented peer-punishment opportunities – *no-punishment/punishment*, see, e.g., Fehr and Gächter, 2000). We complement these asymmetric treatments with two (baseline) treatments that vary transparency of actions while keeping the action space identical for all agents. Therefore, our experiment provides important insights into the interplay between transparency and punishment, and their ability to sustain cooperation in social dilemmas.

It turns out that high transparency without punishment opportunity significantly reduces the cooperation rates. More specifically, we confirm for homogeneous groups the findings by Weimann (1994) and Croson (2000) that transparency of individual contributions does not lead to systematic changes in cooperation rates.[13] However, we obtain a different and surprising effect of high transparency when the action spaces differ and the special agent has the option to take: high transparency without punishment backfires, that is, making individual actions observable leads to a significant increase in the take-out rate by special agents. One can interpret this observation such that there are non-strategic reasons (e.g., social preferences), but also as strategic reasons for special agents not to exploit the public good completely under low transparency, since they do not want to be identified as harming the public good (i.e., standard agents may stop contributing once they discover that the special agent takes from the public good). Since high transparency eliminates any opportunity to hide take-out rates, strategic reasons against large-scale take-out rates vanish and special agents abstain less likely from exploiting the public good in this treatment condition. Therefore, high transparency deteriorates cooperation when punishment is not available.[14]

Introducing punishment options reveals the complementary nature of transparency and punishment. Our results show significantly higher contribution rates to the public good in the low transparency and high transparency treatments than in the non-transparency treatment. These findings thereby extend the literature on the effectiveness of punishment (to name only

12 Notice that all agents are informed about the net sum of contributions (i.e., the sum of contributions minus the takeout by the special agent).

13 A question somehow connected is asked in Nikiforakis (2010). Here, players (who can also punish other players at their own cost) receive different feedback formats. One group receives information about the individual contributions, while the other group is informed about individual payoffs. The experimental results show that information about the earnings in comparison with information about the contributions reduces significantly cooperation.

14 In some sense, our finding puts a caveat to the argument by Bohnet and Frey (1999) that making others observe actions increases the pro-social behavior.

few, Yamagishi, 1986, Ostrom et al., 1992, Fehr and Gächter, 2000, 2002, Gürerk et al., 2006, Herrmann et al., 2008) to settings with heterogeneous action sets and limited information on the contributions of other group members. In line with previous research on players who receive noisy signals about the contributions of other group members (e.g., Grechenig et al., 2010, Ambrus and Greiner, 2011), we find even in the non-transparency treatment players making substantial use of their punishment opportunities. However, this punishment does not induce accountability, since it is not directed at exploiting special agents or at free-riding standard agents. It is distributed rather arbitrarily, while it harms considerably the overall sum of payoffs in the public good game.

In great contrast, both low transparency and high transparency accompanied with punishment sustain cooperation. Comparing their effects on the behavior of special and standard agents shows that they work through different channels. Punishment combined with low transparency (i.e. a stigmatization of special players while precise actions are unknown) disciplines special agents sufficiently, since they reduce their take-out rates relative to the treatments without punishment. The reason for this is that even when actions are not fully transparent, punishment points are targeted primarily at them: standard agents punish special agents whenever they observe a small aggregate provision level of the public good and believe that this results from misbehavior (i.e., taking) of the special agent. In other words, special agents can be made accountable for their misbehavior by means of statistical discrimination. Statistical discrimination due to stigmatization (Arrow, 1972, 1973, Phelps, 1972) has recently received much interest in the literature (e.g., Altonji and Pierret, 2001, List, 2004, Gneezy et al., 2012, Zussman, 2012).[15] In order to reduce this (sometimes undeserved) punishment, the special agent increases contributions such that punishment leads to higher provision levels of the public good even though contributions are not fully transparent. On the other hand, high transparency breaks the stigma effect such that both standard and special players are made accountable through being punished based on their true contributions rather than their type. Therefore, high transparency increases the average cooperation rates of standard agents by allowing them to conditionally cooperate among their peers, while it does not further affect the behavior of special agents. As a consequence, cooperation is significantly greater in the high transparency treatment with punishment than in the low transparency treatment with punishment.

15 The concept of statistical discrimination contrasts taste-based discrimination according to Becker (1957).

Overall, our experimental design combines several research streams in the literature on voluntary giving. We employ a modification of the public good game with giving-and-taking action sets introduced in chapter 2 of this dissertation and motivated by Andreoni (1995). Their results with symmetric action sets for players show that the taking option significantly decreases cooperation levels in public goods as it seems that players choose actions relative to given action sets,[16] while in our giving-and-taking framework, we focus on the heterogeneity in action sets. As such, our experiment contributes to the public good literature with heterogeneous endowments. Technically speaking, allowing only special agents to take from the public good is strategically equivalent to providing higher endowments to special agents than to the standard agents. However, as we will show below, this type of heterogeneity has severe behavioral consequences, while existing results with respect to the effect of endowment heterogeneity are inconclusive: some studies report lower contributions (van Dijk et al., 2002, Cherry et al., 2005), some higher contributions (Buckley and Croson, 2006), and some report no effect on aggregate provision levels (Chan et al., 1999, Sadrieh and Verbon, 2006, Reuben and Riedl, 2013). Cox et al. (2013) are the first ones combining the giving-and-taking action sets with heterogeneity of agents. While they do not analyze the effect of transparency and accountability, they show that giving and taking produces comparable contribution under symmetry, but asymmetry leads to significantly lower contributions in the taking game than in the giving game.

A related issue is the behavioral effect of heterogeneous marginal benefits that subjects draw from the public good. In the experimental setting of Fisher et al. (1995), groups consisting of four players, two with a substantially larger marginal benefit from the public good than the other two. Likewise, Palfrey and Prisbrey (1997) randomly vary the marginal benefit for the money that is not contributed to the public good. Again, no systematic effect on the contribution rates can be found. If at all, there is a positive effect of heterogeneity for contributions of players once they are paired with players whose marginal benefit are such that investments in the public good is a dominant strategy for them (Glöckner et al., 2011). Along this line of research, Reuben and Riedl (2009) draw attention to individual differences in marginal benefits under (high) transparency and punishment. Here, the results do not show any significant difference in terms of the sum of contributions compared to the homogeneous case. However, contributions are very heterogeneous within the groups: players who receive higher marginal benefits contribute almost fully while the others contribute significantly less.

16 This finding complements evidence by List (2007) and Bardsley (2008) and supports Rabin (1993)'s theory of fairness.

Following up, Reuben and Riedl (2013) focus on the interplay between heterogeneity both concerning endowment and marginal benefits from the public good on the one hand, and the contribution norm that trigger punishment on the other. In this setting, heterogeneity has a positive effect on contributions (though, the size of the effect differs for diverse types of heterogeneity), while the enforced contribution norms relate to fairness standards of equity regarding the contributions but not regarding the earnings. With those results in mind, in our experiments we will focus on heterogeneity in the action space allowing special agents to take from the public good, regardless of the fact that all agents have nonetheless identical options in the giving domain.

The remainder of this work is organized as follows; Section 2 describes the experimental design and procedures. Section 3 discusses behavioral predictions. In section 4 we report experimental results. Section 5 provides a concluding discussion of the insights of this work and their implications for economic theory and public policy.

## 3.2 Experimental Design

The starting point of our experiment is a linear public good game (e.g., Isaac et al., 1985) where each of $n$ players is endowed with $w$ Taler, which they may contribute to the public good. All contributions within a group of players are summed, multiplied with some (exogenously given) productivity and distributed evenly among all group members (irrespectively of their contributions). The game is played repeatedly for a commonly known number of periods.

We vary this game along three important dimensions: most importantly, we alter the symmetry of the players' action space. For this purpose, we allow one player within the group, the *special agent*, to deplete the initially existing public good account. That is, there is some initial endowment $E$ in the public good, which special agents may or may not extract for their own benefits. Notwithstanding, instead of extracting the initial endowments, they may or may not increase the public good by their contributions. The others, *standard agents,* may or may not increase the public good by their contributions. Standard agents cannot deplete $E$, nor can special agents extract the contributions of the other group members. In two (baseline) treatments without special agents (SYM), the action set is the same for all players: agents are able to contribute to the public good by reducing their initial endowment, i.e. $a_i^{\sim S} \in [0, w]$. In treatments ASYM $n - 1$ of $n$ group members face the same action set as agents in the symmetric treatments: these standard group members are able to contribute to the public good: $a_i^{\sim S} \in [0, w]$. One of *n* group members, however, holds the ability to both give to and take from the public good $a_i^S \in [-E, w]$.

For the second dimension, we consider either a game with (_PUN) or without peer punishment following the "standard" approach by, e.g., Fehr and Gächter (2002). Here, each player *i* receives an additional endowment $p$ that they may use to buy punishment points $p_{i\rightarrow j}$ for player *j*, such that $p_{i\rightarrow j} \in \{0,\ldots, p\}$, $\Sigma_j p_{i\rightarrow j} \leq p$). The marginal cost of $p_{i\rightarrow j}$ for *i* is 1 Taler, the marginal destruction of $p_{i\rightarrow j}$ for *j* is 3 Taler (i.e., each Taler *i* invests into *j*'s punishment destroys 3 Taler of *j*'s earnings). Any amount of the additional endowment $p$ that remains unspent adds to *i*'s payoff in that period.

The third dimension varies transparency. This is introduced by changes in feedback information agents receive on their group members' decisions while deciding upon their punishment points (if they have to) and at the end of a period. Notice that each player in the experiment receives a random identification number. Under low transparency (denoted as _L) the identification number of the special player in common knowledge within the group, the same is true under high transparency (denoted as _H). Moreover, in this treatment condition players receive detailed information on the decisions by each individual group member at the punishment stage (if there is one) and the end of the period, while under non-transparency (denoted as _NT) neither the identification number of the special player nor the detailed information on the decisions by each individual group member are common knowledge. In all treatments, information on the sum of contributions is provided at the punishment stage (if there is one) and the end of the period (i.e., $\sum_{i=1}^{n-1} a_i^{\sim S} + a_i^S$).

Overall, we present seven treatments, two symmetric treatments (SYM_L and SYM_H) and two asymmetric treatments (ASYM_L and ASYM_H) without punishment.[17] Finally, we have three asymmetric treatments with punishment ASYM_NT_PUN, ASYM_L_PUN, and ASYM_H_PUN varying the magnitude of transparency. Therefore, in the treatments without punishment, SYM_L, SYM_H, ASYM_L, and ASYM_H, the payoff to an agent *i* in the respective treatments is given by

[17] We refrain from running symmetric punishment treatments, since this case is well documented and lies beyond the scope of this paper's focus. Likewise, we refrain from running symmetric and asymmetric treatments under non-transparency without accountability, since standard agents do not have any mean to discipline the special agent only.

$$\pi_i = w - a_i + h\left(E + \sum_{j=1}^{n} a_j\right)$$

where $h$ denotes the per capita return to the public good with $1/n < h < 1$, $a_i$ denotes *i*'s transfer to the public good account. In the experiment, we choose $n = 4$, $h = 0.4$, $w = 12$ and $E = 32$.

In contrast, in the treatments with punishment, ASYM_NT_PUN, ASYM_L_PUN, and ASYM_H_PUN, the payoff to an agent *i* in the respective treatments is given by

$$\pi_i = w + p - a_i + h\left(E + \sum_{j=1}^{n} a_j\right) - \sum\nolimits_{j \neq i} p_{i \to j} - 3\sum\nolimits_{j \neq i} p_{j \to i}$$

In the experiment, we choose $p = 5$;[18] Table 1 summarizes the experimental design.

**Table 1.** Experimental Design.

| | | One Special (Give-and-Take) Agent? | | |
|---|---|---|---|---|
| | | No | Yes | |
| Transparency of Individual Contributions | None *(neither identity nor actions)* | | | ASYM_NT_PUN |
| | Low *(identity but not actions)* | SYM_L | ASYM_L | ASYM_L_PUN |
| | High *(both identity and actions transparent)* | SYM_H | ASYM_H | ASYM_H_PUN |
| | | No | | Yes |
| | | Punishment Stage? | | |

All games are played repeatedly for ten periods. All parameters and payoff functions are common knowledge. At the end of each period, participants receive feedback: in the non-transparency and low transparency treatments they are shown information about their

18 Notice that one may argue that participants in the PUN treatments play the game differently because they are endowed with 5 additional Taler. In contrast, we believe that it is important to endow participants in the punishment treatments additionally so that they do not save Taler from their initial endowment (that they may contribute to the public good otherwise) for consecutive punishment.

earnings and the final amount of Taler in the group account (i.e., the sum of Taler in the public good net of contributions to and extraction from the group account). In contrast, in the high transparency treatments subjects are also able to identify *individual* contributions to and extractions from the group account. Participants know that the experiment terminates after ten periods; the composition of the group remains constant throughout the entire 10 periods of the experiment (partner design).

**Table 2.** Summary of Experiment Sessions.

| Session | Number of groups | Number of participants | Treatment |
|---|---|---|---|
| 1 | 5 | 20 | ASYM_L |
| 2 | 5 | 20 | SYM_L |
| 3 | 6 | 24 | ASYM_L |
| 4 | 5 | 20 | SYM_L |
| 5 | 5 | 20 | ASYM_H |
| 6 | 4 | 16 | ASYM_H |
| 7 | 5 | 20 | SYM_H |
| 8 | 4 | 16 | ASYM_H |
| 9 | 5 | 20 | SYM_H |
| 10 | 4 | 16 | ASYM_L |
| 11 | 6 | 24 | ASYM_H_PUN |
| 12 | 7 | 28 | ASYM_L_PUN |
| 13 | 7 | 28 | ASYM_L_PUN |
| 14 | 6 | 24 | ASYM_H_PUN |
| 15 | 5 | 20 | ASYM_NT_PUN |
| 16 | 5 | 20 | ASYM_NT_PUN |
| 17 | 5 | 20 | ASYM_NT_PUN |

Note: Numbers of groups across treatments are not equal due to some registered subjects not showing up.

Once the participants were seated and logged into the terminals, a set of instructions was handed out and read out loud by the experimenter.[19] In order to ensure that subjects

[19] We mainly followed the instructions of Fehr and Gächter (2000), but slightly changed the wording. For instance, instead of 'contributions to a project', instructions asked participants to divide tokens between a private and a group account. English translations of the German instructions are enclosed in appendix A.

understood the respective game, experimental instructions included several numerical examples and participants had to answer nontrivial control questions via their computer terminals.[20] At the beginning of the experiment subjects were randomly assigned to groups of four. In the ASYM treatments, one member per group is randomly determined to be the special agent for all ten periods. All experimental sessions were conducted in the computer laboratory of the Faculty of Economic and Social Sciences, University of Hamburg, Germany in March to April 2011, April and October 2012. Each session lasted approximately one hour. We used z-Tree (Fischbacher, 2007) to program and ORSEE (Greiner, 2004) for recruiting. In total, 356 subjects participated in the experiment, providing 10 independent observations for SYM_L and SYM_H, 15 independent observations for ASYM_L, 13 independent observations for ASYM_H, 15 independent observations for ASYM_NT_PUN, 14 independent observations for ASYM_L_PUN, and 12 independent observations for ASYM_H_PUN.[21] Each subject participated only once. Subjects were students a variety of academic backgrounds, including economics. At the end of the experiment, one of the periods was randomly selected as the period that determined earnings with an exchange rate of 1 Taler equal 30 Euro Cents. Including a show-up fee of 4 Euro, the average payment over all treatments was 11.50 Euro. Table 2 summarizes the information for all 17 sessions.

## 3.3 Predictions

Of course, the standard game theoretic prediction with respect to pure money maximization for our finitely repeated public good game in all treatment conditions is zero contributions of initial endowment and – if possible – the maximal transfer to the private account. We thus predict $a_i^{\sim S} = 0$ and $a_i^S = -E$, even when punishment is possible (notice that punishment is a second order public good: the person who distributes a punishment point bears its cost, while the entire group benefits from disciplining free-riders from the public good).

Despite the equilibrium prediction of zero contributions, an enormous number of experiments have demonstrated positive contributions by players from the first to the very last period of experiments (Ledyard, 1995, Chaudhuri, 2011). Punishment combined with

---

[20] In case a participant did not answer the questions correctly, she was given a help screen that explained the correct sample answers in detail. We believe this might further reduce experimenter demand effects compared to individual talks with subjects. See Zizzo (2010) for more information on experimenter demand effects.

[21] Unequal number of independent observation is due to non-show-ups of participants for sessions.

transparency turned out to sustain cooperation very successfully in symmetric group compositions (e.g., Fehr and Gächter, 2000). Social preferences (Meier, 2007), warm-glow of giving (Andreoni, 1990) and strategic considerations might explain such departures from the standard game theoretic prediction.

Transparency is predicted to play an important role, in particular in asymmetric settings as it allows players to conditionally cooperate (cf., e.g., Fischbacher et al., 2001, Frey and Meier, 2004, and Fischbacher and Gächter, 2011).[22] Notice that the special player can calculate the average contributions of the standard players from the sum of contributions, since she knows her take-out rate. Yet, the information of standard players does not allow them to conclude very little from the sum of contributions.[23] As such, only high transparency facilitates conditional cooperation (at least) among standard agents and allows them to delink their actions from those of the special player. Moreover, one may hypothesize that positive contributions by the special agent may over-proportionally trigger cooperativeness of standard players as the same (positive) contribution may be perceived as more kind when coming from the special than from a standard player as the special agent could have extracted the public good instead.[24] Summarizing both effects, one would expect that cooperation rates in ASYM_H are at least as high as in ASYM_H, joint with a larger in-group conditional cooperation among standard players.

To the contrary, following broad experimental evidence we assume that punishment under high transparency sustains cooperation, even in the case of asymmetrically endowed agents (see, e.g., Reuben and Riedl, 2013). We would like to interpret our claim such that punishment under high transparency creates accountability. That is, standard agents can make special agents accountable for their depletion of the initial public good, while both standard and special agents may make non-cooperative standard agents accountable for not-contributing to the public good. The more accountable agents are for their degree of cooperation, the more cooperatively they act.

---

22 Recall that transparency has no effect on contribution rates in symmetric settings (e.g., Weimann, 1994), since one can easily compute average contributions from the sum of contributions in the non-transparent setting. Thus, we do not expect a significant difference in terms of contributions between SYM_L and SYM_H.

23 At the very most, extreme take-outs or contributions are revealed: (almost) full contributions by all standard and special players (almost) maximize the sum of contributions, whereas sums of contributions below zero indicate take-outs by the special player.

24 Some contemporary reciprocity models (e.g., Rabin, 1993, Dufwenberg and Kirchsteiger, 2004) stress the importance for the action space for the perceived kindness of an agent, i.e. actual actions are evaluated against the range of possible actions.

The crucial question is how much accountability can be implemented with punishment under less transparency. Let us start with the non-transparent treatment. One may argue that agents may abstain from any punishment on the basis of such poor information. However, there is substantial evidence that subjects punish disregarding their poor information concerning others' behavior (see, Ambrus and Greiner, 2011; Grechenig et al. 2010). Given that agents punish, one can ask, which punishment may create some accountability for non-cooperation. Recall that the special agent has an informational advantage in ASYM_NT_PUN since she can compute the average contributions of the standard players. Suppose the special player seeks to increase standard players' contributions and standard players' contributions are sufficiently homogenous. In this case, only the special player can create some degree of accountability by punishing all standard players equally. However, the probability that this happens is rather low, even more, since special players are not accountable at all for their behavior, as standard players may punish arbitrary players.

On the other hand, low transparency allows some stigmatization of the special player. That is, the special player attracts punishment due to statistical discrimination so that ASYM_L_PUN allows standard agents to make special players to some degree accountable for taking from the public good (as mentioned earlier, the sum of contributions reveals to behavior of the special player to some degree). This may increase contributions by standard agents, although standard players and the special player cannot make other standard players clear-cut accountable for their non-cooperation. Nonetheless, the overall effect of stigmatization and punishment for cooperation rates and payoffs in the public good game remains unclear. It may increase special agents' and standard agents' contributions, but it could trigger special agent's revenge, that is, punishment towards arbitrary standard agents in subsequent periods.[25] Yet, it seems that overall the degree of cooperation increases with the degree of transparency, since increasing transparency increases the degree of accountability. Therefore, we hypothesize higher contributions in ASYM_L_PUN than in ASYM_NT_PUN and higher contributions in ASYM_H_PUN than in ASYM_L_PUN, and expect important differences in the channels through which punishment interacts with transparency when establishing accountability for misbehavior. We discuss our experimental results on the effects of punishment and transparency and in their relations in the next section.

[25] Several authors analyze the effect of punishment against contributors in symmetric public good games, "anti-social punishment," (e.g., Herrmann et al., 2008), and revenge, "counter-punishment" (e.g., Nikiforakis, 2008).

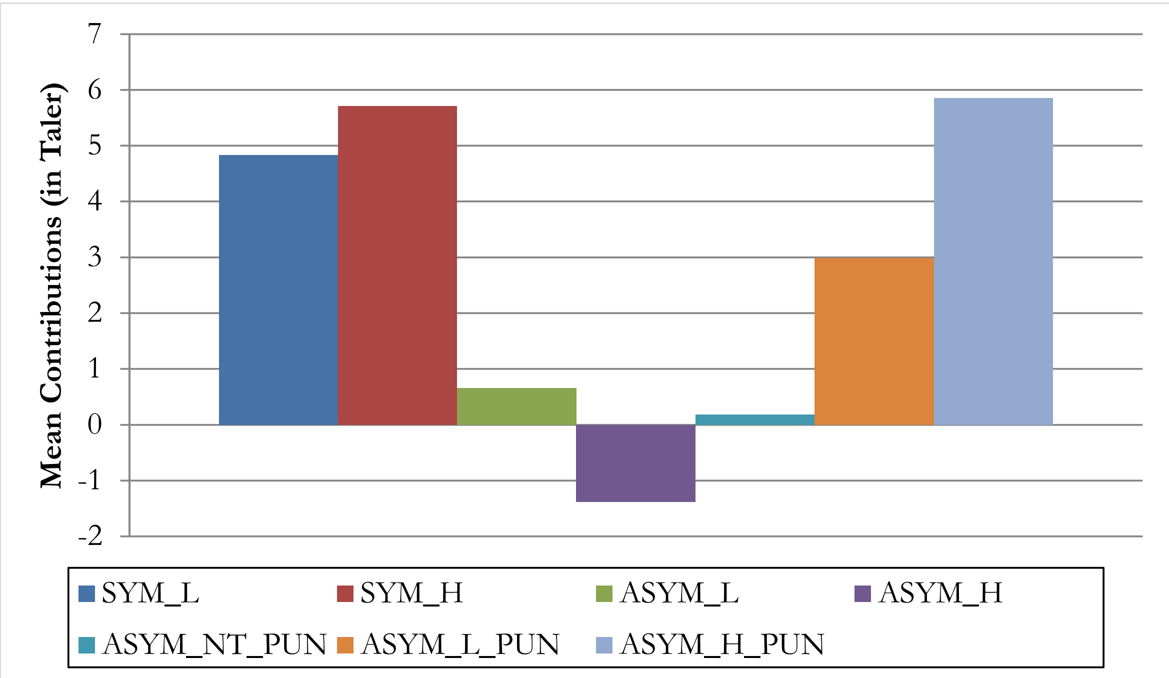
## 3.4 Results

We craft our results by first comparing the public good provision levels in the different treatments. In a second step we take a closer look at the underlying decisions of special and standard agents. For this purpose, we apply several non-parametric and parametric methods to disentangle the data. Third, we analyze the mechanism of when and how subjects use punishment, and under which treatment conditions punishment influences behavior of standard and special agents.

### *3.4.1 Public Good Provision*

Table 3 reports summary statistics and Figure 1 illustrates the mean contributions decisions for all treatments. Recall that the contribution decisions range from 0 to 12 Taler in the SYM treatments and in the ASYM treatments for standard agents, while special agents' decisions in the ASYM treatments and from -32 to 12 Taler. Therefore, the mean contribution decision in ASYM lies potentially between -8 and 12 Taler.

**Figure 1.** Decisions in all Treatments, averaged across all Types of Agents.



In the treatments with symmetric action sets, SYM_L and SYM_H, mean contributions over all ten periods amount to 4.84 and 5.71 Taler, respectively. A two-tailed Mann-Whitney (hereafter MW) test does not reject the null hypothesis of equal contributions in the two symmetric treatments over all ten periods (p = 0.4963) as well as in the final period (p = 0.1605).[26] We thus confirm the result by Weimann (1994) and Croson (2000) that transparency does not significantly influence contributions in symmetric settings. These treatments serve as a baseline.

The introduction of the extraction option to a special agent significantly reduces these averages to 0.66 Taler in ASYM_L and -1.38 Taler in ASYM_H. MW tests comparing SYM_H and SYM_L with ASYM_H and ASYM_L over all ten periods and in period 10 yield p-values ranging from 0.0006 to 0.0244. Interestingly, high transparency in ASYM_H tends to reduce the public good provision level compared to ASYM_L (MW tests over 10 periods p = 0.1172, final period p = 0.0335). A downward trend appears particularly detrimental to the contributions in these two treatments. Last period decision average at -2.73 Taler in ASYM_L and -5.65 Taler in ASYM_H such that the total provision of the public good is less than in the initial condition. Here more transparency does not yield more cooperation and higher public good provision; contrarily, transparency tends to backfire. We further explore drivers of this finding below.

Let us now turn to the three punishment treatments. We aim at tracing out the transparency condition that is needed to turn the punishment opportunity into true accountability. While the contribution level of 0.18 Taler in ASYM_NT_PUN does not differ from ASYM_L and ASYM_H, punishment increases the contribution levels to 2.98 Taler in ASYM_L_PUN and 5.85 Taler in ASYM_H_PUN, respectively. Given the punishment opportunity, the higher the level of transparency is, the higher the public good provision levels are. Comparing ASYM_NT_PUN and ASYM_H_PUN over all ten periods a MW test reports a difference at p = 0.0147.[27] Note again that the effect obtained without punishment tends to be reversed. Figure 2 further illustrates the time-trends across the diverse treatments.

---

26 For all Mann-Whitney tests one observation is a group's mean contribution in a time interval. For instance, comparing SYM_L and SYM_H we analyze ten groups per treatment yielding ten observations for each treatment.

27 Comparing ASYM_L_PUN and ASYM_H_PUN, transparency has a minor increasing impact on average contributions (MW test over all periods, p = 0.1648; final period ASYM_H_PUN > ASYM_L_PUN, p = 0.0883).

**Table 3.** Summary Statistics.

| | Treatment | | | | | | |
|---|---|---|---|---|---|---|---|
| Statistic | SYM_L | SYM_H | ASYM_L | ASYM_H | ASYM_NT_PUN | ASYM_L_PUN | ASYM_H_PUN |
| Mean decision | 4.84 (4.42) | 5.71 (4.44) | 0.66 (11.62) | -1.38 (14.78) | 0.18 (11.52) | 2.98 (7.32) | 5.85 (9.41) |
| Mean decision of *standard* agents | 4.84 (4.42) | 5.71 (4.44) | 4.38 (4.27) | 4.81 (4.79) | 4.23 (4.19) | 3.89 (4.05) | 7.49 (5.18) |
| Mean decision of *special* agents | - | - | -10.51 (17.89) | -19.95 (18.59) | -11.97 (16.81) | 0.25 (12.49) | 0.93 (15.59) |
| Mean pun. points | - | - | - | - | 0.87 (1.46) | 0.99 (1.99) | 1.07 (2.38) |
| Mean pun. points to *special* agents | - | - | - | - | 0.69 (1.29) *(player types are unknown here)* | 2.30 (3.30) | 2.21 (3.79) |
| Mean pun. points to *standard* agents | - | - | - | - | 0.93 (1.51) *(player types are unknown here)* | 0.55 (0.95) | 0.69 (1.49) |
| Mean income | 27.70 (3.59) | 28.23 (3.75) | 25.19 (10.75) | 23.97 (13.96) | 26.44 (12.07) | 27.65 (7.32) | 29.03 (8.47) |
| Mean income of *standard* agents | 27.70 (3.59) | 28.23 (3.75) | 21.47 (8.03) | 17.78 (8.28) | 22.22 (9.58) | 28.07 (7.41) | 28.41 (8.45) |
| Mean income of *special* agents | - | - | 36.37 (10.14) | 42.55 (10.67) | 39.11 (9.65) | 26.38 (6.91) | 30.87 (8.31) |

Note: Standard deviations in parentheses.

**Figure 2.** Public Good Provision in all Six Treatments, by Period.

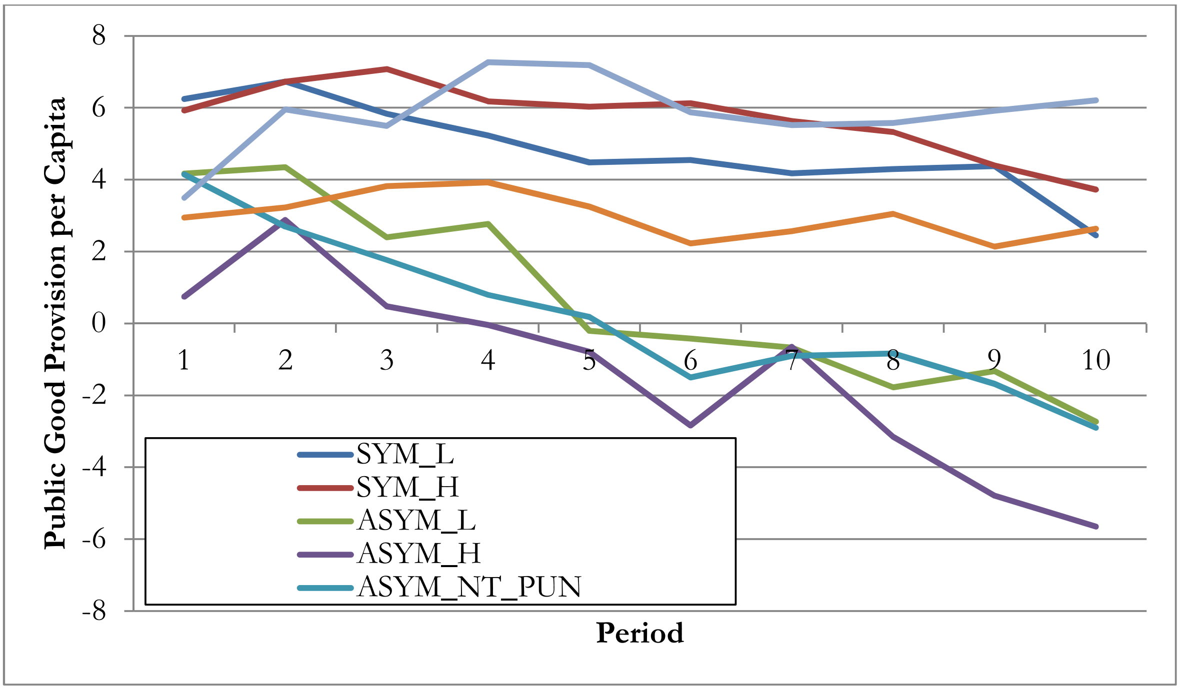
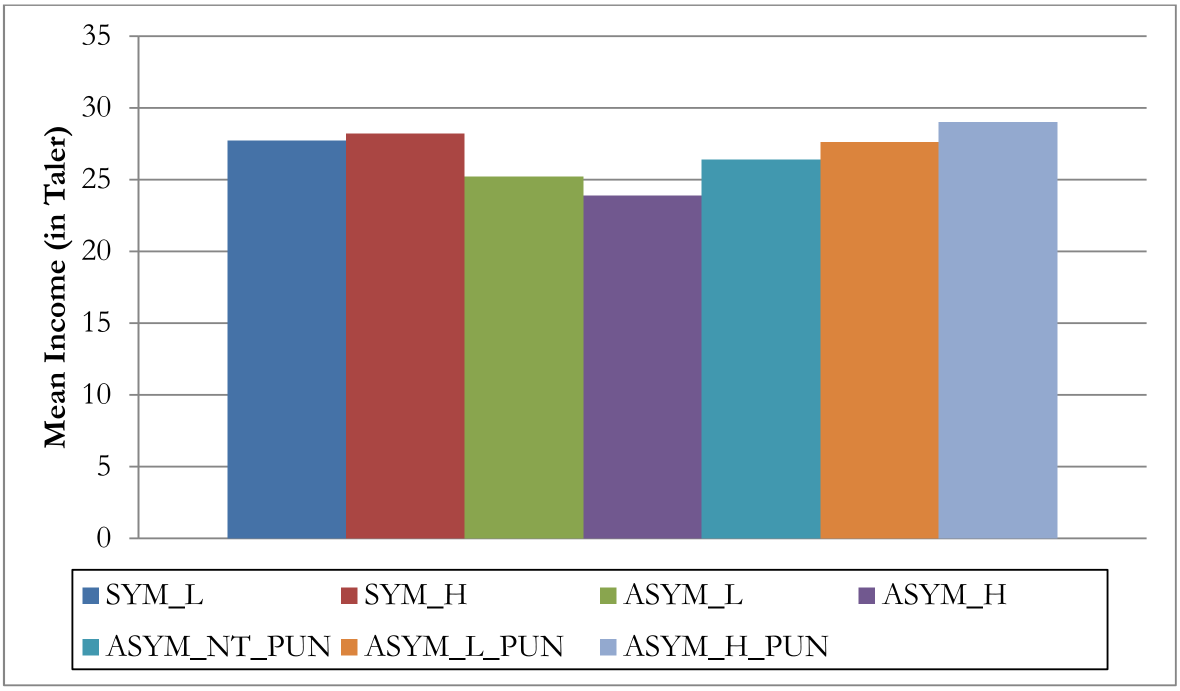



**Figure 3.** Mean Income in all Treatments, averaged across all Types of Agents.




A linear regression analysis with random-effects specification that we report in Table 4 confirms these findings. Model I reveals that the combination of punishment and high transparency catalyzes contributions to the public good: contribution levels are significantly higher than in reference treatment of model I, ASYM_L. Conversely, the large negative (yet insignificant) coefficient of ASYM_H shows that contributions in that treatment are equal to and tend to be lower than in ASYM_L. That is, transparency without punishment does not lead to greater public good provision.

Taking a look at mean income in the different treatments tells a similar story.[28] Table 3 reports mean income for all treatments; they range from 23.97 Taler in ASYM_H to 29.03 Taler in ASYM_H_PUN. Note that these are the two asymmetric treatments with high transparency – only that the punishment opportunity turns the impact of transparency from being detrimental to mean income to being beneficial for income. MW tests with mean incomes per group as independent observations reveal incomes differences such that on average agents in ASYM_H earn significantly less compared to SYM_L ($p = 0.0013$), SYM_H ($p = 0.0008$), ASYM_NT_PUN ($p = 0.0322$), ASYM_L_PUN ($p = 0.0369$) and ASYM_H_PUN ($p = 0.0502$), while agents in ASYM_L earn significantly less compared to SYM_L ($p = 0.0198$) and SYM_H ($p = 0.0047$); all other differences between treatments are insignificant (i.e., $p>0.05$). Figure 3 illustrates mean income for all agents and separately for standard and special agents.

### *3.4.2 Disentangling Standard and Special Agents*

We now discuss the differences between the underlying decisions of special and standard agents in more detail. Figure 4 separates the contribution decisions by special and standard agents. Obviously, special agents in ASYM_H extract more Taler from the public good than special agents in ASYM_L (-19.95 vs. -10.51). Using an individual's mean contribution over the ten periods as the unit of observation, a MW test shows that this difference is significant ($p = 0.0421$). That is, higher transparency leads special agents to contribute substantially *less* to the public good. Conversely, contributions of standard agents in both ASYM_L (4.38 Taler) and ASYM_H (4.81 Taler) do not differ significantly (MW test, $p = 0.3941$). They also do not differ from the average contributions in the symmetric treatments. We thus deduce that a

[28] Notice that comparing payoffs of public good games with and without punishment is by no means a trivial thing: Punishment devastates efficiency, which depends crucially on marginal destruction rate of punishment points, while it (usually) facilitates efficiency by means of higher contributions. In addition, players receive an additional endowment for punishment in our experiment for good reasons (see footnote 8). Hence, evaluating income effects between non-punishment and punishment treatments is very sensitive with respect to parameters. Since this issue is not at the focus of our study, we simply compare incomes including the additional endowment. Of course, studies interested in efficiency effects of punishment (in asymmetric settings) have to take into considerations a number of crucial parameters like the marginal destruction rate of punishment or the time length of play (see, e.g., Gächter et al., 2008).

negative effect of transparency on public good provision is caused by lower contribution levels of special agents.

Regression model IV in Table 4 supports this finding. Additionally controlling for time trends by including the dummy variable Period 6_10 and its interactions with treatment dummies, model V shows no different time trend between ASYM_H and the reference treatment ASYM_L. We therefore formulate the following result:

**Result 1.** *High transparency without punishment backfires: special agents exploit the public good to a larger extent when their actions are transparent.*

Although the result for transparency surprises at the first glance, there are plausible reasons for it: special agents, paired with three standard agents in ASYM_L, may believe that their extraction behavior remains undetected if standard agents cannot perfectly identify it from the average contributions. For example, if a standard agent contributes $a_i^{\sim S}$ and observes a public good provision level of $E + a_i^{\sim S}$, she cannot infer whether no other agent has contributed or the special player took exactly the amount that the other two standard agents contributed. If standard agents are conditional cooperators,[29] they may hesitate to punish all other group members by reducing their contributions in subsequent periods as a reaction to low average contributions of group members in the present period. As a consequence, special agents may attempt to exploit a "moral wiggle room" (see, e.g, Dana et al., 2007) in the low transparency treatment. With high transparency, exploiting special agents cannot hide their taking such that their decision may no longer be restrained by this fear of being explicitly detected. Consequently, taking may increase.

It is interesting to note, however, that this additional taking by the special agent does not induce standard players to reduce their contributions significantly (see Figure 4). In line with the non-parametric results reported above, the regression in Table 4, model II and III, do not show a significant difference between ASYM_L as a baseline and ASYM_H or the symmetric treatments.

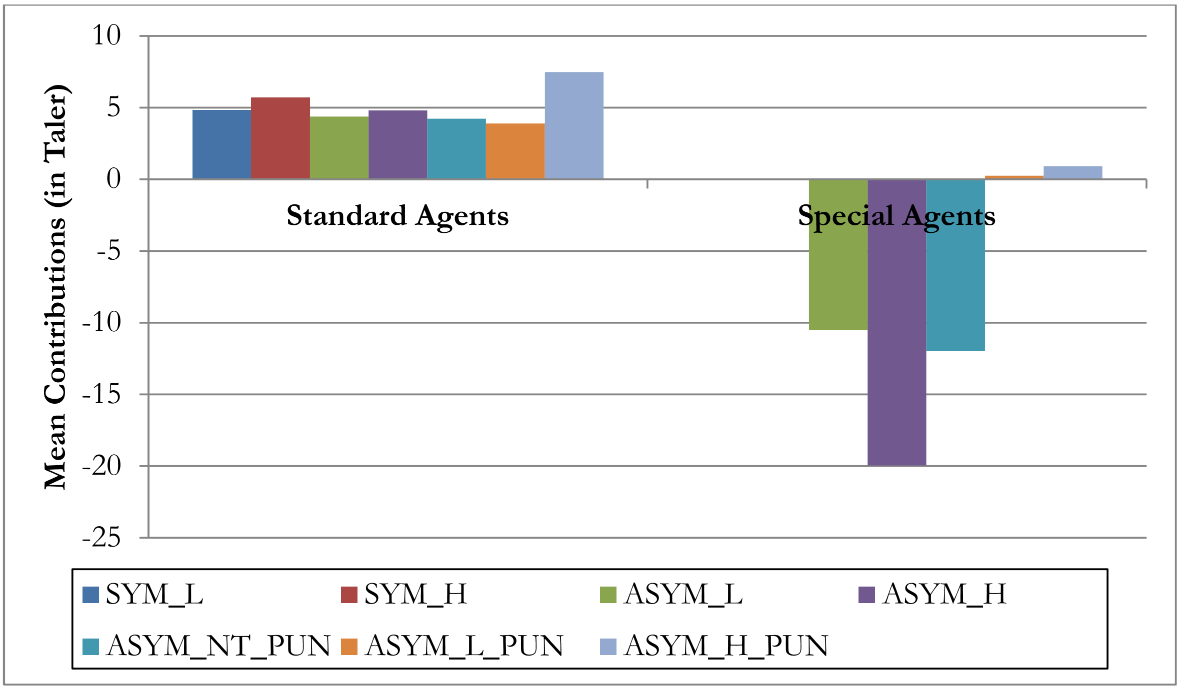
[29] The downward trend of contributions to the public good in the data at hand, like reported in uncountable other public good experiments, suggests the presence of conditional cooperation (see, e.g., Fischbacher and Gächter, 2010). In the first period a conditional cooperator contributes a positive amount with the anticipation that her cooperation will be matched by others in order to overcome the social dilemma. If her anticipation is not met and her cooperation is exploited by group members, she will decide to reduce her contribution in the second period.

**Table 4.** Linear Regressions of Contributions to the Public Good.

| Independent Variable | Dependent Variable: Contribution | | | | |
|---|---|---|---|---|---|
| | I<br>all treatments<br>all agents | II<br>all treatments<br>only *standard* agents | III<br>all treatments<br>only *standard* agents | IV<br>only asymmetric treatments<br>only *special* agents | V<br>only asymmetric treatments<br>only *special* agents |
| SYM_L | 4.177***<br>(1.247) | 0.453<br>(0.962) | 0.704<br>(1.075) | | |
| SYM_H | 5.054***<br>(1.234) | 1.330<br>(0.944) | 1.389<br>(0.975) | | |
| ASYM_H | -2.037<br>(1.569) | 0.431<br>(0.821) | 0.969<br>(0.895) | -9.441**<br>(4.812) | -11.062**<br>(5.441) |
| ASYM_L_PUN | 2.324*<br>(1.344) | -0.489<br>(0.859) | -0.581<br>(0.970) | 10.763***<br>(3.932) | 4.700<br>(4.338) |
| ASYM_H_PUN | 5.194***<br>(1.850) | 3.112**<br>(1.344) | 2.193*<br>(1.326) | 11.438**<br>(4.527) | 6.167<br>(4.599) |
| ASYM_NT_PUN | -0.480<br>(1.440) | -0.156<br>(0.728) | 0.204<br>(0.860) | -1.453<br>(4.415) | -3.720<br>(4.405) |
| Period 6_10 | | | -1.227**<br>(0.473) | | -12.627***<br>(3.715) |
| SYM_L x Per6_10 | | | -0.503<br>(0.794) | | |
| SYM_H x Per6_10 | | | -0.118<br>(0.626) | | |
| ASYM_H x Per6_10 | | | -1.076<br>(0.912) | | 3.242<br>(4.787) |
| ASYM_L_PUN x Per6_10 | | | 0.184<br>(0.582) | | 12.127***<br>(4.214) |
| ASYM_H_PUN x Per6_10 | | | 1.838**<br>(0.764) | | 10.543*<br>(5.521) |
| ASYM_NT_PUN x Per6_10 | | | -0.720<br>(0.652) | | 4.533<br>(4.190) |
| Constant | 0.658<br>(0.946) | 4.382***<br>(0.515) | 4.996***<br>(0.630) | -10.513***<br>(3.008) | -4.200<br>(2.883) |
| Observations | 3560 | 2870 | 2870 | 690 | 690 |
| Individuals | 356 | 287 | 287 | 69 | 69 |
| Groups | 89 | 89 | 89 | 69 | 69 |

Note: Random effects estimation with robust standard errors clustered at group level. ASYM_L is the baseline in all estimations. Standard errors in parentheses, significance: *p < 0.10, **p ≤ 0.05, ***p < 0.01.

**Figure 4.** Decisions in all Treatments, by Types of Agents.




The non-parametric results already indicated the potential benefits from punishment options. Depending on the magnitude of transparency, punishment makes agents accountable for their actions. As discussed above and shown by the results across Table 4, punishment in ASYM_NT_PUN is not able to foster contributions and results are very similar to ASYM_L. Table 4 however reveals quite different effects for the respective types of agents when there is low or high transparency: for special agents, we find significant positive coefficients, both for 'ASYM_L_PUN' and 'ASYM_H_PUN' in model IV and for 'ASYM_L_PUN x Period 6_10' and 'ASYM_H_PUN x Period 6_10' in model V, respectively. In other words, in comparison with the reference treatment (ASYM_L), punishment leads to significantly larger contributions with this difference even becoming more pronounced in later periods. However, there is no significant difference between the coefficient for ASYM_L_PUN and ASYM_H_PUN (F-test $p=0.873$ in model IV) or the respective time trends implying that stigmatization under low transparency allows for accountability of special agents' behavior and similar contribution rates as punishment under high transparency.

**Result 2.** *Punishment under low and high transparency improves the cooperativeness of special agents.*

In contrast, we find a different picture for standard agents. Figure 4 shows the mean contributions for standard agents only. Here, punishment with non- or low transparency does *not* lead to increased contributions. Punishment and high transparency, however, complement each other in the sense that they also create accountability for standard agents and contribution levels increase when both apply (i.e., in ASYM_H_PUN). Table 4, model II and III confirm these results: only coefficients for ASYM_H_PUN and 'ASYM_H_PUN x Period 6_10' are significantly positive. In consequence, the treatment condition with punishment and transparency is the only one where contributions of standard agents are stable over time, while they follow the typical decreasing pattern in all other treatment conditions.

**Result 3.** *Transparency complements punishment for standard agents: if punishment is introduced under high transparency standard agents increase their contributions to the public good.*

In order to gain additional insights into the channels through which transparency and punishment affect the contribution decisions of standard and special agents, we now have a closer look at the temporal nature of decisions. As mentioned earlier, one obvious channel is conditional cooperation: higher degrees of transparency allow agents to condition their own contributions more specifically on the contributions of other group members – this channel could apply even without punishment. With punishment, the fear of being punished could provide another channel: without transparency punishment could only be conditioned on average contributions of others and the type of the group member, while punishment can be better targeted if actions are transparent.

To study conditional cooperation, we first concentrate on the actions of standard players. Table 5 reports results from regression models which, for each treatment, analyze the relation between contributions by standard agents and their group members' choices in the previous period ("t-1"). Along the variable "individual's own contribution in t-1", we introduce the variable "group contribution in t-1" denoting in the symmetric treatments and the asymmetric treatments with no or low transparency the total contributions of all other three group members. In ASYM_H and ASYM_H_PUN, however, this variable denotes the total contributions by the other two standard agents in the group, while we separate the effect of the decision of the special agent by incorporating the separate variable "the special agent's contribution in t-1."

We find that standard agents positively reciprocate on the other agents' contributions: the larger the average contribution from other players in the previous period is, the more the agent contributes. However, when transparency allows to condition actions on previous contributions from fellow standard agents and the special agents, we see that the impact from previous period contributions of the special player tends to be smaller than the one from other standard agents.[30] That is, agents appear to be influenced more strongly by agents of their own type. As a consequence, the larger exploitation of the public good by the special agent in ASYM_H relative to ASYM_L may not induce standard agents to reduce their own contributions too much as transparency allows them to delink their actions from the special agent by comparing their actions with those fellow agents of their own type. We summarize this discussion in the following result:

**Result 4.** *Conditional cooperation (in particular) of standard agents tends to focus on agents of the same type.*

**Table 5.** Linear Regressions of Contributions to the Public Good, Individual Behavior of *Standard* Agents (by Treatment).

| | Dependent Variable: Contribution | | | | | | |
|---|---|---|---|---|---|---|---|
| | VI | VII | VIII | IX | X | XI | XII |
| Independent Variable | SYM_L | SYM_H | ASYM_L | ASYM_H | ASYM_L_PUN | ASYM_H_PUN | ASYM_NT_PUN |
| Group contribution in t-1 (excluding *i*'s own contribution) | 0.0938*** (0.025) | 0.083** (0.039) | 0.025*** (0.007) | 0.118*** (0.034) | 0.052*** (0.009) | 0.088* (0.048) | 0.058*** (0.013) |
| Individual's own contribution in t-1 | 0.539*** (0.072) | 0.645*** (0.100) | 0.604*** (0.070) | 0.478*** (0.048) | 0.645*** (0.050) | 0.776*** (0.084) | 0.407*** (0.082) |
| Period 6_10 | -0.254 (0.276) | -0.587** (0.263) | 0.051 (0.221) | -0.516 (0.432) | -0.279 (0.329) | -0.455** (0.218) | -0.148 (0.296) |
| The special agent's contribution in t-1 | - | - | *-Not observable for standard agents-* | 0.044** (0.018) | *-Not observable for standard agents-* | -0.054 (0.059) | *-Not observable for standard agents-* |
| Constant | 0.633 (0.476) | 0.704* (0.413) | 1.459*** (0.248) | 2.156*** (0.767) | 0.944*** (0.252) | 0.753** (0.297) | 2.383*** (0.550) |
| Observations | 360 | 360 | 405 | 351 | 378 | 324 | 405 |
| Individuals | 40 | 40 | 45 | 39 | 42 | 36 | 45 |
| Groups | 10 | 10 | 15 | 13 | 14 | 12 | 15 |

Note: Random effects estimation with robust standard errors clustered at group level. Standard errors in parentheses. Significance: *p < 0.10, **p < 0.05, ***p < 0.01.

[30] This tendency is not statistically significant. The coefficient "group contribution in t-1" is larger than the coefficient "the special agent's contribution in t-1", p=0.106 in model IX, p=0.186 in model XI.

**Table 6.** Linear Regressions of Contributions to the Public Good, Individual Behavior of *Special* Agents (by Treatment).

| Independent Variable | Dependent Variable: Contribution | | | | |
|---|---|---|---|---|---|
| | XIII<br>ASYM_L | XIV<br>ASYM_H | XV<br>ASYM_L_PUN | XVI<br>ASYM_H_PUN | XVII<br>ASYM_NT_PUN |
| Group contribution in t-1 (excluding *i*'s own contribution) | 0.214<br>(0.131) | 0.325*<br>(0.168) | 0.123<br>(0.084) | 0.160<br>(0.106) | 0.147<br>(0.134) |
| Individual's own contribution in t-1 | 0.865***<br>(0.128) | 0.538***<br>(0.100) | 0.559***<br>(0.154) | 0.621***<br>(0.064) | 0.662***<br>(0.092) |
| Period 6_10 | -2.000<br>(2.284) | -3.070<br>(2.626) | -0.483<br>(1.276) | -3.624**<br>(1.704) | -1.176<br>(2.117) |
| Constant | -7.227***<br>(2.488) | -13.644***<br>(2.774) | -0.699<br>(1.123) | -0.568<br>(2.869) | -6.833**<br>(2.907) |
| Observations | 135 | 117 | 126 | 108 | 135 |
| Individuals | 15 | 13 | 14 | 12 | 15 |
| Groups | 15 | 13 | 14 | 12 | 15 |

Note: Random effects estimation with robust standard errors clustered at group level. Standard errors in parentheses. Significance: *p < 0.10, **p < 0.05, ***p < 0.01.

Result 4 demonstrates that standard agents in their conditional cooperation, i.e., implicit punishment of other players through reducing their own contributions, focus more on agents of the own type. That is, the pure introduction of taking options for only a subset of agents generates some in-group behavior. Result 4 receives support when we analyze conditional cooperation of special agents. Table 6 presents similar specification as Table 5, but this time for special agents. In four out of five asymmetric treatments we do not find conditional cooperation of special agents with regard to the contributions of standard agents; in the remaining case the coefficient is significant at the 10 percent level. These findings further support Result 4 in that conditional cooperation tends to focus on agents of the same type, i.e., with the same action space. We will now analyze in the next step how the use of explicit punishment is affected by the transparency of the actions.

### *3.4.3 The Use of Punishment*

As mentioned above, we do not expect that players distribute substantially less punishment points in ASYM_NT_PUN than in the other two treatments despite the considerable lack of information. Indeed, there is no significant difference between the numbers of punishment points between treatment conditions (mean in ASYM_NT_PUN is 0.86, in ASYM_L_PUN 0.99, in ASYM_H_PUN is 1.07, for all MW tests $p > 0.1$), although there is tendency that higher transparency leads to more punishment. Interestingly, against our prediction that special players use their informational advantage in ASYM_NT_PUN to punish standard players on average if the sum of their contributions is low, there is no significant difference between the numbers of punishment points for special players between treatment conditions (mean in ASYM_NT_PUN is 0.88, in ASYM_L_PUN 1.04, in ASYM_H_PUN is 0.74, all MW test $p > 0.1$) nor a significant difference between special and standard players in ASYM_NT_PUN (mean for standard players is 0.87; for comparison, mean for standard players in ASYM_L_PUN and ASYM_H_PUN is 0.97 and 1.18, respectively; again, for all MW tests $p > 0.1$).

Who receives punishment? Table 3 provides summary statistics of punishment points assigned to standard and special agents in ASYM_NT_PUN, ASYM_L_PUN and ASYM_H_PUN. In ASYM_NT_PUN standard agents and special agents on average receive 0.69 and 0.93 punishment points, respectively. This difference is not statistically significant. Hence, non-transparency both with regard to identity and actions makes it impossible to sanction special agents even though the punishment opportunity is available.

Standard agents in ASYM_L_PUN receive 0.55 punishment points on average and 0.69 punishment points in ASYM_H_PUN. Special agents receive on average 2.30 punishment points in ASYM_L_PUN and 2.21 punishment points in ASYM_H_PUN. Special agents thereby receive substantially higher punishment than standard agents (MW test, $p = 0.0007$), while the change from low to high transparency neither affects the magnitude of punishment of special nor of standard agents.

This result indicates an important stigma (or statistical discrimination) effect: even though individual actions are not observable in ASYM_L_PUN, primarily special agents receive punishment. This holds even when controlling for the individual and group contribution level as is shown in Table 7. Estimation IIXX report results from a linear regression with random-effects specification that controls the punishment points received by an agent on her type (standard being the baseline), the treatment condition, an interaction term of both the agent type and the treatment condition, her own contribution decision, the group contribution (excluding her own contribution), and time effects. The agent's own contribution has a significant and negative effect on the number of punishment points received and the higher the contributions of others are, ceteris paribus, the more punishment is given to the agent. In the baseline treatment of specification IIXX, ASYM_L_PUN, special agents attract more punishment than standard agents. Hence, punishment and behavior are directly related even under low transparency, as expected.

**Table 7.** Analysis of Received Punishment.

| Independent Variable | IIXX<br>**Random-Effects Regression; dependent variable: # of Received Punishment Points (continuous)** | IXX<br>**First Hurdle; dependent variable: Being punished (dummy)** | XX<br>**Second Hurdle; dependent variable: # of Received Punishment Points (continuous)** |
|---|---|---|---|
| Individual's own contribution | -0.108***<br>(0.020) | -0.046***<br>(0.009) | -0.250***<br>(0.029) |
| ASYM_H_PUN | 0.485**<br>(0.206) | -0.011<br>(0.283) | 2.665***<br>(0.993) |
| ASYM_NT_PUN | 0.483**<br>(0.210) | 0.262<br>(0.241) | 2.804***<br>(1.012) |
| Special Agent | 1.339***<br>(0.403) | 0.554***<br>(0.185) | 2.939**<br>(1.267) |
| ASYM_H_PUN x Special Agent | -0.568<br>(0.747) | -0.437*<br>(0.258) | -2.285<br>(1.789) |
| ASYM_NT_PUN x Special Agent | -3.417***<br>(0.692) | -1.665***<br>(0.393) | -10.141***<br>(2.077) |
| Sum of the group contribution (excluding *i*'s own contribution) | 0.006<br>(0.006) | 0.002<br>(0.005) | 0.020<br>(0.024) |
| Period 6_10 | -0.388***<br>(0.082) | -0.407***<br>(0.128) | -0.119<br>(0.445) |
| Constant | 1.113***<br>(0.128) | -0.066<br>(0.183) | -1.836<br>(1.241) |
| Observations | 1640 | 1640 | |
| Individuals | 164 | 164 | |
| Groups | 41 | 41 | |

Note: Regression IIXX is a random effects estimation with robust standard errors clustered at the group level. Regressions IXX and XX are the two parts of a double hurdle analysis. The first hurdle is estimated using a Probit specification with robust standard errors clustered at group level. The second hurdle is estimated using truncated linear regression with robust standard errors clustered at group level. We used STATA 11.1 and Bill Burke's command *craggit* for double hurdle models (Cragg 1971, Burke 2009).
Standard agents in the ASYM_L_PUN treatment are the baseline. Standard errors in parentheses. Significance: *p < 0.10, **p < 0.05, ***p < 0.01.

This result may be attributed to a stigma, i.e. given that standard agents are not sure about the individual contributions of group members, they focus their punishment on the special agent. Here, high transparency changes the assignment of punishment points in two manners. While transparency generally increases the number of received punishment, the interaction of special agents in ASYM_H_PUN shows that punishment of special agents is reduced. Hence, transparency shifts a share of the punishment away from the (stigmatized) special agent to all agents and the stigma is lifted. This finding is supported by the fact that the joint effect of the special agent dummy (1.339, p = 0.001) and the interaction effect of special agents and high transparency ("ASYM_H_PUN*special agent") (-0.568, p = 0.447) is not significantly different from zero (Chi square test, p = 0.1876). We formulate the following result:

**Result 5.** *With stigmatization but without transparency of individual actions, special agents are discriminated against by receiving significantly more punishment points.*

Result 5 shows statistical discrimination against special agents. Their stigma that – just because they can take from the public good – they will (probably) be responsible for an apparent low total provision level, leads them to receive more punishment points. Thus punishment based on stigmatization creates accountability for special players under low transparency. As a consequence, they try to reduce this punishment by contributing more and thereby increasing the provision level of the public good. This punishment channel combined with stigmatizing special agents is therefore a channel that increases special agents' contributions under punishment even if actions are not transparent.

We finally take a closer look at the drivers of getting punished and the punishment magnitude separately. We employ a double hurdle model as this method is able to let us understand the drivers of punishment more thoroughly. In our experiment the idea underlying Cragg (1971)'s approach is as follows: the decision process of whether or not to assign punishment points to group members may not be the same process determining how many punishment points to assign.[31] Hence the first hurdle, a Probit model, analyzes the drivers of the *probability* of being punished. The second hurdle, a linear regression truncated at zero (i.e. only observations with positive amounts of punishment points are taken into consideration), estimates the drivers of the punishment *intensity*. The estimation results are also reported in Table 7, so that they are easily comparable to the random-effects estimation results.

[31] This analysis is commonly used in the literature, e.g. by Nikiforakis (2008) and Nikiforakis (2010). We employ Burke (2009)'s STATA command *craggit*.

We again confirm that higher own contribution to the public good reduces both the punishment likelihood and the magnitude of punishment. Interestingly, the first hurdle reports that the probability of punishment is not higher in ASYM_NT_PUN and ASYM_H_PUN compared to ASYM_L_PUN; the magnitude of punishment increases however. This result indicates that punishment will get more severe due to the fact that agents are certain about group members' contributions in ASYM_H_PUN. Hence, transparency of actions appears to clear the doubt about whether a punishment is just or not and thereby increases intensity. Like in the random-effects estimation the hurdle model provides further evidence that special agents both have a higher likelihood of getting punished and higher punishment intensity – yet only when there is low transparency. In ASYM_H_PUN this effect disappears by the interaction effect of special agent and ASYM_H_PUN. The results from the hurdle model thereby provide evidence that transparency shifts the focus of punishment from punishing mainly special agents to including defecting standard agents as well. In summary, we find

**Result 6.** *Transparency of identity and actions altogether allows for targeted punishment for both special and standard agents such that the intensity of punishment increases, while in the presence of transparency of actions, the stigma of special agents is lifted.*

Taken together, Results 5 and 6 indicate an important channel through which transparency of identity and actions affect individual contribution decisions: when individual actions are not transparent, punishment is targeted towards those agents that are most likely to have contributed less, i.e., the special agents who had the option to take from the public good. This statistical discrimination against special agents has two effects: (i) special agents contribute more in the hope of being punished less, (ii) standard agents do not have to fear punishment and therefore contribute not differently than without punishment. Transparency of actions eliminates the stigma of special agents such that they receive punishment for their actions not differently from standard agents. In consequence, in ASYM_H_PUN standard agents are disciplined to contribute more to the public good as well.

Finally, comparing punishment in ASYM_NT_PUN with punishment in ASYM_L_PUN and ASYM_H_PUN reveals a weakness of the peer-punishment mechanism that has been studied intensely over more than a decade (e.g., Fehr and Gächter, 2000, 2002; Nikiforakis, 2008, 2010). The large body of literature so far takes a substantial level of transparency of decisions as given. By introducing ASYM_NT_PUN, we show in line with previous studies (e.g., Grechenig et al., 2010) that transparency of identity is a necessary condition for punishment to achieve the desired increase in contributions and to overcome the dilemma. Consequently punishment in ASYM_NT_PUN loses its bite and decisions are comparable to ASYM_L.

## 3.5 Conclusion

Transparency is often seen as crucial in order ensure a functioning provision of public goods by (self-interested) individuals. In this chapter, we explored the interplay of transparency and punishment for voluntary contributions to public goods. More specifically, we analyze which degree of transparency creates sufficient accountability that sustains the cooperation within groups. For this purpose, we study situations where agents differ with respect to their available actions, i.e., when one agent may contribute to or reduce the public good, while all other agents can only contribute. Examples for such asymmetries in action space reach from environmental problems to the problem of corruption or tax evasion where some agents have better access to privately beneficial options to the expense of the public.

Interestingly, we find that *transparency is a double-edged sword* – that is, it can be useful to improve welfare if combined with a sanction mechanism. Without this mechanism, however, transparency causes a reduction in the public good provision: special agents with the taking option extract more when actions are transparent. One reason may be that without transparency of actions they attempt to avoid negative reciprocal actions from standard agents by limiting their taking to an extent that cannot be identified. With high transparency (of identity and actions), taking is imminently identified such that these special agents may no longer feel restrained. While transparency alone does not significantly change the level of contributions by standard agents, it allows them to compare their behavior with agents of their own type. Our results indicate that a within-group effect is generated such that agents reciprocate primarily on the decisions of their own type. This within-group comparison essentially decouples the contribution of the special agents from those by the other agents and allows the special agent to increase taking.

A sanction mechanism in form of a peer-punishment stage significantly changes the effects of transparency: without transparency punishment loses its bite. Yet, when low transparency (only of identity, not of actions) is present special agents are stigmatized as potential extractors and therefore are more likely to receive punishment. It seems that stigmatization creates sufficient accountability for special agents' actions. As a consequence, they contribute more to the public good. Finally, high transparency eliminates this statistical discrimination. It allows targeted punishment such that agents of all types are punished based on their actual actions and not based on their type. High transparency thereby particularly increases punishment of standard agents and, consequentially, their contributions. As such, we find transparency and punishment to serve as complements in generating higher voluntary contributions to the public good. Hence, both transparency and punishment are necessary conditions to overcome the social dilemma of the provision of public goods.

Our results are important on several fronts. First, our results on the impact of the action space show the importance of studying agents' heterogeneities that go beyond varying endowments or benefits from public goods. Second, punishment without or with low transparency may lead to wrong punishment of some agents just based on perceived differences in actions available to them. As such, transparency and punishment may not only be complements in increasing the provision of public goods, but also be effective in reducing the negative effects of stigma. Third, and most important, the findings indicate the importance of calls by policy-makers, NGOs or consumer organizations to increase transparency of publicly relevant actions (e.g., by civil servants, or managers): high transparency is a necessary condition to make agents accountable for their actions. Movements like the open government campaign in Australia, Europe and the US serve the cooperativeness and well-being of the entire society (see for this the memorandum of the US president Obama, 2009).

It will be interesting to extend the scope of our results. For example, we exogenously imposed transparency and punishment conditions. It would be worthwhile studying how different individuals may voluntarily make their actions transparent or increase their accountability in order to provide signals of pro-social intentions. We leave investigations of such voluntary disclosure and their interaction with exogenously imposed transparency or sanction mechanisms for further research.

## 3.6 Appendix A: English Translation of the Experimental Instructions

These are the English translations of the German instructions for the ASYM_L_PUN. The instructions for other treatments are available in German from the authors upon request.

*General explanations for participants*

You are taking part in an economic experiment. You can earn a significant sum of money, depending on the decisions and the decisions of other participants. It is therefore very important that you pay attention to the following points.

The instructions you have received from us are intended solely for your private information. *During the experiment, you will not be allowed to communicate with anyone***.** Should you have any questions, please direct them directly to us. Not abiding by this rule will lead to exclusion from the experiment and from any payments.

All decisions in the experiment are made anonymously. Only the experimenter knows your identity, while we cannot match your decisions with your identity.

For your participation in this experiment, you receive an initial income of 6 Euros. FYour additional income depends on your decisions. In this experiment, we calculate in Taler, rather than in Euro. Your entire income will therefore initially be calculated in Taler. The total sum of taler will later be calculated in Euro as follows:

*1 Taler = 0,30 Euros*

You will be paid *in cash* at the end of the experiment.

The experiment is divided into separate periods. It consists of a total of 10 periods. In each period you will play the same game. Each period consists of two steps. Participants are randomly assigned to groups of four. Each group, thus, has three other members, apart from you. During these 10 periods, the constellation of your group of four will remain unaltered. *You will therefore be in the same group for 10 periods.* Please note that you and the other group members decide anonymously. That is, other group members cannot match your decision with your identity.

At the end of the experiment, we will randomly determine one out of the ten periods to be decisive for your payoff. Therefore, the outcome of every period determines potentially your payoff.

The following pages outline the exact procedure of the experiment. As mentioned earlier, each of the ten periods contains two steps.

*Exact procedure for step one*

At the beginning of each period, each participant faces the same decision problem. Your decision (as well as the decision of all other group members) is to divide Taler between a private account and a group account. At the beginning, each group member is allotted 12 Taler on your private account. Furthermore, there are 32 Taler on the group account.

At the beginning of the experiment, we will randomly determine one member of your group to be player “1”. The player will be the same throughout the entire 10 periods. Player 1 has to decide whether to transfer Taler from her private account to the group account **or** from the group account to the private account. Therefore, player 1’s transfers range between -32 and 12 (only integers); positive numbers imply transfers from the private to the group account, negative numbers imply transfers from the group account to the private account.

Each of the remaining three players (player “2”, “3”, and “4”) has to decide whether to transfer Taler from her private account to the group account. Transfers range between 0 and 12 (only integers).

Your income in step one consists of two parts, namely:

(1) the Taler you have kept or transferred on your private account,

(2) the *"income gained from the group account"*. Your income from the group account is calculated as follows:

Income from the group account = .4 * total sum of Taler on the group account

Thus your income in step one equals:

(12 – your transfers) + .4*(total sum of Taler on the group account)

The income gained from the group account is calculated using the same formula for each member of the group. That is, each group member (irrespectively whether it be “1” or any other group member) gains the same income from the group account.

If, for example, the sum of the transfers from all group members adds up to 28 Taler, that is, the group account holds 32 + 28 = 60 Taler, you and all other members each gains an income from the group account of .4x 60 = 24 Taler. If the sum of your and the transfers from all other group members adds up to 9 Taler, that is, the group account holds 32 + 9 = 41 Taler, you and all other members each gains an income from the group account of .4x 41 = 16.4 Taler. For each Taler you keep on your private account, you earn an income of 1 Taler.

*Exact procedure for step two*

After all group members have made their decisions in step one, you will proceed to step two.

Here, every group member receives an additional endowment of 5 Taler, which you may use to distribute points. You can decide in every row (in the blue box) on how many points you want to distribute; at most, you can distribute five points. Each point costs you one Taler. For each point, you distribute to another player, three Taler are deduced from her income. While you do not receive information on the specific transfers of players in step one, you will receive information on the sum of Taler on the group account. All Taler from the additional endowment that is not used for distributing points are added to your total income of this period.
An example: If you assign 2 points to another player in your group, you reduce your additional endowment from 5 to 3 Taler. If you do not assign any other points nor receive any points

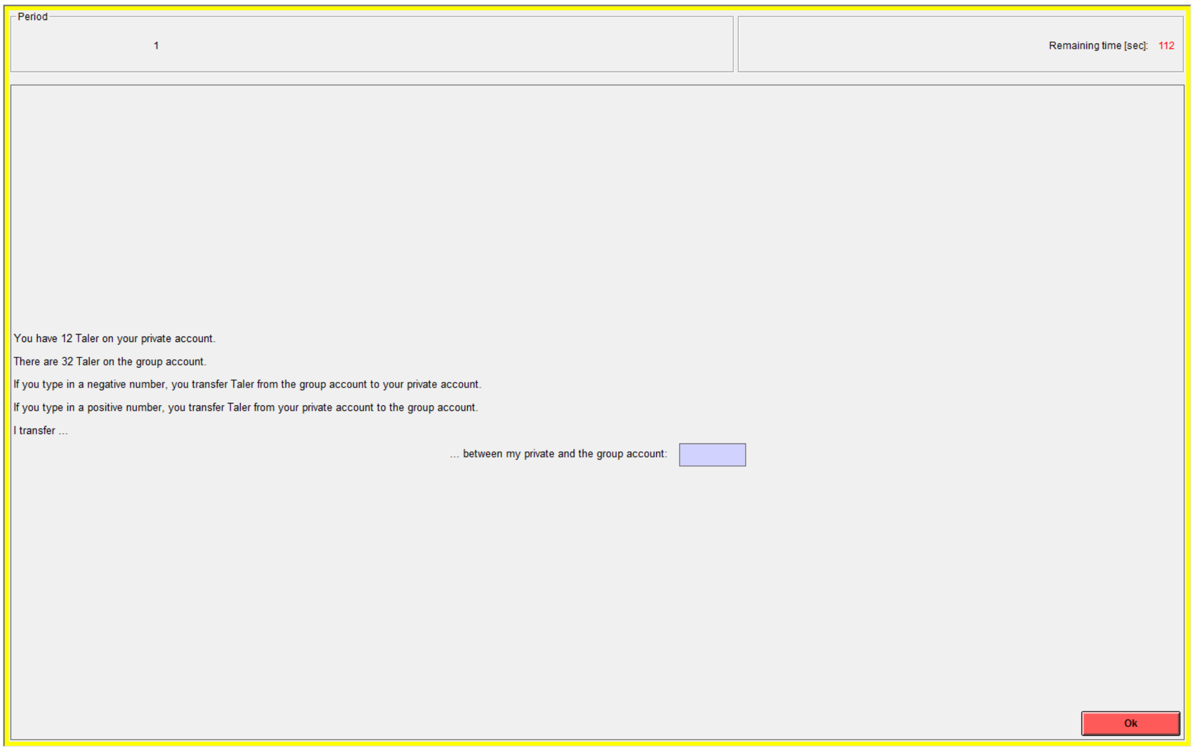
from other group members, your total income in this period increases by the remaining 3 Taler. The total income of the player who received 2 points from you decreases by 2*3 = 6 Taler.

Your total income from step one and two (in Taler) in each period equals
income in step one + 5 – distributed points – 3*received points

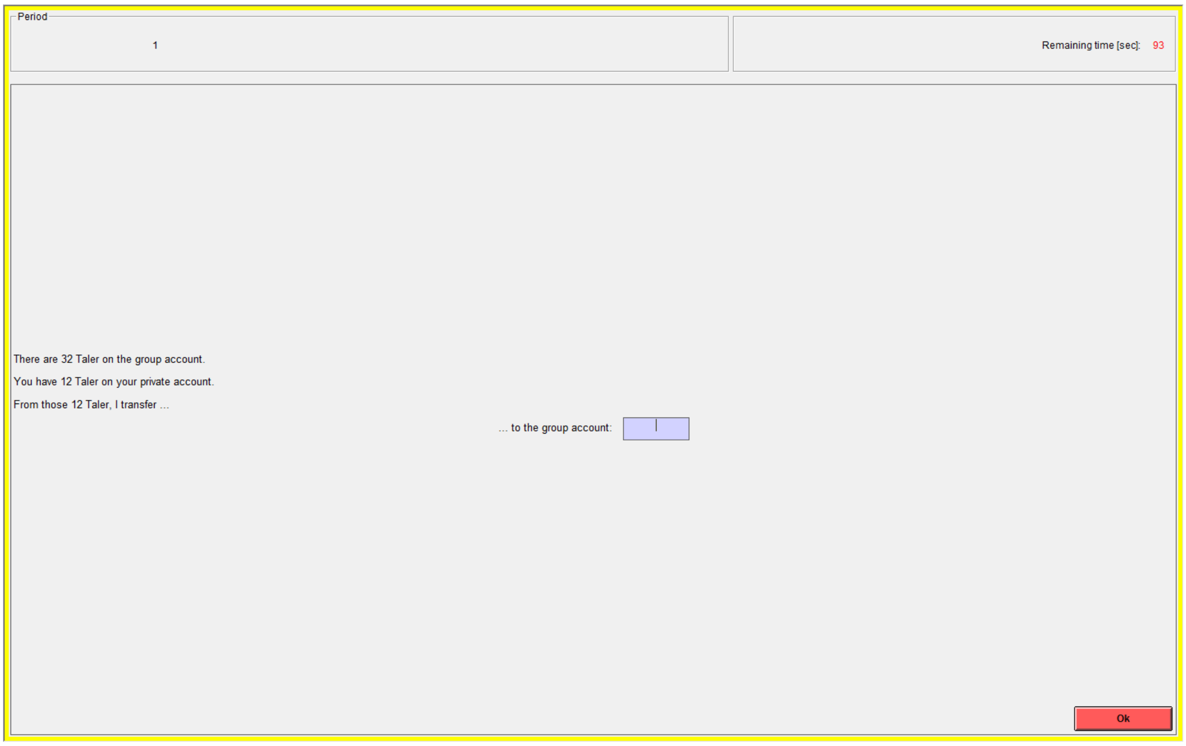
*Exact procedure for the course of the experiment*

You will see the input screen at the beginning of each period. In the left upper corner of the screen you will find the period number. In the right upper corner you will find the remaining time for your decision in seconds.

The input screen for player 1 in step one look like this:

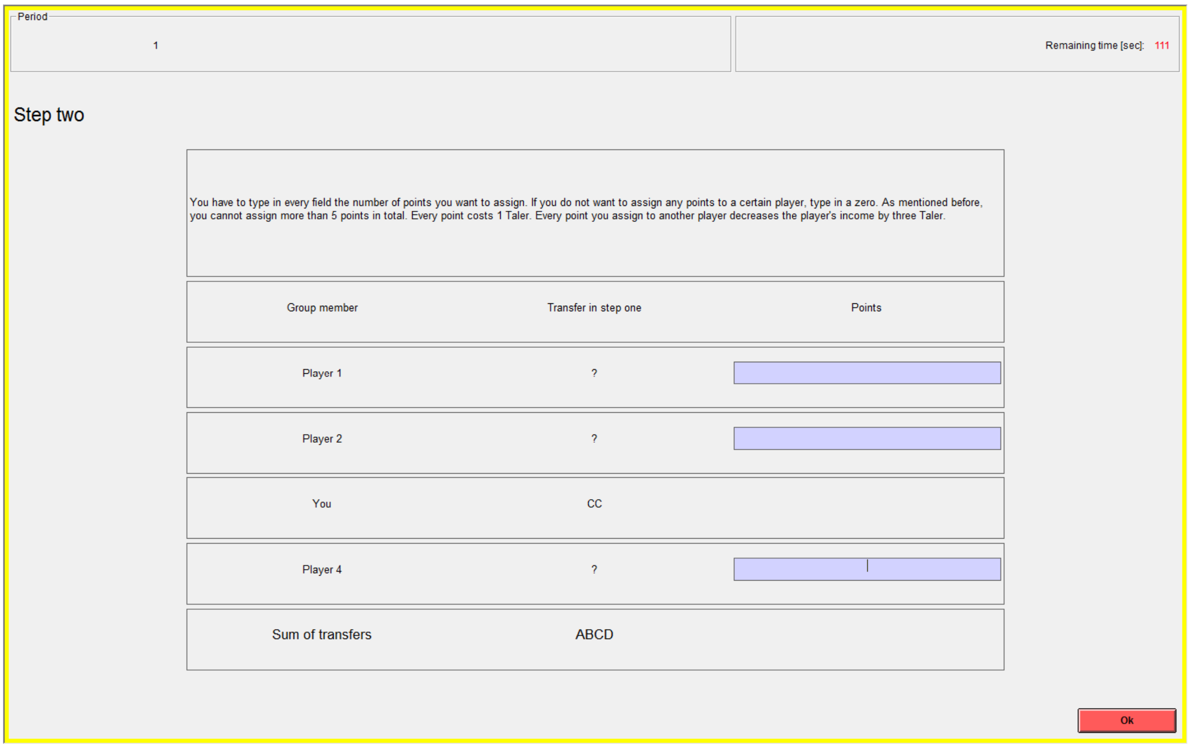


The input screen for players 2, 3, and 4 in step one looks like this:




There are 32 Taler on the group account in every period. You make a decision on your transfers on the group account by typing any one whole number between 0 and 12 (player 2, 3, and 4), and between -32 and 12 (player 1), respectively, into the appropriate field on your screen. This field can be accessed using the mouse. Once you have typed in your contribution, please click on *OK*, again using the mouse. Once you have done this, your decision for this period is irreversible.

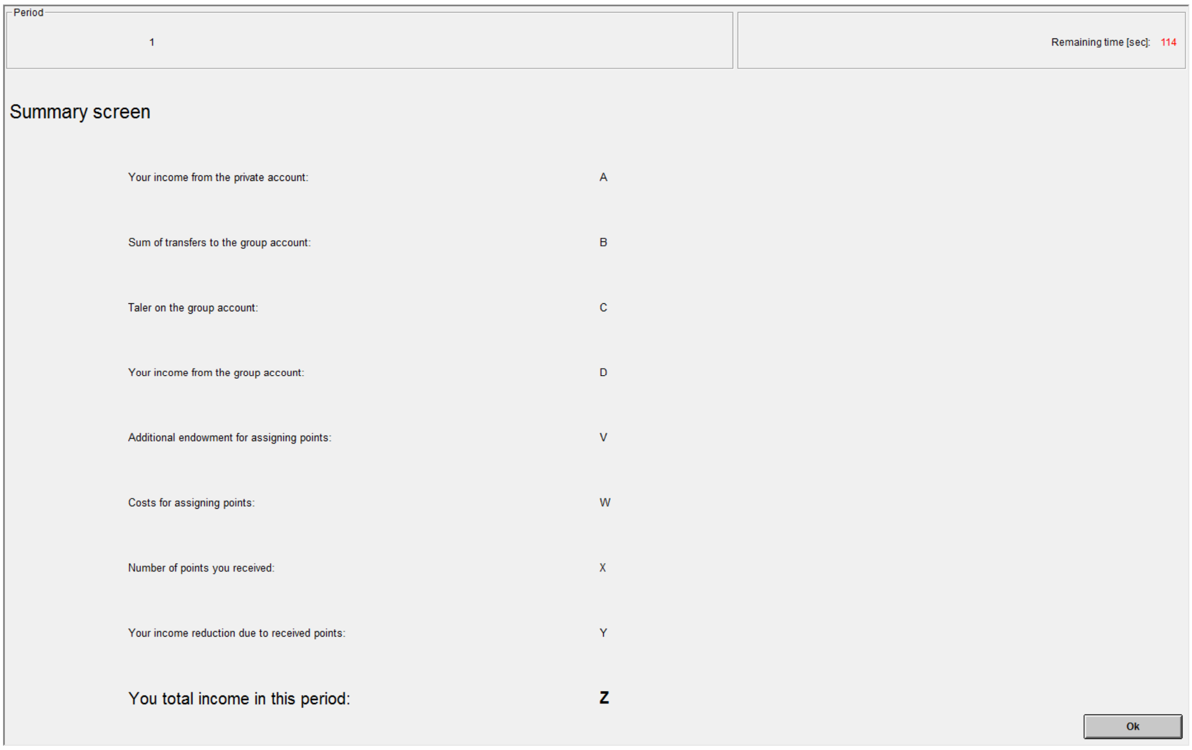
In step two, you will see the sum of all transfers in your group. The input screen on step two looks like this (here for player 3):




Recall: While you do not see the transfers of the other group members (there you will see "?"), you will see in row five the sum of all transfers in the group.

You may assign points you want to distribute. You can operate within the fields by using the mouse. If you do not wish to alter a certain group member's income, please enter 0. As mentioned earlier, the sum of points distributed cannot exceed 5.

Once all members of the group have made their decisions and have distributed their points, you will be informed about the sum of transfers on the group account, your income from the private account, the number of points you received, as well as the resulting reduction in income. Finally, you will see your period income.




As mentioned earlier, your income (in Taler) in step one is:

**(12 – your transfers) + .4*(total sum of Taler on the group account)**

And your **total income** equals

**income in step one + 5 – distributed points – 3*received points**

Before we proceed with the experiment, all participants have to answer some control questions on the computer screen. The control questions will help you to understand the rules of the game.

Do you have any further questions?

# 4 Setting the Bar – An Experimental Investigation of Immigration Requirements[32]

*"I'm very inspired by him – it was my father who taught us that an immigrant must work twice as hard as anybody else, that he must never give up."*
Zinedine Zidane

## 4.1 Introduction

Several European Union (EU) countries have shifted towards more restrictive immigration policies in response to changing economic conditions and increasing public sensitivity on migration issues since 2010 (OECD 2012a). Even before the recent economic downturn in many EU countries, several governments set more restrictive requirements for immigration, settlement and citizenship (Brubaker 2001, Joppke 2004). Anti-immigration sentiment is often seen as a response to increased immigration and ethnic diversity levels in recent decades, especially in Western European countries (OECD 2012b).

Examples of a restrictive trend in immigration policy include income or employment requirements (economic) as well as language and civic knowledge requirements (sociocultural) for immigration, settlement and naturalization. Immigrants have to prove that they own a certain level of economic resources[33] in order to legally immigrate to most EU countries for work purposes, to reunite with family members, to be permitted permanent residence

[32] This chapter is co-authored by Jasper D. Tjaden.

[33] Mostly measured by one or a combination of attributes like income levels, employment record, tax records, social benefit uptake, employment offers.

(settlement) or to be eligible for naturalization (see Goodman 2010 for an overview, Reichel 2013).

In the case of labor migration, the EU blue card – an attempt to attract highly skilled workers from third countries by harmonizing entry and residence conditions throughout the EU (Council Directive 2009/50/EC) – sets relative income requirements. The Directive lays out general economic requirements for migrant worker as 'a work contract or binding job offer with a salary of at least *one and a half times* the average gross annual salary paid in the Member State concerned'.[34]

In addition to income requirements, the restrictive trend in immigration policy in the EU extends to sociocultural immigration requirements: In the last decade, the EU has also seen a trend to formalize language requirements and introduce civic knowledge tests for settlement and citizenship. Some argue that these changes were introduced to reduce immigration levels rather than promoting social integration in host societies (van Oers et al. 2010, Joppke 2004, 2007, Groenendijk 2011, Carrera et al. 2009, Bauböck and Joppke 2010). The introduction or tightening of these socioeconomic immigration requirements has sparked debate about the appropriateness of such measures (Bauböck and Joppke 2010). Certain requirements have been perceived as unfair because they exceed host society averages (e.g. annual mean income and common civic knowledge).

Due to a continuous demographic decline, many EU countries are faced with an increasing need for labor migration.[35] Given the 'competition for global talent', immigration requirements do not only pose normative questions of fairness ('Are certain levels justified or appropriate?'), they also pose the question of efficiency ('Will requirements deter migration and thereby reduce welfare?'). Research has so far contributed little to either question. This is surprising in the light of salient media and political debate on the issue. Policy makers are faced with the 'uncomfortable' challenge of reconciling increasing labor demand with public opinion that does not generally favor immigration.

---

34 Member states may lower the salary threshold to a factor of 1.2 for certain professions where there is a particular need for third-country workers. Instead of relative requirements, some member states have introduced absolute thresholds. For example, regarding family migration eight EU member states have adopted rules fixing the amount of resources required from the sponsor in order to be eligible for family reunification (Pascouau et al 2011). Only recently has the UK substantially increased income requirements for family reunions (Home Office 2012).

35 In the case of Germany, a McKinsey study estimates a labor shortage of two million workers in 2020. The Prognos Institute estimates a labor force gap of 5.2 million workers in 2030 (see Bundesagentur für Arbeit 2011, Kolodziej 2012).

In this context, our research aims to contribute to the theoretical discussion on immigration policy by analyzing two general questions: (1) What are the drivers behind different immigration policy settings (i.e. why are immigration requirements higher in some countries than in others)?, and (2) Does the level of (economic) immigration requirements have an impact on immigration patterns and welfare (via public goods contributions)?

Possible reasons for the scarcity of comparative academic work on the issue of immigration requirements are the complexity of different policies and the constant change of policy which makes most empirical research on this issue impossible to generalize. We circumvent these issues by using experimental economic methods. To our knowledge, this has not been done in migration policy studies. By employing experimental economics methods, we are able to provide insights into the fundamental group decision making processes that underlie immigration requirements; as such our approach is novel and supplements evidence provided by sociological or political studies of immigration requirements which are specific with regard to geographical, political and cultural contexts (e.g. Massey and Espinosa 1997, Scheve and Slaughter 2001, Mayda 2006).

Our design enables us to identify which migration scenarios and societal factors determine more restrictive or more liberal immigration requirements. In order to create social groups of subjects in the laboratory, we randomly selected our subjects into two societal groups, that is citizens (red players) and migrants (blue players). The seminal works by Tajfel et al. (1971) and Billig and Tajfel (1973) show that such a random assignment of roles is sufficient to create feelings of in-group affiliation and in-group favoritism. We let the citizens play a giving-and-taking public good game (see chapter 2), but restricted migrants' opportunities to contribute to or receive returns from a public good.[36] Periodically, citizens set an immigration requirement in the form of a minimum contribution requirement for migrants. Our treatments mimic different migration streams in which migrants either *must enter* the group and contribute accordingly or hold the liberty to *decide to* immigrate or not.

To judge the fairness and efficiency of immigration requirements in our experiment, the citizens' population can be used as a comparison reference group. Income requirements for labor migration and family migration are often related to average income levels of the host population, minimum wage levels or the eligibility level for social assistance. What seems legitimate to expect from newcomers is often drawn from what is expected from the host society itself. As such, requirements for immigrants' contributions to public goods such as

36 See Ledyard (1995) and Chaudhuri (2011) for surveys on public good games.

paying a certain minimum amount of taxes can be seen as relative to citizens' contributions to public goods.

From a public choice perspective, it is informative to analyze the effect of different in-group decision procedures on immigration policy. Different treatments enable us to examine the impact of debate among in-group members and different migration scenarios on individual and policy outcomes. Voting on an issue, which has not been discussed, may result in the establishment of a different policy compared to voting on a policy that is preceded by a debate of in-group members. Furthermore, it is not clear whether voting for a policy that sets a threshold for immigrant contributions establishes contribution norms for the in-group, as well. Such social norms may help to overcome the social dilemma associated with public good provision by private actors. Conversely, in-group members may decrease their contributions or even exploit the public good while out-group members are bound to contribute. The establishment of a 'bar' (required contribution level) may deter potential immigrants although they forfeit payoff gains.

Our results suggest that immigration policy-makers would be well-advised to design economic requirements for labor migration that are in appropriate relation to average performance of the population and based on public debate. In our experiment, greater *migration policy coherence* – which we define as the relationship between requirements and population averages and which is based on active public debate – leads to greater perceived fairness of the policy, greater individual contributions to the public good by both citizens and migrants, and the greatest overall welfare. Moreover, our results point to the importance of debate on immigration. In a labor demand context, negative frames of immigrants' potential in the debate can result in more restrictive immigration requirements regardless of the immigrants' actual potential to contribute to the public good.

Only a limited number of experimental economic studies on public good provision by private actors include processes of endogenous group formation so far. We review papers with endogenous group formation based on other subjects' characteristics and actions more closely.[37] The literature refers to Ehrhart and Keser (1999) as the first experimental study to allow for endogenous re-grouping. They allow for free (but costly) regrouping in each period. In a stage before the public good game itself, subjects received information on public good

37 Note that there is a strand of literature that examines endogenous group formation in public good games with self-selection into groups with pre-set institutions (e.g. Brekke et al. 2011, Gürerk et al. 2006, Gürerk et al. 2011).

contributions in all existing groups. They were then able to migrate to another group or form a new one. Erhart and Keser (1999) find that subjects who contributed high amounts to the public good were 'chased' by low contributors. Coricelli et al. (2004) present a similar pattern. Instead of costly allowed migration, they let subjects bid for the right to choose group members. Subjects learned about the contribution history of potential partners. Again, low contributors tried to chase high contributors by submitting high bids.

Cinyabuguma et al. (2005) employ an expulsion mechanism. After each period, subjects received information on the contribution levels of all fellow group members. They could then vote to expel group members. If a majority of group members voted to expel another group member, this subject became part of an outsider-group. This outsider-group might again invest in its own public good; however, outsider-group members received a lower initial endowment each period. Cinyabuguma et al. (2005) find that the presence of their expulsion mechanism fostered cooperation compared to a standard linear public good game.

Page et al. (2005) allow for endogenous group formation by letting subjects rank their preferred group members in a set interval. An algorithm then regrouped the subjects by optimizing matching preferences. They find that such a regrouping mechanism was able to greatly enhance contributions to the public good. A combination of their regrouping mechanism and a punishment mechanism with a 1:4 fee-to-fine ratio increased contributions further. Efficiency by the regrouping mechanism was greater than by the punishment mechanism (due to the costs of punishment). Gunnthorsdottir et al. (2010) also use a matching procedure based on the amount contributed to the public good in the previous period instead of preferred group members.

Closer to our research question, Ahn et al. (2008) investigate endogenous group formation with entry and exit mechanisms: both entry and exit were free or one of them could be permitted by the group members with a majority voting rule while the other was free.[38] Voting was based on individual subjects who might enter the group, given their contribution history in their present group. In this setup, all groups of size 1 to *n* might contribute to and consume a public good. They find that restricted entry increases contributions to the public good.

Our approach differs from Ahn et al. (2008) along several dimensions. First and most importantly, we use predefined groups of insiders and outsiders, what we refer to as citizens

38 Note also that a companion paper Ahn et al. (2009) investigates endogenous group formation when the public good is congestible.

and migrants, to mirror the setting of individuals born in different countries. Second, citizens do not select certain migrants, but they set a policy that applies for all migrants. While Ahn et al.'s setup is highly informative to identify avenues of high public good provision, our approach is more in line with countries' immigration policies, which cannot arbitrarily select certain individuals. Third, migrants differ with respect to their initial private endowment, but there is no contribution history to inform citizens about the expected contribution of a migrant. This again mirrors reality, as citizens in countries cannot ex ante identify pro-social migrants. Fourth, in our design migrants may hold bargaining power and reject the immigration requirement set by citizens. Fifth, our design includes debate on the requirement which is not available in Ahn et al (2008). Based on all these factors, we consider our design novel and informative for the literature on endogenous user groups in public good games.

The remainder of this chapter is structured as follows; section two lays out the experimental design, including predictions and information on experimental procedures. The results are presented in section three. Section four discusses the implications of our results for policy and section five presents our conclusions.

## 4.2 Experimental Design

In this section, we will first introduce the two dimensions of our 2x2 experimental design. Next, we will formalize our design and develop predictions that explain how behavior may change depending on the existence of other-regarding social preferences. In the last part of this section, we will describe the procedures of the experiment.

### *4.2.1 Two Dimensions of Immigration Policy*

We designed our experiment to resemble a Western welfare-state setting. Our baseline scenario employs a non-satiated public good. As mentioned above, most Western welfare states are characterized by an ageing population and an associated increased labor demand. We apply the generalized giving-and-taking framework to the public good game. This framework was first introduced in chapter 2 of this dissertation. The giving-and-taking framework represents the fundamental distribution mechanism common in Western welfare states. 'Giving' to the public good equals paying taxes and 'taking' from the public good equals receiving social transfers.

For our study, we randomly select subjects to be in two sub-groups: 'citizens' and 'migrants'.[39] We use a partner matching that is consistent with the analogy of citizenship and is useful for our analysis of behavior over time. Citizens are always 'in the country' and are able to enjoy the consumption of a public good (with an initial public good endowment). They need to decide how much to give to the public good (analogy: donate, contribute) or take (analogy: receive a social transfer). Initially, all migrants are 'outside of the country' and thus do not profit from the public good. To keep our design simple there is no second public good outside the country that migrants may profit from. One reason may be a lack of good governance. The fact that migrants have no public good introduces an (economic) hierarchy between the two groups and this defines a direction of the migration flow. As potential payoffs are higher for citizens, migrants have an incentive to migrate.

Different initial endowments introduce within-group socioeconomic stratification of both citizens and migrants. We added this to our calibration in order to analyze how the initial endowment or socioeconomic status determines contributions to the public good and the setting of the immigration policy. One can also think of this endowment as a proxy for achievement (productivity) potential or human capital.[40]

Our 2x2 experimental design varies the freedom of migrant choice and the opportunity for debate among citizens about the immigration requirement. An overview of our design is provided in Table 1. The first dimension, the freedom of migrant choice, may have two different conditions: The two dictator treatments (Dict_) do not give migrants the choice to stay outside the country if their endowments are sufficient to permit entry (i.e. if their endowments are greater than the immigration requirement). Conversely, the ultimatum treatments (Ulti_) provide the opportunity for migrants to accept or reject the immigration requirement set by citizens, even if they could meet the requirement.

---

39 Note that the vocabulary we use in this chapter (e.g. 'citizens', 'migrants', 'country', 'giving', 'taking', etc.) does not match the language of the instructions and programs of the experiment. For example, we called in-group players 'red players' and out-group players 'blue players'. For the instructions, see appendix B.

40 Note that there is a strand of literature on heterogeneous endowments in public good games, including Chan et al. (1999), van Dijk et al. (2002), Cherry et al. (2005), Buckley and Croson (2006), Sadrieh and Verbon (2006).

**Table 1.** 2x2 Experimental Design.

| | | Migrant choice / Perceived migration | |
|---|---|---|---|
| | | Ultimatum treatment / High perceived migrant potential | Dictator treatment / Low perceived migrant potential |
| Ex ante Debate? *(via a chat of red in-group members)* | yes | Ulti_chat | Dict_chat |
| | no | Ulti_NOchat | Dict_NOchat |

We designed the dictator vs. ultimatum treatment manipulation to mirror different migration flows in the real world. In the ultimatum treatment, migrants have more bargaining power as they can refuse immigration despite eligibility. Higher bargaining power may originate from a greater choice between different destination countries. Thus, ultimatum treatments indicate citizens' *perception* of high-potential migration. Conversely, in the dictator treatments, migrants have less bargaining power which implies fewer choices. This indicates citizens' *perception* of low-potential migration. The difference in migration scenarios can be understood as a signaling mechanism of the kind of migration flow.[41] However, the signal itself does not provide information about actual contribution (productivity) differentials between migrants in the ultimatum (perceived high-potential) and dictator (perceived low-potential) treatments. Both groups hold similar contribution potential.

If one were to look for real life examples of both categories, we can broadly distinguish countries that have traditionally attracted low skilled migration (Western Europe without the UK)[42] and countries that have traditionally attracted high skilled migration (USA, Canada, UK) (OECD 2012b).

The second dimension varies the availability of a free-form text debate among citizens via a chat screen. In the chat treatments (_chat) citizens (but not migrants) are able to debate freely about the game and the height of the bar before voting for it in private. This treatment equals

41 Note that our design does not include a competition among countries for migrants. Such an extension would have complicated our design considerably. In this chapter we are primarily interested in the consequences of migrants' bargaining power on citizens' decision on the magnitude of the 'bar', their contributions to the public good and associated the fairness. Like any other directions, we however regard a game of competition among countries as a fruitful avenue for future research.

42 See guest worker programs and higher share of family and humanitarian migration in continental Europe.

public debate in society. Citizens can debate the advantages and disadvantages of liberal vs. restrictive immigration requirements. Analogous to the real world, citizens can exchange viewpoints, present evidence and argue in favor or against a certain policy. The public debate treatment allows us to analyze the effect that it has on citizens' contributions to the public good and on the level of the immigration requirements (will debate lower the requirement?). Most importantly, we will analyze how debate interacts with different perceptions of potential migration flows. While debate may lead to a more restrictive policy in one case, it could lead to a more liberal one in another. This feature is important as it may reveal motivational channels of subjects playing as citizens in our experiment.

### *4.2.2 Formalization*

In our experiment, we match three citizens and three migrants (i.e. $n = 6$) in a group. By definition, citizens are beneficiaries of the public good of the 'country', while migrants initially remain outside and may *decide* to immigrate into the country or not (in the Ulti_ treatments) or *have to* immigrate into the country given that they are able to fulfill the minimum contribution requirements (in the Dict_ treatments). We defined the payoff of an individual $i$ 'residing within the country' as

$$\pi_i = w_i - c_i + h\left(E + \sum_{j=1}^{n} c_j\right)$$

with the private endowment $w_i$, initial public good condition $E$, the marginal per capita return from the public good $h < 1 < hn$ and, in principle, private contribution $c_i \in [-\frac{E}{n}, w_i]$. For a migrant $j$ residing outside the country, the payoff is $\pi_j = w_j$. Note that in our calibration, we set $h = 0.5$, $E = 60$ Taler so that $\frac{E}{n} = 10$ Taler; 'Taler' is the artificial currency in our experiment. The initial private endowments $w_j$ either amount to $w^{low} = 5\,Taler$, $w^{mid} = 10\,Taler$ or $w^{high} = 15\,Taler$ so that for every endowment level, there is exactly one citizen and one migrant. Table 2 illustrates the setup for a given partner-group. We include heterogeneous endowments in our design to learn more about the motivation of different types of citizens. For instance, low-endowed citizens may vote for lower or higher bars compared to high-endowed citizens. They may show distinct sympathy with low-endowed migrants and aim to set a low bar. Conversely, low-endowed citizens may vote for substantially higher bars in order not to lose their relative position in the income rank.

The decision stages in our experiment are: (1) citizens set a requirement for migrants (all treatments), (2) if eligible, migrants individually decide to accept or reject the requirement (only Ulti_ treatments), and (3) citizens and migrants simultaneously decide how much to contribute to the public good (all treatments).[43]

**Table 2.** Grouping.

| Random grouping of members and random allocation of initial endowment | | Citizens (instructions: "red players") | Migrants (instructions: "blue players") |
|---|---|---|---|
| Initial Endowment | 5 | 1 Player | 1 Player |
| | 10 | 1 Player | 1 Player |
| | 15 | 1 Player | 1 Player |

Hence, the decision problem of migrant $j$ includes decision stages two and three in the Ulti_ treatments and stage 3 in the Dict_ treatments. In the Ulti_ treatments, in stage two, $j$ needs to decide whether to accept the requirement and immigrate (i.e. $e_j = 1$) or not ($e_j = 0$). The third stage is the decision on the private contribution to the public good $c_j$. Conversely, citizen $i$ always faces decision stages 1 and 3: she needs to vote on the immigration requirement $r$ and decide on her contribution to the public good $c_i$. In the following, we formulate predictions regarding the requirement $r \in [-\frac{E}{n}, w^{high}]$, i.e. in our calibration between -10 and 15 Taler, set by citizens for migrants. Hence, the action set of a citizen $i$ always reads $c_i \in [-\frac{E}{n}, w_i]$. The action set of a migrant $j$ reads $c_j \in \left[e_j r, e_j w_j\right]$, with $e_j \in \{0,1\}$. In the Dict_ treatments $e_j = 1$ if $r \leq w_j$, else $e_j = 0$. In the Ulti_ treatments $e_j = 1$ if $r \leq w_j$ and the migrant *accepts* the requirement, else $e_j = 0$.

### *4.2.3 Predictions for Payoff Maximization*

By solving $\max_{c_i} \pi_i = w_i - c_i + h\left(E + \sum_{j=1}^{n} c_j\right)$ we get the standard solution for the linear public good game $\frac{\partial \pi_i}{\partial c_i} = -1 + h < 0$, which translates to the prediction that citizens appropriate as much as possible, and that migrants contribute the minimal amount or appropriate as much as they can. We summarize

[43] In order to summarize our design, Figures A.1a and A.1b depict the two or three decision stages in Dict_ and Ulti_ treatments respectively in a simplified 2-player case.

**Prediction 1a.** *Citizens will appropriate the maximal amount* $c_i = -\frac{E}{n}$, *and migrants contribute the minimum requirement* $c_j = r$.

With this prediction, we can turn to stages one and two of the game including the setting of the bar by citizens and acceptance or rejection of $r$ by migrant in Ulti_ or direct inclusion or exclusion of migrants in Dict_. Let us first predict $r$ for the Dict_ cases. Note again that there are three migrants with endowments $w^{low} = 5\,Taler$, $w^{mid} = 10\,Taler$ or $w^{high} = 15\,Taler$ such that for every endowment level there is exactly one migrant. Payoff maximizing citizens aim at maximizing migrants' contributions to the public good by choosing the optimal $r$; however as the requirement increases migrants are excluded automatically, starting with the migrant endowed with $w^{low} = 5\,Taler$. More formally, the sum of contributions by migrants $C_{migrants}$ is

$$C_{migrants} = \begin{cases} 3r & if \quad r \leq 5 \\ 2r & if \quad 5 < r \leq 10 \\ r & if \quad r > 10 \end{cases}$$

and $C_{migrants} = 20$ is maximal with $r = 10$. Note that we chose the parameters to generate this interior solution. Payoff-maximizing citizens therefore set the optimal requirement $r^*_{Dict} = 10$ in the Dict_ treatments. The calibration of our design allows for a second straightforward prediction: payoff-maximizing migrants accept *any* requirement.[44] Hence, based on payoff-maximizing citizens and migrants, we formulate

**Prediction 2a.** *Citizens will vote for requirements of* $r^*_{Dict} = r^*_{Ulti} = 10$. *Migrants will either immigrate automatically in Dict_ or accept this requirement voluntarily in Ulti_.*

Standard game theory is also straightforward when it comes to the chat opportunity among citizens. That is, it regards promises and non-binding contracts as cheap talk. Consequently, there should be no difference between citizens' decision making after a debate and citizens' decision making without a debate:

---

[44] To clarify this point, consider the extreme case of a migrant $j$ with $w_j = 15$ and $r = 15$. Further, assume that all three citizens behave purely selfish and contribute $-3\frac{E}{n}$ (i.e. -30 Taler) in sum. Migrant $j$ will still receive an income of $15 - 15 + 0.5(60 - 30 + 15) = 22.5$ Taler which is greater than staying outside and receiving 15 Taler. Likewise, a migrant $k$ with $w_k = 5$ who accepts a requirement of $r = 5$ will receive an income of $5 - 5 + 0.5(60 - 30 + 5) = 17.5$ Taler 'inside the country' compared to only 5 Taler 'outside'.

**Prediction 3a.** *Debate is cheap talk. There are no differences in decision making between _chat and _NOchat treatments.*

### *4.2.4 Behavioral Economic Predictions*

Let us now discuss alternative predictions inspired by insights from behavioral economic research. Surveys on the standard public good game by Ledyard (1995) and Chaudhuri (2011) as well as results on the giving-and-taking public good game introduced in chapter 2 show that individuals frequently and voluntarily contribute to public goods. As discussed earlier in chapter 1, other-regarding, social preferences such as altruism, warm-glow from giving and conditional cooperation may motivate such behavior (Meier 2007). Therefore, the first behavioral economic prediction reads

**Prediction 1b.** *On average, citizens do not appropriate the maximal amount* $c_i > -\frac{E}{n}$, *and migrants contribute more than the minimum requirement* $c_j > r$.

Further, let us consider stage two in the Ulti_ treatments (i.e. the decisions of migrants to accept or reject the requirement $r$). Compared to migrants, citizens are privileged based on two characteristics: (1) the privilege to reside within the country at all times and (2) the privilege to contribute to and appropriate the public good freely. Let us assume that migrants value their social status, represented by their action set, and that they care about intentions of others. The idea is closely related to models of reciprocity (e.g. Rabin 1993, Dufwenberg and Kirchsteiger 2004, Falk and Fischbacher 2006). Citizens may then anticipate and incorporate migrants' preferences for reciprocity when setting the bar. The intuition is that if migrants get too disadvantaged, they will lose utility due to what they perceive as an unfair policy. Consequently, migrants will respond by rejecting a sufficiently unfair policy. Hence, in Ulti_ citizens may need to set the requirement *below* the optimal requirement $r^*_{Dict} = 10$ in order to get migrants to accept it (and thereby secure some payoff from migrants' contributions). On the contrary, migrants cannot reject an 'unfair' requirement in Dict_ and citizens can disregard migrants' preferences for reciprocity. As a result the requirement in Ulti_ may be lower than the requirement in Dict_:

**Prediction 2b.** *If migrants are sufficiently reciprocating, then citizens set* $r^*_{Ulti} < r^*_{Dict}$.

Finally, let us turn to the chat opportunity. Indeed the behavioral economic literature suggests that a debate among citizens in the _chat treatments potentially changes public good contributions and votes on the requirement. As suggested by the findings of Ostrom et al.

(1992) and Brosig et al. (2003), the chat opportunity may offer a mechanism for citizens to coordinate the social dilemma situation of the public good game. We predict:

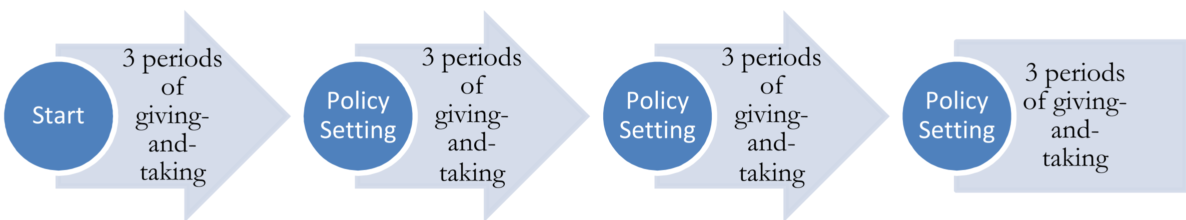
**Prediction 3b.** *Debate is a coordination mechanism. Public good provision is greater in _chat than in _NOchat treatments.*

With regard to the requirement, it is even harder to formulate predictions without a set of additional assumptions. Debate might change the requirements based on some group norm; if such a group norm evolves, it may be a norm of fairness or equality which could develop in a lower requirement or a norm of in-group exclusiveness which could develop in a higher requirement. These are just two examples of how debate may play a role. While we regard the possibility of such norms to evolve as a valuable feature that we will investigate with our design, we refrain from extended speculations on this matter.

### *4.2.5 Course of Events & Procedures*

Figure 1 provides an overview of the timeline of our experiment. After three periods of public good game interaction (of citizens only), citizens were able to implement a migration policy (i.e. to 'set a bar'). After a bar had been set, all players within the country (i.e. citizens potentially joined by (some) migrants) again played the public good game for three periods. The process was then repeated after period 6 and a new policy was implemented. The game was played for another three periods. After period 9, citizens had one last opportunity to adapt their policy. The median requirement of the three citizens (majority rule) was implemented as policy result after every 'policy setting' procedure. We incorporated repeated voting in our design to be able to identify adjustment (and possibly convergence) of the minimum contribution requirement *r* over time. This rule resembles real circumstances where certain polices are usually debated at separate recurring instances (elections etc.).

**Figure 1.** Timeline of our Experiment.



All nine sessions were conducted in the computer laboratory of the University of Hamburg in May 2012. Each session lasted approximately one hour. We used z-Tree (Fischbacher 2007) for programming and ORSEE (Greiner 2004) for recruitment. In the four treatments depicted in Table 2, 240 subjects participated in groups of six, with five observations per session. In addition, we collected data on baseline contributions in the 'standard' treatment in one session (i.e. when three citizens played a public good game and there were no migrants, no chats and no setting of any bars). This session yielded ten independent observations. Hence, we analyze the behavior of a total of 270 subjects in five treatments. All subjects were students with different academic backgrounds and no subject participated in the experiment more than once.

Once the participants were seated, a set of instructions was distributed and read out loud by the experimenter. In order to ensure that subjects understood the respective game, experimental instructions included several numerical examples and participants had to answer control questions via their computer terminals. After all periods were played, one out of the twelve periods was randomly selected for payment. Average payment over all treatments was 12.43 EUR.

## 4.3 Results

We will analyze the results of our experiment in four steps. First, we will examine treatment effects with respect to the public good provision. Second, we will investigate the drivers of the treatment effects. Third, we will analyze differences in the scales of the migration requirement and migration decisions. Fourth, we will have a closer look at the arguments in the chat debates of in-group members and ex-post questionnaire answers concerning the setting of the bar.

Considering all six players of a group and all periods, average contributions were the highest in Ulti_chat with an average of 3.589 Taler. This public good provision is higher than in the standard (average of -3.958 Taler, difference significant at $p = 0.0005$, Mann Whitney (MW) test), Ulti_NOchat (0.907 Taler, $p = 0.0588$, MW test) and Dict_NOchat (0.496 Taler, $p = 0.0343$, MW test) treatments. There is no significant difference between Ulti_chat and the Dict_chat treatment in which average public good provision amounts to 2.121 Taler. Figure 2 depicts public good provision averaged over all 12 periods and for all player types and Figure

3 depicts time trends. Table 3 provides descriptive statistics.[45] In accordance with earlier findings on social dilemmas, e.g. by Ostrom et al. (1992) on common pool resources and Brosig et al. (2003) on the public good game, we find that communication helps to foster and coordinate contributions. We report:

**Result 1.** *Considering all periods and player types, contributions to the public good were greater in _chat than in _NOchat treatments. Our data supports Prediction 3b.*

**Table 3.** Descriptive Statistics.

| | Treatment | | | | |
|---|---|---|---|---|---|
| | Standard | Ulti_chat | Ulti_NOchat | Dict_chat | Dict_NOchat |
| Individuals | 30 | 60 | 60 | 60 | 60 |
| Groups | 10 | 10 | 10 | 10 | 10 |
| Mean contribution (all players, *in Taler*) | -3.958 | 3.589 | 0.496 | 2.121 | 0.907 |
| Mean contribution (citizens, *in Taler*) | | 2.536 | -3.328 | -0.242 | -2.575 |
| Mean *voluntary* contribution of migrants (*in Taler*) | | 2.757 | 1.176 | 1.662 | 2.215 |
| Mean migration requirement (*in Taler*) | | 7.030 | 5.720 | 8.410 | 5.967 |
| Percent of migrants who accepted a requirement (in %) | | 95.7 | 100.0 | | |
| Mean *accepted* requirement (*in Taler*) | | 6.250 | 5.720 | | |
| Mean *rejected* requirement (*in Taler*) | | 8.667 | | | |
| Mean income of all player types (*in Taler*) | 38.021 | 40.155 | 35.717 | 36.539 | 35.859 |
| Mean income of citizens (*in Taler*) | | 48.231 | 44.815 | 46.604 | 45.296 |
| Mean income of migrants (*in Taler*) | | 32.079 | 26.619 | 26.475 | 26.424 |

[45] Table A.1 in appendix A provides an overview of pairwise Mann-Whitney tests with one observation per group averaged over all 12 periods (i.e. ten independent observations per treatment).

**Figure 2.** Public Good Provision, Citizens and Migrants, Average over all Twelve Periods.

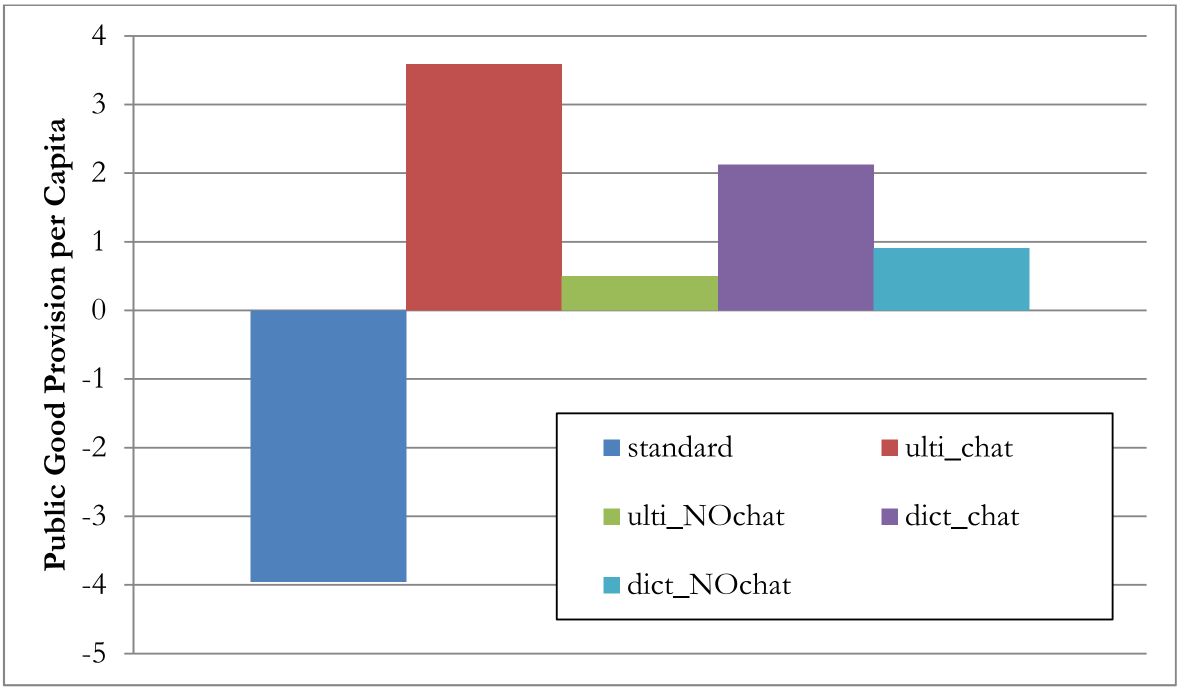
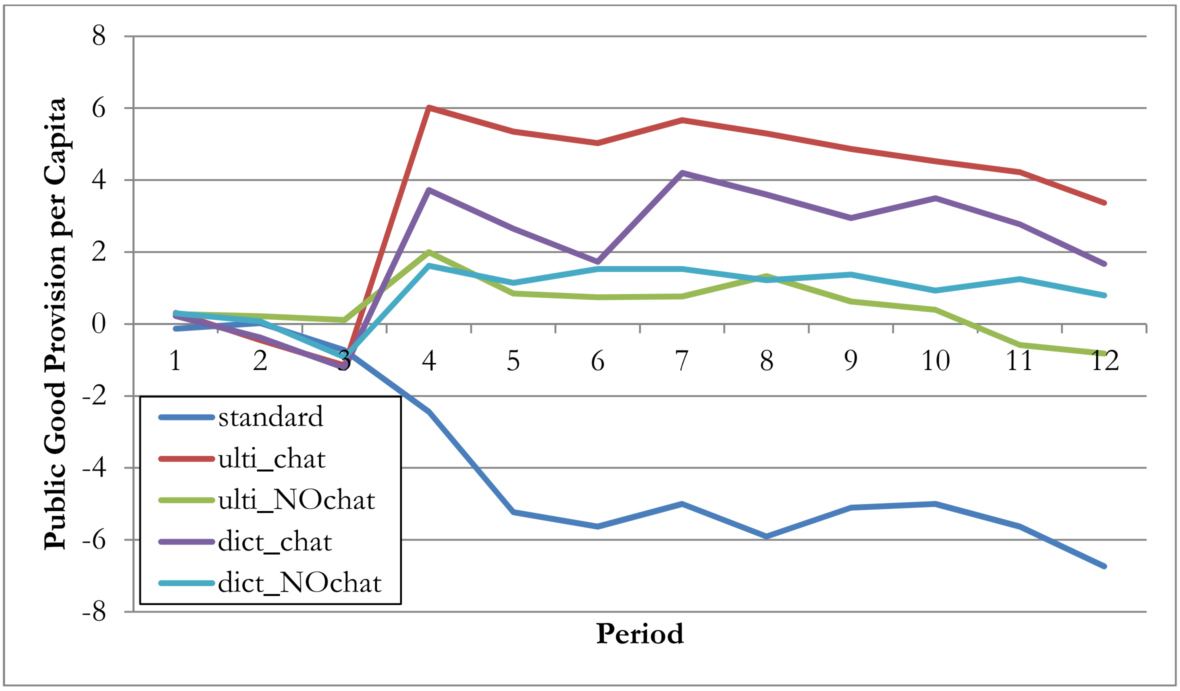



**Figure 3.** Public Good Provision in the Five Treatments over Time: Citizens and Migrants.




The confirmation of Prediction 3b is a product of our experimental design in which we aim to identify a potential interaction effect of public debate (chat) and different migration scenarios (ultimatum vs. dictator). After we have had a glance at the data at most aggregated level, we turn to the *contributions of citizens* (red players) in our treatments. Descriptive statistics in Table 3 report mean contributions of citizens over all twelve periods. These contributions amount to 2.536 Taler in Ulti_chat, -0.242 Taler in Dict_chat, -3.328 Taler in Ulti_NOchat and -2.575 Taler in Dict_NOchat. These mean contributions again hint at contribution differences between _chat and _NOchat treatments. Furthermore, we observe a nearly-3-Taler-difference of contributions between Ulti_chat and Dict_chat; this is a first indication that the *interaction* of the debate with the policy decision rule may play a role. We employ a series of regressions to further investigate similarities and differences. Table 4 reports five specifications providing evidence for important differences in contribution behavior of citizens. Figure 4 depicts contributions of citizens in our treatments over time. While specifications I to IV in Table 4 provide the reader with a better feel for the data, the full model in specification V controls for time effects, initial private endowments and treatment effects for periods four to twelve.[46] We argue that specification V is most useful for our analysis as it focuses on the periods in which treatment differences of institutions come to play a role for subjects. In periods one, two and three subjects already receive differing information, yet debate and decision rule institutions do not play a role yet, i.e. there are no treatment effects (all treatment dummies are not significantly different from zero, i.e. with $p > 0.1$). This circumstance is highlighted by results in specification I; here we observe no treatment differences in the first period by using a simple OLS specification. Considering all periods in specifications III and IV we find more evidence that a chat of citizens fosters contributions to the public good. Citizens' contributions in Ulti_chat and Dict_chat are greater than in the two _NOchat treatments. We remove periods one, two and three from our model in specification V and find evidence that citizens in Ulti_chat contribute greater amounts to the public good compared to all other treatments, including Dict_chat (all coefficients are negative and significantly different from zero at $p < 0.05$). Hence, the interaction of a debate opportunity via chat and the migration scenario indeed makes a difference. We report

**Result 2.** *High bargaining power of migrants in interaction with a debate among citizens fosters contributions by citizens. Considering periods in which institutions begin to matter, citizens' contributions to the public good are significantly greater in Ulti_chat than in all other treatments.*

---

[46] For all the specifications, we observe that citizens with endowments of 10 and 15 Taler contribute higher amounts to the public good. This finding is in accordance with earlier findings in the literature by Cherry et al. (2005). While it appears to be necessary to control for this characteristic in our analysis of contributions to the public good, we are mainly interested in observing how endowment levels change votes for *setting the bar*. A discussion of this will follow below.

**Figure 4.** Public Good Provision in the Five Treatments over Time: Citizens only.

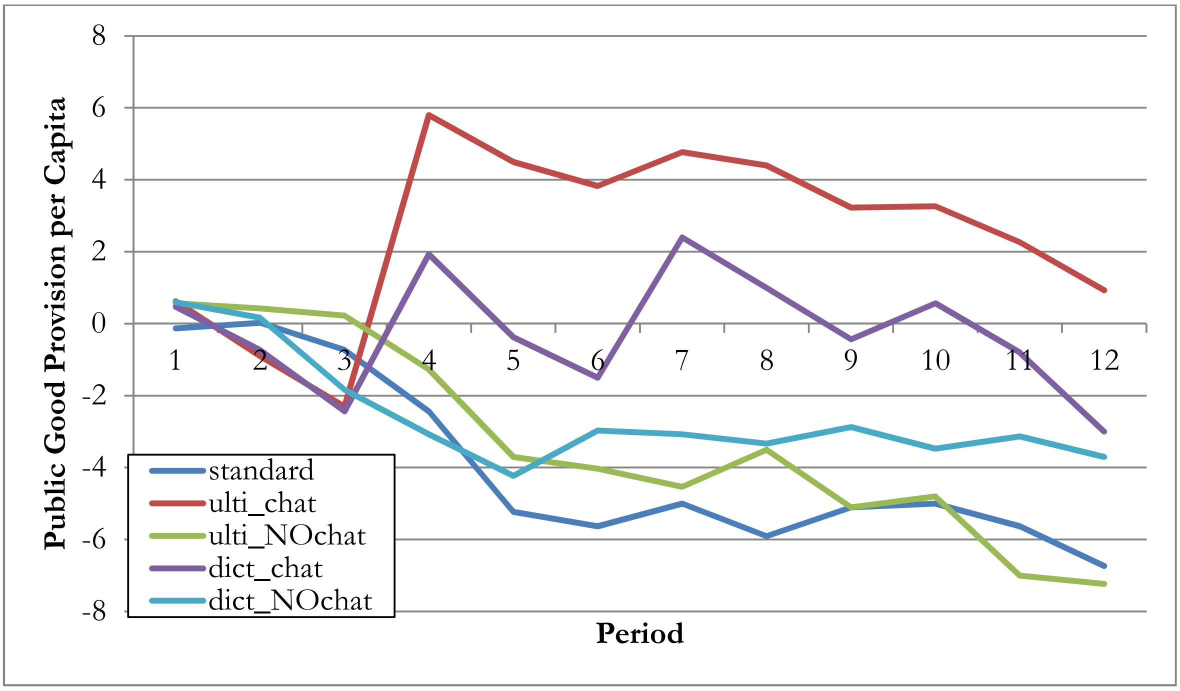



**Table 4.** Linear Regressions of Contributions to the Public Good, Individual Behavior of Citizens.

| Independent Variable | Dependent Variable: Contribution | | | | |
|---|---|---|---|---|---|
| | I<br>First period (OLS) | II<br>All periods | III<br>All periods | IV<br>All periods | V<br>Periods 4 to 12 |
| Endowment15 | 5.250***<br>(1.819) | 3.125**<br>(1.506) | 3.125**<br>(1.399) | 3.125**<br>(1.404) | 2.725*<br>(1.495) |
| Endowment10 | 2.900*<br>(1.588) | 2.713**<br>(1.266) | 2.713**<br>(1.193) | 2.713**<br>(1.198) | 2.483*<br>(1.284) |
| Ulti_NOchat | -0.067<br>(2.010) | | -5.864***<br>(1.509) | -5.864***<br>(1.514) | -8.241***<br>(1.654) |
| Dict_chat | -0.167<br>(2.073) | | -2.778<br>(1.691) | -2.778<br>(1.698) | -3.689**<br>(1.872) |
| Dict_NOchat | -0.033<br>(1.964) | | -5.111***<br>(1.647) | -5.111***<br>(1.653) | -6.981***<br>(1.816) |
| Period Dummies | - | *No* | *No* | *Yes* | *Yes* |
| Constant | -2.083<br>(1.534) | -2.848***<br>(0.795) | 0.590<br>(1.244) | 2.059<br>(1.351) | 3.842***<br>(1.423) |
| Observations | 120 | 1440 | 1440 | 1440 | 1080 |
| Individuals | 120 | 120 | 120 | 120 | 120 |
| Groups | 40 | 40 | 40 | 40 | 40 |

Note: Random effects estimation with robust standard errors (except specification I: OLS). The 'standard' treatment is excluded. The baseline are 'Ulti_chat' for treatment effects, 'Endowment5' for effects with regard to the size of the initial endowment and Period 1 (specification IV) or Period 4 (specification V). Standard errors in parentheses. Significance: *p < 0.10, **p < 0.05, ***p < 0.01.

Now we consider the drivers of this result. Before we have a closer look at the contributions of migrants to the public good, we follow the chronology of the experiment in order to better understand the dynamics of the game. Next, we analyze the results of the setting-the-bar stage of our design. Note that before periods four, seven and ten, citizens (red players) had to decide on an undisclosed vote on the migration requirement. Table 3 reports descriptive statistics on mean minimum contribution requirements. They are 7.030 Taler in Ulti_chat, 5.720 Taler in Ulti_NOchat, 8.410 Taler in Dict_chat and 5.967 Taler in Dict_NOchat. To analyze endowment and treatment effects, we employ three OLS regressions for the three decision rounds before periods four (specification VI), seven (specification VII) and ten (specification VIII); Table 5 reports estimation results and Figure 5 provides a graphic overview. Ceteris paribus, all three models do not reject the null hypotheses that votes by citizens with endowments of 5, 10 or 15 Taler were equal. Hence, we neither find any noticeable solidarity nor any discrimination of low-endowed or high-endowed citizens with their migrant counterparts:

**Result 3.** *Votes on minimum contribution requirements were independent from citizens' endowments.*

**Figure 5.** Minimum Contribution Requirements for Migrants to Join the Group.

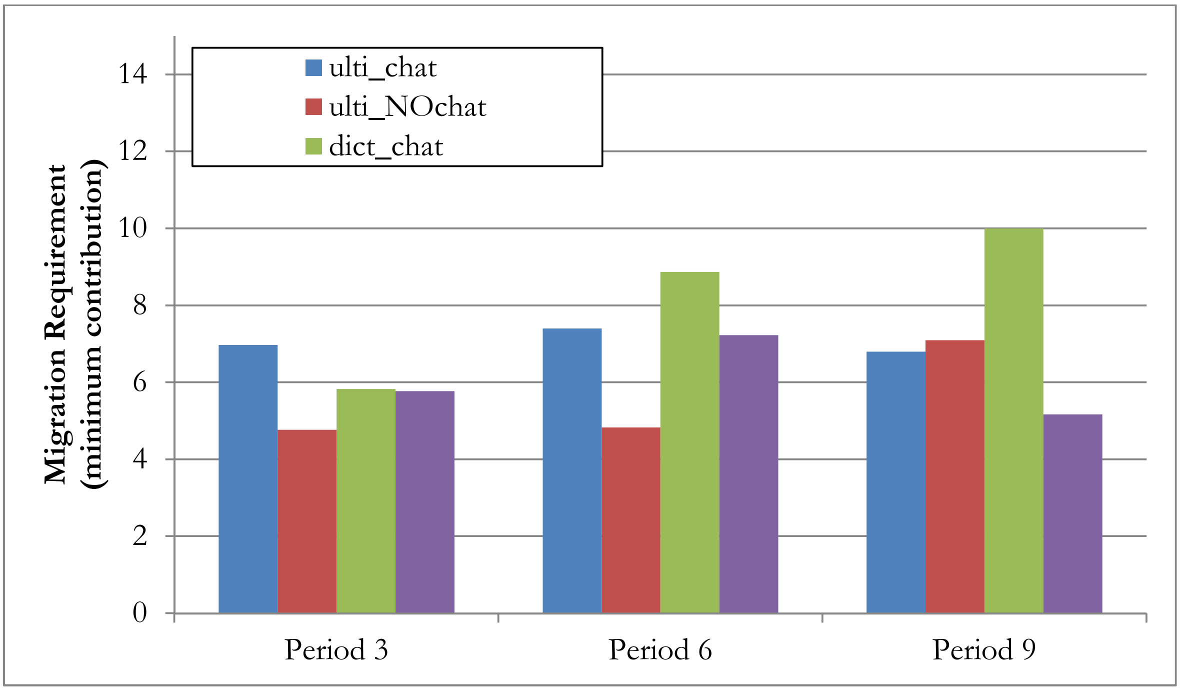


**Table 5.** Linear Regressions of Migration Requirements, Individual Behavior of Citizens.

| Independent Variable | Dependent Variable: Migration Requirement | | |
|---|---|---|---|
| | VI<br>First Voting, Period 4 | VII<br>Second Voting, Period 7 | VIII<br>Third Voting, Period 10 |
| Ulti_chat | 1.133<br>(1.867) | 1.467<br>(1.633) | -3.200***<br>(0.862) |
| Ulti_NOchat | -1.067<br>(1.794) | -4.033***<br>(1.437) | -2.900***<br>(0.946) |
| Dict_NOchat | -0.067<br>(1.641) | -1.633<br>(1.446) | -4.833***<br>(1.429) |
| Endowment15 | 0.375<br>(1.181) | 0.250<br>(1.267) | 0.250<br>(1.033) |
| Endowment10 | 0.250<br>(1.006) | 1.200<br>(0.875) | -0.800<br>(1.440) |
| Constant | 5.625***<br>(1.402) | 8.383***<br>(1.223) | 10.183***<br>(1.440) |
| Observations | 120 | 120 | 120 |
| Individuals | 120 | 120 | 120 |
| Groups | 40 | 40 | 40 |

Note: OLS estimation with robust standard errors. The 'standard' treatment is excluded. The baselines are 'Dict_chat' for treatment effects and 'Endowment5' for effects regarding the size of the initial endowments. Standard errors, clustered at the group level, in parentheses. Significance: *p < 0.10, **p < 0.05, ***p < 0.01.

Turning to treatment effects, we find a peculiar time trend for the minimum contribution requirement for migrants in specifications VI to VIII (Table 5). While the requirements do not differ across treatments in the first voting, subsequent requirements evolve to be especially high in Dict_chat. By the third (and last) voting, the requirement in Dict_chat has evolved to be significantly greater than in the three other treatments (all at least at the 5 percent level). In fact, it supports *Prediction 2b*, i.e. at least in Dict_chat (but not in Dict_NOchat) the bar is set optimally at $r^*_{Dict} = 10$ Taler and higher than in the Ulti_ treatments. We formulate

**Result 4.** *While minimum contribution requirements are comparable and stable for all other treatments over time, the requirement in Dict_chat increased over time and was significantly higher compared to all other treatments in the final voting.*

This difference suggests that the driver of a lower bar in the Ulti_ treatments may be due to citizens anticipating migrants' preferences for reciprocity (a strategic reason). Altruism of citizens is a less likely explanation; else the bar in Dict_chat should not be different. Note that this finding is only possible because our experimental design employed multiple voting periods. Taking Result 2 and Result 4 together, they suggest that the interaction of public debate with different migration scenarios, i.e. Ulti_chat vs. Dict_chat, motivates different behavior of citizens in two dimensions. First, citizens contributed significantly more resources to the public good in perceived high-potential migration settings (Ulti_). Second, citizens in perceived low-potential migration scenarios (Dict_chat) used the debate opportunity to set a bar which maximizes their payoffs. This is neither the case in Ulti_chat nor in the two treatments without chat.

How do migrants react to the requirements described above? In principle, migrants have either one or two channels to display their preferences. Given that a migrant's endowment is sufficient to allow entry into the country, she can voice her preferences via her contributions in Dict_. In Ulti_ she can (1) accept or reject the requirement, and, given that she has accepted it, (2) voice her preferences via contributions (like in Dict_). The descriptive statistics in Table 3 reveal that 95.7 and 100 percent of all migrants accepted the given requirement in Ulti_chat and Ulti_NOchat respectively. These numbers may not appear surprising, given the strong incentive to migrate in order to earn a higher income. They are however also a result of the relatively low requirements in the Ulti_ treatments. The mean accepted requirement in Ulti_chat is 6.25 Taler while the mean rejected requirement is 8.66 Taler. Recall that the mean requirement in Ulti_NOchat is 5.72 Taler and all migrants in Ulti_NOchat accepted the requirements. A Mann-Whitney test reports that the difference between all accepted and rejected requirements in the two Ulti_ treatments is statistically significant at the five percent level (p=0.0229). We find:

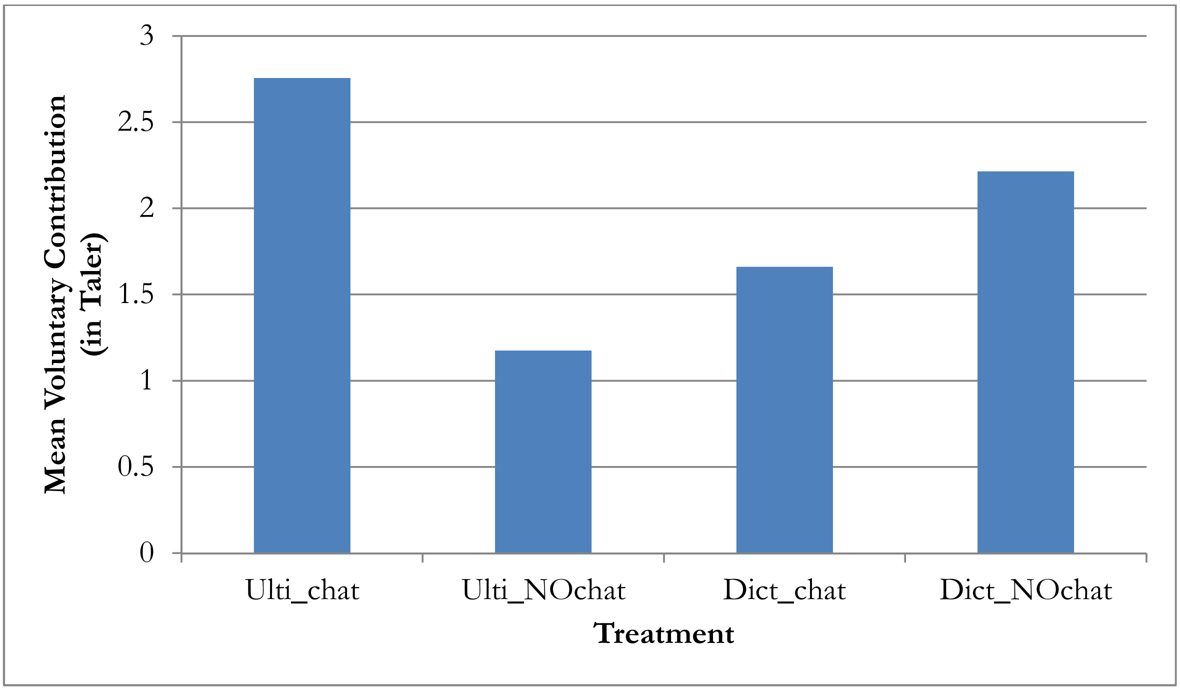
**Result 5.** *The great majority of migrants accept the requirements in the Ulti_ treatments. Those migrants who rejected requirements reacted to significantly higher requirements than those who accepted requirements.*

**Table 6.** Linear Regressions of Voluntary Contributions, Migrants for Periods 4 to 12.

| Independent Variable | Dependent Variable: Voluntary Contribution | |
|---|---|---|
| | X<br>Excluding time effects | XI<br>Including time effects |
| Ulti_NOchat | -1.297*<br>(0.778) | -1.357*<br>(0.784) |
| Dict_chat | -0.576<br>(0.783) | -0.669<br>(0.782) |
| Dict_NOchat | -0.406<br>(0.779) | -0.419<br>(0.804) |
| Period dummies | *No* | *Yes* |
| Constant | 2.593***<br>(0.609) | 3.083***<br>(0.752) |
| Observations | 597 | 597 |
| Individuals | 80 | 80 |
| Groups | 40 | 40 |

Note: Random-effects estimations with robust standard errors clustered at the individual level. An observation is the voluntary contribution of a migrant in a period, i.e. $abs(a_l - \bar{r})$. Migrants outside of groups in a given period, i.e. if $w_l < \bar{r}$, do not yield information and are excluded from this analysis. Further, we exclude migrants who are not able to contribute more Taler voluntarily (i.e. $w_l = \bar{r}$). The 'standard' treatment is excluded. Baseline: 'Ulti_chat'. Robust standard errors in parentheses. Significance: *p < 0.10, **p < 0.05, ***p < 0.01.

**Figure 6.** Mean Voluntary Contribution by Migrants for Periods Four to Twelve.




Note: Here we only consider migrants who indeed have a choice to contribute more than the bar, i.e. $w_l > \bar{r}$.

Next, we investigate the impact of our treatments on voluntary contributions of migrants that have entered a group. We define the size of a voluntary contribution as the absolute value of the difference between the contribution of migrant $c_j$ and a set minimum contribution requirement $\bar{r}$, i.e. $abs(c_j - \bar{r})$. For an adequate comparison, we need to exclude certain migrants from the analysis. First, we exclude all migrants who did not join a group in a given period. Second, we exclude all migrants who faced $\bar{r}$ such that $w_j = \bar{r}$ , because these migrants have to contribute exactly the minimum requirement and cannot contribute more Taler voluntarily. Consequently, we are left with 597 observations of 80 migrants from all 40 groups. Table 3 reports and Figure 6 depicts mean voluntary contributions by treatment. These are greatest in Ulti_chat with 2.757 Taler, followed by 2.215 Taler in Dict_NOchat, 1.662 Taler in Dict_chat and 1.176 Taler in Ulti_NOchat. To measure the statistical significance of these differences, we employ two random-effects regressions that included between three and nine observations per migrant. The results are reported in Table 6. For most comparisons, we cannot reject the null hypothesis of equal voluntary contributions of migrants, except for Ulti_chat > Ulti_NOchat at the 10 percent level. We generally have to be cautious about these comparisons, as treatments did differ significantly with respect to other characteristics, such as cooperation by citizens, as discussed above. Note, however, that the minimum contribution requirements in Ulti_chat were not greater than in Ulti_NOchat and Dict_NOchat. Hence, the described effect is not solely based on a larger action space of migrants.

**Result 6.** *Voluntary contributions of migrants are the greatest in Ulti_chat. However they do not differ statistically except in one marginal case.*

What impact do these treatment effects have on income of citizens and migrants and ultimately on welfare? Figure 7 provides mean income levels over all periods split by treatments and player identities respectively and Table 3 reports descriptive statistics on income. Over all periods and identities, mean income was the highest in Ulti_chat with 40.155 Taler, followed by 38.021 Taler in the Standard treatment, 36.539 Taler in Dict_chat, 35.859 Taler in Dict_NOchat and 35.717 Taler in Ulti_NOchat. Further, we observed that both citizens and migrants in Ulti_chat earned the highest income compared to all other treatments (see Table 3). To investigate the statistical significance of these income differences, we employed two OLS estimations with one observation per individual (i.e. average income over all periods). Table 7 reports the results of these estimations. In both specifications, we tested Ulti_chat against the other treatments. In specification IIX, we included the Standard treatment to get a first picture of the results, while we exclude it in specification IX that controls for endowments and the group status (that does not exist in the Standard treatment). Specification IIX reports that income is marginally lower (at the 10 percent level) in Ulti_NOchat and Dict_NOchat compared to Ulti_chat. The income levels in Standard and Dict_chat do not differ significantly from Ulti_chat in this first estimation. The second specification (IX) paints a more elaborate picture: being a citizen on average increases income by 15.283 Taler compared to migrants; endowments of 10 and 15 Taler result in 2.288 Taler and 6.875 Taler higher income respectively compared to an endowment of 5 Taler. Controlling for initial endowments and group membership, in specification IX we indeed find that average income in Ulti_chat is significantly greater than in Ulti_NOchat (at the 5 percent level), Dict_chat (at the 10 percent level) and Dict_NOchat (at the 5 percent level). We therefore report

**Result 7.** *Ulti_chat achieves the greatest welfare of all treatments. Both citizens and migrants are better off compared to all other treatments.*

**Table 7.** Linear Regressions of Income, all Players and Periods.

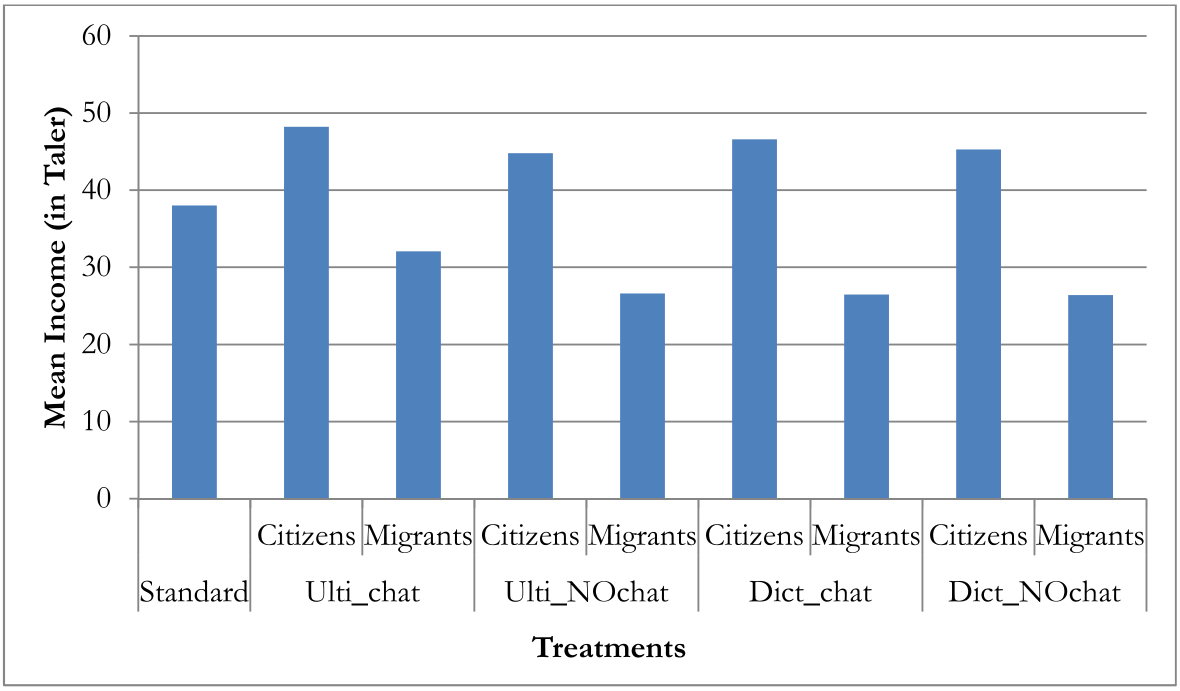
| Independent Variable | Dependent Variable: Average Income | |
|---|---|---|
| | IIX<br>All treatments | IX<br>Without Standard treatment |
| Standard | -2.134<br>(2.048) | |
| Ulti_NOchat | -4.438*<br>(2.359) | -4.438**<br>(1.733) |
| Dict_chat | -3.615<br>(2.632) | -3.615*<br>(2.023) |
| Dict_NOchat | -4.295*<br>(2.456) | -4.295**<br>(1.820) |
| Citizen | | 15.283***<br>(1.472) |
| Endowment15 | | 6.875***<br>(1.109) |
| Endowment10 | | 2.288*<br>(1.183) |
| Constant | 40.155***<br>(1.816) | 30.986***<br>(1.855) |
| Observations | 270 | 240 |
| Individuals | 270 | 240 |
| Groups | 50 | 40 |

Note: OLS estimations with robust standard errors. An observation is the average income of an individual over all twelve periods. The 'standard' treatment is excluded in specification IX. The baselines are 'Ulti_chat' for treatment effects and 'Endowment5' for effects regarding the size of the initial endowments. Standard errors in parentheses. Significance: *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Before we discuss implications of our results for policy, we further investigate our experiment data. We will take a closer look at the chat entries in Ulti_chat and Dict_chat and ex-post survey answers to explore insights into the emergence of differences between these treatments. This analysis holds the potential to explain or substantiate observations from the quantitative analysis. Moreover, such evidence could be useful to construct hypotheses about the reasons behind certain behavioral patterns.

Our quantitative analysis shows that the chat function has a significant effect on the minimum requirement setting for migrants to join the group in Dict_chat. In the debates before voting on the bar, we find notions of normative drivers for policy setting. Some subjects referred to equality and solidarity as reasons for specific policy setting levels. Few subjects also made reference to 'punishment' of the migrants or protection from (financial) exploitation as the basis for increased immigration requirements.

**Figure 7.** Mean Income over all Periods, by Citizens and Migrants.




We closely examine the chat data for each group over the course of the experiment to explore the arguments leading to the minimum requirement setting and to identify potential differences in reasoning between Ulti_chat and Dict_chat. The analysis outlines different argumentative frames. Here, we are interested in all the arguments for or against high or low immigration requirements. Further, we are interested in the most dominant arguments in group consensus building. The profit maximization argument was the most dominant argumentative frame across most groups and both treatments. We also observe some instances of intuitive reasoning ('the smaller the group the better, right?', 'this feels right'), especially in the first chat.

In Ulti_chat, we additionally find evidence for discussions on the potential deterrence effects of high immigration requirements. High requirements make it impossible for some migrants to join and even some that can join may be deterred by high requirements. Similarly, we find occasional arguments about fairness, solidarity and generosity in Ulti_chat. However, our hypothesis is that the deterrence effect by itself is not perceived to be a significant factor in setting the policy. This is suggested by the quantitative similarities in setting the bars between Ulti_chat, Ulti_NOchat and Dict_NOchat. Rather, when migrants hold some bargaining power in Ulti_chat in combination with a debate, this motivates citizens to contribute more to the public good (i.e. in accordance with the norm they set for migrants and different from all other treatments) and refrain from setting a high bar (compared to Dict_chat).

**Table 8.** Regressions Explaining Perceived Fairness of the First Bar.

| Independent Variable | Dependent Variable: Perceived Fairness (on a five-point scale) | |
|---|---|---|
| | XII<br>OLS | XIII<br>Tobit |
| Size of first Bar (in Taler) | -0.093**<br>(0.035) | -0.137***<br>(0.047) |
| Ulti_NOchat | -0.692**<br>(0.327) | -1.083**<br>(0.485) |
| Dict_chat | -0.536*<br>(0.293) | -0.751*<br>(0.440) |
| Dict_NOchat | -0.317<br>(0.275) | -0.541<br>(0.412) |
| Citizen | 0.405*<br>(0.232) | 0.663**<br>(0.335) |
| Endowment15 | 0.095<br>(0.237) | 0.156<br>(0.351) |
| Endowment10 | 0.039<br>(0.269) | -0.048<br>(0.386) |
| Constant | 3.808***<br>(0.237) | 4.179***<br>(0.423) |
| Observations | 235 | 235 |
| Individuals | 235 | 235 |
| Groups | 40 | 40 |

Note: OLS and Tobit estimates with robust standard errors clustered at the group level. We employ the Tobit estimation with a lower limit of 1 and an upper limit of 5. The 'standard' treatment is excluded. The baselines are 'Ulti_chat' for treatment effects and 'migrant' for group membership effects. Five observations are missing because of incomplete survey answers. Robust standard errors in parentheses. Significance: *p < 0.10, **p < 0.05, ***p < 0.01.

Furthermore, these findings are supported by the ex-post survey answers. We asked citizens and migrants to rate how fair they considered the first minimal transfer on a five-point scale[47]. We only asked for the *first* bar in order to avoid confounding answers on the average of all bars which would have been harder to analyze. Figure 8 depicts the mean perceived requirement fairness levels for the four treatments with migration requirements for both citizens and migrants. Overall fairness levels amounted to 3.379 points in Ulti_chat, 2.915 points in Ulti_NOchat, 2.932 points in Dict_chat and 3.186 points in Dict_NOchat. In order to separate drivers of perceived fairness, we employ OLS and (as a robustness check) Tobit estimations. Table 8 reports the results of these estimations in which we control for the size of the first bar in the respective group, treatment effects, whether the answer comes from a citizen or a migrant, and endowment effects. As may be expected, ceteris paribus a higher bar (minimum contribution level) yielded lower fairness perceptions of the bar. We found that, on

[47] Throughout the experiment we used language that was as neutral as possible, e.g. 'minimum transfer' instead of 'immigration requirement'. See appendix B for experiment instructions.

average, a one Taler increase in the minimum requirement yielded a reduction of 0.09 fairness points (significantly different from zero at the 5 percent level, OLS estimation). Furthermore we observe that, on average, citizens regarded the same bar 0.4 points fairer compared to migrants (significantly different from zero at the 10 percent level, OLS estimation). More surprisingly, we find that the perceived fairness about the first bar was significantly greater in Ulti_chat compared to Ulti_NOchat (at the 5 percent level) and Dict_chat (at the 10 percent level). Perceived fairness of the bar was also lower in Dict_NOchat compared to Ulti_chat, but not significantly so. We summarize

**Result 8.** *The chat in Ulti_chat contained debate about fairness towards and possible deterrence of migrants. Perceived fairness of the first bar (minimum contribution level) was greatest (and in 2 of 3 cases significantly) in Ulti_chat compared to the other three treatments.*

What drives Result 8? The result does not appear to be motivated by a lower mean requirement (more liberal immigration policy) in Ulti_chat (see Figure 5 and Table 5). Rather, we compare a citizen's contribution in a given period and the vote she casts in the decision-making on the requirement valid in that period. Naturally, these two decisions were made by the same individual and are therefore not independent from each other. We employ Wilcoxon signed-rank tests to compare whether the contributions and demanded requirements differed from each other. This test is depicted by the difference between solid and the dashed lines in Figure 9. Examining periods 4 to 12 and each treatment individually, we find that in Ulti_NOchat, Dict_chat and Dict_NOchat, citizens' contributions and citizens' demands differed in the vast majority of periods and at the 1 percent level. The only exceptions were period 4 (no significant difference) and period 5 (significant at the 5 percent level) in Dict_chat. Conversely, in periods 4 to 10 in Ulti_chat, there was no significant difference between citizens' contributions and their individual demands. Only in periods 11 and 12, we find differences at the 5 and 1 percent level respectively, probably driven by the experiment coming to an end.

This final analysis is key for understanding the results of our experiment. Only in Ulti_chat, citizens showed coherence between what they demand from migrants to contribute to the public good and their own contributions. This finding is the fundamental insight that drives Results 2, 6 and 8 presented above. Compared to all other treatments, in Ulti_chat, citizens contributed more to the public good in order to achieve fairness of the requirement. In appreciation of this fairness, voluntary contributions by migrants tended to be higher in Ulti_chat. Consequently, the public good got nurtured and yielded higher payoffs for both citizens and migrants. When asked about the fairness of the requirement, it was rated most fair in Ulti_chat as a result of this coherence. We shall call this mechanism *migration policy coherence*. In the next section, we will discuss the implications of our experimental results for immigration requirements.

**Figure 8.** Mean Perceived Fairness of the *First* Bar, by Citizens and Migrants.

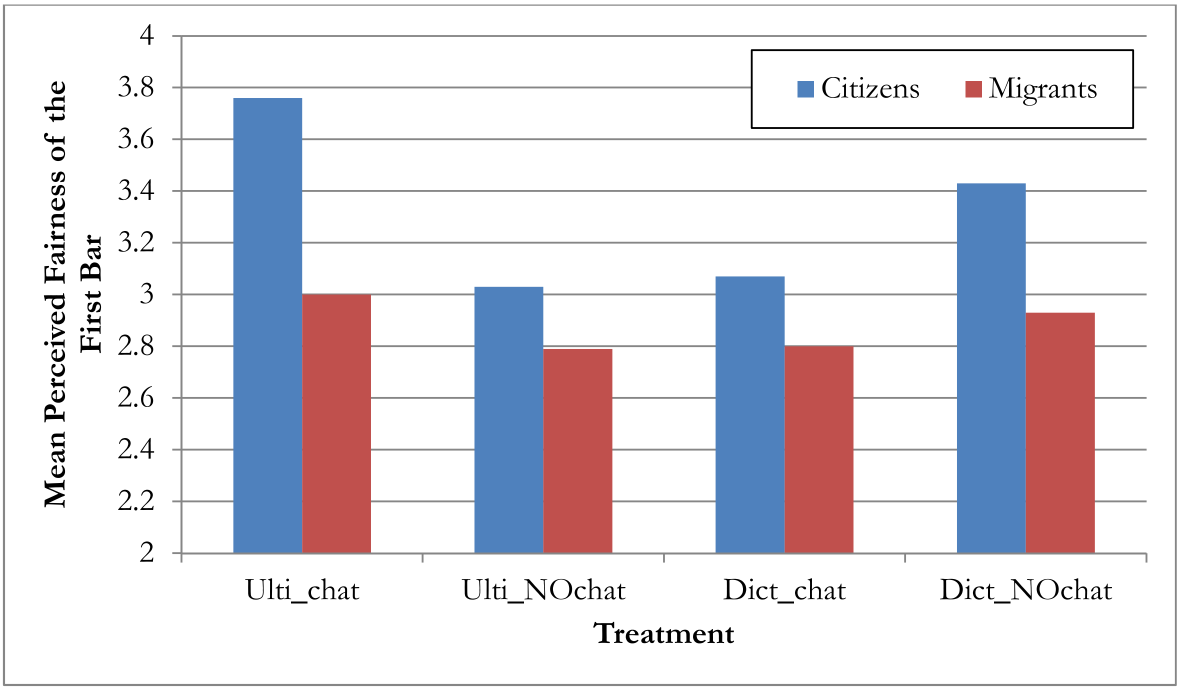
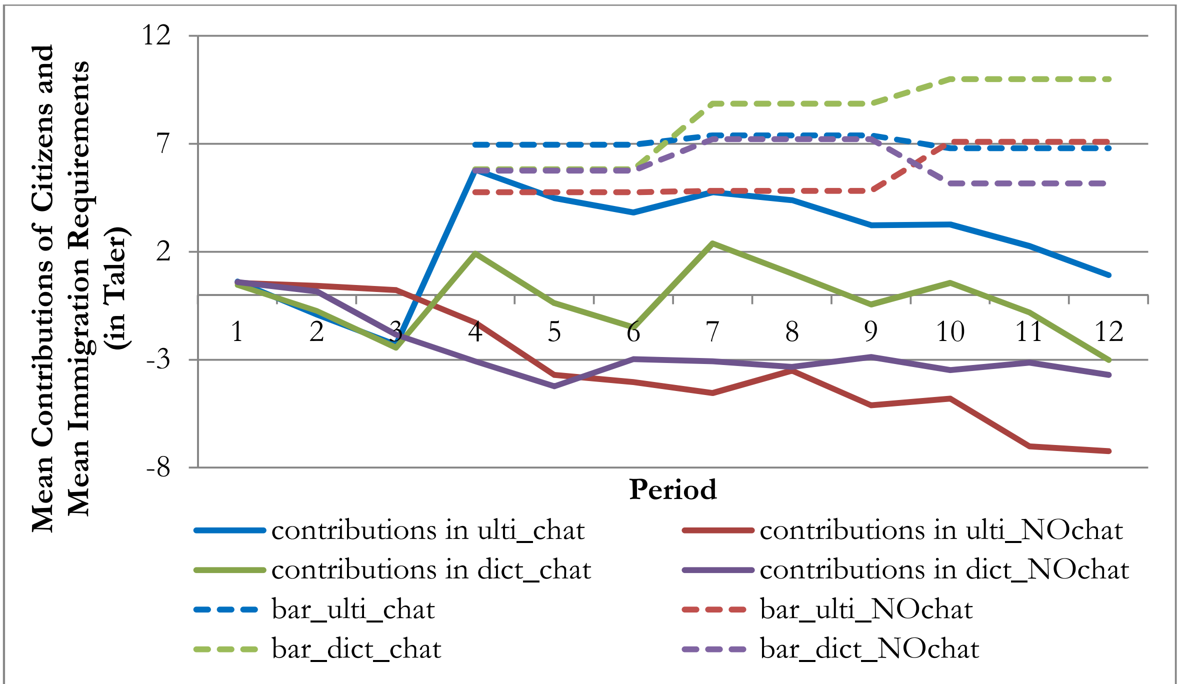



**Figure 9.** Public Good Provision over Time: Requirements and Citizens only.



## 4.4 Discussion

Policy makers across EU countries will have an increasing interest in how to reconcile future labor demand and growing public opposition to migration. In this section, we discuss our main results and suggest potential implication for immigration policy making. How does public debate and varying migration scenarios influence economic immigration requirements in a labor demand context?

We find that immigration requirements are generally higher than mean contributions of citizens. Citizens expect more than they deliver, except in Ulti_chat. This finding could be interpreted as a form of in-group favoritism (see Tajfel et al. 1971, Billig and Tajfel 1973). In real-world policy, we observe that immigration requirements, such as income or sociocultural requirements (civic knowledge) often exceed population averages (e.g. mean income).

Public debate did not lead to lower immigration requirements from fairness or solidarity group preferences. We do not observe a lower immigration requirement in perceived high-potential migration scenarios (Ulti_ treatments). In fact, immigration requirements were higher in cases with perceived low-potential migration flows in combination with public debate (Dict_chat). This suggests that citizens use public debate to coordinate a strategy to exploit immigrants that have less bargaining power. Public debate as such can be a double-edged sword. The opportunity to debate can increase citizens' contributions to the public good and facilitate in-group cooperation and solidarity (e.g. in Ulti_chat). However, the findings also suggest that it can be used to set disproportionately restrictive immigration requirements in order to maximize profits while keeping contribution levels low.

This finding can also be interpreted in the context of recent trends among EU countries that are setting more restrictive immigration requirements in response to public opposition to traditionally low-skilled migration. Hence, policy makers aiming to stimulate labor demand should consider the implications of framing the potential of future migration in discourse, especially in light of persisting negative perceptions of past migration flows.

In order to understand the results, we also have to challenge the concept of fairness. 'Fair' immigration requirements are intuitively measured by the degree of exclusion of migrants. Hence, absolute higher immigration requirements indicate a more 'unfair' policy. However, if we look at fairness as a relative concept that incorporates individual contributions in relation to the immigration requirement, this perspective leads to some interesting insights.

In sum, we conclude that perceived high-potential migration flows in combination with public debate induce the greatest *migration policy coherence*. We assume that higher bargaining power of migrants leads to a debate among citizens, increasing their own contributions and adhering to self-imposed standards set for migrants. We argue that this relative fairness induces higher fairness perception, higher voluntary contributions of migrants and, in return, higher overall welfare of citizens and migrants.

Overall, our analysis shows the varying impact of public debate and different migration scenarios on immigration policy. First, policy makers have to be aware of the potential of debate. Different perceived potential of migration flow can lead to more restrictive immigration policy or to a 'fairer' and more efficient immigration policy defined by high *migration policy coherence*. Consequently, how a certain debate about immigration is framed in public discourse can have serious effects on immigration policy. Stressing the benefits and contributions of migration, facilitating high skilled migration and boosting socioeconomic integration of migrants could be strategies to promote a more positive debate on migration.

## 4.5 Conclusion

Social science has documented a restrictive trend in setting immigration requirements across EU countries. Some EU countries, particularly Germany, already face labor demand in certain sectors. As this demand grows as a result of population ageing, policy makers have to look for ways to reconcile public opposition to immigration with a growing need to attract more immigrants. Given the political salience of immigration policy, it is surprising that the issue of immigration requirements (in particular, the consideration of legal migration) has been characterized by a lack of empirical research.

Sociologists and political scientists often focus on the historic, cultural, political and social context in which immigration policy is developed. Results are rarely comparable due to legal complexity, ongoing government changes, and most importantly, differences between countries on several (i.e. historic, cultural, geographical etc.) dimensions. For these reasons, we apply experimental methods to gain some context-independent insights into the drivers behind immigration policy setting. To the best of our knowledge, this has not been done before. Our novel approach has the potential to reveal fundamental dynamics of group decision making that could be interesting for policy makers across different geographical, cultural or political contexts.

We decided to select two dimensions of immigration policy for our experiment. First, the perception of migrants and their status in comparison to citizens of a country is a critical factor to consider. Whether migrants are regarded as equals or as inferior is expected to play a great role. Second, group decision making is likely to change if members of a group (in this case, citizens in a country) are able to communicate and exchange points of view. Voter opinion on a policy may converge to a compromise when arguments of fairness and payoff maximization are exchanged.

Our results show the varying impact of public debate and different migration scenarios on immigration policy. First, policy makers have to be aware of the double-edged potential of public debate. Different perceived potential of migration flow can lead to more restrictive immigration policy or to a 'fairer' and more efficient immigration policy defined by higher citizens' contributions, high voluntary contributions of migrants and high overall welfare. As mentioned, we defined this indicator as *migration policy coherence*.

Consequently, how a certain debate about immigration is framed within a public discourse can have serious effects on immigration policy. We proposed certain mechanisms that could explain this finding: Citizens in a context of perceived high-potential immigration consider immigrants their equals. Thus, citizens themselves comply with the immigration requirements that they enforce for migrants. Conversely, we have seen that discounting the potential of migration can lead to more restrictive immigration requirements and lower contributions from both citizens and migrants.

Stressing the benefits and contributions of migration, facilitating high skilled migration and boosting socioeconomic integration of migrants could be strategies to promote a more positive perception of migration. Most importantly, immigration requirements should be proportionate to population means. This will lead to a greater perception of fair policy and, in return, higher contributions from both citizens and migrants. Our study has shown that policy 'fairness' is not measured by the degree of restrictiveness of requirements or the potential exclusion of immigrants, but in relative terms. Rather, high level *migration policy coherence* is a good indicator for fairness and efficiency of immigration requirements.

Naturally, our study does not reflect nearly all facets of immigration policy. There is a need for further research and we hope to motivate further experimental studies in this area. There are numerous set-up alternatives that may be considered in the future. For example, one may consider a debate between citizens and migrants, countries which compete for high-endowed migrants, and naturalization of migrants after a certain period of time. Such factors could provide further valuable insights into the drivers of immigration policy.

## 4.6 Appendix A: Supplementary Figures and Tables

**Figure A.1a.** Extensive Form of our Game in a Simplified 2-player Case, Ulti_ treatments.

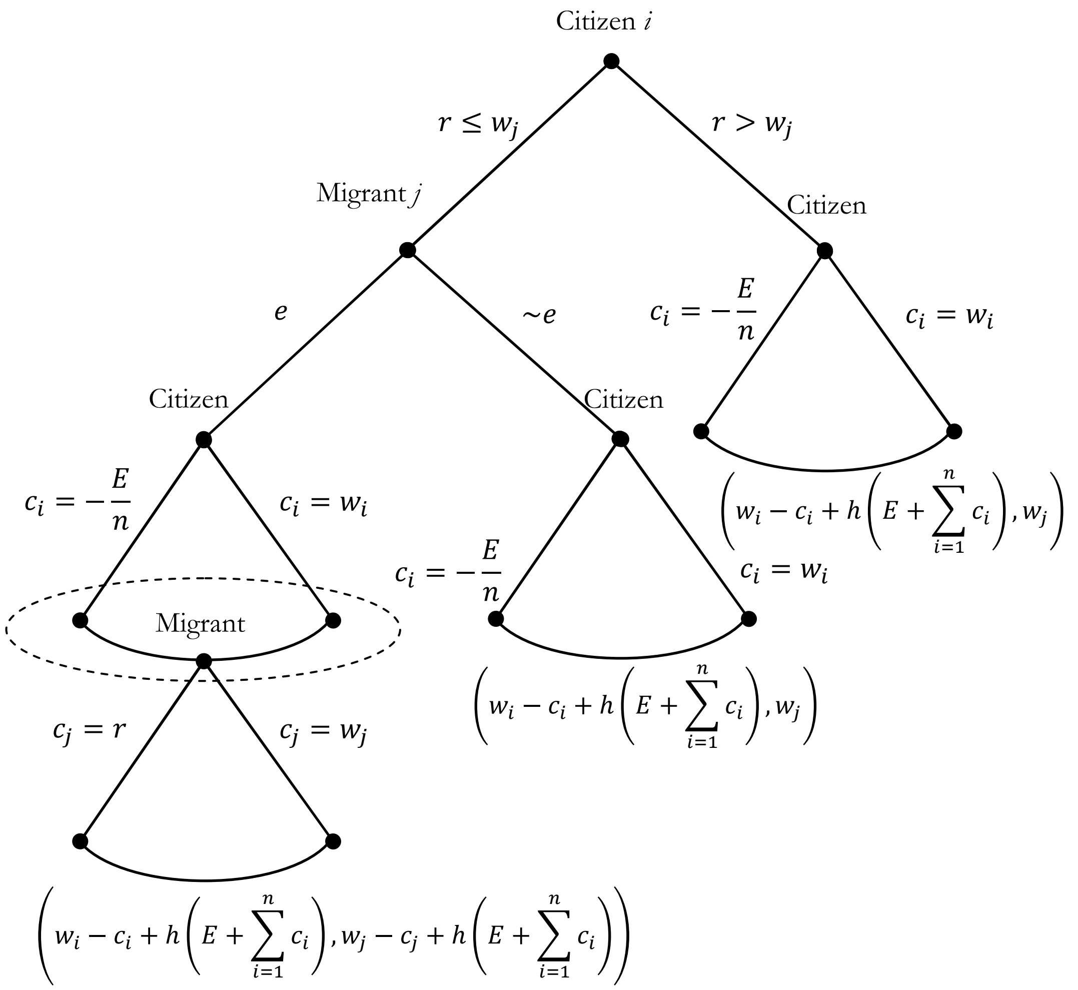


**Figure A.1b.** Extensive Form of our Game in a Simplified 2-player Case, Dict_ treatments.

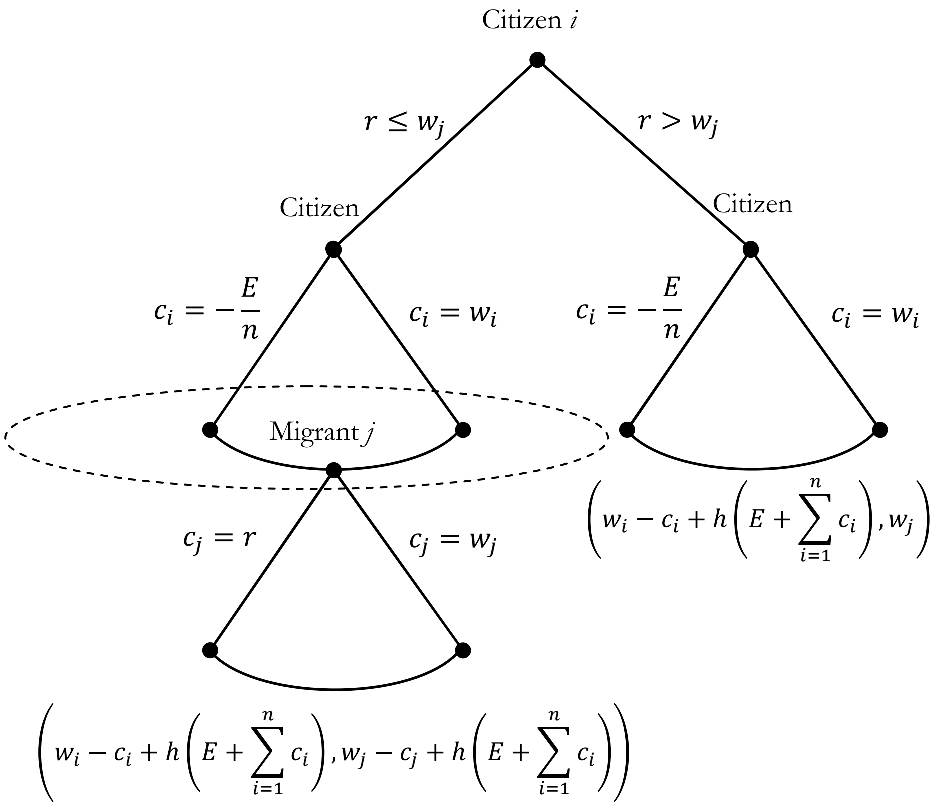


**Table A.1.** Test Statistics for Comparison of Group Contributions, all Players and Periods.

| (row vs. column comparison) | | Treatment | | | |
|---|---|---|---|---|---|
| | | Dict_NOchat | Standard | Ulti_chat | Ulti_NOchat |
| Treatment | Standard | < (p = 0.0019) | | | |
| | Ulti_chat | > (p = 0.0343) | > (p = 0.0005) | | |
| | Ulti_NOchat | = (p = 0.7624) | > (p = 0.0009) | < (p = 0.0588) | |
| | Dict_chat | = (p = 0.4057) | > (p = 0.0003) | = (p = 0.4497) | = (p = 0.1736) |

Note: All test statistics are (nonparametric) Mann Whitney tests. One observation is the average group contribution of a group over twelve periods, yielding 10 independent observations per treatment. The table is to be read row vs. column. For instance, group contributions are significantly greater in *Dict_NOchat* compared to *Standard*.

## 4.7 Appendix B: English Translation of the Experimental Instructions

General Instructions for Participants

*[Ulti_chat treatment instructions, translated from German. Original German instructions for all treatments are available from the authors upon request.]*

**Welcome to the Experiment Laboratory!**

You are now taking part in an economic experiment. You will be able to earn a considerable amount of money depending on your decisions and the decisions of others. It is therefore important that you read these instructions carefully.

The instructions which we have distributed to you are solely for your private information. **It is prohibited to communicate with other participants during the experiment.** Should you have any questions, please raise your hand and an experimenter will come to answer them. If you violate this rule, we will have to exclude you from the experiment and from all payments.

During the experiment, you will make decisions **anonymously**. Only the experimenter knows your identity and your personal information is confidential. Your decisions will not be traceable to your identity.

In any case, you will earn 5 Euros for your participation in this experiment. The additional earnings depend on your decisions. During the experiment, your earnings will be calculated in tokens. At the end of the experiment, your earned tokens will be converted into Euros at the following exchange rate:

**1 Taler = 0,20 €,**

and they will be paid to you in **cash**.

The experiment consists of **12 periods** in which you always play the same game. The participants are divided into cohorts of 6 with two colors: 3 RED and 3 BLUE players. Hence, you will interact with 5 other participants. The composition of the cohort will remain the same for all 12 periods. Please note that you and all other participants decide anonymously. Therefore, cohort members will not be identifiable over the periods.

At the end of the experiments, you will receive your earning from **one out of the twelve periods** converted in Euros (according to the exchange rate above) in addition to the 5 Euros for your participation. The payout period will be determined **randomly**. You should therefore take the decision in **each** period seriously, as it may be determined to be the payout period.

The following pages describe the course of the experiment in detail.

## Assignment to Colors, Endowments and Group-Membership

At the beginning of the experiment, colors will be randomly assigned to all players. You are either a RED or a BLUE player. This color will remain the same throughout the 12 periods of this experiment.

You will receive an endowment to your private account which is the same for all 12 periods. Your endowment can be 5, 10 or 15 Taler. This assignment is also done at random by the computer program. A 6-player cohort will be set up as follows:

| In each cell there is always **exactly one player** of a 6-player cohort.<br>The random assignment at the beginning of the experiment will remain the same for all 12 periods. | | Color | |
|---|---|---|---|
| | | RED player<br>*(always group members)* | BLUE player<br>*(may become group members)* |
| Endowment | 5 Taler | 1 Player | 1 Player |
| | 10 Taler | 1 Player | 1 Player |
| | 15 Taler | 1 Player | 1 Player |

**All RED players are always members of the group.** BLUE player are not members in the beginning of the experiment, but may become members depending on their own decision and the decisions of RED players. Further explanations will follow on the next pages.

## Rules of the Experiment

Each group member (at the beginning, each RED player) needs to make the following decision. Your task (and the task of all other group members) is to decide on the transfer between your private account and a group account.

At the beginning of each of the 12 periods, each player receives an endowment of 5, 10 or 15 Taler in a **private account**. Moreover, there are **60 Taler** in a **group account**.

**Each group member has to decide** how many Taler to transfer from the private account to the group account **or** from the group account to the private account. Transfers may be between -10 and the respective initial endowment (either 5, 10 or 15 Taler, only whole numbers). If a group members types in a positive number, then she transfers Taler from her private account the group account. If the typed number is negative, then she transfers Taler from the group account to her private account.

All other players who are not group members (all BLUE players in the beginning of the experiment) make **no decisions** concerning the transfer between the private and the group account.

**The Income of Group Members**

The complete income of group members is constituted by two parts:

(1) the Taler which are in the private account after the transfer

(2) the **income from the group account**. The income from the group account is calculated as follows:

| |
|---|
| Your income from the group account =<br>0.5 *times* The sum of Taler in the group account |

The income of a group member (in Taler) therefore is

**(your initial endowment – your transfer) + 0.5*(the sum of Taler in the group account).**

The income from the group account of all other group members is calculated using the same formula so that each group member receives the same income from the group account.

For instance, if the sum of transfers of all group members equals 10 Taler, then the group account holds 60 + 10 = 70 Taler. Accordingly, you (in case you are a group member) and all other group members receive an income from the group account of 0.5*70 = 35 Taler each. If you and the other group members transfer the sum of -3 Taler to the group account, then the group account holds 60-3= 57 Taler. You and all other group members receive 0.5*57=28.5 Taler each as income from the group account. For each Taler that you hold in your private account, you will receive 1 Taler.

**The Income of Players who are not Group Members**

Players who are not group members (in the beginning, all BLUE players) in a given period receive their initial endowment (either 5, 10 or 15 Taler) as income in that period. Beyond this income, such players receive **no income from the group account**.

## Voting on the Minimal Transfer by BLUE Players

At the end of periods 3, 6 and 9 RED players are able to vote on the **minimal transfer** by BLUE players. This minimal transfer represents a **requirement for the group membership of BLUE players** and can be between -10 and 15 Taler. **It is not a requirement for RED players**. The group membership requirement is set such that a majority of RED players agrees with it (at least two out of three RED players). Before RED players vote on the requirement, they will be able to communicate for three minutes with each other via a chat.

Thereafter, each BLUE player may **decide** whether she accepts the minimal transfer and **becomes a group member**. Only BLUE players who hold initial endowments which are greater than or equal to the minimal transfer get the chance to decide. BLUE players with endowments that are lower than the minimal transfer **do not** get the chance to decide to become group members. If a BLUE player becomes a group member, in the following period, she is able to transfer Taler between her private account and the group account while observing the minimal transfer requirement. The requirement is **not binding** for RED players who will always be group members for all 12 periods.

Example 1:
Imagine that the three RED players decide on the following votes for the minimal transfer:
Player 1: -7 Taler Player 2: 12 Taler Player 3: -1 Taler.

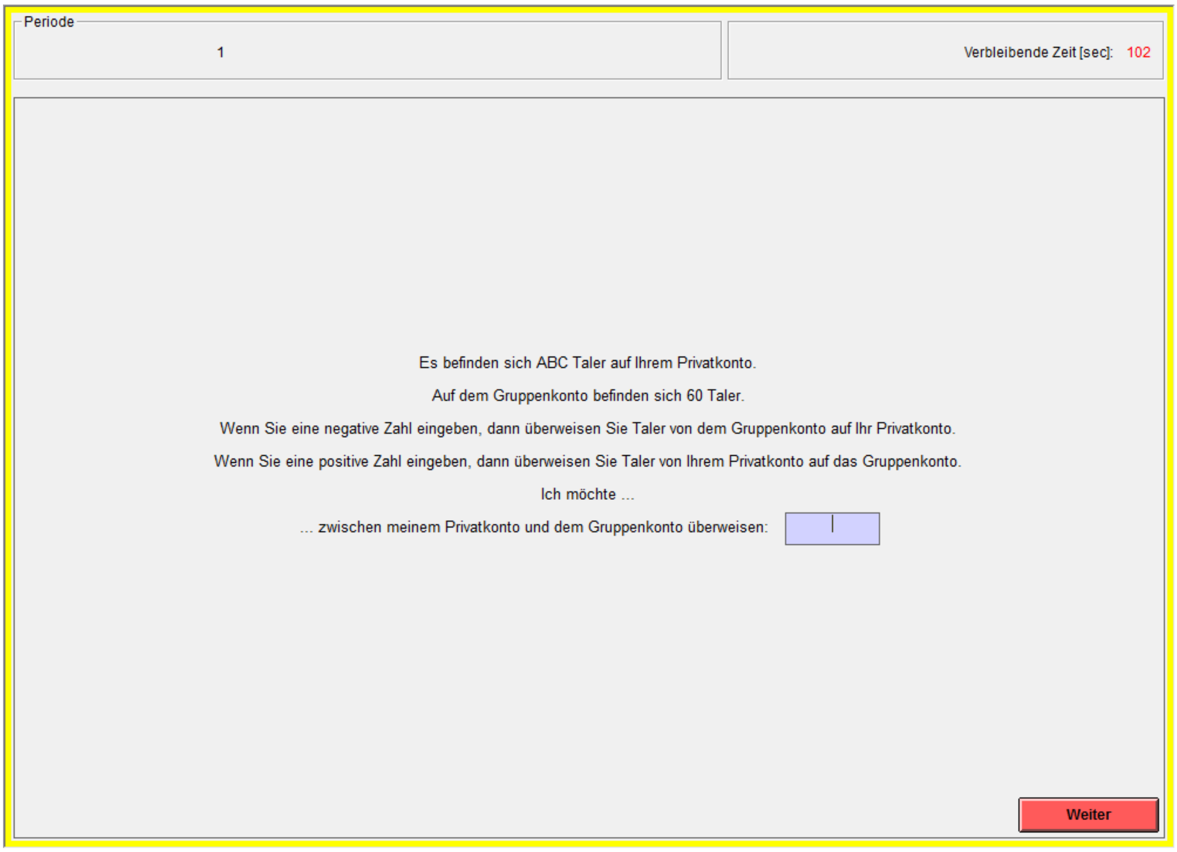
This means that player 1 wants to introduce a minimal transfer of -7 Taler for BLUE players, while players 2 and 3 want to set the minimal transfer at 12 Taler and -1 Taler, respectively. In this case, the **minimal transfer** will be set at **"-1 Taler"** as two out of three RED players regard a minimal transfer of -1 Taler as acceptable. **All three BLUE players are then able to decide individually and in private** whether they want to accept the minimal transfer requirement. After their decisions, all players receive information on the number of members the group has from there on.

Example 2:
Imagine that the three RED players decide on the following votes for the minimal transfer:
Player 1: 11Taler Player 2: -10 Taler Player 3: 15 Taler.

In this example, the **minimal transfer** would be **"11 Taler"** as two out of three RED players regard 11 Taler as acceptable. In this example, the **two BLUE players** with endowments of 5 Taler and 10 Taler **cannot decide** to join the group as their endowments are **lower** than the minimal transfer requirement. The BLUE player with the endowment of 15 Taler can still decide to join the group while observing the minimal transfer requirement. After this decision, all players receive information on the number of members the group has from there on.

## Information on the Course of Events of the Experiment

At the beginning of each period, all **group members** see a **decision screen**. Players who are **not** group members (at the beginning, all BLUE players) are **not able to make a decision**. The **period number** is shown in the left upper corner. The remaining **time** to make a decision is shown in the right upper corner.
The decision screen for group members is the following:

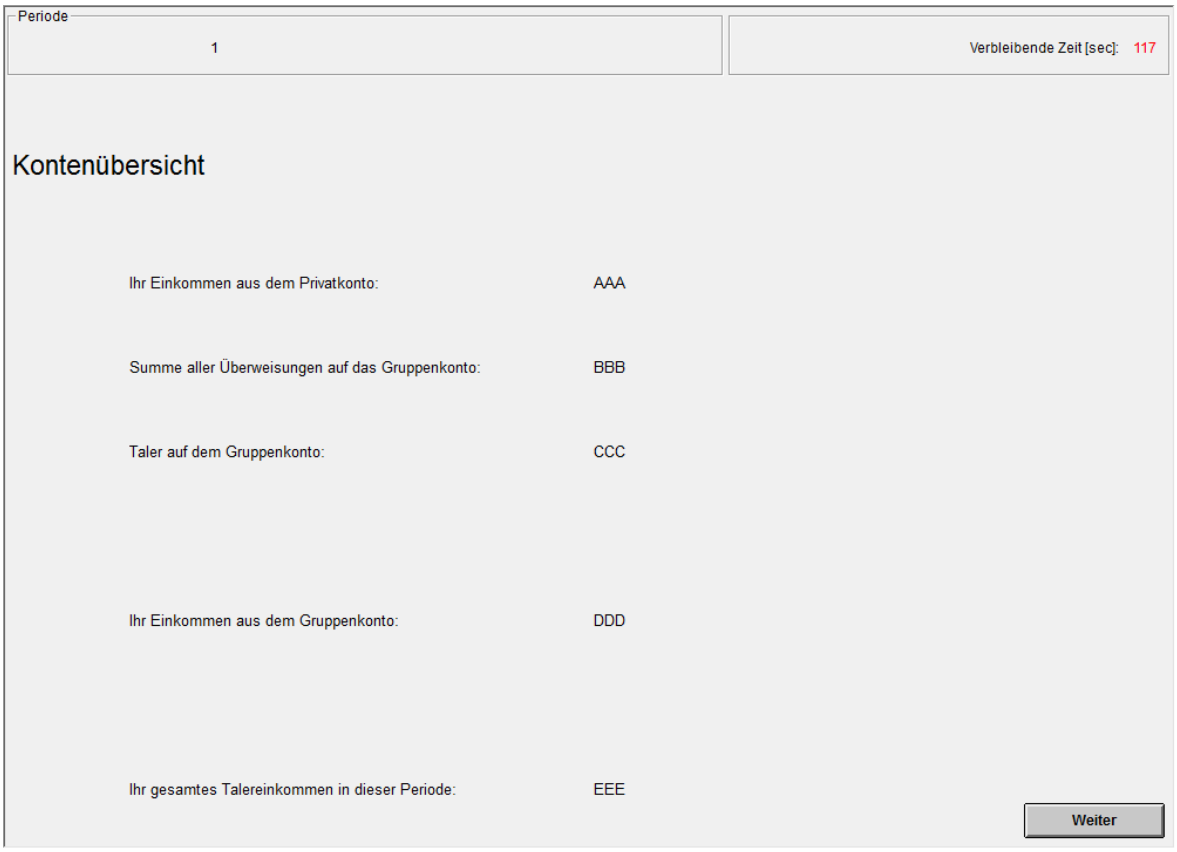



As described above, the group account holds **60 Taler** at the beginning of **every period**. You decide on your transfer by entering a whole number between -10 and your private endowment (5, 10 or 15 Taler) into the input window. You can click on this window with the computer mouse.

When you have entered your decision, you need to click the **continue button** (*German "Weiter-Taste"*) by use of the mouse. After you have clicked the continue button, your decision is final for that period.

After all group members have made their decisions, the account screen will inform you about the total number of Taler in the group account and your private account. Furthermore, you will be informed about your total income in this period.

**The account screen:**

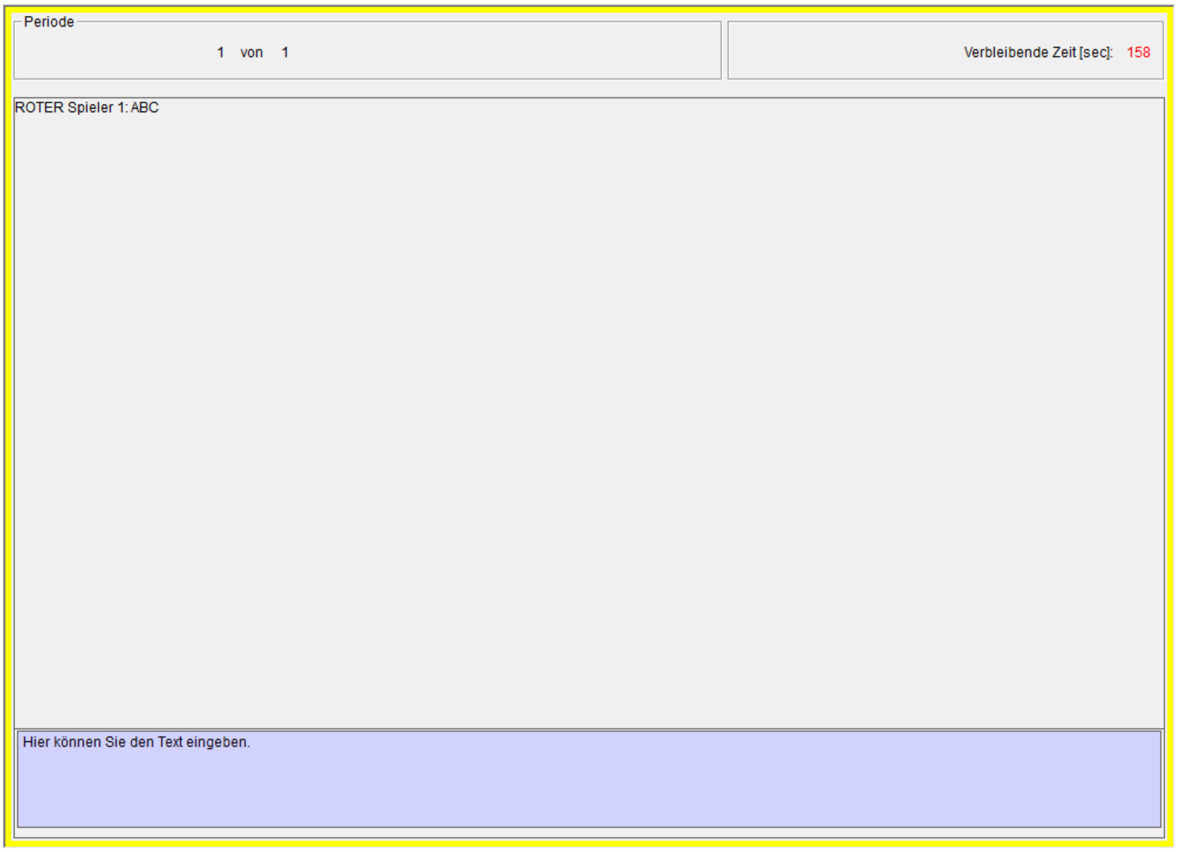



As described above, the total income of a group member consists of

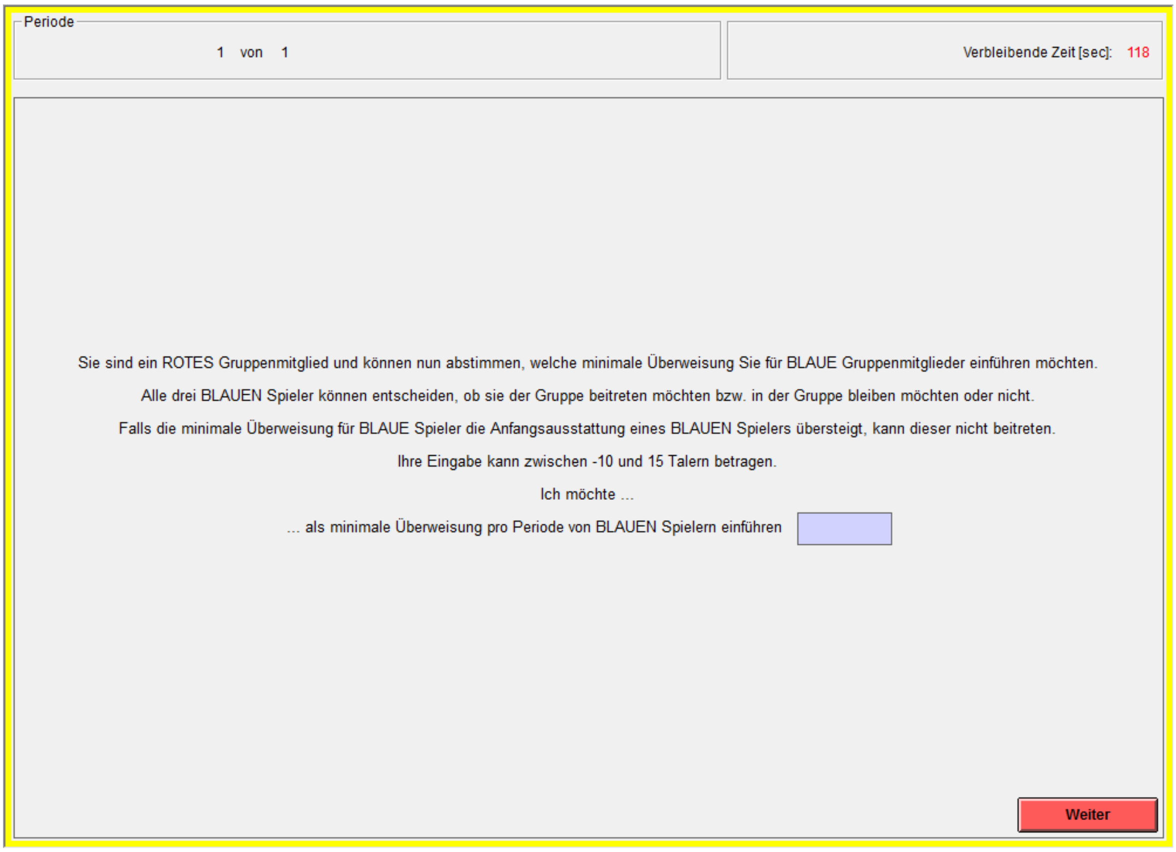
**(your initial endowment – your transfer) + 0.5*(the sum of Taler in the group account).**

Players who are **not group members** will receive their initial **private endowments** (that is 5, 10 or 15 Taler) in that period. Beyond the private endowment, those players do not receive any income from the group account.

At the end of periods 3, 6 and 9, RED players vote on a minimal transfer that BLUE players have to observe if they want to become or stay group members. Before RED players decide on their votes in private, they are able to chat with each other for three minutes. BLUE players do not participate in this chat and will not be able to read the messages. The chat screen:



The voting screen of RED players is the following:




Thereafter, BLUE players get the opportunity to become group members while observing the minimal transfer. Only BLUE players whose initial private endowment is equal or greater than the minimal transfer get the chance to become group members.

After BLUE players have made their decisions, all players, RED and BLUE, will receive information on how many players are members of the group from now on. The experiment then continues with the next period.

Before the experiment begins, all participants have to answer some control questions on the computer screen. These questions are designed to familiarize you with the rules of the experiment.

Do you have any questions concerning the experiment?

# 5 Deterrence Works for Criminals[48]

*"There is no distinctly native American criminal class except Congress."*

Mark Twain

## 5.1 Introduction

Since Gary Becker's (1968) seminal paper, criminal law and economics assume that welfare loss from crime can be minimized by optimal deterrence policy. The underlying assumption that individuals react to incentives and changes in expected payoffs constitutes the deterrence hypothesis. In his analysis, as in virtually all subsequent work, the major inputs for deterrence are punishment size (e.g. the magnitude of a fine) and probability of conviction (e.g. quality of public prosecutors, size of police force). Garoupa (1997) and Polinsky and Shavell (2007) offer overviews of theoretical extensions of Becker's work.[49] Surveys by Cameron (1988) and Levitt and Miles (2007) discuss several empirical economic studies on the identification of the relative deterrence benefits of punishment size and probability of conviction. Many empirical studies however have been criticized for their estimation techniques and issues of simultaneity, data collection and aggregation, and possible incapacitation bias (see Levitt and Miles 2007).

[48] I am grateful to the JVA für Frauen in Vechta and especially Petra Huckemeyer and Elsbeth Lübbe for dedicated cooperation and to the Kriminologischer Dienst in Lower Saxony for its permission to conduct this study.

[49] Examples are issues of marginal deterrence (Stigler 1970), the optimal use of the inputs (Polinsky and Shavell 1979), fines versus imprisonment (Polinsky and Shavell 1984), and the interdependence of inputs due to reasonable-doubt decision making (Andreoni 1991).

In fact the issue of obtaining individual level data, which is most useful to test theory of individual decision making, remains tedious.[50]

It is for this reason that this study aims at investigating individual (stealing) decision making in a ceteris paribus analysis, i.e. by application of experimental techniques to criminal law and economics, and more specifically to the deterrence hypothesis. A majority of recent experimental work provides evidence that an increase in expected punishment (i.e. a decrease in expected returns from crime) decreases criminal activity for students (e.g. Abbink et al. 2002, Anderson and Stafford 2003, DeAngelo and Charness 2012, Friesen 2012, Harbaugh et al. 2011, Nagin and Pogarsky 2003, Rizzolli and Stanca 2012, and chapter 6 of this dissertation).[51]

This study takes one step out of the conventional student lab. It is the first attempt to directly investigate stealing decisions of prison inmates in a lab-in-the-field experiment (also referred to as artefactual field experiment, Harrison and List 2004), and to compare them with those of students.[52] This innovation is important as criminals and inmates are a major target group of deterrence and habilitation policy. Hence, this study builds a bridge from findings in conventional laboratory experiments to the target group. Anderson (2002) argues that there are two necessary conditions for the deterrence hypothesis to hold for criminals: (1) sufficient information about probabilities and punishments and (2) rational behavior. Experimental methods allow us to create the case of complete information while varying the subject pool, so that we are able to test whether inmates' preferences differ systematically from a frequently studied subject pool, university students, with respect to their perception or processing of deterrence incentives. To the best of our knowledge, this chapter is the first study to disentangle these two conditions and yield insights into potential biases of criminals.

---

50 Some notable exceptions are Witte (1980) who solves this issue by using individual level data of former prisoners in North Carolina, USA, and Levitt (1997), Corman and Mocan (2000), as well as Fisman and Miguel (2007) who employ empirical techniques to disentangle simultaneity issues. Their findings support the deterrence hypothesis.

51 Criminal activity is defined as any activity that constitutes an action of harm against a non-consenting victim or breach of a set of explicitly defined rules or laws.

Note that Harel and Segal (1999) and Baker et al. (2004) are closely connected to DeAngelo and Charness (2012)'s investigation on the effect of uncertainty on deterrence.

52 Block and Gerety (1995) also analyze decision making under risk of prison inmates. Their auction experiment however does not involve stealing.

The second innovation analyzes crowding out of pro-social behavior. Textbook economics assumes that preferences and decisions of individuals are well-behaved in that they directly react to incentives, such as changes in prices for goods and expected punishment for crime, in a purposeful manner. There are many examples, especially in market environments, in which this assumption holds. Yet, over the course of the last decades behavioral economics has provided ample evidence that extrinsic incentives may backfire by crowding out intrinsic (pro-social) behavior (for surveys see Frey and Jegen 2001, Gneezy et al. 2011, Bowles and Polanía-Reyes 2012).[53]

In the case of deterrence incentives crowding out may occur as follows: Player 1 can steal money from a defenseless player 2. In situation A there are no institutions which set external, extrinsic incentives to curb stealing (i.e. to punish crime). Player 1 considers the moral implications[54] of stealing and may refrain from doing so. That is, in situation A player 1's focus is on her interaction with a defenseless player 2, such that player 1 internalizes the moral implications of her actions. Conversely, in situation B there is an institution that sets extrinsic incentives, i.e. if player 1 steals from player 2, she will be convicted and punished with some probability and fine. In situation B player 1 focuses on this institution and her strategy to escape punishment – the moral implications of her action rest on this institution. If the institution fails to punish player 1 for stealing, then player 1 is happy to get away with stealing and the institution is to blame. Hence, player 1 does not or to a lesser extent consider the moral implications of her actions, as the responsibility to protect the defenseless player 2 rests on the external institution. Heyman and Ariely (2004) refer to the two situations as a social (situation A) and a monetary market (situation B) while Meier (2007) refers to individuals as being either in a moral mode (A) or in an exchange mode (B).

Findings by Schulze and Frank (2003) and Schildberg-Hörisch and Strassmair (2012) with university students suggest that deterrence incentives indeed crowd out pro-social behavior. It is a common method in behavioral economics to measure pro-social behavior via dictator

53 For instance, Falk and Kosfeld (2006) provide experimental evidence for the hidden costs of control in a principle-agent type game, while Bénabou and Tirole (2006) formulate theory that finds one reason for crowding out of pro-social behavior in reputational concerns. In a field experiment Gneezy and Rustichini (2000a) show how extrinsic incentives by setting a fine permanently increase the frequency of violations of a pick-up rule in 10 day-care centers in Haifa, Israel.

54 Depending on player 1's moral concept, the moral implication of stealing may take different forms. For instance, if player 1 is Kantian and her moral sentiment is the categorical imperative, she will want to take an action such that this action should become a universal norm.

game giving. Our lab-in-the-field experiment includes dictator game giving as a later decision of students and inmates. Here we are able to test whether deterrence incentives in the earlier stealing game influence later dictator game giving. If crowding out of pro-social behavior occurs, we should find dictator game giving to be lower if following a situation with extrinsic deterrence incentives. Therefore the second innovation of this study is another bridge from the students' lab to the prisoners' lab in the field with regard to crowding theory.

Our experiment yields two major findings on the two innovations discussed above: with regard to the decisions analyzed in this chapter, we find that inmates and students do not behave systematically differently from each other, neither (1) with regard to deterrence incentives in the stealing game nor (2) with regard to crowding out of pro-social behavior in the dictator game. That is, we confirm that deterrence incentives work for the target group of crime policy and that this target group's pro-social behavior is crowded out in the dictator game. On the one hand, such findings are good news as they provide external validity of former results on students in the literature. The results of this chapter do not support Anderson's (2002) concerns that inmates may be systematically irrational. On the other hand, crowding out of pro-social behavior is not merely an issue with students, but equally with inmates.

The remainder of this work is divided into three parts. Section 2 reports the experimental design, behavioral hypotheses and procedures of the experiment. Results are presented in section 3. A concluding discussion completes the study.

## 5.2 Experimental Design

In this work we employ a 2x2 experimental design. The first dimension of interest is the comparison of economic behavior of students (_S treatments) and prisoners (_P treatments) in two games. The first game is the stealing game and the subsequent, second game is the standard dictator game. Both games will be described in detail below. The second dimension of interest is the deterrence scheme; we examine whether the behavioral patterns of these two groups change in the two games when (1) stealing is not fined (NoDeter_ treatments) and (2) stealing is fined in fifty percent of the time with some known fine (DeterFine_ treatments). Table 1 depicts the experimental design. Let us turn to the two games.

**Table 1**. Experimental Design.

| | | Subject Population | |
|---|---|---|---|
| | | Students | Prisoners |
| Deterrence Scheme in the Stealing Game | No Deterrence | NoDeter_S | NoDeter_P |
| | Deterrence by an Increasing Fine | DeterFine_S | DeterFine_P |

### *5.2.1 The Stealing Game*

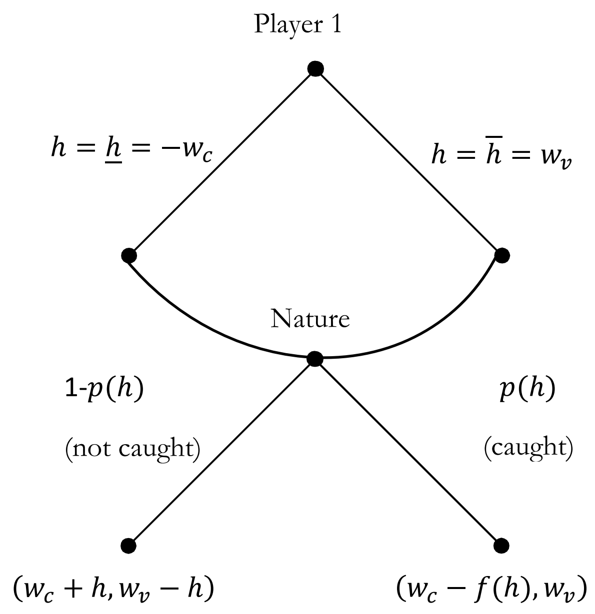
The structure of our stealing game is depicted in Figure 1. It closely corresponds to the game of Schildberg-Hörisch and Strassmair (2012), yet differs from their calibration in some important aspects (that will be described below). In the remainder of this chapter we will always refer to player 1 as *criminal* and player 2 as *victim*; note that our chapter applies this wording for easy reading while our experiment instructions (which can be found in the appendix) and programs use a strictly neutral language. The endowments of potential criminals and victims are unequal: each criminal is endowed with $w_c$ while each victim is endowed with $w_v$, with $w_v > w_c$. Potential criminals can then decide to take from or give to potential victims by choosing haul *h*, with $h \in \{-w_c, \dots, w_v\}$. Potential criminals hence have the opportunity to choose no allocation change or to give to potential victims. All giving and neutral decisions, i.e. $h \leq 0$, are certain. Depending on the deterrence scheme, a taking decision, i.e. $h > 0$, triggers a positive probability of conviction $p(h)$ and a possible fine $f(h)$ if convicted.[55] After a criminal has made her decision, the deterrence scheme at hand either punishes her with fine *f* (with probability *p*) or lets her keep *h*.

In our experimental design we set the endowments at $w_c = 2$ Taler and $w_v = 10$ Taler, and the stealing increment is 1 Taler. We employ two deterrence schemes to test the deterrence hypothesis. Scheme *NoDeter_* resembles a risk free environment and is comparable to a manipulation of the dictator game. Hence, in *NoDeter_* $f = 0$ and $p = 0$. *NoDeter_* is needed as a baseline against all risky decision making. In the *DeterFine_* treatments, criminals face the risk of being fined for stealing, i.e. when choosing some $h > 0$. The probability of conviction in *DeterFine_* reads $p(h) = \frac{1}{2}\ if\ h > 0$ and $p(h) = 0\ if\ h \leq 0$. Fine *f* is increasing in *h*, that is: $f = 1.25 - 0.25h + 0.1h^2$. The probability of conviction *p* is held constant at

55 Note that either *p* or *f* needs to be a function of and increasing in *h*, so that no issues of marginal deterrence arise. Stigler (1970) discusses marginal deterrence issues: marginal deterrence theory shows that the combination of *p* and *f*, i.e. the expected punishment, should fit the crime. For instance, if the expected punishment is the same for stealing a bike and stealing a car, a potential criminal might as well go for the car.

$p_{DeterFine} = 0.5$. Note that our stealing game includes two innovations from Schildberg-Hörisch and Strassmair's setup in that (1) we allow for stealing *and* giving so as to avoid suggestiveness of the action space (motivated by findings of List 2007 and Bardsley 2008) and (2) that our calibration offers an *interior* solution for stealing under risk (i.e. in *DeterFine_*).

**Figure 1.** The Stealing Game.




Corresponding to the description and Figure 1 above, a potential criminal *i* maximizes expected utility

$$\max_{h_i} U_i = \big(1 - p(h_i)\big)u_i(w_i + h_i) + p(h_i)u_i\big(w_i - f(h_i)\big)$$

where $u_i(\cdot)$ denotes the Bernoulli utility function. Note that in *NoDeter* criminals do not face any risk, such that the maximization problem collapses to $\max_{h_i} u_i = w_i + h_i$. Therefore, the standard game theoretic prediction, independent from any kind of risk preferences a criminal may hold, is that she chooses the maximum haul, $h^*_{NoDeter} = \overline{h} = w_v = 10$ Taler. In *DeterFine* a criminal's risk preferences play a role. Table 2 reports the deterrence incentives, fine if convicted *f* and probability of conviction *p*, for every action *h*. If the criminal is risk neutral and does not hold other-regarding preferences, she will maximize her expected payoff by stealing $h^*_{DeterFine} = \frac{25}{4}$ which in the increments available rounds to 6 Taler. It is an important feature of our design that we do not predict no stealing in *DeterFine*. Recall also that the optimal haul in *NoDeter* is $h^*_{NoDeter} = \overline{h} = 10$ Taler. We can formulate

**Hypothesis 1.** *The deterrence hypothesis holds, i.e.* $h_{NoDeter} > h_{DeterFine}$.

**Table 2.** Overview of the Deterrence Schemes.

| *h* | Deterrence Scheme | | | | | |
|---|---|---|---|---|---|---|
| | *NoDeter* | | | *DeterFine* | | |
| | *f* in EUR | *P* | $\pi$ in EUR | *f* in EUR | *p* | E($\pi$) in EUR |
| -2 | 0 | 0 | -2 | 0 | 0 | -2 |
| -1 | | | -1 | | | -1 |
| 0 | | | 0 | | | 0 |
| 1 | | | 1 | 1.10 | 0.5 | -0.05 |
| 2 | | | 2 | 1.15 | | 0.425 |
| 3 | | | 3 | 1.40 | | 0.80 |
| 4 | | | 4 | 1.85 | | 1.075 |
| 5 | | | 5 | 2.50 | | 1.25 |
| 6 | | | 6 | 3.35 | | *** 1.325 *** |
| 7 | | | 7 | 4.40 | | 1.3 |
| 8 | | | 8 | 5.65 | | 1.175 |
| 9 | | | 9 | 7.10 | | 0.95 |
| 10 | | | *** 10 *** | 8.75 | | 0.625 |

Note: Fine *f* is a function of stealing *h*, that is: $f = 1.25 - 0.25h + 0.1h^2$. Accordingly, an individual *i* with risk neutral, narrowly self-interested preferences maximizes her utility by choosing *h**=6.25≈6.

Meier (2007) provides an overview of other-regarding, social preferences theories. For instance, theories of inequality aversion (Fehr and Schmidt 1999, Bolton and Ockenfels 2000) provide predictions for behavior in *NoDeter*. Here equal splits are important: in inequality aversion models the equal split constitutes a decision of individuals who hold a sufficiently large preference for equality. For instance, in *NoDeter* we predict that if a criminal holds a sufficiently large preference for equality, then she will steal a haul *h* such that $\pi_{criminal} = \pi_{victim} = \frac{w_c + w_v}{2} = 6$ Taler.[56] It is not straightforward to formulate hypotheses on the interaction of social preferences and risk preferences (for instance, see Cappelen et al. 2013

[56] Additionally, both players were endowed with 7 Taler (which were needed to rule out negative payoffs). These could neither be stolen nor given away.

and Brock et al. 2013 on this matter) for treatment *DeterFine*. In accordance with Frey and Jegen (2001), Gneezy et al. (2011), Bowles and Polanía-Reyes (2012) we assume that external, extrinsic incentives may crowd out intrinsic, pro-social behavior. If this is the case, then crowding out of pro-social behavior works against Hypothesis 1. The intuition is that $h_{NoDeter}$ may be smaller than expected as a criminal (i.e. a player 1) may steal *less* from a victim due to intrinsic pro-social motivation. If we nevertheless identify evidence in support of Hypothesis 1, then deterrence works despite crowding out of pro-social motivation in the stealing game. Consequently, to identify whether crowding out of intrinsic, pro-social motivation occurs at all, we need a second decision without a strategic or risky environment: the dictator game.

### *5.2.2 The Dictator Game*

In our experiment the dictator game succeeded the stealing game (about 10 minutes later) and was set up in standard terms.[57] Originally designed by Kahneman et al. (1986), the dictator game is commonly employed "[…] to provide insights into the prevalence of social preferences." (List 2007, p. 482). That is, a dictator was endowed with $w_d = 10$ Taler and able to give $g$ Taler to the recipient, with $g \in \{0, \dots, w_d\}$. The recipient's endowment was $w_r = 0$ and 1 Taler was the action set increment. As strategic considerations do not matter in dictator game giving, any $g > 0$ is associated with other-regarding, pro-social preferences. These may originate from inequality aversion, altruism (Andreoni and Miller 2002, Levine 1998) or warm-glow of giving (Andreoni 1990). List (2007) and Bardsley (2008) extend the dictator game literature by providing evidence that giving may not be easily interpretable. In this study we are not concerned with the identification of the kind of intrinsic motivation and need but one assumption on dictator game giving: we assume that *ceteris paribus* crowding out of pro-social behavior can be detected by a *change* in dictator game giving. If this is the case and if extrinsic deterrence incentives in the earlier stealing game indeed crowd out intrinsic pro-social behavior, we should see that on average dictator game giving is smaller if followed by *DeterFine* compared to *NoDeter*. We formulate

**Hypothesis 2**. *Pro-social behavior is crowded out by deterrence incentives, i.e.* $g_{NoDeter} > g_{DeterFine}$.

Let us now turn to the procedures of our experiment.

---

57 Feedback on the outcomes of the decisions was only provided after decision making for all stages had finished. Subjects only received information on the content of the present stage. The games and decision making of later stages were unknown. Please consult section 2.3 for more information on experiment procedures.

### *5.2.3 Procedures*

We employ a design to analyze stealing decisions and crowding out of pro-social behavior within and between the two groups of our experiment, i.e. female university students and female inmates. The stealing game and the dictator game discussed in this chapter are part of a sequence of games that we analyze in a series of companion papers. All decisions analyzed in this chapter were made one-shot without feedback. For simple handling all decisions in the experiment were made with the click of a mouse and subsequent confirmation of the choice. The sequence of games was the following: 1) the stealing game with or without deterrence incentives (*discussed in this chapter*), 2) either a simultaneous or sequential Prisoner's Dilemma (Khadjavi and Lange 2013), 3) a lottery (based on Dave et al. 2010), 4) a dictator game (*discussed in this chapter*), and 5) a task to measure individual discount rates.[58] Subjects learnt about the content of the next stage only after the current stage was over. All stages were paid out.

The five sessions in the field were conducted at *JVA für Frauen* (*penitentiary for women*) in Vechta, Germany, on June, 14 2012.[59] It is the central women's prison in the German state of Lower Saxony (in North-Western Germany, bordering the City of Hamburg) with a capacity for about 150 detained women in its main section. Inmates serve sentences ranging from a couple of days to life sentences, with an average detention time period of 6 months. It offers facilities for mother-and-child care and social therapy as well as secondary school education and apprenticeship opportunities such as cooking and painting. Inmates are able to earn money for their release savings and phone accounts by sewing for a company that produces bags within the prison grounds. About fifty percent of women are convicted for drug-related crimes.

Three weeks before the lab-in-the-field experiment we distributed invitations for participation in the study. The invitation included information on the nature of the experiment, that participation is strictly voluntary, and that we guarantee absolute anonymity. In German prisons it is not possible to earn cash, so we needed to apply an alternative payment vehicle. The state institution which was responsible for permitting our study suggested using coffee, tobacco or phone credit as alternatives. Consequently participants were able to earn either a

---

58 An analysis of order effects with respect to the different games that were played without feedback cannot reject the null hypothesis that the treatment effects in the dictator game (game 4) were independent from the second and third game. We are therefore able to safely analyze our data.

59 Further information (in German) on the JVA für Frauen in Vechta: www.jva-fuer-frauen.niedersachsen.de.

jar of instant coffee (200 grams) or a pack of tobacco (40 grams) as show-up fees. This choice was made at the very end of the experiment, individually and unobserved by other subjects. The monetary value of both show-up fee options was approx. 5 EUR. We ordered coffee and tobacco from the prison's delivery company and both the brand of coffee and of the tobacco were the most commonly used in the prison. Participants were informed that they will receive individual phone credit depending on their decisions in the study. This way we were able to create a quasi-continuous payoff space for the experiment decisions. The day before the experiment we set up a mobile computer lab with 20 terminals, i.e. laptops and separation walls, in the prison's gymnasium.

The four conventional lab sessions were conducted in the experimental economic laboratory of the Faculty of Economic and Social Sciences at the University of Hamburg three weeks later, in the first week of July, 2012. Female students for our lab sessions were recruited from the subject pool via ORSEE (Greiner 2004). Our emphasis was on including a number of subjects with no and little experiment experience. Similar to the inmates in the prison, subjects in the conventional lab received information that they could receive earnings by making anonymous decisions on computer terminals. Sessions lasted about 45 minutes each. The payment vehicle in the lab was the usual: a sum of cash consisting of a 5-EUR show-up fee and money determined by the decisions made in the experiment.

Both in the conventional lab and the prison lab, the sessions proceeded as follows. Once the participants were seated, a set of general instructions was handed out and read out loud by the experimenter. Participants were able to raise their hand and ask questions in private. Then instructions for task 1 (the stealing game) were distributed. The instructions were again read out loud and questions were answered in private. All participants then made their decisions on computer screens. Note that all participants made their decisions as player 1 (the player who can steal from the other player). After all decisions were made, instructions for task 2 followed. This process continued until the end of a session. It was only at the end of the experiment that subjects were informed about the outcome of the games and randomly selected as being either player 1 or player 2. All games were paid out. We programmed the experiment in a way that decisions could be made with a couple of simple clicks of the computer mouse. Participants had to click the check box of their choice and then confirm it by clicking an 'Okay-Button'. We deliberately aimed for simplicity to ensure user-friendliness.

90 female inmates participated in our lab-in-the-field sessions, while 92 female students participated in our conventional lab sessions, yielding a total of 182 subjects. No subject

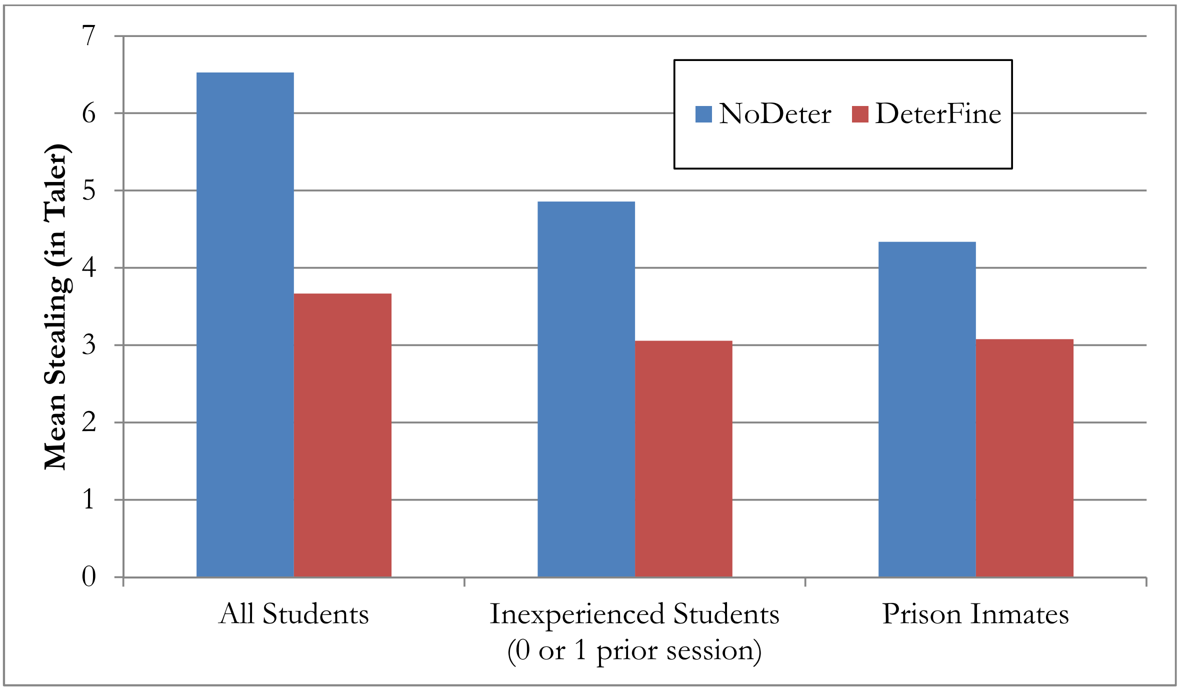
participated in the experiment twice. We used *Taler* as the experimental currency, with 1 EUR equal to 5 Taler. Average earnings in the stealing game were 12.76 Taler in the lab and 12.79 Taler in the field. Including the show-up fees, average earnings in the field were approx. 14.40 EUR compared to approx. 13.20 EUR in the lab. Note that these earnings depend on decisions in all five games of the experiment. We used z-Tree (Fischbacher 2007) to program and run our experiment.

## 5.3 Results

We had 40 students and 38 inmates in our *NoDeter* treatment and 52 students and 52 inmates in the treatment with deterrence, *DeterFine*. Table 3 reports summary statistics and Figure 2 provides an overview of stealing and dictator game giving in the conventional lab and the prison lab. Students decided to steal an average amount of 6.53 Taler in *NoDeter_S* compared to 3.67 Taler in *DeterFine_S*. A Mann-Whitney test rejects the null hypothesis with $p < 0.0000$, i.e. we confirm findings of the literature that deterrence works in the conventional lab with (female) students. In the prison inmates on average took 4.34 Taler in the *NoDeter_P* treatment and 3.08 Taler in *DeterFine_P*. This difference is also statistically significant at $p = 0.0643$ (Mann-Whitney test). Hence, we formulate that

**Result 1.** *Deterrence works in the conventional lab and in the prison lab, i.e. both students and inmates steal significantly less with punishment incentives in place.*

**Figure 2.** Mean Stealing of Students and Prison Inmates.



On the one hand, this finding is assuring for experimental economists who employ student subjects. It shows that deterrence works with the target group of such policy in similar ways like in the conventional lab. Hence, on average inmates appear to process the information like students and as expected by Becker (1968). On the other hand, comparing stealing of students and inmates, i.e. 6.53 Taler vs. 4.34 Taler in *NoDeter* and 3.67 Taler vs. 3.08 Taler in *DeterFine*, we also find differences (Mann-Whitney tests, p = 0.0188 in *NoDeter* and p = 0.1103 in *DeterFine*).

Do these results show that students are more prone to stealing? We try to answer this question by employing linear regressions that include socio-demographic information (reported in Table 4).

**Table 3.** Summary Statistics of the Lab-and-Field Experiment.

| Variable | Lab | Prison[a] |
|---|---|---|
| N | 92 | 90 |
| NoDeter | 40 | 38 |
| DeterFine | 52 | 52 |
| **NoDeter** | | |
| Mean Stealing | 6.53 Taler | 4.34 Taler |
| Mean Dictator Game Giving | 2.95 Taler | 3.92 Taler |
| **DeterFine** | | |
| Mean Stealing | 3.67 Taler | 3.08 Taler |
| Mean Dictator Game Giving | 2.52 Taler | 3.25 Taler |
| Mean Age (in years) | 23.35 | 31.43 |
| Mean Education[b] | 6 (all in top category) | 1.47 |
| Mean Experiment Experience (in sessions) | 8.33 | -none- |
| Rate of Married Subjects (in %) | 6.52 | 16.67 |
| Rate of Subjects with Children (in %) | 4.35 | 59.72 |

Note: (a): Ex-post demographic questionnaire answers are missing for 18 subjects in the prison due to an IT problem at the end of one session ('fortunately' only after the decision-making part of the experiment).
(b): The German education system is rather complicated and comparing education quantitatively is tedious. We transferred questionnaire data on education into qualitatively ascending categories which are roughly in accordance with years of schooling: no school diploma = 0, 9-year high school diploma = 1, 10-year high school diploma = 2, 13-year high school diploma = 3, high school diploma plus a completed apprenticeship = 4, completed master craftsman diploma = 5, (some) academic education = 6.

While regression I reports that subjects in the prison steal less, regression II shows that the prison impact turns insignificant when including control variables. Especially 'experiment experience', i.e. the number of times a subject had participated in other experiments, appears to have a positive influence on stealing.[60] The finding that experiment experience influences behavior may not be surprising (see e.g. Smith 1994); yet it is important that we control for it.[61] Note that the joint effect in regression I of the coefficients "DeterFine" and "Prison x DeterFine" is significantly different from zero and negative at F (1, 178) = 3.64, p = 0.0581. This test statistic offers further robustness for Result 1, namely that deterrence works both for students and inmates.

**Table 4.** OLS Regressions Estimating Drivers of Stealing and Giving by Students and Inmates.

| Independent Variable | Dependent Variable: Stealing | | Dependent Variable: Dictator Giving | |
|---|---|---|---|---|
| | I | II | III | IV |
| Prison (dummy) | -2.183*** (0.704) | -1.992 (1.754) | 0.356 (0.514) | 0.619 (1.036) |
| *DeterFine* (dummy) | -2.852*** (0.654) | -2.727*** (0.558) | -2.247*** (0.756) | -2.345*** (0.850) |
| Prison x *DeterFine* (dummy) | 1.587* (0.931) | 1.723 (1.134) | 0.371 (0.673) | 0.899 (0.789) |
| Stealing (continuous) | | | -0.282*** (0.079) | -0.299*** (0.096) |
| Stealing x *DeterFine* (continuous) | | | 0.276** (0.107) | 0.293** (0.121) |
| Age (continuous) | | -0.038 (0.042) | | 0.003 (0.023) |
| Education (continuous) | | 0.023 (0.243) | | 0.171 (0.137) |
| Married (dummy) | | -0.364 (0.962) | | -0.708 (0.576) |
| Mother (dummy) | | 0.914 (0.852) | | 0.219 (0.546) |
| Experiment Experience (continuous) | | 0.059** (0.027) | | 0.009 (0.029) |
| Constant | 6.525*** (0.491) | 6.698*** (1.548) | 4.788*** (0.641) | 3.754*** (1.013) |
| N | 182 | 164 | 182 | 164 |

Note: *NoDeter* and 'student lab' are the baselines in all estimations. Robust standard errors in parentheses, significance: *p < 0.10, **p < 0.05, ***p < 0.01.

60 Note that nearly 30 percent of our student subjects had no or little (0 or 1 prior sessions) of experiment experience. Hence, we do not predict out-of-sample.

61 Another explanation may be a selection effect, such that subjects who participated in experiments more often are different from subjects who participated infrequently. See Falk et al. (2011) and Cleave et al. (2013) for more information self-selection into experiments.

Let us now turn to the decisions in the dictator game. Our aim is to test Hypothesis 2, i.e. whether deterrence incentives in the earlier stealing game (treatment *DeterFine*) impact dictator game giving negatively. Indeed we find that dictator giving is smaller both in the lab (2.95 Taler in *NoDeter_S* compared to 2.52 Taler in *DeterFine_S*) and the prison (3.92 Taler in *NoDeter_P* compared to 3.25 Taler in *DeterFine_P*). Overall, dictator game giving following *DeterFine* is lower than following *NoDeter*, (Mann-Whitney test, p = 0.0775). Regressions III and IV analyze dictator game giving further. Here we also include stealing and an interaction term of stealing and the *DeterFine* dummy as independent variables. Both estimations do not reject the null hypothesis that dictator game giving of students and prison inmates is significantly different. We find that the coefficient for *DeterFine* is significantly different from zero at the 1 percent level and negative (-2.247 Taler in specification III). An F-test confirms that the joint effect of the coefficients "DeterFine" and "Prison x DeterFine" remains significantly different from zero at F (1, 176) = 9.32, that is p = 0.0026. This analysis allows us to formulate

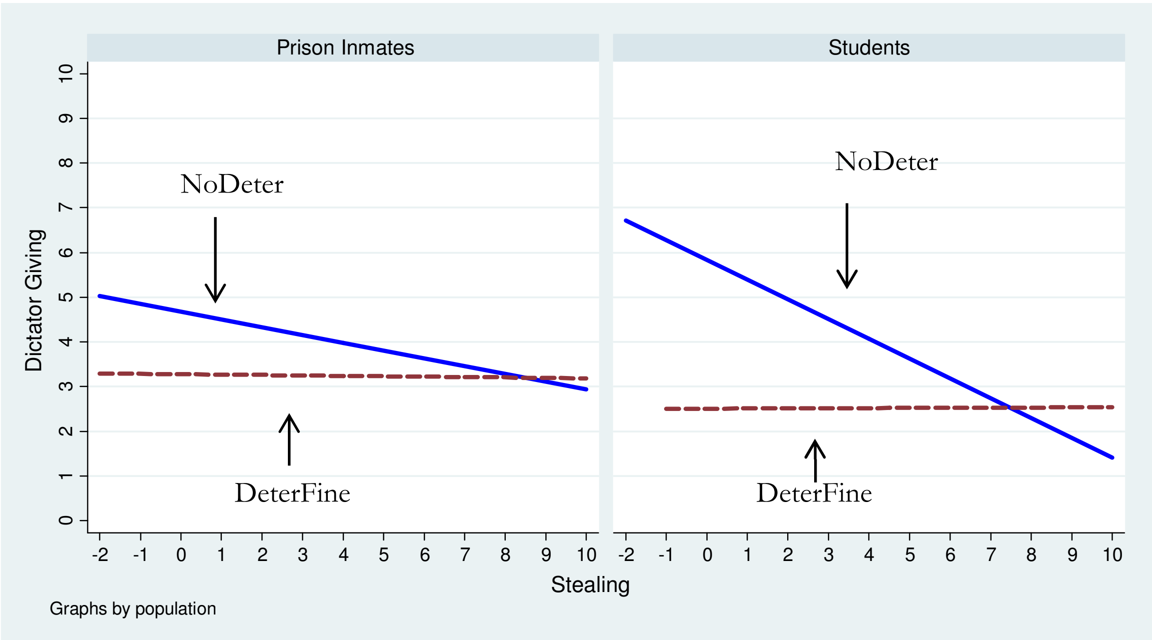
**Result 2.** *Crowding out of pro-social behavior occurs in the conventional and the prison lab, i.e. both students and inmates give less in the dictator game with deterrence incentives in place in the earlier stealing game.*

Further, these estimations help us to find a peculiar correlation: stealing and dictator game giving are inversely related, yet only in treatment *NoDeter*. This finding refutes the idea that an income effect might be present between the two tasks. Rather, the less a subject steals in the stealing game, the more she will give in the dictator game later. This correlation suggests consistent social preferences between the two games. In *DeterFine* this effect is nullified as shown by the interaction term. That is, we find consistent decision making when comparing pro-social behavior in the stealing game and in the dictator game in *NoDeter*, i.e. in both games in which the criminal and the dictator do not face incentives trying to curb selfish behavior. Conversely this consistency is absent in *DeterFine*, i.e. when deterrence incentives are in place. To get a clearer picture of the results, we plot linear fits of correlations between mean stealing and dictator game giving for *NoDeter* and *DeterFine* in Figure 3.

These graphs reflect the estimations for dictator game giving in Table 4. Note that dictator game giving is not absent in *DeterFine*. A possible explanation for this finding is that there is one part of intrinsic, pro-social motivation that is crowded out by the deterrence incentives. There appears to be yet another part of pro-social motivation that is robust against crowding out. In the restricted time and organizational frame of our lab-in-the-field experiment with prison inmates it was not possible to elicit pro-social behavior in order to sort subjects into groups corresponding to different theories of social preferences. We therefore refrain from speculating about the nature of different parts of pro-social motivation. Such an investigation may serve as an avenue for future research.

Returning to Result 2, i.e. that crowding out is present for both students and prison inmates, this study delivers good news for experimental economics: it provides external validity of findings in studies like Schulze and Frank (2003) and Schildberg-Hörisch and Strassmair (2012) who also find crowding out of pro-social behavior in the lab with students. Yet, it also calls for a more complex law and economic policy; it implies that policies and laws regarding criminal activity need to consider that incentives may backfire.

**Figure 3.** Linear fit of Correlations between Stealing and Dictator Giving, by Subject Population.




## 5.4 Conclusion

This study presents two innovations: first, in line with current research on the external validity of conventional laboratory experiments (e.g. Alevy et al. 2007, Benz and Meier 2008, Stoop et al. 2012, Stoop 2013), it tests the effectiveness of deterrence incentives for prison inmates. This aim is of central importance for law and economic policy. If this target group of policy does not react to incentives in the desired way and reduces criminal activity – a central assumption in all deterrence policy – then this policy may be ineffective to reduce crime. Second, it examines whether deterrence incentives crowd out pro-social behavior of students and inmates. Here we examine whether deterrence incentives may carry a detrimental effect.

Our results provide insights into the deterrence hypothesis and crowding out of pro-social behavior. We confirm that deterrence incentives mitigate stealing behavior of inmates. Anderson (2002) discusses that both lack of information on the incentive structure and systematical biases may cause the deterrence hypothesis to fail for criminals. This is one of the reasons why the randomized, controlled and (sometimes) counterfactual approach of experimental research is useful here. In our experiment inmates receive complete information about the deterrence incentives – and they react to these incentives. Our results therefore do not support the conjecture of criminals' decision making to be systematically biased and ignorant concerning deterrence incentives.

Like Chmura et al. (2010) and Birkeland et al. (2011) who examine dictator game giving of male inmates, we find that female inmates in our lab-in-the-field experiment exhibit similar dictator game giving compared to students. Hence, we provide additional evidence that criminal behavior is not a consequence of a self-selection of individuals who are structurally more selfish. Our experiment further reports that deterrence incentives backfire for both students and prison inmates. Consequently, the impact of laws to regulate and punish certain behavior may crowd out intrinsic incentives to behave lawfully. This finding is important as it suggests a more careful practice with new laws and their impact on behavior. We hope that this study is able to motivate further research on the specific behavior of important groups for policy, such as prison inmates, as well as on extrinsic and intrinsic motivation and decision making.

## 5.5 Appendix A: English Translation of the Experimental Instructions

| General Instructions for Participants |
|---|

**Welcome to the Study!**

Thank you for participating in our study today. You will be able to earn a considerable amount of money. It is therefore important that you read these instructions carefully.

**It is prohibited to communicate with other participants during the study.** Should you have any questions please raise your hand and an instructor will come to answer them. If you violate this rule, we will have to exclude you from the study and from all payments.

During the experiment you will make decisions **anonymously**, other participants will not learn about your decisions.

In any case you will earn *[Field: a pack of coffee or a pack of tobacco; Lab: 5 Euros]* for participation in this experiment. The additional earnings depend on your decisions *[Field: and will be paid to your phone account]*. During the study your earnings will be calculated in Taler. At the end of the experiment your earned Taler will be converted into Euros at the following exchange rate:

**1 Taler = 0,20 €.**

**The study consists of five independent tasks.** Your decision in a task does not have any impact on the other tasks.

The instructions for the five tasks will be handed out one after another. You will first receive instructions for task 1 and then make your decision at the computer terminal. After this task 1 is done.

Thereafter you will receive instructions for task 2 and again make your decision at the computer terminal. This procedure continues until the end of the study. In the end we will also ask you to answer some general questions.

At the end of the study you will receive your payment. Your payment is the sum of payments from all five tasks. All earned Taler will be converted to Euros and paid to you *[Field: in addition to a pack of coffee or tobacco]*. Hence you will get

**Your total payment**

**=**

**Your payment from task1, 2, 3, 4 and 5 + *[Field: a pack of coffee or tobacco, Lab: 5 EUR]***

All payments will be done separately, without any other participant being able to see what you have earned. Apart from the instructor nobody will know what you have earned.

*[We do not include instructions for tasks 2, 3 and 5 in this appendix. They will be included in companion papers on our prison study and are available upon request. Readers interested in the original German instructions may also contact us.]*

## Instructions for Task 1

*[The Stealing Game, NoDeter Treatment:]*

In task 1 you need to decide how many Taler to transfer between your account and the account of another participant who is randomly matched with you. You will not receive information on the identity of the other participant. Neither will the other participant receive information on your identity.

At the end of the experiment one half of the participants will be randomly selected to be person 1, the other half will be person 2. Your own decision will count if you are selected as person 1. If you are selected as person 2, you will receive your payment in accordance with the decision of a person 1 who is matched with you.

**There are 9 Taler in your account while the account of the other person holds 17 Taler.**

**You are able to decide freely how many Taler you would like to transfer between the account of the other person and your own account.** That is, you are able to transfer up to 2 Taler from your account to the account of the other person (transfer “-2”) or up to 10 Taler from the account of the other person to your account (transfer “10”).

For example, if you decide to transfer -2 Taler, you will receive 7 Taler in this task while the person who is matched with you will receive 19 Taler. If you decide to transfer 7 Taler, then you will receive 16 Taler in this task and the matched person will receive 10 Taler.

In summary, your transfer between the account of the other person and your own account may be between “-2” and “10”.

Please make a decision on the screen to decide how many Taler to transfer to your account from the account of person 2.

Instructions for Task 1

*[The Stealing Game, DeterFine Treatment:]*

In task 1 you need to decide how many Taler to transfer between your account and the account of another participant who is randomly matched with you. You will not receive information on the identity of the other participant. Neither will the other participant receive information on your identity.

At the end of the experiment one half of the participants will be randomly selected to be person 1, the other half will be person 2. Your own decision will count if you are selected as person 1. If you are selected as person 2, you will receive your payment in accordance with the decision of a person 1 who is matched with you.

**There are 9 Taler in your account while the account of the other person holds 17 Taler.**

**You are able to decide freely how many Taler you would like to transfer between the account of the other person and your own account.** That is, you are able to transfer up to 2 Taler from your account to the account of the other person (transfer "-2") or up to 10 Taler from the account of the other person to your account (transfer "10").

If you decide to transfer Taler from the account of the other person to your own account, this transfer will only be successful half of the times, i.e. the probability of success is 50 %. If the transfer is not successful, you will have to pay a fee that is shown on your screen.

For example, if you decide to transfer -2 Taler, you will receive 7 Taler in this task while the person who is matched with you will receive 19 Taler. If you decide to transfer 7 Taler and your transfer is successful, then you will receive 16 Taler in this task and the matched person will receive 10 Taler. If your transfer is not successful, then you will receive the initial 9 Taler minus the fee and the person who is matched with you keeps 17 Taler.

In summary, your transfer between the account of the other person and your own account may be between "-2" and "10".

Please make a decision on the screen to decide how many Taler to transfer to your account from the account of person 2. Mind the respective fee in case your transfer is not successful.

## Instructions for Task 4

*[The Dictator Game:]*

In task 4 you need to decide how many Taler to transfer between your account and the account of another participant who is randomly matched with you. You will not receive information on the identity of the other participant. Neither will the other participant receive information on your identity.

At the end of the experiment one half of the participants will be randomly selected to be person 1, the other half will be person 2. Your own decision will count if you are selected as person 1. If you are selected as person 2, you will receive your payment in accordance with the decision of a person 1 who is matched with you.

**There are 10 Taler in your account while the account of the other person holds no Taler.**

You need to decide how many Taler you would like to transfer between the account of the other person and your own account. You can choose the transfer freely, i.e. at the extremes you can keep all or transfer all. For example, if you decide to transfer 2 Taler, you will receive 8 Taler in this task while the person who is matched with you will receive 2 Taler. If you decide to transfer 8 Taler, then you will receive 2 Taler in this task and the matched person will receive 8 Taler.

You can decide freely how many Taler (between 0 and 10) to transfer between your account and the account of the other person.

Please click on the screen how many Taler you want to transfer between your account and the account of the other person.

# 6 On the Interaction of Deterrence and Emotions

*"The advantage of the emotions is that they lead us astray."*
Oscar Wilde

## 6.1 Introduction

Prevention, detection and prosecution of crime involve major expenses worldwide. In the United States, the total expenditure for the legal system amounted to over 260 billion US dollars in 2010 (Kyckelhahn and Martin 2013).[62] Law and economics deterrence theories commonly build upon the seminal contribution by Gary Becker (1968) that incentives can be put to work in order to reduce crime. This so-called deterrence hypothesis has been tested and mostly supported based on evidence by empirical studies (e.g. Corman and Mocan 2000, Fisman and Miguel 2007, Levitt 1997, Witte 1980), lab experiments with students (e.g. Abbink et al. 2002, Anderson and Stafford 2003, DeAngelo and Charness 2012, Friesen 2012, Harbaugh et al. 2011, Nagin and Pogarsky 2003, Rizzolli and Stanca 2012, Schildberg-Hörisch and Strassmair 2012, Schulze and Frank 2003, and chapter 5 of this dissertation) as well as in a lab-in-the-field experiment with criminals (in chapter 5 of this dissertation).

However there is also a dark side to such external and extrinsic deterrence incentives. Gneezy and Rustichini (2000a), Schildberg-Hörisch and Strassmair (2012), Schulze and Frank (2003) and chapter 5 of this dissertation find evidence for detrimental effects of (deterrence) fines in

[62] These estimates were reported on July, 1 2013 and are preliminary. Split into federal, state and local governments' expenditures, the shares of total expenditures were 19 percent, 30.5 percent and 50.5 percent respectively.

laboratory, lab-in-the-field and field experiments. Surveys by Bowles (1998), Frey and Jegen (2001), Gneezy et al. (2011) and Bowles and Polanía-Reyes (2012) discuss situations in which extrinsic incentives hold undesired side effects: they change the perceptions or feelings of individuals for the situation they are in. Consequently, intrinsic pro-social motivation of some individuals is crowded out and part of the desired effect is offset. Heyman and Ariely (2004) describe this process as individuals feeling to be either in a 'social market' (no extrinsic incentives) or in a 'monetary market' (when such extrinsic incentives are present). Meier (2007) labels these two situations 'moral mode' and 'exchange mode' respectively.

This chapter aims at investigating one possible underlying mechanism of crowding out: changes of emotions. While it is well established that crowding out occurs, the reasons why exactly individuals change their behavior remain unclear. Is it that individuals simply put on their 'moral mode hat' in some situations and replace it with an 'exchange mode hat' in others? If so, is this transition frictionless and purely rational, or is it accompanied by a change in emotions?

In this chapter we analyze whether subjects whose intrinsic motivation is crowded out by extrinsic incentives exhibit different emotions compared to subjects who do not face extrinsic incentives. To this end we employ a laboratory experiment and use a combined within and between subjects design. Our workhorse is the two-player stealing game (see also Schildberg-Hörisch and Strassmair 2012): player 1 is the potential criminal and makes the decision whether, and if so how much, to steal from player 2, or to abstain from stealing. Player 2 does not make a decision and only receives information on the stealing decision of player 1 and her success. There may or may not be an external deterrence institution present.[63] If this institution is present, it will detect and punish stealing by player 1 with some probability and fine in accordance with a deterrence scheme which is known to both players.

The procedure of our experiment is as follows: first, depending on the treatment, player 1 either faces extrinsic deterrence incentives or she does not. If present and depending on their bite, these incentives should work to reduce player 1's stealing magnitude (or force her to refrain from stealing entirely). Yet these deterrence incentives may also make player 1 change from the 'moral mode' into the 'exchange mode'. Second, we elicit subjects' emotions via self-reports. Here we may find that different treatments invoke different emotions. Third, subjects

[63] If the external deterrence institution is not present, this game is similar to the dictator game in the taking domain. See Bardsley (2008) and List (2007) for dictator games with giving and taking domains.

face a regime change (from no deterrence to deterrence or vice versa) which tests whether crowding out is present. We then test whether certain clusters of emotions, invoked by extrinsic incentives, explain the stealing decision in the old regime and whether crowding out is present in the new regime.

Motivated by previous works such as Elster (1998), Frank (1988), Loewenstein (2000) and Rick and Loewenstein (2008), economic research increasingly engages in a deeper understanding of emotional motivation in decision making. Recent literature suggests that emotional motivation is able to significantly impact economic decision making of individuals in a set of different games (see Andrade and Ariely 2009, Bosman and van Winden 2002, Cubitt et al. 2011, Hopfensitz and Reuben 2009, Reuben and van Winden 2010).

Van Winden and Ash (2012) provide theory on behavioral determinants of criminal activity and include emotions as one component of decision making. The aim of our study is to provide evidence for their work with regard to the emotional motivation of crime. Such motivation likely depends on the institutions that are present and how they activate and deactivate pro-social emotions like shame and guilt and negative emotions like anger and gloating.

Furthermore, Bowles and Polanía-Reyes (2012) discuss a number of factors that may influence crowding of pro-social behavior. Amongst others, they consider *moral disengagement* as one factor of crowding out. We find this explanation likely to apply to our experiment: the focus of a criminal in our study may rest dominantly on the victim in our (risk-free) no-deterrence treatment. Conversely when deterrence incentives are present this focus shifts from the victim to the scheme. The criminal then concentrates on the optimal amount to steal, while the source of this amount, the victim's pocket, loses its focus. Consequently the moral implications are decoupled from the decision. Our hypothesis is that a criminal's emotions are informative with regard to her state of mind. If emotions differ depending on the treatment (deterrence scheme) in place, they are also likely to affect behavior.

Our results indeed suggest that emotions motivate stealing. Interestingly, criminals' emotions in our experiment depend on whether a deterrence scheme is in place and motivate behavior. In the treatment without any deterrence incentives ceteris paribus a class of so-called pro-social emotions *decreases* stealing. Conversely, in face of potential punishment self-centered emotions *increase* stealing. Hence, the implemented policy does not only affect behavior directly via incentives. Rather, it also (de)activates different clusters of emotions of potential criminals which again cause pro-social behavior to crowd out. Our study therefore contributes to the interaction of deterrence incentives and emotions and to the understanding of emotional mechanisms of crowding out.

The remainder of this work is organized as follows; section 2 presents the experimental design, including the stealing game, hypotheses and procedures. The results of our study are discussed in section 3. Finally, this chapter provides a concluding discussion.

## 6.2 Experimental Design

In this section we first introduce the workhorse of this chapter: the stealing game. Next, we discuss how we make use of it, we explain our treatments and their aims. Based on our treatments, we formulate hypotheses and explain how we test them. Last, we provide information on the procedures of our experiment.

### *6.2.1 The Stealing Game and our Treatments*

The stealing game is a two-person game with 'nature' as a stochastic player. Player 1 is the only active player making a decision, and player 2 solely receives information about player 1's action and her success. Hence, there is no strategic interaction between the two players. The decision problem of player 1 is to decide how much to give to or take (steal) from player 2's endowment. We denote the initial private endowments of player 1 and player 2 as $w_1$ and $w_2$, with $w_1 < w_2$ in our experiment. The action set of player 1 is the entire range from giving all of her endowment to player 2 to taking (stealing) all of player 2's endowment, i.e. for her haul we set $h_1 \in \{-w_1, \ldots, w_2\}$. This action set calibration is important, since player 1 should have the opportunity to choose no allocation change or to give to player 2. The calibration thus avoids a suggestive action set that might influence results significantly (List 2007, Bardsley 2008). This calibration is new to the experimental literature on stealing and a slight innovation of this study. Figure 1 depicts the stealing game.

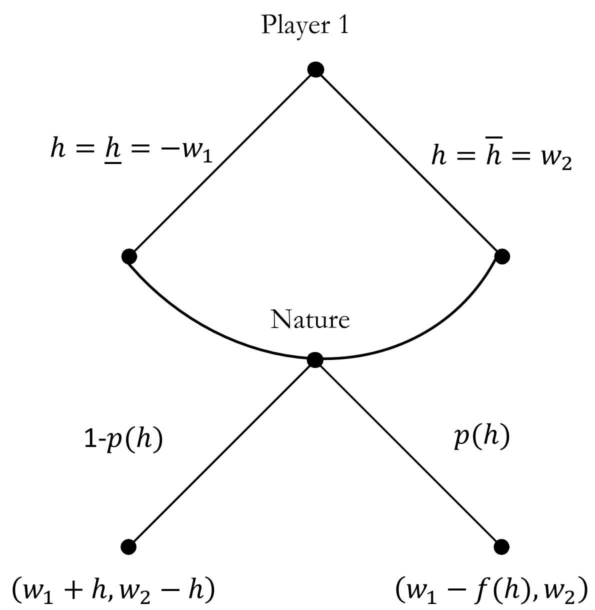
As is usual with giving gifts to others, there is no external institution which aims to prevent voluntary (charitable) giving. On the contrary, taking from others without their consent is the definition of the criminal act of stealing.[64] Accordingly, if player 1 decides to steal some monetary amount from player 2, she will face some fine $f$ with probability $p$. Corresponding

[64] Merriam-Webster's online dictionary defines to steal as "to take without permission".

to the description of the stealing game above, a player 1 with utility function $u_1$ faces the maximization problem of expected utility

$$\max_{h_1} EU_1 = \left(1 - p(h_1)\right)u_1(w_1 + h_1) + p(h_1)u_1\left(w_1 - f(h_1)\right).$$

**Figure 1.** The Stealing Game.




In our experiment, we calibrate $w_1 = 2$ EUR and $w_2 = 10$ EUR. Note that either $f$ or $p$ or both may be a function of the amount $h_1$ player 1 tries to steal. In fact, Stigler (1970) points out that it is crucial for the punishment to fit the crime. Else if the expected punishment was to remain constant, player 1's potential gains would increase in her amount stolen and would likely result in a corner solution. With this in mind, we proceed to elaborate our treatments.

The three treatments of our experiment resemble three deterrence schemes. These schemes systematically vary fine $f$ and probability $p$. In the treatment *NoDeter* there is no external institution that executes the extrinsic deterrence incentives. Hence, we implicitly set $f = 0$ and/or $p = 0$. This treatment allows us to collect information on the behavior of player 1 when she wears her 'moral mode hat'. Accordingly, player 1's maximization problem reduces to $\max_{h_1} EU_1 = u_1(w_1 + h_1)$. A rational, non-satiated and narrowly self-interested player 1 solves this problem by stealing the maximum amount, i.e. all of player 2's endowment, $h_1^* = w_2 = 10$ EUR.

As noted above, especially in *NoDeter* subjects may wear their 'moral mode hats'. It may therefore be the case that player 1 holds social, other-regarding preferences or certain moral concepts, and acts on them. One prominent form of social preferences includes inequality aversion with regard to final payoff $\pi$ (Fehr and Schmidt 1999, Bolton and Ockenfels 2000), such that $u_1(\pi_1, \pi_2)$. A sufficiently inequality averse[65] player 1 will therefore not steal player 2's entire endowment, but rather prefers to achieve ex post payoff-equality. Another motivation to abstain from stealing (or stealing the maximum amount) is a coherent self-image (e.g. Brekke et al. 2003) motivated by moral concepts like Kantianism. If such social preferences or self-image concerns are present, we will expect to find some average level of stealing which is somewhat below the standard prediction of $h_1^* = w_2$.

In the second and third treatment of our experiment, called *DeterFine* and *DeterProb*, the external deterrence institution is in place. This deterrence institution only gets activated if stealing takes place. Giving actions remain permitted. In *DeterFine* player 1 faces a probability of prosecution of

$$p_{DeterFine}(h) = \begin{cases} 0 & if \quad h \leq 0 \\ 0.5 & if \quad h > 0 \end{cases}$$

and a fine $f$ which is increasing in $h_1$. The function reads

$$f_{DeterFine}(h) = \begin{cases} 0 & if \quad h \leq 0 \\ 1.25 - 0.25h + 0.1h^2 & if \quad h > 0 \end{cases}.$$

This function is useful to calibrate treatment *DeterFine* in order to achieve an interior solution for player 1's optimal stealing. If she is risk neutral, her expected payoff is maximal at $h_{DeterFine}^* = \frac{25}{4} = 6.25$ EUR. Treatment *DeterProb* was designed to offer the exact same solution $h_{DeterProb}^* = \frac{25}{4} = 6.25$ EUR and, importantly, also corresponds to *DeterFine* with regard to the expected value for all other $h$ choices. *DeterProb* meanwhile varies the probability of prosecution

65 For instance, in the standard linear Fehr-Schmidt model sufficiently inequality averse means $\beta_1 > \frac{1}{2}$.

$$p_{DeterProb}(h) = \begin{cases} 0 & if \quad h \leq 0 \\ 0.25 + 0.05h & if \quad h > 0 \end{cases}$$

while the fine is constant at

$$f_{DeterProb}(h) = \begin{cases} 0 & if \quad h \leq 0 \\ 2.50 & if \quad h > 0 \end{cases}.$$

Note again that these calibrations are in accordance with Stigler (1970), i.e. the expected punishment increases in the amount stolen. Table 1 summarizes the three treatments and their probabilities and fines for all possible stealing actions. We include both *DeterFine* and *DeterProb* in our experiment to test for differences in deterrence from increasing fines and increasing probabilities. We can therefore directly test their relative usefulness to deter stealing when expected punishment is constant across the two treatments *for every possible action*. To the best of our knowledge this is the first study with such a calibration and thereby contributes to theory on optimal deterrence (e.g. Becker 1968, Polinsky and Shavell 1979, Garoupa 1997, Garoupa 2001).

**Table 1.** Treatment Overview.

| $h$ | Treatments (Deterrence Schemes) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *NoDeter* | | | *DeterFine* | | | *DeterProb* | | |
| | $f$ in EUR | $p$ | $\pi$ in EUR | $f$ in EUR | $p$ | E($\pi$) in EUR | $f$ in EUR | $p$ | E($\pi$) in EUR |
| -2 | | | -2 | | | -2 | | | -2 |
| -1 | | | -1 | 0 | 0 | -1 | 0 | 0 | -1 |
| 0 | | | 0 | | | 0 | | | 0 |
| 1 | | | 1 | 1.10 | | -0.05 | | 0.3 | -0.05 |
| 2 | | | 2 | 1.15 | | 0.425 | | 0.35 | 0.425 |
| 3 | | | 3 | 1.40 | | 0.80 | | 0.4 | 0.80 |
| 4 | 0 | 0 | 4 | 1.85 | | 1.075 | | 0.45 | 1.075 |
| 5 | | | 5 | 2.50 | 0.5 | 1.25 | 2.50 | 0.5 | 1.25 |
| 6 | | | 6 | 3.35 | | 1.325 | | 0.55 | 1.325 |
| 7 | | | 7 | 4.40 | | 1.3 | | 0.6 | 1.3 |
| 8 | | | 8 | 5.65 | | 1.175 | | 0.65 | 1.175 |
| 9 | | | 9 | 7.10 | | 0.95 | | 0.7 | 0.95 |
| 10 | | | 10 | 8.75 | | 0.625 | | 0.75 | 0.625 |

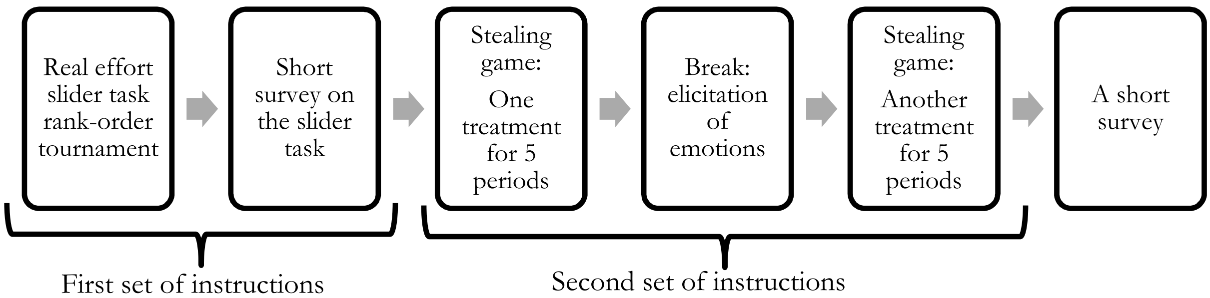
### 6.2.2 *Procedures*

All sessions were conducted in the computer laboratory of the University of Hamburg in October and November 2011 as well as January and February 2012. Each session lasted approximately one hour. We used z-Tree (Fischbacher 2007) for programming and ORSEE (Greiner 2004) for recruitment. In total 408 subjects participated in our experiment. All subjects were students from various academic backgrounds and 50.7 percent of the participants were female. No subject participated in the experiment more than once. We conducted fourteen experimental sessions of two parts; the first part was a real-effort task to endogenously determine endowments and positions for the second, stealing-game part.

Once the participants were seated, a set of instructions was handed out and read out loud by the experimenter. In order to ensure that subjects understood the respective game, experimental instructions included several numerical examples and participants had to answer control questions via their computer terminals.

Let us now consider the course of the experiment more closely. Figure 2 depicts the timeline of our experiment. As outlined above, studies by Schulze and Frank (2003), Schildberg-Hörisch and Strassmair (2012) and the preceding chapter "Deterrence Works for Criminals" find evidence for crowding out of pro-social behavior due to deterrence incentives. These three studies randomly assign subjects to be player 1 and player 2. Studies however show that the mechanism which determines a player's position in a game and her endowment may affect her decision making (see e.g. List 2007, List and Cherry 2008). Established mechanisms are the exogenous, random mechanism (by the experimental program) or 'earned' positions and endowments in a preceding endogenous mechanism. For instance, giving in a dictator game with earned endowment is usually less generous compared to giving when endowments are randomly provided (List and Cherry 2008).

**Figure 2.** Timeline of the Experiment.



For this reason, participants of our experiment engaged in a preceding real effort task before playing the stealing game. The task we use is the slider task of Gill and Prowse (2011, 2012)[66] which was part of one common rank-order tournament scheme. The more successful half of subjects received endowments of $w_2 = 10$ EUR each and later became player 2, while the less successful half of subjects received endowments of $w_1 = 2$ EUR each and later became player 1. In addition, all subjects received an initial endowment of 7 EUR each.

Subjects knew beforehand that the experiment would be split into parts and that the earned endowment from the rank-order tournament would be transferred to later parts. Also, they were informed that the final payout would depend on the events of *all* parts of the experiment. After the end of the tournament but before subjects learnt about their endowment and the content of the latter part of the experiment (i.e. the stealing game), subjects answered three questions regarding the task.

Hence, the slider task led to unequal endowments based on a rank-order tournament scheme: 'rich' player 2 subjects (potential victims) and 'poor' player 1 subjects (potential criminals). This design features carries two advantages. First, it increases identification with endowments and social positions and thereby potentially increases feelings of entitlement. This is important because we aim at analyzing changes in emotions and emotional involvement is crucial for this to work. Second, the competitive nature of the task may already let subjects put on their 'exchange mode hat' and crowd-out pro-social behavior *before* the actual stealing decision. Hence, this procedure works against our aim to detect and explain crowding out. If we nevertheless find crowding out to occur in our experiment, we can rule out that this pro-social behavior stems from random roles.

Subsequently, subjects received information on the latter course of the experiment that was again distributed and read out loud. Player 1 was then able to give up to 2 EUR to or steal up to 10 EUR from player 2. The stealing game was split into two halves of five periods each. We employed an absolute stranger matching. Note that due to the rank-order tournament in part

[66] The slider task is programmed in z-tree and offers a number of useful features (each effort unit is approx. equally costly, no prior knowledge is needed, etc.). Subjects see a screen of 48 sliders. All sliders have a range between 0 and 100. The default position is 0 for all sliders. Subjects then need to put the sliders exactly midway of the range, i.e. on the '50' mark, to receive a point. Subjects were only able to use the computer mouse for this task and time was limited and equal for all subjects, so that no subject could collect all points. Subjects were able to conduct a test round of the task (without rank feedback). See the instructions in appendix B for further information.

1 that included all subjects in a session, player 1 subjects could be sure that the player 2 subjects they were matched with owned higher endowments because they realized more effort in the slider task. At the end of every period player 1 and player 2 received information about the amount stolen and the payoffs of both matched subjects of this period. Hence, both subjects knew whether stealing was successful. To keep player 2 involved we offered her 1 EUR extra payoff in case she correctly guessed the decision of her matched player 1 in that period.[67]

After five periods player 1 subjects were asked to take their time to fill in a survey on their emotions when making their decisions. Player 2 subjects were meanwhile asked to state their emotions when seeing the decisions of their matched player 1.

**Table 2.** Session Overview.

| Session # | # of Subjects | Av. Age | % male | Part 1 | Part 2 1st half incentives | Part 2 1st half Chat? | Break | Part 2 2nd half incentives | Part 2 2nd half Chat? |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 28 | 24.1 | 64.3 | Real-Effort Slider Task | NoDeter | No | Self-Report of Emo-tions | DeterFine | No |
| 2 | 28 | 25.1 | 39.3 | | NoDeter | No | | DeterProb | No |
| 3 | 30 | 24.9 | 40.0 | | DeterFine | No | | DeterProb | No |
| 4 | 30 | 23.0 | 30.0 | | DeterFine | No | | DeterProb | No |
| 5 | 30 | 26.1 | 53.3 | | DeterProb | No | | DeterFine | No |
| 6 | 30 | 22.2 | 50.0 | | DeterProb | No | | DeterFine | No |
| 7 | 30 | 23.3 | 46.7 | | DeterFine | No | | NoDeter | No |
| 8 | 28 | 22.5 | 60.7 | | DeterProb | No | | NoDeter | No |
| 9 | 30 | 25.2 | 66.7 | | NoDeter | No | | NoDeter | Yes |
| 10 | 30 | 23.8 | 36.7 | | DeterFine | No | | DeterFine | Yes |
| 11 | 30 | 24.5 | 60.0 | | DeterProb | No | | DeterProb | Yes |
| 12 | 28 | 25.0 | 46.4 | | NoDeter | Yes | | NoDeter | No |
| 13 | 28 | 25.7 | 46.4 | | DeterFine | Yes | | DeterFine | No |
| 14 | 28 | 24.3 | 71.4 | | DeterProb | Yes | | DeterProb | No |
| Overall | 408 | 24.3 | 50.7 | | | | | | |

Note: The number of subjects per session is not perfectly equal over the fourteen sessions due to some registered individuals not showing up.

After the survey on emotions, the experiment continued for another five periods with some treatment change. As an alternative to a treatment change, additional sessions tried to trace out whether change in judicial procedure, i.e. simple ex-post communication between the player 1 and 2 (via an anonymous chat), is able to deter criminal activity. Such communication is a first step towards testing the effect of alternative procedure on deterrence, as suggested by

[67] This opportunity was private information of player 2.

proponents of restorative justice (e.g. Braithwaite 2002, Umbreit et al. 1994). Depending on the interpretation of means to ends, restorative justice can be regarded as an alternative for or improvement of retributive justice. Here the aim was to examine alternative ways of deterrence. In the results section we will briefly discuss the impact of a forced chat opportunity between player 1 and player 2.

After all periods were played, one out of the ten periods was randomly selected for payment. Average payment over all treatments was 11.70 EUR. Table 2 summarizes the information for all 14 sessions.

### *6.2.3 Hypotheses*

The central aim of this study is to investigate the interaction of deterrence incentives and emotions. First, we use the stealing game and the three treatments *NoDeter*, *DeterFine* and *DeterProb* to establish the deterrence effect. The underlying hypothesis is that stealing will be lower when an external deterrence institution is present than when it is not. This resembles Gary Becker's deterrence hypothesis. Accordingly, we formulate

**Hypothesis 1.** *Deterrence incentives cause lower average stealing, i.e.* $\bar{h}_{NoDeter} > \bar{h}_{DeterFine}$ *and* $\bar{h}_{NoDeter} > \bar{h}_{DeterProb}$.

Second, there has been a long-standing on the relative deterrence effectiveness and probabilities and fines (Garoupa 1997). On the one hand, Schildberg-Hörisch and Strassmair (2012) provide evidence that the both probabilities and fines work to reduce stealing and hold similar relative effectiveness and can be regarded as substitutes. On the other hand, Harbaugh et al. (2011) suggest that fines hold a stronger relative effectiveness compared to probabilities. One explanation may be risk aversion (Ehrlich 1973). To provide further evidence on this question, we include the two treatments *DeterFine* and *DeterProb* in order to investigate the relative effectiveness of increasing probabilities and fines as a function of stealing. Our design includes the novel feature that for the two treatments *expected punishment is equal for every stealing action.*

**Hypothesis 2.** *Probabilities and fines are substitutes for deterrence of stealing, i.e.* $\bar{h}_{DeterFine} = \bar{h}_{DeterProb}$.

Third, as indicated by the results of Schulze and Frank (2003), Schildberg-Hörisch and Strassmair (2012) and chapter 5, we assume that while deterrence incentives reduce average stealing, they also crowd out intrinsic motivation to act pro-socially. To test this, we employ the combination of between- and within-subject design laid out above. Note that we aim to test whether deterrence incentives in *DeterFine* and *DeterProb* make subjects put on their 'exchange mode hat'. If this is the case, then we should see that stealing in *NoDeter* is greater in the second five periods of the experiment if it is preceded by stealing in *DeterFine* or *DeterProb* beforehand. Hence, we formulate

**Hypothesis 3.** *Deterrence incentives crowd out intrinsic motivation for pro-social behavior. That is, stealing in NoDeter is higher if preceded by stealing in DeterFine and/or DeterProb.*

Fourth, if we are able to establish that crowding out indeed occurs, then we may find evidence for the 'moral mode hat' or 'exchange mode hat' in the self-reported emotions of player 1 subjects before the regime change. We borrow the emotions we elicit in our experiment from Reuben and van Winden (2010). Amongst others, these emotions are shame and guilt. Gilbert (2003) and Haidt (2003) discuss and classify shame and guilt as moral, pro-social emotions. Hence, if deterrence incentives 'morally disengage' (Bowles and Polanía-Reyes 2012) player 1 subjects, then they should feel less shame and guilt when engaging in stealing in *DeterFine* and *DeterProb* compared to *NoDeter*. We hypothesize

**Hypothesis 4.** *Crowding out of intrinsic motivation occurs via changes in emotions when making decisions. In the act of stealing player 1 subjects in NoDeter exhibit stronger emotions such as shame and guilt compared to those in treatments DeterFine and DeterProb.*

Van Winden and Ash (2012) present theory on behavioral criminal law and economics. One of their contributions is the addition of emotions as a motivational factor of criminal activity. Our research tests their hypothesis. If incentives shape emotions of subjects, we should find differences in self-reported emotions, especially between *NoDeter* and *DeterFine/Prob*. Further, if these emotions then affect stealing, the regime changes should be able to pick up decision differences in the data.

## 6.3 Results

Before we formulate results concerning the hypotheses we aim at testing for the stealing game, we need to ensure comparability of sessions. For this purpose we use two Kruskal-Wallis tests. The first test analyses equality of populations with regard to effort levels in the real-effort slider task. The test does not reject equality of all fourteen sessions ($p = 0.4227$, for histograms see Figure A.1 in appendix A). That is, subjects in all fourteen sessions exerted similar effort in the task. The second Kruskal-Wallis test examines the perceived fairness of the slider task, which was reported on a five-point scale before subjects received information on their relative position and on the stealing game. This second test also does not reject the null hypothesis of equality of populations across all fourteen sessions ($p = 0.4985$, for histograms see Figure A.2 in appendix A). Hence, we are able to conduct our analysis with respect to stealing in our three treatments *NoDeter*, *DeterFine* and *DeterProb*.

### *6.3.1 The Deterrence Hypothesis and Crowding Out of Pro-Social Behavior*

Let us next analyze the magnitudes of stealing in a between-subject comparison of deterrence schemes in the first five periods without ex-post chat opportunity. Using average stealing for Mann Whitney tests (yielding one observation per individual), we find significantly higher stealing of 7.71 EUR in *NoDeter* ($n = 43$) compared to both *DeterFine* (4.17 EUR, $n = 60$) and *DeterProb* (4.75 EUR, $n = 59$). Both deterrence treatments yield lower stealing at $p < 0.0000$. Hence, our results firmly support Becker's deterrence hypothesis (our Hypothesis 1).

Comparing *DeterFine* and *DeterProb* (with the same expected punishment for every stealing decision), we find no treatment effects ($p = 0.2114$). The fact that we cannot reject the null hypothesis that stealing is equally pronounced in *DeterFine* and *DeterProb* supports Hypothesis 2, i.e. substitutability of probabilities and fines.[68] Figure 3 depicts average stealing in the three treatments in the first five periods (and decision distributions are illustrated in Figure A.3 in appendix A). We summarize

**Result 1.** *Our experiment supports the deterrence hypothesis and supports the general result in the literature that deterrence works. Further, we find that punishment size and probability deter stealing equally effectively.*

Disentangling the deterrence effect from time effects and demographic information does not change these results. Table 3 reports random-effects regressions for the first five periods. We include the sessions with chat opportunity and control for individual periods, availability of the chat, age, gender, and subjects being students of business and economics or of law. None

[68] Note however that our treatments only consider fines and probabilities in a medium range. An investigation on extremely high or low probabilities and fines may be an avenue for future research.

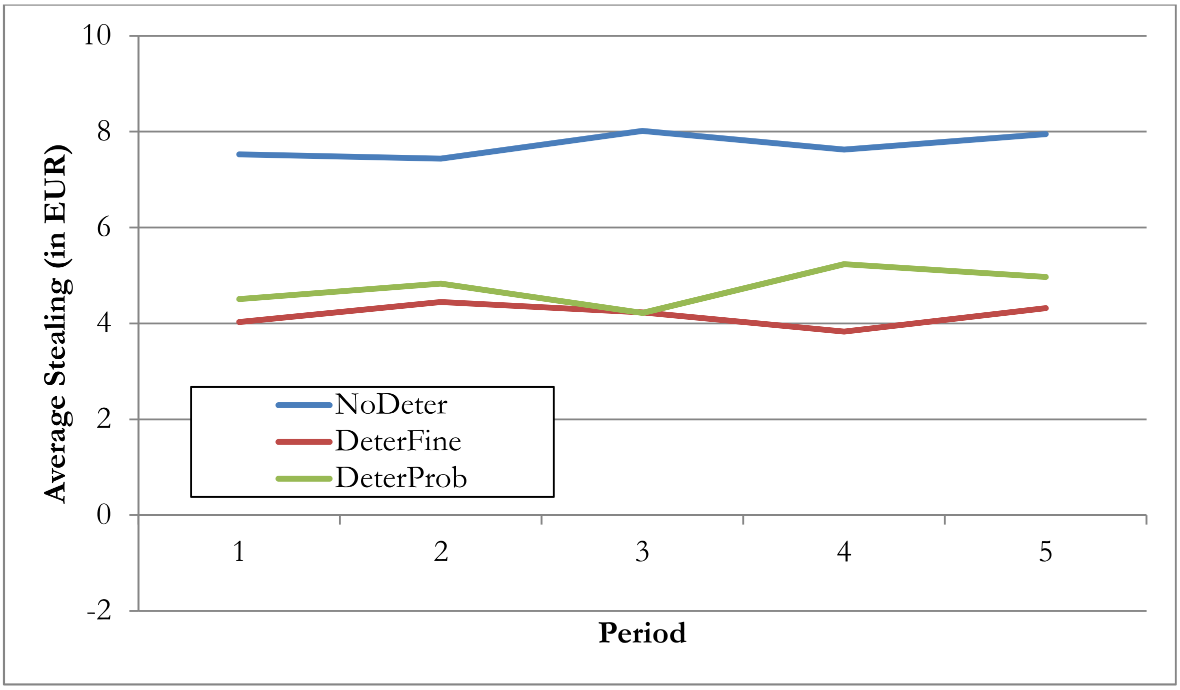
of these control variables reports significant influence on stealing behavior.[69] Note that there is no chat effect in our treatments. Hence, the hypothesis that ex post communication alone deters and mitigates a part of criminal activity does not hold. Consequently we can pool the data of chat and no-chat treatments in our later analysis.

In specification V we also include a time-lag of the magnitude of stealing and a dummy variable that indicates when subjects were caught stealing in t-1. Indeed subjects appear to be consistent in their choice of stealing over time; while the time-lag of stealing is highly significant, subjects do not show a change in behavior when caught in the former period. Hence, subjects do not appear to update their perception of risk. We formulate the two findings concerning the chat and the update of risk attitudes:

**Result 2.** *An ex-post chat opportunity of matched player 1 and player 2 does not change stealing behavior.*

**Result 3.** *Player 1 subjects do not change behavior if caught stealing in a previous period. Hence, there appears to be only minor or no updating of risk attitudes.*

**Figure 3.** Average Stealing, First Five Periods without Chat.




[69] There is one minor exception: specification IV reports higher stealing in period 5 (at the 10 percent level). This effect disappears when employing a time-lag of a subject's stealing in t-1.

So far, we have reported results from a between-subject-design analysis. When decision making in the first five periods was over, the experiment took a break to ask subjects for their self-reported emotions when making the stealing decision.[70] Before we have a closer look at these emotions we examine the effects of regime change. Here we take advantage of a within-subject design feature of our experiment. It may well be that decision making in a treatment for five periods establishes a norm that carries over to another treatment. Therefore, we are able to test two relationships: first, we test whether the deterrence hypothesis holds when there is no prior punishment. That is, we test whether subjects steal less in *DeterFine* and *DeterProb* after being in *NoDeter* for five periods. Second and conversely, we also test whether deterrence continues to work after the punishment incentives have been removed.

We find that our deterrence effects are the same whether or not player 1 subjects have a history of no deterrence for both *DeterFine* and *DeterProb* (sessions # 1 and 2 respectively, Wilcoxon signed-ranks tests, both tests yield $p < 0.0000$).[71] For the inverted case in which criminals face deterrence and these incentives are then removed, stealing increases significantly (sessions # 7 and 8, Wilcoxon signed-ranks test, $p < 0.0000$). Figures 4 and 5 provide graphs of these effects of regime change. Figure 5 also includes graphs of mean stealing in *NoDeter* in the first five periods (sessions # 1, 2, 9 and 12) and the second five periods without prior deterrence (sessions # 9 and 12). We find that stealing in the treatments with deterrence history is significantly *greater* than without such history (Mann Whitney test, $p = 0.0495$). This finding provides evidence for the crowding out of pro-social behavior by deterrence incentives and supports Hypothesis 3. We summarize this result as follows:

**Result 4.** *Deterrence incentives crowd out pro-social behavior. That is, stealing in NoDeter is significantly greater after facing deterrence institutions than without such a history.*

---

[70] Player 2 subjects were similarly asked for their emotions in the moment they received information on the stealing decisions of matched player 1 subjects. As this chapter focuses on stealing decisions of player 1, the player 2 data is not discussed in this chapter.

[71] In accordance with our data structure, we compare mean stealing of a player 1 before and after the regime change. Hence, we keep two observations (means) per individual for our analysis.

**Figure 4.** Average Stealing, Sessions *NoDeter_DeterFine* and *NoDeter_DeterProb*.

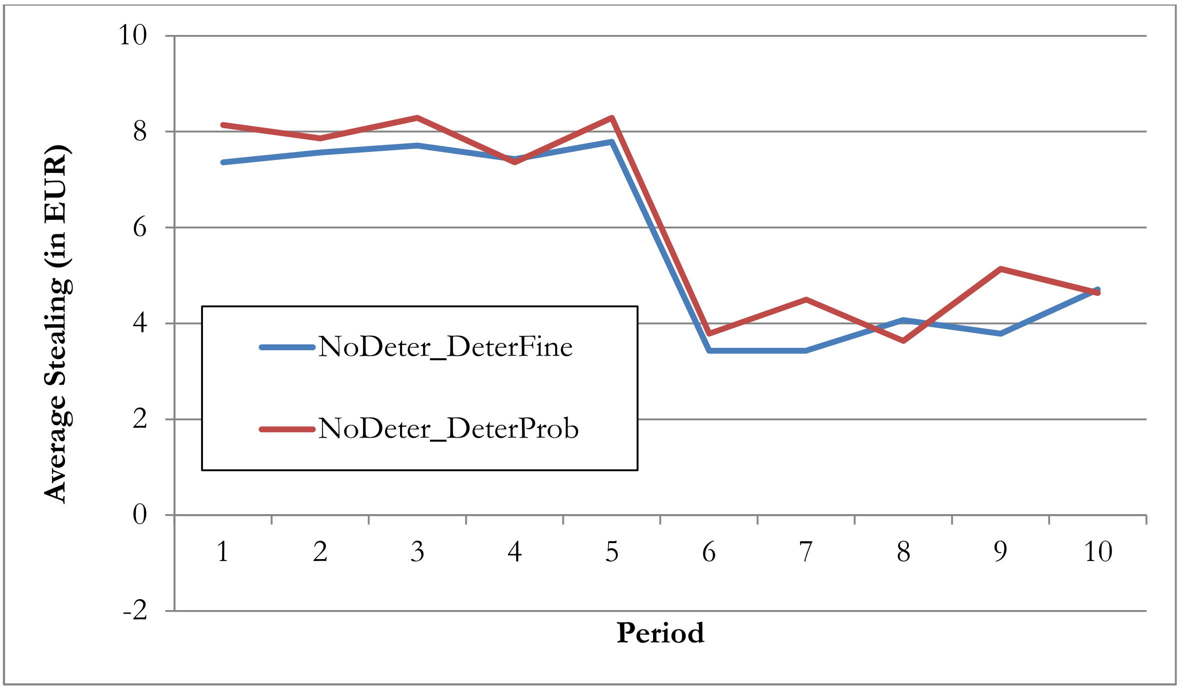



**Figure 5.** Average Stealing, Sessions *DeterFine_NoDeter* and *DeterProb_NoDeter*, Compared to *NoDeter* Stealing in other Sessions.

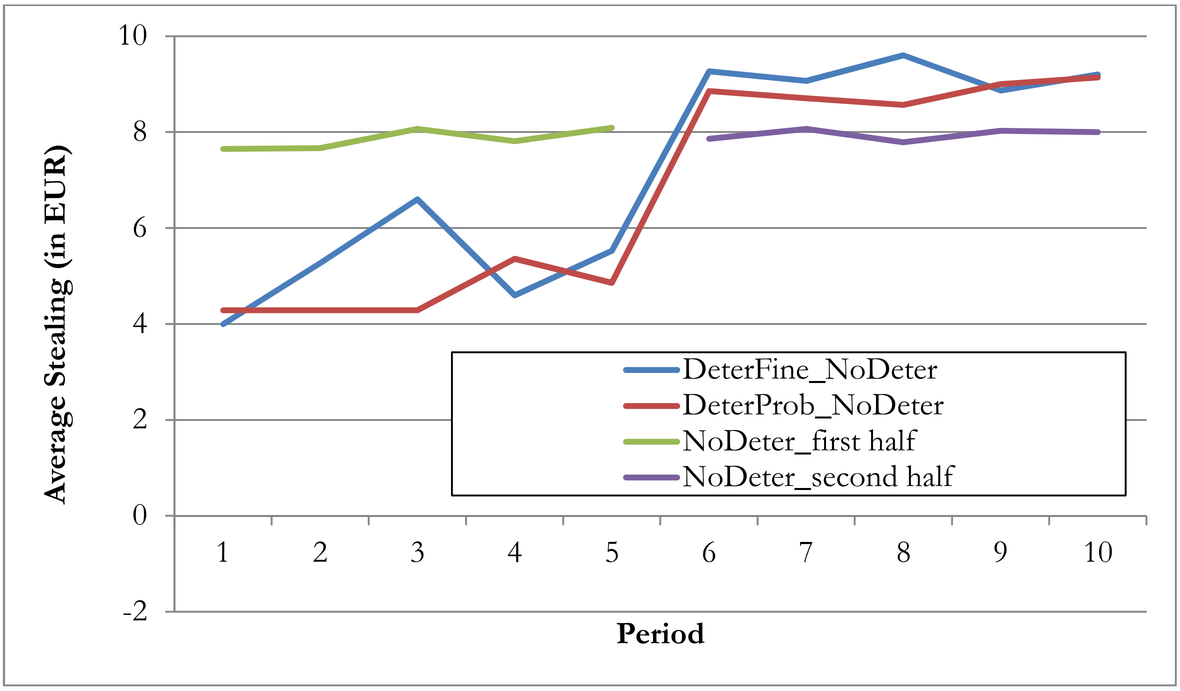


**Table 3.** Linear Random-Effects Regressions: Testing for Treatment Effects in Criminals' Behavior, First Five Periods.

| Independent Variable | Dependent Variable: *h* (magnitude of stealing) | | | | |
|---|---|---|---|---|---|
| | I (*without chat*) | II | III | IV | V |
| $h_{i,t-1}$ (continuous, time lag) | | | | | 0.482*** (0.050) |
| Caught stealing in previous period (dummy, time lag) | | | | | 0.223 (0.255) |
| DeterFine (dummy) | -3.543*** (0.467) | -3.656*** (0.400) | -3.796*** (0.396) | -3.796*** (0.397) | -2.089*** (0.339) |
| DeterProb (dummy) | -2.964*** (0.496) | -3.038*** (0.432) | -3.195*** (0.438) | -3.195*** (0.438) | -1.684*** (0.328) |
| Chat (dummy) | | 0.289 (0.387) | 0.259 (0.386) | 0.259 (0.386) | 0.127 (0.213) |
| Age (continuous) | | | -0.027 (0.044) | -0.027 (0.044) | -0.007 (0.024) |
| Male (dummy) | | | 0.511 (0.348) | 0.511 (0.349) | 0.148 (0.201) |
| Business or economics student (dummy) | | | 0.326 (0.358) | 0.326 (0.358) | 0.145 (0.209) |
| Law student (dummy) | | | 0.685 (0.691) | 0.685 (0.692) | 0.400 (0.410) |
| Period2 (dummy) | | | | 0.272 (0.220) | |
| Period3 (dummy) | | | | 0.153 (0.192) | -0.251 (0.301) |
| Period4 (dummy) | | | | 0.124 (0.222) | -0.227 (0.215) |
| Period5 (dummy) | | | | 0.436* (0.238) | 0.097 (0.258) |
| Constant | 7.716*** (0.365) | 7.785*** (0.328) | 8.159*** (1.108) | 7.962*** (1.113) | 4.292*** (0.793) |
| Individuals | 162 | 204 | 202 | 202 | 202 |
| Observations | 810 | 1020 | 1010 | 1010 | 808 |

Note: An observation is a subject's magnitude of criminal activity in a period. Treatment *NoDeter* is the baseline. Two subjects did not enter their demographic information, so that they are excluded from analysis with demographics (regressions III to V). Robust standard errors (clustered at the individual level) in parentheses, significance: *p < 0.10, **p < 0.05, ***p < 0.01.

### *6.3.2 Incentives, Emotions and Behavior*

We now turn to the emotions in order to investigate their impact on behavior. Note again that after the first five periods, and before a treatment change was introduced, player 1 subjects were asked to self-report emotions they exhibited *when making the decision*. Self-reports were done on a seven-point scale.

The results show that whether or not a deterrence scheme is in place (*NoDeter* vs. *DeterFine* and *DeterProb*) indeed influences emotions of criminals. That is, emotion differences mostly exist between *NoDeter* and *DeterFine* as well as *NoDeter* and *DeterProb*. Meanwhile there are few to no differences between emotions in *DeterFine* and *DeterProb*. Descriptive statistics and pairwise Mann-Whitney test for treatment differences are report in Table 4.

For instance, on average player 1 subjects in *DeterFine* and *DeterProb* feel more anger (2.63 and 2.71 points respectively) and more envy (3.28 and 3.36 points respectively) when making their decisions compared to player 1 subjects in *NoDeter* (on average 1.56 points of anger and 1.79 points of envy). On contrary, player 1 subjects in *NoDeter* on average feel more guilt (2.72 points) and shame (2.95 points), but also more gratitude (3.16 points) compared player 1 subjects in *DeterFine* (1.42 points of guilt, 1.93 points of shame, and 2.02 points of gratitude) and *DeterProb* (2.00 points of guilt, 1.97 points of shame, and 2.19 points of gratitude). All these differences are statistically significant at conventional levels based on pairwise Mann Whitney tests. For more information, see Table 4.

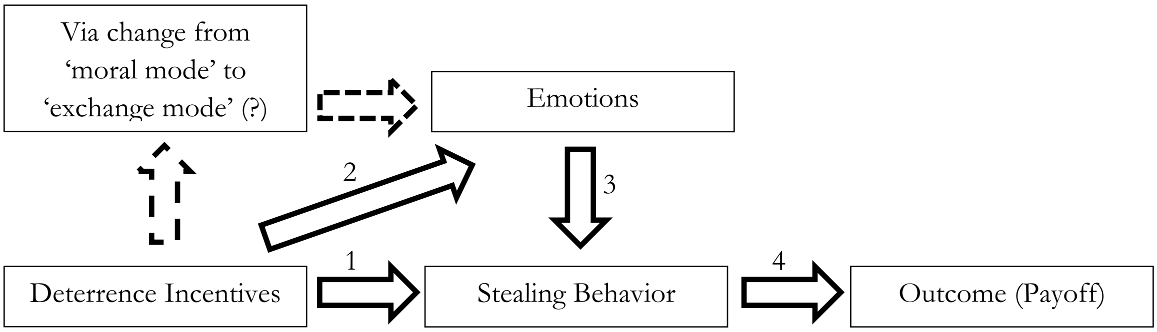
So far, we have established a number of links in our experiment. Figure 6 provides a graphic illustration. The first link that we can support is (the classic one) that incentives influence behavior. Naturally then, behavior influences outcomes, as they are directly deterministically or stochastically related. These links are labeled '1' and '4' in Figure 6. The self-reported emotions of player-1 subjects in the three treatments demonstrate differences with respect to both pro-social emotions (like shame, guilt and gratitude) and self-centered emotions (like envy). Note again that these subjects were asked about the emotions they exhibited *when making their decisions*. Accordingly, we are able to establish link '2' between the treatments and emotions. The interpretation we favor is that there is also an intermediate link via the change from player 1 being in the 'moral mode' to being in the 'exchange mode' (Meier 2007). As we have no way to directly elicit subjects' 'modes', we postpone further explanations of this matter to future research.

**Table 4.** Descriptive Statistics and Pairwise Mann-Whitney Tests for Equality of Emotion Intensity of Player 1 Subjects.

| Emotions | Descriptive Statistics: Mean & Std. Dev. in parentheses | | | Mann-Whitney Test Results: p-values in parentheses | | |
|---|---|---|---|---|---|---|
| | *NoDeter* | *DeterFine* | *DeterProb* | NoD vs. DFi | NoD vs. DPr | DFi vs. DPr |
| Anger | 1.56 (1.22) | 2.63 (1.80) | 2.71 (1.81) | NoD < DFi (p = 0.001) | NoD < DPr (p = 0.000) | DFi = DPr (p = 0.911) |
| Shame | 2.95 (1.88) | 1.93 (1.45) | 1.97 (1.27) | NoD > DFi (p = 0.003) | NoD > DPr (p = 0.007) | DFi = DPr (p = 0.585) |
| Pride | 2.16 (1.53) | 2.37 (1.67) | 1.90 (1.49) | NoD = DFi (p = 0.523) | NoD = DPr (p = 0.309) | DFi > DPr (p = 0.071) |
| Disappoint-ment | 2.05 (1.59) | 3.97 (2.23) | 3.98 (1.95) | NoD < DFi (p = 0.000) | NoD < DPr (p = 0.000) | DFi = DPr (p = 0.891) |
| Surprise | 2.84 (2.06) | 2.73 (1.97) | 2.90 (1.93) | NoD = DFi (p = 0.986) | NoD = DPr (p = 0.752) | DFi = DPr (p = 0.646) |
| Joy | 4.14 (2.02) | 3.35 (1.89) | 3.15 (1.92) | NoD > DFi (p = 0.047) | NoD > DPr (p = 0.016) | DFi = DPr (p = 0.541) |
| Contempt | 1.51 (1.01) | 1.57 (1.16) | 1.93 (1.62) | NoD = DFi (p = 0.890) | NoD = DPr (p = 0.269) | DFi = DPr (p = 0.289) |
| Envy | 1.79 (1.47) | 3.28 (2.01) | 3.36 (2.00) | NoD < DFi (p = 0.000) | NoD < DPr (p = 0.000) | DFi = DPr (p = 0.901) |
| Irritation | 1.53 (1.10) | 3.40 (2.12) | 3.44 (2.07) | NoD < DFi (p = 0.000) | NoD < DPr (p = 0.000) | DFi = DPr (p = 0.967) |
| Guilt | 2.72 (1.91) | 1.42 (1.00) | 2.00 (1.52) | NoD > DFi (p = 0.000) | NoD > DPr (p = 0.046) | DFi < DPr (p = 0.024) |
| Regret | 2.70 (1.54) | 2.88 (1.87) | 3.15 (1.97) | NoD = DFi (p = 0.829) | NoD = DPr (p = 0.311) | DFi = DPr (p = 0.535) |
| Admiration | 1.79 (1.42) | 1.33 (0.95) | 1.69 (1.24) | NoD > DFi (p = 0.030) | NoD = DPr (p = 0.789) | DFi < DPr (p = 0.041) |
| Pity | 2.93 (1.86) | 1.43 (0.96) | 1.59 (1.07) | NoD > DFi (p = 0.000) | NoD > DPr (p = 0.000) | DFi = DPr (p = 0.360) |
| Gratitude | 3.16 (1.98) | 2.02 (1.55) | 2.19 (1.58) | NoD > DFi (p = 0.003) | NoD > DPr (p = 0.008) | DFi = DPr (p = 0.633) |
| Gloating | 1.70 (1.17) | 1.68 (1.28) | 1.73 (1.34) | NoD = DFi (p = 0.583) | NoD = DPr (p = 0.820) | DFi = DPr (p = 0.723) |
| Sadness | 1.81 (1.48) | 2.57 (1.85) | 2.44 (1.58) | NoD < DFi (p = 0.016) | NoD < DPr (p = 0.006) | DFi = DPr (p = 0.894) |
| Observations | 43 | 60 | 59 | | | |

Note: Subjects self-reported emotions on a seven-point scale. The list of emotions corresponds to Reuben and van Winden (2010).

**Figure 6.** Links of Incentives, Emotions and Behavior.




In the last part of this results section, we will show that there is also a link between emotions and stealing behavior, i.e. link '3'. In combination with evidence that stealing in *DeterFine* and *DeterProb* does not differ significantly, we pool subjects in *NoDeter* (n = 57) and subjects in *DeterFine* and *DeterProb* (n = 147) in two groups. Principle component analyses (PCAs) use these emotion self-reports to construct emotion clusters.[72]

Table 5 reports the clusters for *NoDeter* and Table 6 reports emotion clusters for *DeterFine* and *DeterProb*. All of the analyzed clusters feature eigenvalues of about 1 and above; this feature commonly defines the usefulness of a cluster. The clusters relate to different families of emotions: 'negative emotions', 'positive emotions', 'pro-social emotions', 'self-centered emotions', and 'vicious emotions'.

Interestingly, almost the same clusters are constructed by the two PCAs. We use these clusters to explain stealing in the first five periods of our experiment. Table 7 reports two OLS regressions, specification VI to explain stealing in treatment *NoDeter* and specification VII to explain stealing in treatments *DeterFine* and *DeterProb*. In these regressions we include all five clusters of emotions (while controlling for 'chat' in both specifications and '*DeterFine*' in specification VII).

[72] This step follows Reuben and van Winden (2010). It is motivated by the large number of pairwise correlations of emotions as shown in tables A.1 and A.2 in appendix A. The aim is to organize these emotions in clusters.

**Table 5.** Principal Emotion Components of Player-1 Subjects, Treatment *NoDeter*.

| Emotions | Components for NoDeter | | | | |
|---|---|---|---|---|---|
| | 1st Eigenvalue: 6.82 Proportion: 0.43 | 2nd Eigenvalue: 2.15 Proportion: 0.13 | 3rd Eigenvalue: 1.25 Proportion: 0.08 | 4th Eigenvalue: 1.19 Proportion: 0.07 | 5th Eigenvalue: 1.03 Proportion: 0.06 |
| Anger | 0.2626 | -0.2002 | 0.3569 | -0.2586 | -0.0888 |
| Shame | 0.3090 | 0.0755 | -0.0464 | -0.2269 | -0.2165 |
| Pride | 0.1209 | 0.2396 | 0.6652 | 0.1095 | 0.0243 |
| Disappointment | 0.3068 | -0.1381 | -0.1333 | 0.2277 | 0.0916 |
| Surprise | 0.2379 | 0.1324 | -0.2756 | 0.2507 | -0.3102 |
| Joy | 0.0340 | 0.5235 | 0.2698 | 0.0674 | -0.0386 |
| Contempt | 0.2979 | -0.0660 | 0.0209 | 0.1739 | 0.4604 |
| Envy | 0.2801 | -0.2288 | 0.2913 | -0.1350 | -0.0477 |
| Irritation | 0.3176 | -0.3016 | 0.0771 | 0.1482 | -0.0482 |
| Guilt | 0.2666 | 0.1643 | 0.0217 | -0.3541 | -0.1737 |
| Regret | 0.2550 | 0.1102 | -0.3209 | -0.2520 | 0.1127 |
| Admiration | 0.2955 | -0.0313 | -0.1051 | 0.3671 | 0.0836 |
| Pity | 0.1528 | 0.3610 | -0.2191 | -0.3853 | -0.0405 |
| Gratitude | 0.1583 | 0.3738 | -0.0258 | 0.4481 | -0.3300 |
| Gloating | 0.1575 | 0.3169 | -0.0369 | -0.0480 | 0.6734 |
| Sadness | 0.3231 | -0.1671 | -0.0753 | -0.0695 | -0.0897 |
| | | | | | |
| Family of emotions | negative | positive | self-centered | pro-social | vicious |
| Mean | <0.00 | <0.00 | <0.00 | <0.00 | <0.00 |
| Std. dev. | 2.61 | 1.47 | 1.12 | 1.09 | 1.01 |

Note: Principal Component Analysis of criminals' emotions in treatment *NoDeter* with and without chat. The five components account for 77.73 percent of variation.
n = 57. KMO: 0.8092.

When analyzing the impact of these clusters on stealing behavior, we find that, on average, the more pronounced the 'pro-social emotions' of player 1, the *smaller* is the amount she wants to steal in *NoDeter* (specification VI). The four other emotion cluster do not impact her decision significantly. On the contrary, in *DeterFine* and *DeterProb* the more pronounced 'self-centered emotions' the *greater* is the amount player 1 wants to steal (specification VII). Thus not only do our treatments influence the emotions of player-1 subjects, i.e. link '2' in Figure 6. In addition, we find evidence that the existence of an external deterrence institution nullifies the explanatory power of pro-social emotions while it activates the explanatory power of self-centered emotions. We therefore provide evidence for the existence of link '3' in Figure 6, i.e. that emotions impact stealing behavior. We summarize:

**Result 5.** *The impact of incentives on emotions is twofold. First, deterrence incentives change the intensity of player 1's emotions. Second, incentives (de)activate the emotions which are relevant for stealing behavior.*

**Table 6.** Principal Emotion Components of Player-1 Subjects, Treatments *DeterFine* and *DeterProb*.

| Emotions | Components for *DeterFine* and *DeterProb* | | | | |
|---|---|---|---|---|---|
| | 1st<br>Eigenvalue: 4.48<br>Proportion: 0.28 | 2nd<br>Eigenvalue: 2.30<br>Proportion: 0.14 | 3rd<br>Eigenvalue: 1.65<br>Proportion: 0.10 | 4th<br>Eigenvalue: 1.01<br>Proportion: 0.06 | 5th<br>Eigenvalue: 0.96<br>Proportion: 0.06 |
| Anger | 0.2782 | -0.2432 | 0.1805 | 0.1649 | -0.1811 |
| Shame | 0.2518 | 0.0959 | -0.3925 | -0.0084 | 0.0618 |
| Pride | 0.1399 | 0.3319 | 0.4073 | 0.0211 | 0.0242 |
| Disappointment | 0.3413 | -0.1689 | 0.1339 | -0.2749 | 0.0676 |
| Surprise | 0.2486 | 0.1526 | -0.0279 | -0.4632 | -0.2715 |
| Joy | 0.0530 | 0.4897 | 0.3199 | -0.1500 | 0.1724 |
| Contempt | 0.2867 | 0.0090 | -0.0970 | 0.2763 | -0.2970 |
| Envy | 0.2973 | -0.1242 | 0.2239 | 0.1141 | 0.0924 |
| Irritation | 0.3426 | -0.2850 | 0.2123 | -0.0134 | 0.1109 |
| Guilt | 0.2706 | 0.0581 | -0.4395 | 0.0964 | 0.1439 |
| Regret | 0.3157 | -0.0227 | -0.1592 | -0.2092 | 0.2762 |
| Admiration | 0.1976 | 0.1635 | -0.1527 | 0.1938 | -0.6600 |
| Pity | 0.1119 | 0.3040 | -0.3484 | 0.1669 | 0.3825 |
| Gratitude | 0.0920 | 0.4889 | 0.0207 | -0.2363 | -0.1674 |
| Gloating | 0.1362 | 0.2508 | 0.2247 | 0.6268 | 0.1543 |
| Sadness | 0.3451 | -0.0905 | 0.0876 | -0.0550 | 0.1226 |
| | | | | | |
| Family of emotions | negative | positive | self-centered | vicious | pro-social |
| Mean | <0.00 | <0.00 | <0.00 | <0.00 | <0.00 |
| Std. dev. | 2.12 | 1.52 | 1.29 | 1.01 | 0.98 |

Note: Principal Component Analysis of criminals' emotions in treatments *DeterFine* and *DeterProb* with and without chat. The five components account for 64.99 percent of variation.
n = 147. KMO: 0.7790.

Result 5 supports illustrations, e.g. by Rick and Loewenstein (2008), that incentives influence behavior both directly and indirectly: directly via incentives to steal less and indirectly via emotions that turn the focus away from a pro-social, other-regarding orientation to a self-centered orientation. Or in Meier (2007)'s terms from the 'moral mode' to the 'exchange mode'. Hence, we find that one explanation for crowding out of pro-social behavior is the (de)activation of different emotions. To the best of our knowledge this study is the first to provide evidence on interaction of deterrence, crowding out and emotions.

**Table 7.** OLS Regressions: Principal Emotion Components' Impact on Stealing.

| | Dependent variable: *h* (magnitude of stealing) | |
|---|---|---|
| | VI | VII |
| Independent Variable | *NoDeter* | *DeterFine & DeterProb* |
| Negative emotions | 0.059<br>(0.119) | 0.016<br>(0.093) |
| Positive emotions | 0.146<br>(0.214) | -0.043<br>(0.130) |
| Self-centered emotions | -0.007<br>(0.280) | 0.344**<br>(0.156) |
| Pro-social emotions | -0.621**<br>(0.292) | 0.142<br>(0.204) |
| Vicious emotions | 0.363<br>(0.320) | 0.289<br>(0.196) |
| Chat | 0.110<br>(0.766) | 0.320<br>(0.514) |
| *DeterFine* | | -0.723*<br>(0.399) |
| Constant | 7.829***<br>(0.362) | 4.794***<br>(0.297) |
| Individuals | 57 | 147 |

Note: Each observation is the *average of the magnitude* of stealing of a player 1 subject over the first five periods. Standard errors in parentheses, significance: *p < 0.10, **p < 0.05, ***p < 0.01.

## 6.4 Concluding Discussion

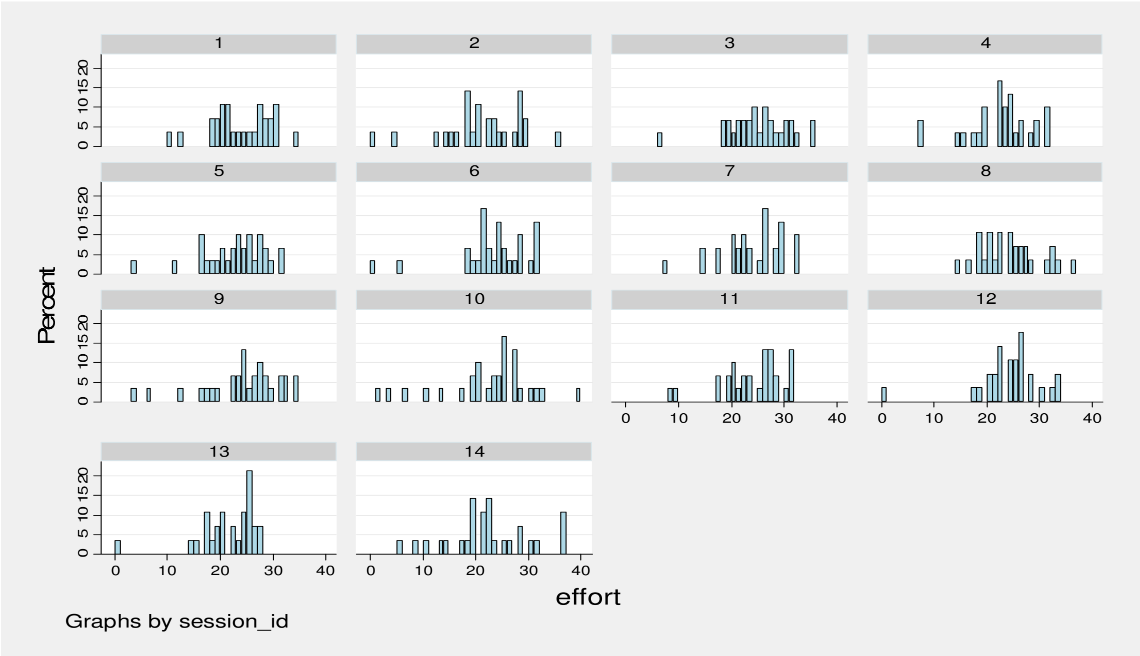
Beginning with Gary Becker's seminal work in 1968, the beneficial feature of deterrence incentives to reduce social welfare loss from crime has been well documented over the last fifty years. We employ experimental economic methods to provide new insights into their detrimental effect of crowding out of pro-social behavior. Further, we investigate the role of emotions for stealing decisions and the interaction of incentives with emotions. Our experiment asks whether crowding out acts through emotional motivation.

Our results support the (direct) beneficial feature of deterrence incentives. Yet our results also suggest that crowding out of pro-social behavior is an undesired by-product. Our analysis suggests that incentives indeed (de)activate different clusters of emotions that then motivate behavior. Consequently, we offer support for prior findings and theories of individual decision making that integrate experienced, integral emotions. Van Winden and Ash (2012) provide theory on behavioral criminal law and economics that incorporates emotions. Emotional motivation appears to be a vehicle of crowding out of pro-social behavior. Our study therefore connects two important strands of behavioral economic literature: we regard this finding as quantitative and measureable support for Bowles and Polanía-Reyes' (2012) illustration of *moral disengagement* as a cause of crowding out.

In the spirit of Gneezy and Rustichini (2000b)'s title "Pay enough or don't pay at all", we invert this suggestion and propose to 'punish enough or don't punish at all'. This idea finds support in results of Schildberg-Hörisch and Strassmair (2012) who find that deterrence works for high fines but not for low fines (likely due to crowding out). One caveat of our study is that we analyze a two-player game. Moral implications and emotions of an action may change behavior in the way we document it in this study. Whether the same is true for an interaction of an individual with a group or firm where no direct, single victim can be identified, pro-social emotions may be less relevant for behavior. There is clearly room for future research on this matter. We also hope that this study is able to motivate further research on the interaction of emotions, extrinsic and intrinsic motivation and decision making.

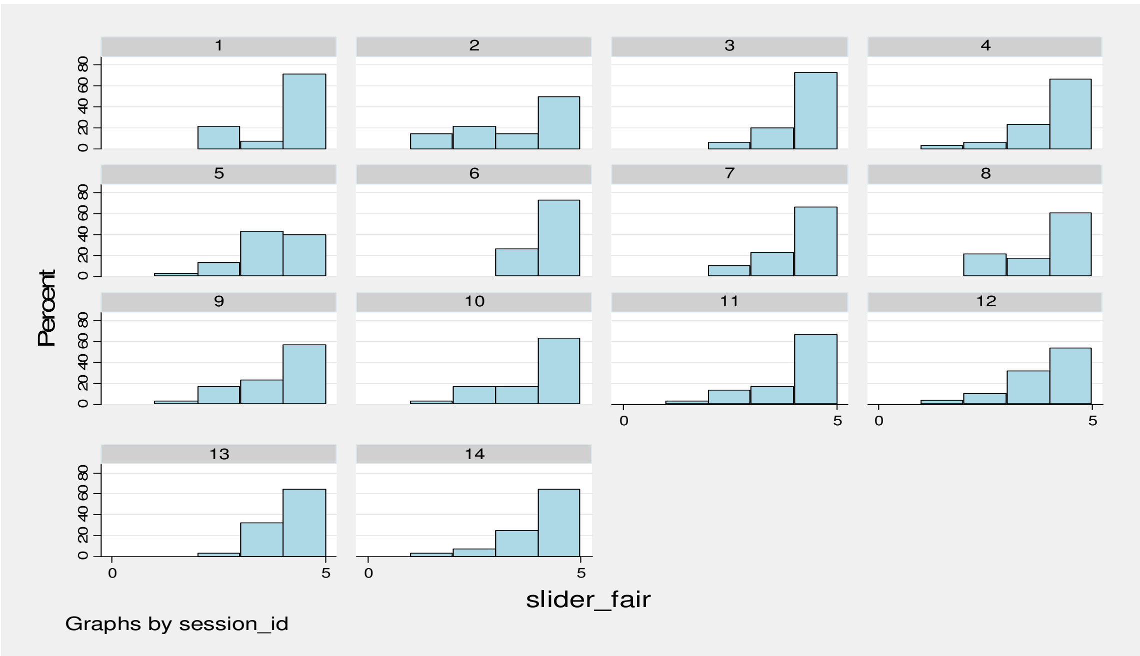
## 6.5 Appendix A: Supplementary Figures and Tables

**Figure A.1.** Histograms of Effort in the Real-Effort Slider-Task, by Session.



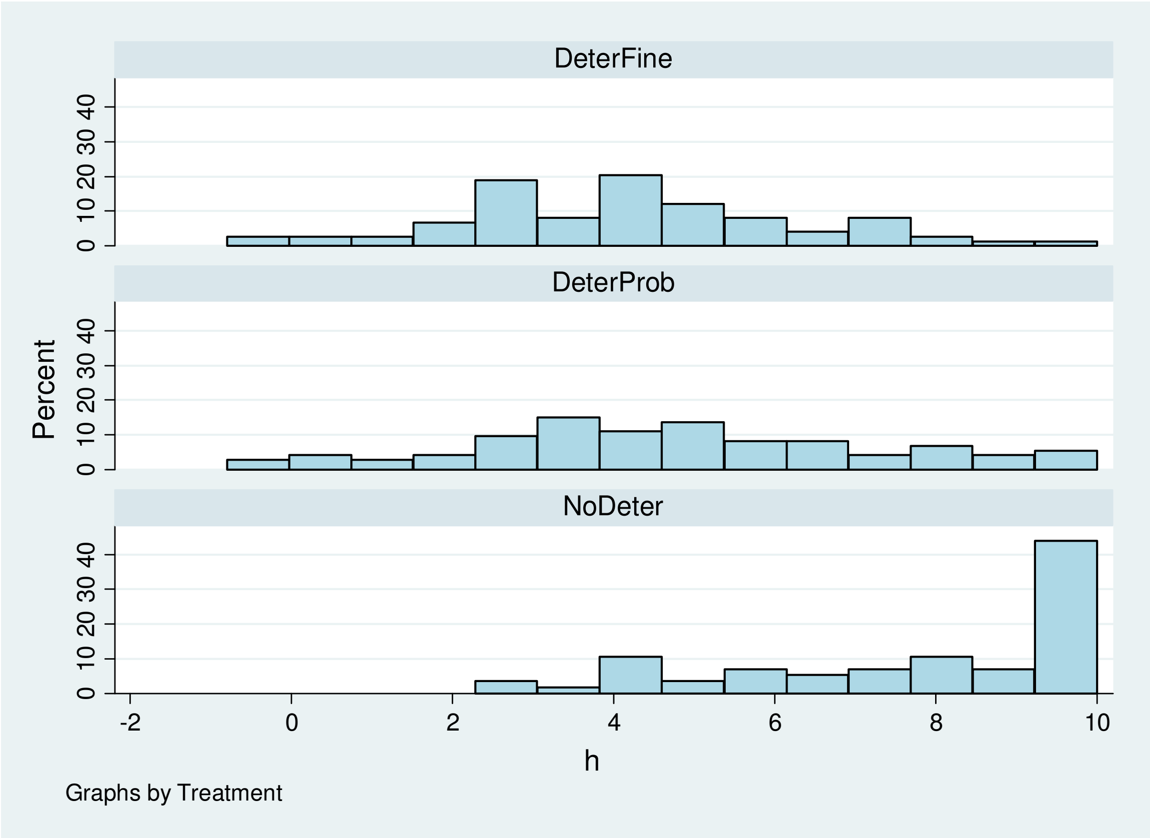

Note: A Kruskal-Wallis test cannot reject the null hypothesis of equality of populations (p = 0.4227). Mean effort over all sessions is 22.90.

**Figure A.2.** Histograms of Perceived Fairness of the Slider Task, by Session.




Note: A Kruskal-Wallis test cannot reject the null hypothesis of equality of populations (p = 0.4985). The answers were given on a five-point scale directly after the task and before subjects received information on their earnings in part 1. Mean perception of fairness over all sessions is 3.73.

**Figure A.3.** Histograms of Stealing in the First Five Periods, by Treatment.




Note: The above histograms display average stealing over the first five periods by treatment. It includes the average amount stolen by a player 1, such that there is one observation per player-1 subject.

**Table A.1.** Pairwise Correlation of Criminals' Emotions, Treatment *NoDeter*.

| Emotions | Anger | Shame | Pride | Disappointment | Surprise | Joy | Contempt | Envy | Irritation | Guilt | Regret | Admiration | Pity | Gratitude | Gloating |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Anger | 1.000 | | | | | | | | | | | | | | |
| Shame | **0.578** | 1.000 | | | | | | | | | | | | | |
| Pride | **0.309** | 0.187 | 1.000 | | | | | | | | | | | | |
| Disappointment | **0.444** | **0.581** | 0.110 | 1.000 | | | | | | | | | | | |
| Surprise | **0.280** | **0.478** | 0.095 | **0.431** | 1.000 | | | | | | | | | | |
| Joy | -0.066 | 0.135 | **0.369** | -0.084 | 0.107 | 1.000 | | | | | | | | | |
| Contempt | **0.492** | **0.485** | 0.258 | **0.661** | **0.390** | -0.015 | 1.000 | | | | | | | | |
| Envy | **0.717** | **0.503** | **0.266** | **0.573** | 0.243 | -0.047 | **0.512** | 1.000 | | | | | | | |
| Irritation | **0.689** | **0.587** | 0.135 | **0.743** | **0.457** | -0.166 | **0.715** | **0.775** | 1.000 | | | | | | |
| Guilt | **0.467** | **0.719** | **0.267** | **0.398** | **0.424** | 0.230 | **0.360** | **0.439** | **0.398** | 1.000 | | | | | |
| Regret | **0.279** | **0.547** | 0.053 | **0.528** | **0.385** | 0.145 | **0.474** | **0.360** | **0.382** | **0.558** | 1.000 | | | | |
| Admiration | **0.368** | **0.468** | 0.204 | **0.717** | **0.547** | 0.060 | **0.660** | **0.489** | **0.689** | **0.395** | **0.507** | 1.000 | | | |
| Pity | 0.185 | **0.416** | 0.124 | 0.134 | **0.309** | 0.218 | 0.176 | 0.175 | 0.030 | **0.381** | **0.456** | 0.154 | 1.000 | | |
| Gratitude | 0.022 | **0.366** | **0.318** | **0.331** | **0.486** | **0.401** | 0.220 | 0.085 | 0.193 | 0.247 | 0.147 | **0.384** | **0.342** | 1.000 | |
| Gloating | 0.128 | **0.287** | 0.223 | **0.265** | 0.204 | **0.289** | **0.538** | 0.113 | 0.092 | **0.319** | **0.309** | **0.277** | **0.388** | 0.213 | 1.000 |
| Sadness | **0.585** | **0.680** | 0.157 | **0.669** | **0.486** | -0.183 | **0.617** | **0.655** | **0.781** | **0.558** | **0.535** | **0.578** | **0.262** | 0.247 | 0.209 |

Note: Significant correlation between two emotions at the 5% level is noted in **bold** numbers.

**Table A.2.** Pairwise Correlation of Criminals' Emotions, Treatments *DeterFine* and *DeterProb*.

| Emotions | Anger | Shame | Pride | Disappointment | Surprise | Joy | Contempt | Envy | Irritation | Guilt | Regret | Admiration | Pity | Gratitude | Gloating |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Anger | 1.000 | | | | | | | | | | | | | | |
| Shame | **0.206** | 1.000 | | | | | | | | | | | | | |
| Pride | 0.124 | 0.058 | 1.000 | | | | | | | | | | | | |
| Disappointment | **0.489** | **0.270** | 0.109 | 1.000 | | | | | | | | | | | |
| Surprise | **0.194** | **0.285** | 0.160 | **0.331** | 1.000 | | | | | | | | | | |
| Joy | -0.147 | -0.024 | **0.518** | 0.029 | **0.209** | 1.000 | | | | | | | | | |
| Contempt | **0.285** | **0.258** | 0.121 | **0.319** | **0.362** | -0.041 | 1.000 | | | | | | | | |
| Envy | **0.417** | **0.184** | **0.187** | **0.417** | **0.184** | 0.056 | **0.284** | 1.000 | | | | | | | |
| Irritation | **0.631** | **0.178** | 0.112 | **0.682** | **0.241** | -0.084 | **0.312** | **0.537** | 1.000 | | | | | | |
| Guilt | 0.158 | **0.528** | -0.028 | **0.281** | **0.213** | -0.045 | **0.341** | **0.216** | **0.264** | 1.000 | | | | | |
| Regret | **0.201** | **0.308** | 0.111 | **0.441** | **0.333** | 0.005 | **0.293** | **0.373** | **0.408** | **0.495** | 1.000 | | | | |
| Admiration | **0.197** | **0.258** | 0.151 | 0.130 | **0.205** | 0.064 | **0.314** | **0.174** | 0.105 | **0.297** | 0.204 | 1.000 | | | |
| Pity | -0.033 | **0.333** | 0.071 | -0.022 | 0.122 | **0.210** | **0.167** | -0.031 | -0.062 | **0.321** | 0.199 | 0.126 | 1.000 | | |
| Gratitude | -0.098 | **0.171** | **0.315** | 0.060 | **0.288** | **0.488** | 0.045 | -0.009 | -0.147 | 0.138 | 0.075 | **0.270** | **0.243** | 1.000 | |
| Gloating | 0.120 | 0.067 | **0.286** | 0.055 | 0.063 | **0.311** | **0.239** | **0.189** | 0.129 | 0.117 | 0.074 | 0.114 | 0.153 | **0.207** | |
| Sadness | **0.346** | **0.302** | **0.205** | **0.529** | **0.284** | 0.021 | **0.419** | **0.450** | **0.576** | **0.255** | 0.481 | **0.203** | 0.102 | 0.036 | 0.142 |

Note: Significant correlation between two emotions at the 5% level is noted in **bold** numbers.

## 6.6 Appendix B: English Translation of the Experimental Instructions

*[Sample instructions for the treatment with* DeterFine *in the first five periods and* NoDeter *in the second five periods. The spacing is slightly adjusted to make Appendix B more compact.]*

General Instructions for Participants

Welcome to the Experiment Laboratory!

You are now taking part in an economic experiment. You will be able to earn a considerable amount of money, depending on your decisions and the decisions of others. It is therefore important that you read these instructions carefully.
The instructions which we have distributed to you are solely for your private information. **It is prohibited to communicate with other participants during the experiment.** Should you have any questions please raise your hand and an experimenter will come to answer them. If you violate this rule, we will have to exclude you from the experiment and from all payments.
During the experiment you will make decisions **anonymously**. Only the experimenter knows your identity while your personal information is confidential and your decisions will not be traceable to your identity.
For your participation in this experiment you receive an initial endowment of **7 Euros**. The additional calculation of your payment depends on your decisions and the decisions of other participants. At the end of the experiment the payment will be made to you in cash.
The experiment consists of multiple, interrelated parts. All parts are payment relevant.

Information for Part 1

All participants work on the same task. The task is to earn points by locating sliders on the computer screen **exactly at '50'**. The sliders are adjustable between 0 and 100, i.e. 50 is the exact middle. Initially all sliders are located at 0. For each slider that you locate at 50, you will earn **one point**.

After all participants worked on the task, the sum of points per individual will be ranked by size. All participants that belong to the more successful half receive **10 Euros**, while all participants who belong to the less successful half receive **2 Euros**. These incomes will be transferred to the next parts of the experiment. Your final payment depends on all parts of the experiment.

*An example:*
Let us assume that four participants A, B, C and D earn the following points: A – 10 Points, B – 13 Points, C – 7 Points and D – 17 Points. Then the participants will be ranked accordingly:

1. Participant D (17 Points)
2. Participant B (13 Points)
3. Participant A (10 Points)
4. Participant C (7 Points)

Below you see the **slider screen:**


You can first try out the task in a **practice round**. This practice round is not payment relevant. Subsequently the **payment relevant Task** begins. While the practice round lasts 200 seconds, you will get **140 seconds** for the payment relevant task.

Please leave the keyboards turned around. The violation of this rule will lead to exclusion of the experiment. We will also ask you to answer three questions after the task.

Thereafter you will receive information for part 2 of the experiment.

Information for Part 2 *[Circulated after Part 1 was over]*

Please turn around your keyboard so that you can use it in part 2.

The participants will be split into **two groups**: **persons 1** and **persons 2**. If you belong to the more successful half of the participants in part 1 and therefore earned 10 Euros, you are person 1. If you belong to the less successful half of the participants and therefore earned 2 Euros, you are person 2. **In part 2 you will always stay in the same group, that in accordance with the result of part 1 you are always either person 1 or person 2.**

Part 2 consists of **10 periods** in which person 2 always faces the same decision making. In each period a person 1 and a person 2 are **randomly matched**. After a period is over, you will **never** again be matched with the same person in later periods.

Person 1 does **not** make an active decision and is only able to **guess** the decision of person 2.

**The Decision Making of Person 2**

Each person 2 faces the same decision making. Your task is to decide about a transfer between your account and the account of person 1. At the beginning of each period each person 2 has **2 Euros in her/his account from part 1**, while there are **10 Euros in the account of the randomly matched person 1**. The income of part 1 was therefore transmitted to part 2 for all participants.

Each person 2 needs to decide how many Euros she/he wants to transfer between her/his account and the account of person 1. The transfer may be between -2 and 10 Euros (only whole numbers). A negative transfer means that **person 2** would like to transfer money from her/his account to the account of **person 1**. A positive transfer means that he/she wants to transfer money from the account of person 1 to her own account.

*Two examples:*
Assume that you are person 2 and that you want to transfer 2 Euros to your account from the account of person 1. Then your transfer should be “2”.
If you would like to transfer 1 Euro from your account to the account of person 1, then your transfer should be “-1”.

Negative and neutral transfers have a **probability of success of 100 percent**, while positive transfers may **fail** with a certain probability. Failing means that person 2 does not receive the transfer. If a transfer fails, then person 2 will also need to pay a fee. This **fee** will be deducted from the account of person 2 and **not** transferred to the account of person 1.
The probabilities of success and the fees are dependent on the size of the positive transfer. The table below shows how high the probabilities and the fees are for different transfers:

| **Transfer** | **Probability of Success** | **Fee in Case of No Success** |
|---|---|---|
| -2 Euro | 100 % | - |
| -1 Euro | 100 % | - |
| 0 Euro | 100 % | - |
| 1 Euro | 50 % | 1.10 Euro |
| 2 Euro | 50 % | 1.15 Euro |
| 3 Euro | 50 % | 1.40 Euro |
| 4 Euro | 50 % | 1.85 Euro |
| 5 Euro | 50 % | 2.50 Euro |
| 6 Euro | 50 % | 3.35 Euro |
| 7 Euro | 50 % | 4.40 Euro |
| 8 Euro | 50 % | 5.65 Euro |
| 9 Euro | 50 % | 7.10 Euro |
| 10 Euro | 50 % | 8.75 Euro |

The income of person 2 is the sum of 7 Euros for participation, 2 Euros for the less successful slider task performance in part 1 and the consequences of her/his transfer in part 2:

**Income of Person 2 =**
7 Euros + 2 Euros + Transfer (if successful) - fee (if not successful)

The income of person 1 analogously is the sum of 7 Euros for participation, 10 Euros for the more successful slider task performance in part 1 and the consequences of the person 2's transfer:

**Income of Person 1 =**
7 Euros + 10 Euros – Consequences of the Transfer

After 5 periods we will ask you some questions. We ask you to take your time for the questions and think hard about your answers. Thereafter the experiment will continue with a similar decision making for person 2.
After the 10 periods of part 2 are over, the experiment will end. At the end of the experiment you will receive the payment of **one of the ten periods** of part 2 in cash. That period is chosen **randomly**. Therefore each period is potentially payment relevant. All payments are done in private and other participants will not see what you have earned.

Information on the Course of Events of the Experiment

In the beginning of each period you will see an **input screen**. This input screen is the same for all persons 2:

| Überweisung | Erfolgswahrscheinlichkeit | Gebühr bei keinem Erfolg |
|---|---|---|
| -2 Euro | 100 % | - |
| -1 Euro | 100 % | - |
| 0 Euro | 100 % | - |
| 1 Euro | 50 % | 1.10 Euro |
| 2 Euro | 50 % | 1.15 Euro |
| 3 Euro | 50 % | 1.40 Euro |
| 4 Euro | 50 % | 1.85 Euro |
| 5 Euro | 50 % | 2.50 Euro |
| 6 Euro | 50 % | 3.35 Euro |
| 7 Euro | 50 % | 4.40 Euro |
| 8 Euro | 50 % | 5.65 Euro |
| 9 Euro | 50 % | 7.10 Euro |
| 10 Euro | 50 % | 8.75 Euro |

Sie sind Person 2. Sie können zwischen Ihrem Konto und dem Konto einer Ihnen zufällig zugeordneten Person 1 Geld transferieren. Person 1 kann keinen Transfer durchführen.

Ihre Eingabe kann zwischen -2 und 10 Euro liegen.

Eine negative Zahl gibt an, dass Sie Geld von Ihrem Konto auf das Konto von Person 1 transferieren.

Eine positive Zahl gibt an, dass Sie Geld von dem Konto von Person 1 auf Ihr Konto transferieren.

Wieviele Euro möchten Sie zwischen sich und einer Ihnen zufällig zugeordneten Person 1 transferieren?

As described, the account of person 2 holds 2 Euros in each period, while the account of person 1 holds 10 Euros in each period. Person 2 makes a decision on the transfer between the two accounts by entering a whole number between -2 and 10 into the input window. You can click on this window with the mouse. When person 2 has made his/her decision, she/he needs to press the **OK-button**. When you press the OK-button you cannot change your decision in this period.

After all persons 2 have made their decisions, these decisions will be displayed to persons 1. Persons 1 will know how many Euros person 2 wanted to transfer, whether person 2 was successful and the incomes of both persons are at the end of the period.
Persons 2 will also know whether the transfers were successful and the incomes of both persons at the end of the period.

As described the **income of person 2** is

**7 Euros + 2 Euros + Transfer (if successful) - Fee (if not successful)**

Before the experiment will continue all participants need to answer control questions on the screen. These questions aim to familiarize you with the rules of the experiment.
Do you still have question concerning the experiment? In this case please raise your hand.

Additional Information for Part 2

*[Circulated after the first five periods and the emotions questionnaire were past]*

The **course** of part 2 **remains the same**. The decision making of part 2 however **slightly changes** for the next 5 periods: The **probability of success** of transfer from now on is always 100 %. This means that the **fees** for positive transfer are **abolished**. This also means that the **table** of the first 5 periods is not valid any longer and the decisions of person 2 will be **always** executed directly.

In case you have any questions concerning the changes and the next 5 periods, please raise your hand.

# Bibliography

Abbink, K., B. Irlenbusch, and E. Renner (2002). "An Experimental Bribery Game". *Journal of Law, Economics, and Organization*, vol. 18, pp. 428-454.

Abbink, K. and A. Sadrieh (2009). "The pleasure of being nasty". *Economics Letters*, vol. 105, pp. 306-308.

Ahn, T.K., R.M. Isaac and T.C. Salmon (2008). "Endogenous Group Formation". *Journal of Public Economic Theory*, vol. 10, pp. 171-194.

Ahn, T.K., R.M. Isaac and T.C. Salmon (2009). "Coming and going: Experiments on endogenous group sizes for excludable public goods". *Journal of Public Economics*, vol. 93, pp. 336-351.

Alevy, J.E., M.S. Haigh and J.A. List (2007). "Informational Cascades: Evidence from a Field Experiment with Financial Market Professionals". *The Journal of Finance*, vol. 62, pp. 151-180.

Altonji, J.G. and Pierret, C.R. (2001). "Employer Learning and Statistical Discrimination". *Quarterly Journal of Economics*, vol. 116(1), pp. 313-350.

Ambrus, A. and B. Greiner (2011). *Imperfect Public Monitoring with Costly Punishment – An Experimental Study*. Working Paper.

Anderson, D.A. (2002). "The Deterrence Hypothesis and Picking Pockets at the Pickpocket's Hanging". *American Law and Economics Review*, vol. 4, pp. 295-313.

Anderson, L.R. and S.L. Stafford (2003). "Punishment in a Regulatory Setting: Experimental Evidence from the VCM". *Journal of Regulatory Economics*, vol. 24, pp. 91-110.

Andrade, E.B. and D. Ariely (2009). "The enduring impact of transient emotions on decision making". *Organizational Behavior and Human Decision Processes*, vol. 109, pp. 1-8.

Andreoni, J. (1990). "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving". *The Economic Journal*, vol. 100, pp. 464-477.

Andreoni, J. (1991). "Reasonable Doubt and the Optimal Magnitude of Fines: Should the Penalty Fit the Crime?". *The RAND Journal of Economics*, vol. 22, pp. 385-395.

Andreoni, J. (1993). "An experimental test of the public goods crowding-out hypothesis". *American Economic Review*, vol. 83, pp. 1317–1327.

Andreoni, J. (1995). "Warm-Glow Versus Cold-Prickle: The Effects of Positive and Negative Framing on Cooperation in Experiments". *The Quarterly Journal of Economics*, vol. 110, pp. 1-21.

Arrow, K.J. (1972). "Models of job discrimination" in: A.H. Pascal (ed.), *Racial Discrimination in Economic Life*, pp. 83-102, Lexington, MA: Lexington Books.

Arrow, K.J. (1973). "The theory of discrimination" in: O. Ashenfelter and A. Rees (eds.), *Discrimination in Labor Markets*, pp. 3-33, Princeton: Princeton University Press.

Baker, T., A. Harel and T. Kugler (2004). "The Virtues of Uncertainty in Law: An Experimental Approach". *Iowa Law Review*, vol.89, pp. 443-494.

Bardsley, N. (2008). "Dictator game giving: altruism or artifact?". *Experimental Economics*, vol. 11, pp. 122-133.

Bardsley, N., R. Cubitt, G. Loomes, P. Moffatt, C. Starmer and R. Sugden (2010). *Experimental Economics: Rethinking the Rules*. Princeton: Princeton University Press.

Bauböck, R. and C. Joppke (2010). "*How liberal are citizenship tests?"*. EUI Working Papers 41. EUDO Citizenship Observatory.

Becker, G.S. (1957). The Economics of Discrimination, Chicago: The University of Chicago Press.

Becker, G.S. (1968). "Crime and Punishment: An Economic Approach". *The Journal of Political Economy*, vol. 76, pp. 169-217.

Bénabou, R. and J. Tirole (2006). "Incentives and Prosocial Behavior". *The American Economic Review*, vol. 96, pp. 1652-1678.

Benz, M. and S. Meier (2008). "Do people behave in experiments as in the field? – evidence from donations". *Experimental Economics*, vol. 11, pp. 268-281.

Bergstrom, T., L. Blume, and H. Varian (1986). "On the Private Provision of Public Goods". *Journal of Public Economics*, vol. 29, pp. 25-49.

Bernoulli, D. (1738). "Specimen Theoriae Novae de Mensura Sortis". *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, vol. V; pp. 175-192 (translated into English by L. Sommer: "Exposition of a new theory on the measurement of risk". *Papers of the Imperial Academy of Sciences in Petersburg*, vol. V; printed in *Econometrica*, vol. 22, pp. 23-36).

Billig, M.G. and H. Tajfel (1973). "Social categorization and similarity in intergroup behavior". *European Journal of Social Psychology*, vol. 3, pp. 27-52.

Birkeland d.y., S., A.W. Cappelen, E.Ø. Sørensen and B. Tungodden (2011). *Bad guys or bad circumstances? An experimental study of prosocial motivation among criminals*. Norwegian School of Economics, discussion paper SAM 15 2011.

Block, M.K. and V.E. Gerety (1995). "Some experimental evidence on differences between student and prisoner reactions to monetary penalties and risk". *Journal of Legal Studies*, vol. 24, pp. 123-138.

Bohnet, I. and B.S. Frey (1999). "The sound of silence in prisoner's dilemma and dictator games". *Journal of Economic Behavior & Organization*, vol. 38, pp. 43-57.

Bolton, G.E. and A. Ockenfels (2000). "ERC: A Theory of Equity, Reciprocity, and Competition". *American Economic Review*, vol. 90, pp. 166-193.

Bosman, R. and F. van Winden (2002). "Emotional hazard in a power-to-take experiment". *The Economic Journal*, vol. 112, pp. 147-169.

Bowles, S. (1998). "Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions". *Journal of Economic Literature*, vol. 36, pp. 75-111.

Bowles, S. and S. Polanía-Reyes (2012). "Economic Incentives and Social Preferences: Substitutes or Complements?". *Journal of Economic Literature*, vol. 50, pp. 368-425.

Braithwaite, J. (2002). *Restorative Justice and Responsive Regulation*. New York: Oxford University Press.

Brekke, K.A., K.E. Hauge, J.T. Lind and K. Nyborg (2011). "Playing with the good guys: A public good game with endogenous group formation". *Journal of Public Economics*, vol. 95, pp. 1111-1118.

Brekke, K.A., S. Kverndokk and K. Nyborg (2003). "An economic model of moral motivation". *Journal of Public Economics*, vol. 87, pp. 1967-1983.

Brock, J.M., A. Lange and E.Y. Ozbay (2013). "Dictating the Risk: Experimental Evidence on Giving in Risky Environments". *The American Economic Review*, vol. 103, pp. 415-437.

Brosig, J., J. Weimann and A. Ockenfels (2003). "The Effect of Communication Media on Cooperation". *German Economic Review*, vol. 4, pp. 217-241.

Brubaker, R. (2001). "The return of assimilation? Changing perspectives on immigration and its sequels in France, Germany, and the United States". *Ethnic and Racial Studies,* vol. 24 No. 4, 2001 pp. 531-548

Buckley, E. and R. Croson (2006). "Income and wealth heterogeneity in the voluntary provision of linear public goods". *Journal of Public Economics*, vol. 90, pp. 935-955.

Bundesagentur für Arbeit (2011). "Perspektive 2025: Fachkräfte für Deutschland". Available at http://www.arbeitsagentur.de.

Burke, W.J. (2009). "Fitting and interpreting Cragg's tobit alternative using Stata". *Stata Journal*, vol. 9, pp. 584-592.

Cameron, L., A. Chaudhuri, N. Erkal, and L. Gangadharan (2009). "Propensities to engage in and punish corrupt behavior: Experimental evidence from Australia, India, Indonesia and Singapore". *Journal of Public Economics*, vol. 93, pp. 843-851.

Cameron, S. (1988). "The Economics of Crime Deterrence: A Survey of Theory and Evidence". *Kyklos*, vol. 41, pp. 301-323.

Cappelen, A.W., J. Konow, E. Ø. Sørensen and B. Tungodden (2013). "Just Luck: An Experimental Study of Risk Taking and Fairness". *The American Economic Review*, vol. 103, pp. 1398-1413.

Carrera, S., E. Guild and K. Groenendijk (2009). "Illiberal Liberal States. Immigration, Citizenship and Integration in the EU". Surren: Ashgate.

Chan, K.S., R. Godby, S. Mestelman and R.A. Muller (2002). "Crowding-out voluntary contributions to public goods". *Journal of Economic Behavior and Organization*, vol. 48, pp. 305-317.

Chan, K.S., S. Mestelman, R. Moir, and R.A. Muller (1999). "Heterogeneity and the voluntary provision of public goods". *Experimental Economics*, vol. 2, pp. 5-30.

Chaudhuri, A. (2011). "Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature". *Experimental Economics*, vol. 14, pp. 47-83.

Cherry, T.L., S. Kroll, and J. Shogren (2005). "The impact of endowment heterogeneity and origin on public good contributions: Evidence from the lab". *Journal of Economic Behavior & Organization*, vol. 57, pp. 357-365.

Chmura, T., C. Engel, M. Englerth and T. Pitz (2010). At the Mercy of the Prisoner Next Door – Using an Experimental Measure of Selfishness as a Criminological Tool. MPI Bonn 2010/27.

Cinyabuguma, M., T. Page and L. Putterman (2005). "Cooperation under the threat of expulsion in a public goods experiment". *Journal of Public Economics*, vol. 89, pp. 1421-1435.

Cleave, B.L., N. Nikiforakis and R. Slonim (2013). "Is there selection bias in laboratory experiments? The case of social and risk preferences". *Experimental Economics*, vol. 16, pp. 372-382.

Coase, R. (1960). "The problem of social cost". *Journal of Law and Economics*, vol. 3, pp. 1-44.

Cookson, R. (2000). "Framing Effects in Public Good Experiments". *Experimental Economics*, vol. 3, pp. 55-79.

Coricelli, G., D. Fehr and G. Fellner (2004). "Partner Selection in Public Goods Experiments". *Journal of Conflict Resolution*, vol. 48, pp. 356-378.

Corman, H. and H.N. Mocan (2000). "A Time-Series Analysis of Crime, Deterrence, and Drug Abuse in New York City". *The American Economic Review*, vol. 90, pp. 584-604.

Cox, J. C. and V. Sadiraj (2007). "On modeling voluntary contributions to public goods". *Public Finance Review,* vol. 35, pp. 311-332.

Cox, J. C., E. Ostrom, V. Sadiraj, and J. M. Walker (2013). "Provision versus Appropriation in Symmetric and Asymmetric Social Dilemmas". *Southern Economic Journal,* vol. 79, 496-512.

Cragg, J.G. (1971). "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods". *Econometrica*, vol. 39, pp. 829-844.

Croson, R.T.A. (2000). "Feedback in voluntary contribution mechanisms: an experiment in team production". *Research in Experimental Economics*, vol. 8, pp. 85-97.

Cubitt, R.P., M. Drouvelis and S. Gächter (2011). "Framing and free riding: emotional responses and punishment in social dilemma games". *Experimental Economics*, vol. 14, pp. 254-272.

Dana, J., R.A. Weber and J.X. Kuang (2007). "Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness." *Economic Theory*, vol. 33, pp. 1432-1479.

Dave, C., C.C. Eckel, C.A. Johnson and C. Rojas (2010). "Eliciting risk preferences: When is simple better?". *Journal of Risk and Uncertainty*, vol. 41, pp. 219-243.

DeAngelo, G. and G. Charness (2012). "Deterrence, expected cost, uncertainty and voting: Experimental evidence". *Journal of Risk and Uncertainty*, vol. 44, pp. 73-100.

Dohmen, T., A. Falk, D. Huffman and U. Sunde (2010). "Are Risk Aversion and Impatience Related to Cognitive Ability?". *The American Economic Review*, vol. 100, pp. 1238-1260.

Dufwenberg, M., S. Gächter and H. Hennig-Schmidt (2011). "The framing of games and the psychology of play". *Games and Economic Behavior*, vol. 73, pp. 459-478.

Dufwenberg, M. and G. Kirchsteiger (2004). "A theory of sequential reciprocity". *Games and Economic Behavior*, vol. 47, pp. 268-298.

Ehrhart, K.-M. and C. Keser (1999). "Mobility and Cooperation: On the Run". CIRANO Working Paper.

Ehrlich, I. (1973). "Participation in illegitimate activities: a theoretical and empirical investigation". *Journal of Political Economy*, vol. 81, pp. 521-565.

Elster, J. (1998). "Emotions and Economic Theory". *Journal of Economic Literature*, vol. 36, pp. 47-74.

European Union (2009). "Council Directive 2009/50/EC of 25 May 2009 on the conditions of entry and residence of third-country nationals for the purposes of highly qualified employment". Available at http://eur-lex.europa.eu/.

Falk, A. and U. Fischbacher (2006). "A theory of reciprocity". *Games and Economic Behavior*, vol. 54, pp. 293-315.

Falk, A. and M. Kosfeld (2006). "The Hidden Costs of Control". *The American Economic Review*, vol. 96, pp. 1611-1630.

Falk, A., S. Meier and C. Zehnder (2011). "Do Lab Experiments Misrepresent Social Preferences? The case of self-selected student samples". *Journal of the European Economic Association*, forthcoming.

Fehr, E. and K.M. Schmidt (1999). "A Theory of Fairness, Competition and Cooperation." *Quarterly Journal of Economics*, vol. 114, pp. 817-868.

Fehr, E. and S. Gächter (2000). "Cooperation and Punishment in Public Goods Experiments". *The American Economic Review*, vol. 90, pp. 980-994.

Fischbacher, U. (2007). "Z-Tree: Zurich Toolbox for Ready-made Economic Experiments". *Experimental Economics*, vol. 10, pp. 171-178.

Fischbacher, U. and S. Gächter (2010). "Social preferences, beliefs, and the dynamics of free riding in public goods experiments". *American Economic Review* vol. 100, pp. 541-556.

Fischbacher, U., S. Gächter and E. Fehr (2001). "Are people conditionally cooperative? Evidence from a public goods experiment". *Economics Letters*, vol. 71, pp. 397-404.

Fisher, J., R.M. Isaac, J.W. Schatzberg, and J.M. Walker (1995). "Heterogeneous demand for public goods: Behavior in the voluntary contributions mechanism". *Public Choice*, vol. 85, pp. 249-266.

Fisman, R. and E. Miguel (2007). "Corruption, Norms, and Legal Enforcement: Evidence from Diplomatic Parking Tickets". *Journal of Political Economy*, vol. 115, pp. 1020-1048.

Frank, R.H. (1988). *Passions within Reason – The Strategic Role of the Emotions*. New York: Norton.

Frey, B.S. and R. Jegen (2001). "Motivation crowding theory". *Journal of Economic Surveys*, vol. 15, pp. 589-611.

Frey, B.S. and S. Meier (2004). "Social Comparisons and Pro-Social Behavior: Testing 'Conditional Cooperation' in a Field Experiment". *The American Economic Review*, vol. 94, pp. 1717-1722.

Friesen, L. (2012). "Certainty of Punishment versus Severity of Punishment: An Experimental Investigation". *Southern Economic Journal*, vol. 79, pp. 399-421.

Gächter, S., E. Renner and M. Sefton (2008). "The Long-Run Benefits of Punishment". *Science*, vol. 322, p. 1510.

Garoupa, N. (1997). "The theory of optimal law enforcement". *Journal of Economic Surveys*, vol. 111, pp. 267-295.

Garoupa, N. (2001). "Optimal magnitude and probability of fines". *European Economic Review*, vol. 45, pp. 1765-1771.

Geraats, P.M. (2002). "Central Bank Transparency". *The Economic Journal*, vol. 112, pp. 533-565.

Gilbert, P. (2003). "Evolution, Social Roles, and the Differences in Shame and Guilt". *Social Research*, vol. 70, pp. 1205-1230.

Gill, D. and V. Prowse (2011). *A Novel Computerized Real Effort Task Based on Sliders*. IZA DP No. 5801.

Gill, D. and V. Prowse (2012). "A Structural Analysis of Disappointment Aversion in a Real Effort Competition". *The American Economic Review*, vol. 102, pp. 469-503.

Glöckner, A., B. Irlenbusch, S. Kube, A. Nicklisch, and H.-T. Normann (2011). "Leading with(out) sacrifice? A public-goods experiment with a privileged player". *Economic Inquiry*, vol. 49, pp. 591-597.

Gneezy, U. and A. Rustichini (2000a). "A fine is a Price". *Journal of Legal Studies*, vol. 29, pp. 1-18.

Gneezy, U. and A. Rustichini (2000b). "Pay Enough or Don't Pay at All". *Quarterly Journal of Economics*, vol. 115, pp. 791-810.

Gneezy, U., J.A. List and M.K. Price (2012). *Toward an understanding of why people discriminate: evidence from a series of natural field experiments*. Working Paper No. 17855, National Bureau of Economic Research.

Gneezy, U., S. Meier and P. Rey-Biel (2011). "When and Why Incentives (Don't) Work to Modify Behavior". *Journal of Economic Perspectives*, vol. 25, pp. 191-210.

Goodman, Sara Wallace (2010), "Naturalisation Policies in Europe: Exploring Patterns of Inclusion and Exclusion". *EUDO Citizenship Observatory*. Available at http://eudo-citizenship.eu/docs/7-Naturalisation%20Policies%20in%20Europe.pdf.

Grechenig, K., A. Nicklisch and C. Thöni (2010). "Punishment Despite Reasonable Doubt – A Public Goods Experiment with Sanctions Under Uncertainty". *Journal of Empirical Legal Studies*, vol. 7, pp. 847-867.

Greiner, B. (2004). "An Online Recruitment System for Economic Experiments", in: Kremer, K. and V. Macho (eds.), *Forschung und wissenschaftliches Rechnen 2003*. GWDG Bericht 63, Göttingen: Ges. für Wiss. Datenverarbeitung, pp. 79-93.

Groenendijk, K. (2011). "From Assisting to Requiring Integration: Selective Citizenship Policies in the Netherlands". In: OECD: *Naturalisation: A passport for better integration of immigrants?*, Chapter 15.

Gunnthorsdottir, A., R. Vragov, S. Seifert and K. McCabe (2010). "Near-efficient equilibria in contribution-based competitive grouping". *Journal of Public Economics*, vol. 94, pp. 987-994.

Gürerk, Ö., B. Irlenbusch and B. Rockenbach (2006). "The Competitive Advantage of Sanctioning Institutions". *Science*, vol. 312, pp. 108-111.

Gürerk, Ö., B. Irlenbusch and B. Rockenbach (2011). "*Voting with Feet – Community Choice in Social Dilemmas"*. Working paper.

Güth, W., R. Schmittberger and B. Schwarze (1982). "An experimental analysis of ultimatum bargaining". *Journal of Economic Behavior & Organization*, vol. 3, pp. 367-388.

Haidt, J. (2003). "The moral emotions". In: R.J. Davidson, K.R. Scherer and H.H. Goldsmith (eds.), *Handbook of Affective Sciences*, pp. 852-870.

Harbaugh, W.T., N. Mocan and M.S. Visser (2011). *Theft and Deterrence*. IZA DP No. 5813.

Harel, A. and U. Segal (1999). "Criminal Law and Behavioral Law and Economics: Observations on the Neglected Role of Uncertainty in Deterring Crime". *American Law and Economics Review*, vol. 1, pp. 276-312.

Harrison, G.W. and J.A. List (2004). "Field Experiments". *Journal of Economic Literature*, vol. 42, pp. 1009-1055.

Healy, C. and D. Reichel (2013). *Earning Rights: Economic Status and Access to Citizenship*. ICMPD Working Paper 05.

Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis and R. McElreath (2001). "In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies". *The American Economic Review*, vol. 91, pp. 73-78.

Herrmann, B., C. Thöni, and S. Gächter (2008). "Antisocial Punishment Across Societies". *Science*, vol. 319, pp. 1362-1367.

Heyman, J. and D. Ariely (2004). "Effort for Payment – A Tale of Two Markets". *Psychological Science*, vol. 15, pp. 787-793.

Holmström, B. (1979). "Moral Hazard and Observability". *Bell Journal of Economics*, vol. 10, pp. 74-91.

Home Office (2012). "Radical immigration changes to reform family visas. Securing borders and reducing immigration". Last accessed at 25.04.2013 at https://www.gov.uk/government/news/radical-immigration-changes-to-reform-family-visas.

Hopfensitz, A. and E. Reuben (2009). "The importance of emotions for the effectiveness of social punishment". *The Economic Journal*, vol. 119, pp. 1534-1559.

Hume, D. (1739-40). *A Treatise of Human Nature*. Books 1-3. Oxford: Clarendon Press, 2007.

Isaac, R.M. and J.M. Walker (1988). "Communication And Free-Riding Behavior: The Voluntary Contribution Mechanism". *Economic Inquiry*, vol. 26, pp. 585-608.

Isaac, R.M., J.M. Walker and A.W. Williams (1994). "Group size and the voluntary provision of public goods". *Journal of Public Economics*, vol. 54, pp. 1-36.

Joppke, C. (2004). "The retreat of multiculturalism in the liberal state: Theory and Policy". *The British Journal of Sociology*, vol. 55, pp. 237-257.

Joppke, C. (2007). "Beyond National Models: Civic Integration Policies for Immigrants in Western Europe". *West European Politics*, vol. 30, pp. 1-22.

Kahneman, D., J.L. Knetsch and R.H. Thaler (1986). "Fairness as a Constraint on Profit Seeking: Entitlements in the Market". *The American Economic Review*, vol. 76, pp. 728-741.

Khadjavi, M. and A. Lange (2013). "Prisoners and their Dilemma". *Journal of Economic Behavior and Organization*, vol. 92, pp. 163-175.

Kolodziej, D. (2012). "Fachkräftemangel in Deutschland. Statistiken, Studien und Strategien." *Wissenschaftliche Dienste des Deutschen Bundestages*, Infobrief WD 6 – 3010-189/11, Berlin 2012.

Kotchen, M.J. (2009), "Voluntary Provision of Public Goods for Bads: A Theory of Environmental Offsets". *Economic Journal*, vol. 119, pp. 883-899.

Kyckelhahn, T. and T. Martin (2013). *Justice Expenditure and Employment Extracts*. Bureau of Justice Statistics, U.S. Department of Justice, Washington DC.

Ledyard, J.O. (1995). "Public Goods: A Survey of Experimental Research". In: Kagel, J.H. and A.E. Roth (eds.), *Handbook of Experimental Economics*. Princeton: Princeton University Press, pp. 111-194.

Levine, D.K. (1998). "Modeling Altruism and Spitefulness in Experiments". *Review of Economic Dynamics*, vol. 1, pp. 593-622.

Levitt, S.D. (1997). "Using Electoral Cycles and Police Hiring to Estimate the Effect of Police on Crime". *The American Economic Review*, vol. 87, pp. 270-290.

Levitt, S.D. and T.J. Miles (2007). "Empirical study of criminal punishment". In: Polinsky, A.M. and S. Shavell (eds.), *Handbook of Law and Economics, Volume 1*. Chapter 7. Amsterdam: Elsevier.

List, J.A. (2004). "The nature and extent of discrimination in the marketplace: evidence from the field", *Quarterly Journal of Economics*, vol. 119, pp. 49-89.

List, J.A. (2007). "On the Interpretation of Giving in Dictator Games". *Journal of Political Economy*, vol. 115, pp. 482-493.

List, J.A. and T.L. Cherry (2008). "Examining the role of fairness in high stakes allocation decisions". *Journal of Economic Behavior & Organization*, vol. 65, pp. 1-8.

Loewenstein, G. (2000). "Emotions in Economic Theory and Economic Behavior". *American Economic Review*, vol. 90, pp. 426-432.

Massey, D.S. and K.E. Espinosa (1997). "What's Driving Mexico-U.S. Migration? A Theoretical, Empirical, and Policy Analysis". *American Journal of Sociology*, vol. 102, pp. 939-999.

Mauro, P. (1995). "Corruption and Growth". *The Quarterly Journal of Economics*, vol. 110, pp. 681-712.

Mayda, A.M. (2006). "Who is against immigration? A cross-country investigation of individual attitudes toward immigrants". *The Review of Economics and Statistics*, vol. 88, pp. 510-530.

Meier, S. (2007). "A Survey of Economic Theories and Field Evidence on Pro-Social Behavior". In: Frey, B.S. and A. Stutzer (eds.), *Economics and Psychology – A Promising New Cross-Disciplinary Field*. Cambridge: MIT Press, pp. 51-87.

Mill, J.S. (1859). *On Liberty*. Oxford: Oxford University Press, 1991.

Mo, P.H. (2001). "Corruption and Economic Growth". *Journal of Comparative Economics*, vol. 29, pp. 66-79.

Nagin, D.S. and G. Pogarsky (2003). "An experimental investigation of deterrence: cheating, self-serving bias, and impulsivity". *Criminology*, vol. 41, pp. 167-194.

Nikiforakis, N. (2008). "Punishment and counter-punishment in public good games: Can we really govern ourselves?" *Journal of Public Economics*, vol. 92, pp. 91-112.

Nikiforakis, N. (2010). "Feedback, punishment and cooperation in public good experiments". *Games and Economic Behavior*, vol. 68, pp. 689-702.

Obama, B. (2009). "Memorandum for the heads of executive departments and agencies - Subject: Transparency and open government". In: *General Services Administration: Transparency and Open Government, Intergovernmental Solutions Division.* GSA Office of Citizens Services and Communications, Washington DC, p. 4.

OECD (2012a). "International Migration Outlook 2012". *OECD Publishing.*

OECD (2012b). "Settling In: OECD Indicators of Immigrant Integration 2012". *OECD Publishing.*

Ostrom, E., J. Walker and R. Gardner (1992). "Covenants with and without a sword: self-governance is possible". *American Political Science Review*, vol. 86, pp. 404-417.

Page, T., L. Putterman and B. Unel (2005). "Voluntary association in public goods experiments: reciprocity, mimicry and efficiency". *The Economic Journal*, vol. 115, pp. 1032-1053.

Palfrey, T.R. and J.E. Prisbrey (1997). "Anomalous Behavior in Public Goods Experiments: How much and Why?". *The American Economic Review*, vol. 87, pp. 829-846.

Park, E.-S. (2000). "Warm-glow versus cold-prickle: a further experimental study of framing effects on free-riding". *Journal of Economic Behavior & Organization*, vol. 43, pp. 405-421.

Pascouau, Y., Labayle, H. (2011). "Conditions for family reunification under strain. A comparative study in nine EU member states". Available at http://www.epc.eu/documents/uploads/pub_1369_conditionsforfamily.pdf.

Phelps, E.S. (1972). "The statistical theory of racism and sexism". *American Economic Review*, vol. 62, pp. 659-661.

Pigou, A.C. (1920). *The Economics of Welfare*. London: Macmillan.

Polinsky, A.M. and S. Shavell (1979). "The Optimal Tradeoff between the Probability and Magnitude of Fines". *The American Economic Review*, vol. 69, pp. 880-891.

Polinsky, A.M. and S. Shavell (1984). "The optimal use of fines and imprisonment". *Journal of Public Economics*, vol. 24, pp. 89-99.

Polinsky, A.M. and S. Shavell (2007). "The Theory of Public Enforcement of Law". In: Polinsky, A.M. and S. Shavell (eds.), *Handbook of Law and Economics, Volume 1*. Chapter 6. Amsterdam: Elsevier.

Prat, A. (2005). "The Wrong Kind of Transparency". *The American Economic Review*, vol. 95, pp. 862-877.

Rabin, M. (1993). "Incorporating Fairness into Game Theory and Economics". *The American Economic Review*, vol. 83, pp. 1281-1302.

Reuben, E. and A. Riedl (2009). "Public Goods Provision and Sanctioning in Privileged Groups". *Journal of Conflict Resolution*, vol. 53, pp. 72-93.

Reuben, E. and A. Riedl (2013). "Enforcement of Contribution Norms in Public Good Games with Heterogeneous Populations". *Games and Economic Behavior*, vol. 77, pp. 122-137.

Reuben, E. and F. van Winden (2010). "Fairness perceptions and prosocial emotions in the power to take". *Journal of Economic Psychology*, vol. 31, pp. 908-922.

Rick, S. and G. Loewenstein (2008). "The Role of Emotion in Economic Behavior". In: Lewis, M., J.M. Haviland-Jones and L. Feldman Barrett (eds.), *Handbook of Emotions*, Chapter 9. New York: Guilford Press.

Rizzolli, M. and L. Stanca (2012). "Judicial Errors and Crime Deterrence: Theory and Experimental Evidence". *Journal of Law and Economics*, vol. 55, pp. 311-338.

Roth, A.E. (1993). "The early history of experimental economics". *Journal of the History of Economic Thought*, vol. 15, pp. 184-209.

Sadrieh, A. and H. Verbon (2006). "Inequality, cooperation, and growth: An experimental study". *European Economic Review*, vol. 50, pp. 1197-1222.

Scheve, K.F. and M.J. Slaughter (2001). "Labor Market Competition and Individual Preferences over Immigration Policy". *The Review of Economics and Statistics*, vol. 83, pp. 133-145.

Schildberg-Hörisch, H. and C. Strassmair (2012). "An Experimental Test of the Deterrence Hypothesis". *Journal of Law, Economics, and Organization*, vol. 28, pp. 447-459.

Schulze, G.G. and B. Frank (2003). "Deterrence versus intrinsic motivation: Experimental evidence on the determinants of corruptibility". *Economics of Governance*, vol. 4, pp. 143-160.

Smith, V.L. (1994). "Economics in the Laboratory". *Journal of Economic Perspectives*, vol. 8, pp. 113-131.

Sonnemans, J., A. Schram and T. Offerman (1998). "Public good provision and public bad prevention: The effect of framing". *Journal of Economic Behavior & Organization*, vol. 34, pp. 143-161.

Stigler, G.J. (1970). "The Optimum Enfocement of Laws". *Journal of Political Economy*, vol. 78, pp. 526-536.

Stoop, J.T.R. (2013). "From the lab to the field: envelopes, dictators and manners". Forthcoming in *Experimental Economics*.

Stoop, J.T.R., C. Noussair and D.P van Soest (2012). "From the Lab to the Field: Cooperation among Fishermen". *Journal of Political Economy*, vol. 120, pp. 1027-1056.

Tajfel, H., M.G. Billig, R.P. Bundy and C. Flament (1971). "Social categorization and intergroup behavior". *European Journal of Social Psychology*, vol. 1, pp. 149-178.

Umbreit, M.S., R.B. Coates and B. Kalanj (1994). *Victim Meets Offender: The Impact of Restorative Justice and Mediation*. Monsey: Criminal Justice Press.

van Dijk, F., J. Sonnemans, and F. van Winden (2002). "Social ties in a public good experiment". *Journal of Public Economics*, vol. 85, pp. 275-299.

van Oers, R., E. Ersbøll and D. Kostakopoulou (2010). "A Re-definition of Belonging? Language and Integration Tests in Europe". Leiden; Boston: Martinus Nijhoff.

van Winden, F. and E. Ash (2012). "On the Behavioral Economics of Crime". *Review of Law & Economics*, vol. 8, pp. 181-213.

Walsh, C.E. (2007). "Optimal economic transparency". *International Journal of Central Banking*, vol. 3, pp. 5-36.

Weimann, J. (1994). "Individual behaviour in a free riding experiment". *Journal of Public Economics*, vol. 54, pp. 185-200.

Witte, A.D. (1980). "Estimating the Economic Model of Crime with Individual Data". *The Quarterly Journal of Economics*, vol. 94, pp. 57-84.

Yamagishi, T. (1986). "The provision of a sanctioning system as a public good". *Journal of Personality and Social Psychology*, vol. 51, pp. 110-116.

Zelmer, J. (2003). "Linear Public Goods Experiments: A Meta-Analysis". *Experimental Economics*, vol. 6, pp. 299-310.

Zizzo, D.J. (2010). "Experimenter demand effects in economic experiments". *Experimental Economics*, vol. 13, pp. 75-98.

Zussman, A. (2012). "Ethnic discrimination: Lessons from the Israeli online market for used cars". *Economic Journal*, forthcoming.