Institutional Consequences of Justice on Cooperation and Redistribution Systems

Universität Hamburg

Fakultät für Wirtschafts- und Sozialwissenschaften

Dissertation

Zur Erlangung der Würde einer Doktorin

der Wirtschafts- und Sozialwissenschaften

"Dr. rer. pol."

(gemäß der PromO vom 24. August 2010)

vorgelegt von

Marina Chugunova

aus Moskau, Russland

Hamburg, den 23. Juli 2018

Vorschlag für die Zusammensetzung der Prüfungskommission:

Vorsitzender:	Prof. Dr. Andreas Lange
Erstgutachter:	Prof. Dr. Kai-Uwe Schnapp
Zweitgutachter:	Prof. Dr. Stefan Traub
Weiteres Mitglied:	Prof. Dr. Andreas Nicklisch

Erstbetreuer:	Prof. Dr. Andreas Nicklisch
Zweitbetreuer:	Prof. Dr. Kai-Uwe Schnapp

Acknowledgements

First of all, I would like to thank Andreas Nicklisch and Kai-Uwe Schnapp for supporting me throughout this winding project and being great co-authors and teammates. I appreciate all your contributions in time, ideas, and funding to make my Ph.D. experience productive and stimulating. The joy and enthusiasm you have for research was very motivational for me.

I am greatly indebted to Wolfgang Luhan for his continuous encouragement and making sure that I stay sane at most of the time.

I was lucky to receive great support both from the faculty and members of the department of economics of the University of Hamburg as well as the University of Zürich. I would especially like to mention Roberto Weber, Eva Ranehill, Lea Heursen, Fanny Brun, Florian Schneider, Ivo Schurtenberger and Vanessa Valero.

In the course of creating this dissertation I greatly benefited from being a part of DFG FOR 2104. Our interdisciplinary discussions have greatly enriched this work.

Last but not least, I would like to thank Andreas Lange and Stefan Traub for their helpful comments and suggestions.

It goes without saying that I would not completed this work without the unconditional love and support of my family and friends.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

> If [justice] is removed, the great, the immense fabric of human society, that fabric which to raise and support seems in this world, if I may say so, have been the peculiar and darling care of Nature, must in a moment crumble into atoms.
>
> Adam Smith
> The Theory Of Moral Sentiments

Empirical evidence to the effect (e.g., Güth et al., (1982)) that people are not as rational in making their decisions as the concept of *homo oeconomicus* implies has revolutionized the field of economics producing a number of seminal works (e.g., Bolton and Ockenfels, 2000; Charness and Rabin, 2002; Fehr and Schmidt, 1999) which in different ways incorporate the role of fairness and justice into the utility function of an individual. The concepts of justice and fairness, however, are far from being universally agreed upon and are relevant for essentially any interaction involving at least two people. These facts explain why while giant strides have been made in recent years in understanding how fairness affects our decisions, we have not fully understood the phenomenon.[1]

This work contributes to the research on the importance of fairness for our societies and considers its institutional consequences. I explore if and to what extent considerations of justice affect the functioning of institutions. In chapters two and three I consider the environment in which people can choose which set of rules (i.e., which institution) will apply to them and whether fairness considerations play a role in this choice. In further chapters I consider how a specific notion of justice

[1] A comprehensive literature review is included in the upcoming chapters.

(i.e., needs-based justice) affects the economic outcomes in a fixed institutional framework.

In the second chapter (with Wolfgang Luhan and Andreas Nicklisch), we study how people migrate between different institutional environments. We apply a framework of open communities to public goods provision. In a laboratory experiment participants face a social dilemma situation (i.e., contradiction between self-interest and social welfare) and can change between communities with different cooperation-enhancing institutions: peer-reward and peer-punishment. Although the peer-reward institution is more profitable, we find that with time players abandon it and join a less profitable punishment institution. We analyze conditions that undermine the support for a reward community and trigger migration to a punishment system. The data suggest that inequality of income within the reward community and missing compensations for cooperators are key factors for the decision to migrate to the less profitable punishment community. That is, players are willing to forgo profits to be in a community with more equal distribution of outcomes and institutions that generate inequality may, therefore, lose its best members.

After establishing that inequality may be an important driver of community choice in the second chapter, in the third chapter I explore whether the institution that promotes cooperation and allows the enforcement of more equal outcomes will be chosen even if it is less efficient. A wide range of literature has compared the effect of centralized and decentralized (i.e., peer) sanctioning mechanisms on cooperation (e.g., Andreoni and Gee, 2012; Markussen et al., 2013). Two mechanisms function differently, and usually the former involves some provision costs and the latter is assumed to be freely available. Recent research by Kriss et al., (2016) and Markussen et al., (2013) has shown that the mere presence of a fixed provision cost may prevent players from using the institution regardless of how effective it is in promoting desirable behavior. The former paper shows that, instead of a centralized institution that comes at a fixed per round cost, players prefer a decentralized institution which does not involve provision costs. The latter paper finds that, due to the introduction of a negligible fixed cost of an institution in a weak link game, people stop using a cooperation promoting device despite the enormous potential gains. In light of these findings, the assumption of a

freely available decentralized institution appears to be problematic, since such an institution can nevertheless require some small fixed costs of staying informed about others' actions, having the sanctioning technology in place and maintaining it. I consider whether the aversion to fixed costs can discourage players from joining a decentralized institution (which is otherwise preferred) in a public goods game. I find that provision costs do not affect the likelihood of players joining a community with a cooperation-enhancing institution. That is, unlike in a weak link game, players are ready to bear the provision costs. Moreover I find that players value cooperation beyond the monetary profits it generates. Namely, they join a peer-punishment community with a provision cost and cooperate, even if it is strictly less profitable than joining an institution free environment where cooperation is unlikely.

Changes in the environment may affect the performance of the existing institutions and thus may be seen as policy levers. A framework of open communities allows us to explore the persistence of preferences, how institutional frameworks function not in isolation but in competition with each other, and what affects migration choices. The results of these studies are highly relevant in the times of ongoing migration and rising mobility. Although applicable for a wide variety of situations, there are environments where "voting with feet" is either not plausible or so costly that it is essentially impossible. If people do not have an opportunity to "vote with their feet" and have to stay in the unequal community, preferences for equality may be expressed in different ways. There is a broad consensus among researchers that increasing inequality may cause severe social and political unrest within a society (e.g., Piketty, 2014; Staab and Nachtwey, 2016; Stiglitz, 2012). Often governments attempt to curb inequality not by equalizing outcomes as such but by making sure that the needs of individuals are met through, for instance, the provision of unemployment benefits, family allowances, health subsidies etc. While largely influential in philosophy (Benbaji, 2006; Frankfurt, 2000; Sen, 1973, 1993, 2000), in economics the concept of needs-based justice has been largely overshadowed by equity and equality principles. Needs-based redistribution is based on the non-comparative understanding of justice, which argues that it is important to satisfy the needs of all members of the society up to a certain level. Thus, redistribution should be in favor of those who fail to meet this level themselves. In the fourth and

fifth chapters together with Andreas Nicklisch and Kai-Uwe Schnapp, we consider an environment where redistribution serves the purpose of satisfying the needs of randomly disadvantaged players who struggle to satisfy their needs. We start by considering redistribution through mandatory payments (i.e., taxes) and consider its effects on the labor supply choices of both parties of redistribution: taxpayers and transfer recipients. We deal with the effect on labor supply due to the fact that it is the main mechanism of generating welfare and it is likely to be affected by the changes in a tax regime. If people are purely selfish then taxation may have negative effects on the labor supply due to the income and substitution effects, and therefore lead to decrease in social welfare. However, if players have some form of other-regarding preferences that are important for their labor supply decisions, redistribution does not have to trigger negative welfare consequences.

In the first study (chapter four), we compare consequences on labor supply of needs-based taxation against non needs-based alternatives. In a real-effort experiment, subject's earned income is taxed and the revenues are spent differently with regard to the cause of redistribution. Since our core interest is the effect of the needs-based taxation we introduce a subsistence income threshold within the experiment. That is, players should meet an exogenously imposed threshold (i.e., satisfy their needs) to take part in the next production phase.We first test the effect of the needs threshold on the labor supply and then address the question whether the cause of taxation matters to the labor supply decisions. We vary the way taxes are spent from wasting tax revenues ineffectively to transferring the money to a systematically disadvantaged player who struggles to pass the needs threshold. Even with the low and non-binding threshold, we find that taxpayers slightly (however, insignificantly) increase the effort in response to the justified and efficient spending of tax revenues. Disadvantaged players (welfare recipients) strongly increase their performance if they expect a transfer.

In chapter five we look into an alternative wealth redistribution mechanism, namely voluntary donation. We consider if varying the timing of the donation decision relative to the production of a disposable income leads to higher voluntarily transfers. In the setup with subsistence income threshold players perform a real-effort task. Keeping the cause of redistribution constant (i.e. needs-based redistribution to a disadvantaged player), we vary the timing of the donation decision. An

advantaged player can assist a disadvantaged one in meeting the threshold by either transferring a chosen share of his earned income at the end of the production stage (when it is certain that the threshold is met) or commit to donate a certain percentage before production takes place. Such a design allows to vary how strong the entitlement towards one's earnings is. In the latter case, one donates from the future earnings and may feel less entitled and more willing to redistribute, while, in the former case, the opposite is true. We find that making a decision about future earnings decreases the entitlement feeling. That is, more players decide to donate in the commitment scheme. Moreover, higher donation rates in the treatment with committing in advance trigger the increase in labor supply of donors. We demonstrate that this result is not the artifact of the experimental design (i.e., players could have decreased their effort and would qualify for a further production even considering individually chosen donation rates) and is not driven by the reference dependence of income, which allows us to conclude that it is the "second-order" manifestation of their pro-sociality.

Thus, this work examines several instances in which considerations of justice may enhance or impede functioning of institutions. Due to the scope of the topic we are addressing, our findings generate new challenging questions and require further research in order to provide conclusive evidence.

Bibliography

Andreoni, James and Laura K Gee (2012). "Gun for hire: delegated enforcement and peer punishment in public goods provision". In: *Journal of Public Economics* 96.11, pp. 1036–1046.

Benbaji, Yitzhak (2006). "Sufficiency or priority?" In: *European Journal of Philosophy* 14.3, pp. 327–348.

Bolton, Gary E and Axel Ockenfels (2000). "ERC: A theory of equity, reciprocity, and competition". In: *The American Economic Review*, pp. 166–193.

Charness, Gary and Matthew Rabin (2002). "Understanding social preferences with simple tests". In: *The Quarterly Journal of Economics* 117.3, pp. 817–869.

Fehr, Ernst and Klaus M Schmidt (1999). "A theory of fairness, competition, and cooperation". In: *The Quarterly Journal of Economics* 114.3, pp. 817–868.

Frankfurt, Harry (2000). "Gleichheit und Achtung". In: *Gleichheit oder Gerechtigkeit. Texte der neuen Egalitarismuskritik*. Ed. by Angelika Krebs. Frankfurt a.M.: Suhrkamp Verlag, pp. 38–49.

Güth, Werner, Rolf Schmittberger, and Bernd Schwarze (1982). "An experimental analysis of ultimatum bargaining". In: *Journal of Economic Behavior & Organization* 3.4, pp. 367–388.

Kriss, Peter H, Andreas Blume, and Roberto A Weber (2016). "Coordination with decentralized costly communication". In: *Journal of Economic Behavior & Organization* 130, pp. 225–241.

Markussen, Thomas, Louis Putterman, and Jean-Robert Tyran (2013). "Self-organization for collective action: An experimental study of voting on sanction regimes". In: *Review of Economic Studies* 81.1, pp. 301–324.

Piketty, Thomas (2014). *Capital in the 21st Century*. Harvard University Press Cambridge, MA.

Sen, Amartya K. (1973). *1973. On Economic Inequality . Clarendon Press.* Oxford: Clarendon Press.

– (1993). "Capability and well-being". In: *The quality of life*. Ed. by Martha C. Nussbaum and Amartya K. Sen. Oxford: Clarendon Press, pp. 30–53.

Sen, Amartya K. (2000). "Social Justice and the Distribution of Income". In: *Handbook of Income Distribution*. Ed. by A. B. Atkinson and F. Bourguignon. Amsterdam: Elsevier, pp. 59–85.

Staab, Philipp and Oliver Nachtwey (2016). "Market and Labour Control in Digital Capitalism". In: *tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society* 14.2, pp. 457–474.

Stiglitz, Joseph E (2012). *The price of inequality: How today's divided society endangers our future*. WW Norton & Company.

Chapter 2

When to Leave Carrots for Sticks: On the Evolution of Sanctioning Institutions in Open Communities

Marina Chugunova, Wolfgang Luhan and Andreas Nicklisch[1]

Abstract

There is substantial evidence that punishment is more effective than rewards to maintain cooperation in social dilemmas. Yet, previous findings suggest that people generally dislike negative sanctioning institutions and avoid them when possible. We take a new perspective by directly comparing the migration between punishment and reward regimes in a social dilemma. In our laboratory experiments participants continuously "vote with their feet" by migrating between punishment and reward communities. In line with previous research, the vast majority of subjects in our experiment opts initially for the reward institution. Over time, however, more subjects start to join the less profitable punishment community. Analyzing the conditions which trigger migration, we show that full contributors are the first to migrate to the non-populated punishment communities. Followers cooperate almost fully without the requirement of actual punishment, thus reducing the welfare costs of the institution. Individual data suggest that income differences within communities and missing compensations for cooperators in the reward community are key factors for the decision to migrate.

Keywords: open communities; endogenous grouping; voting with feet; cooperation; sanctions; public goods

JEL-Classification: C72, C92, H41

[1]Financial support by the German Research Foundation (DFG) for Nicklisch (Ni 1610/1-1) is gratefully acknowledged.

2.1 Introduction

We face a large number of social dilemmas every day, ranging from local to international issues. Mitigation of income inequalities, the fight against terror, battling the spread of deadly diseases, and attempts to stop the progression of climate change are just a few examples of the latter type. The problem of how to promote pro-social behavior in this type of situations is consequently of pivotal importance for the existence of every society.

Among the multiplicity of means available to promote pro-social behavior in such situations, research has shown that (peer) punishment is particularly suited to foster cooperation (e.g., Fehr and Gächter, 2000, 2002; Gächter, Renner, et al., 2008). Indeed, a majority of social dilemma situations feature some sort of punishment institutions (see, e.g., Gardner et al., 1990). Yet, this observation is puzzling in itself. If asked for their preference, people mostly do not opt for a punishment regime, but for the complete opposite – a reward system which allows them to approve cooperative behavior (Gürerk et al., 2014; Sutter et al., 2010).[2] Such resentment towards sanctioning systems might stem from the possibility to misuse sanctions (e.g., Herrmann et al., 2008), the escalation of conflicts (e.g., Nikiforakis and Engelmann, 2011) or simply the efficiency loss due to destruction (e.g., Fischbacher and Gächter, 2010).

This raises the question under which conditions people start to realize the benefits of punishment – if they do at all. To study this question, we analyze how peer-punishment and peer-reward institutions evolve in direct competition with each other in a dynamic setting. Studying this type of migration flows and institutional competition seems to be of particular political and academic importance in times of increased mobility. For this purpose, we design an experiment that differs from previous approaches in that we do not compare simply the outcome of a punishment and a reward system in terms of the degrees of cooperation and the efficiency obtained (e.g., Sefton et al., 2007). Rather, we compare when and why

[2] An exception can be found in the study by Dickinson et al., (2015) who let (among others) police commissioners choose between playing repeatedly a public good game with a rewarding or with a punishment system. In their study the majority of subjects favors rewards over sanctions, but police commissioner are more likely to vote for sanctions.

people migrate between punishment and reward systems. We ask which factors lead to convergence to a choice of punishment systems, and which degrees of cooperation are achieved. This allows us to identify the groups of participants choosing certain environments in the first place, the determinants of migration, and the dynamic analysis of the institutions themselves. To the best of our knowledge, we are the first to analyze subjects' behavior in experimental public goods games, in which they repeatedly "vote with her feet"[3] (at no cost) and migrate between a reward and a punishment system.

Our study follows the literature that explores the effectiveness of either punishment in comparison to rewards or a combination of both mechanisms for solving social dilemmas (Andreoni et al., 2003; Sefton et al., 2007; Rockenbach and Milinski, 2006; Rand et al., 2009; Hauert, 2010; Hilbe and Sigmund, 2010; Milinski and Rockenbach, 2012). Similar to the setting in Gürerk et al., (2006, 2014), we implement migration in a repeated three-stage public goods game, extending it by a direct comparison of punishment and reward institutions. In the first stage, participants freely choose between two communities. In the second stage, participants play a public goods game within their community. In the last stage, participants are informed about the others' contributions and, depending on the community, may or may not sanction (through punishment or reward) other community members. We implemented two treatment conditions: in PN, participants choose between a community with punishment and a community without punishment. In PR, participants choose between a community with punishment and a community with reward. While PN aimed to provide a baseline in terms of cooperation rate, PR addresses our research question.

Our results confirm the previous results in PN (e.g., Gürerk et al., 2014). The vast majority of participants move to the punishment community where almost full cooperation is obtained, while contributions in the non-punishment community remain low throughout the experiment. In PR we find much richer dynamics.

[3]Following a similar intuition, Fehr and Williams, (2017) study which modification of punishment regimes (uncoordinated, coordinated or centralized) is the most attractive. In his seminal paper, Tiebout, (1956) first suggested that people "vote with their feet" for communities with an optimal provision of public goods.

Initially, almost all participants opt for the reward community, where contributions are stabilized at a constant rate.[4] However, half way through the course of the experiment, some participants in our setting start "exploring" the alternative community. In a sense, the existence of the reward community postpones the establishment of the punishment community. This saves punishment expenditures which is potentially of great importance for the overall efficiency of the public good (see, e.g., Gächter, Mengel, et al., 2017). The migrated players start contributing at high levels immediately almost without actual punishment required. High contributors are the first to migrate to the punishment community, retaining their high cooperation level. In the following periods, more participants join the punishment community and contribute at the maximum without previous punishment. A detailed analysis shows that the lack of rewards targeted at cooperators predominantly triggers the migration. When the migration occurs, however, the total profits in the reward community remain higher than those in the punishment community (due to a very favorable rewarding mechanism, details are discussed below), meaning that the players forgo higher profits for the sake of joining the community with a lower income inequality.[5]

Our findings highlight the importance of the concept of income inequality for the rise and fall of the institutional arrangements: people seem to start discovering the benefits of another institution if their current institution substantially harms their fairness sentiments. Participants who do not receive significant rewards for high contributions migrate to the punishment community. Our results show that the migration from the reward to the punishment community spurs cooperation purely based on the mere threat of punishment. That is, the existence of the unfair rewarding system remedies the welfare loss of the punishment system.

[4] As such, our findings complement the results of Dal Bó et al., (2016) who show that majority of subjects opt against a rule that implements direct negative effects, but would help to overcome a social dilemma.

[5] Dal Bó et al., (ibid.) also have one treatment condition with a repeated choice for or against the rule that overcomes the social dilemma. In line with our results, they find increasing approval rates for this rule over the course of the experiment. While they suggest that the gradual emergence of the rule results from the unfamiliarity that needs to be relieved over time, we show that those whose fairness concerns are presumably violated the most, first start exploring the new rule.

2.2 Design, Hypotheses and Procedures

2.2.1 Design

To facilitate the comparison of our results, we implemented a design that is similar in the basic structure and the parametrization of Gürerk et al., (2006, 2014). The experiment consists of three stages: *community selection, public goods game* and *sanctioning,* that are repeated for 30 periods. The experiment is implemented in the groups of 12 players who interact anonymously.

Community selection: We analyze two treatment conditions. In PR players choose between two communities: one with a punishment system and another with a reward system. In PN players choose between a community with a punishment system and a community without any sanctioning system.[6] In each treatment, players only interact within their chosen communities.

Public goods game: After choosing the community each player receives an endowment of x tokens. Players may simultaneously contribute an integer amount of $c_i (0 \leq c_i \leq x)$ to a common project. The player keeps the remaining endowment $(x - c_i)$. The sum of all contributions within one community, S, is multiplied by a factor a, and distributed evenly between all n members of the community (with $1/n < a/n < 1$) irrespective of the individual member's contribution.[7]

Sanctioning: After all the members of the community have made their contribution decision, they move to the sanctioning stage. Each player receives an additional endowment of y tokens as well as feedback on individual contributions of the other community members. In the punishment system players may assign punishment points to other community members reducing their own income by one and reducing the income of targeted community members by p tokens. The reward system functions similarly with assigned reward points reducing the own income by one and increasing the income of targeted community members by r tokens. The sanctioning decisions in both regimes are taken simultaneously by all players.

[6] The terms "reward", "punishment" as well as "sanctioning" were not used in the instructions or in the course of the experiment. See the instructions in the Appendix.

[7] If a community consist of only one member, there is no possibility for a common project and the the entire endowment x is transferred to the private account.

Since identification of group members is possible only via ID numbers and these are randomly reshuffled in every period, our design does not allow for reputation building. Therefore immediate action is required if sanctions are to be imposed. In the community without any sanctioning system members receive an additional endowment of y tokens each, but cannot spend these tokens on sanctions.

After assigning sanctioning points, players receive feedback on their individual performance in the current period and the average performance for the past behavior in both communities. Each player receives feedback on all individual performances for both groups, which includes the contribution, the income after the public good game, received and distributed points – if applicable – and the total payoff of the current period. Additionally, players see the average contribution, average received points and average period scores for both regimes for all previous periods.

The payoffs in tokens in the current period in the community without sanctioning (N) are

$$\pi_{iN} = (x - c_i + \frac{a}{n}S) + y,$$

where n equals the number of players choosing N in the current period ($1 < n \leq 12$). Therefore, there is no advantage of the group size.[8]

In the sanctioning communities the payoffs are altered by the sanctioning points distributed to other members, s_i, and the sanctions received from other members of the community, s_{-i}

$$\pi_{iP} = (x - c_i + \frac{a}{n}S) + (y - s_i - ps_{-i})$$
$$\pi_{iR} = (x - c_i + \frac{a}{n}S) + (y - s_i + rs_{-i})$$

The reward regime offers enormous efficiency gains depending on y and r in comparison to x and a. Following Gürerk et al., (2014), we choose the parameters

[8]The efficiency of investments, $\frac{a}{n}$, is constant and independent of group size. Therefore, the expected return of investment is independent of the groups size and should only be affected by the expected contributions of the other group members. Given that the members of any group are stemming from the same experimental population, the expected payoff is again independent of the size of the group. A small group in our experimental setting is ex ante expected to have the same distribution of high- and low contributions as a large group therefore nobody can expect to gain from being in a large or small group. The effect of the size of the group on punishing or rewarding behavior is equally ambiguous. The per-capita budget for sanctioning is constant and the distribution of willingness to sanction within a group is independent of the group size.

such that $x = 20$, $y = 20$, $a = 1.6$, $r = 3$, and $p = 3$. As a consequence, our experimental setting tests deliberately the extreme case where the potential benefits from only two players mutually rewarding each other (i.e., 60 tokens per player) outweigh almost twice the benefits from full cooperation of those two players in the public good game (i.e., 32 tokens per player). That is to say, we test migration to the punishment regime under utmost unfavorable conditions.

Participants play this game repeatedly for 30 periods. To reveal players' changing institutional preferences over time, we allow them to repeatedly select themselves at the start of the new period – prior to the contribution stage – into communities with different sanctioning systems. Migration between communities is costless.

2.2.2 Hypotheses

The focus of our study lies on the migration between institutions, its causes and the evolutionary dynamics of the systems. Applying existing social preference models does not allow us to produce predictions for these questions. We nevertheless discuss the implications of these models for contributions and sanctions in our setting before using them as well as previous empirical findings to develop our hypotheses.

Standard economic theory assumes perfect egoism and predicts complete inactivity in all of our communities, with no contributions and no sanctions. Fehr and Schmidt, (1999) and Gürerk et al., (2006), however, have shown that inequality aversion can motivate positive contributions in public good games with and without peer punishment opportunities and in the former case also positive punishment expenditures. Moreover, punishment may lead to superior contribution rates as compared to the no sanctioning environment. Inequality aversion models fail to explain positive reward expenditures as these potentially can increase inequality within the community, and therefore decrease the utility of inequality averse players. Positive rewards can be explained by using the model of Charness and Rabin, (2002) that assumes players to be concerned with the maximization of the minimum income and the total welfare in a group (hereafter the 'maximin' model). As shown by Sutter et al., (2010) players contribute fully if they are sufficiently

concerned with the lowest payoff and/or the sum of payoffs in their group[9]. Notice that even selfish players cooperate strategically: they contribute sufficiently such that they are among the players with the lowest payoff and, thus, receive rewards from the intrinsic contributors. In turn, in communities without sanctioning nor rewards subjects with high concerns for the lowest payoff and/or the sum of payoffs in their group will contribute fully, while all other contribute nothing. Yet, according to the model of Charness and Rabin, (2002) the threat of punishment is not credible since this reduces social welfare of a punishing player and the entire group. Therefore, the opportunity to punish does not affect contribution rates in the punishment community.

In summary, both models, the inequality aversion model and the maximin model expect positive contributions either in the punishment community or the reward community. However, neither model predicts cooperation in both communities. In contrast with those predictions, empirical results established that, punishment opportunities promote high contribution rates, while reward opportunities lead to comparable contribution rates—at least in the early periods of an experiment (see, e.g., Sefton et al., (2007)). As the two social preference models cannot explain the empirical observation nor predict any opportunity for a true competition via migration between the sanctioning regimes, we base our hypotheses on previous experimental evidence. In our baseline treatment, PN, we expect to replicate the findings of Gürerk et al., (2014); that is, contributors use the punishment mechanism to "discipline" free-riders in the initial periods of the experiment and participants strongly prefer the sanctioning regime, since contributions are substantially higher in the punishment regime than in the regime with no action.

H_1: In the PN treatment, the majority of players prefer the punishment regime. They actively use punishment points to maintain the cooperation at a high level throughout the experiment.

For our main treatment, PR, the existing literature (e.g., Sutter et al., 2010) suggests that participants—at least initially—favor the reward community.

[9]Technically, Sutter et al., (2010) analyze a dichotomous reward system. However, since both the payoff mechanism and the social utility model are linear, players either abstain from rewards or fully reward (similar arguments hold for contributions and punishment), so that the results of the analysis for the dichotomous rewards hold for every reward point in our design.

H_2: In the PR treatment the majority of participants initially joins the reward community.

Recall that the reward regime offers enormous efficiency gains (depending on y and r in comparison to x and a), but due to people's diminishing willingness to reward, the high contributions and high earnings are usually not sustainable in this regime (Andreoni et al., 2003; Sefton et al., 2007; Gürerk et al., 2014). This suggests that people migrate from the reward to the punishment community as soon as the rewards and subsequently the contributions decline in the rewards community.

H_3: In the PR treatment participants gradually migrate from the rewards to the punishment community.

The crucial questions for our analysis is not only whether, but also why participants will start exploring the alternative institution and eventually migrate to the punishment community. As full contributors suffer most from free-riders in malfunctioning reward communities, we expect them to initiate the migration to the punishment regime. Yet, it is a priori unclear which specific circumstances trigger this process and we aim to provide an empirical answer.

2.2.3 Procedure

Experiments were conducted at the WISO Research Laboratory at the University of Hamburg computerized using z-tree (Fischbacher, 2007). Subject were mostly students of various majors of the University of Hamburg. They were recruited online via hroot (Bock et al., 2014) and randomly assigned to treatments. No subject participated more than once.

We conducted 7 sessions of roughly 2 hours each with 156 participants in total. In all but one session, 24 subjects participated. Subjects were randomly matched into two groups of 12, which remained unchanged throughout the experiment. The two groups played the game simultaneously, but independently, yielding 6 independent observations for the PN treatment and 7 independent observations for PR[10].

[10]Data for one group in PR treatment was lost due to a hardware failure.

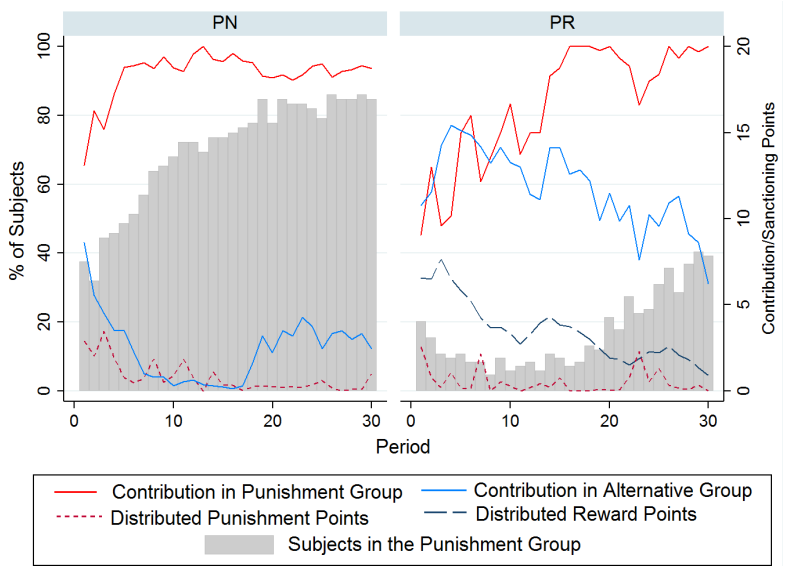


FIGURE 2.1: Community choice, contribution and the use of points over periods

After being seated in cubicles, participants received a copy of the instructions, which were read aloud, followed by a comprehension test. As in the design of Gürerk et al., (2014) subjects were paid no show-up fee but granted a starting endowment of 1000 tokens, equivalent to 10 Euro. The conversion rate (1 token = 10 Eurocent) was common knowledge. At the end of period 30, individual token incomes were converted into Euro. After filling an a short questionnaire, subjects were paid privately and in cash. Payments ranged from 19.10 to 32.60 Euro with an average of 25 Euro.

2.3 Results

2.3.1 Dynamics of the community choice

Figure 2.1 displays the dynamics of the community choice over time. The left panel shows the migration between communities and the level of contributions in the PN treatment. The results for the PN treatment closely mirror the findings from Gürerk et al., (ibid.). Players quickly discover that the level of contributions and profits are higher in the punishing community, migrate there and contribute almost fully under the threat of punishment. Supporting our H_1 and the results from the previous literature, the majority of participants choose the punishment

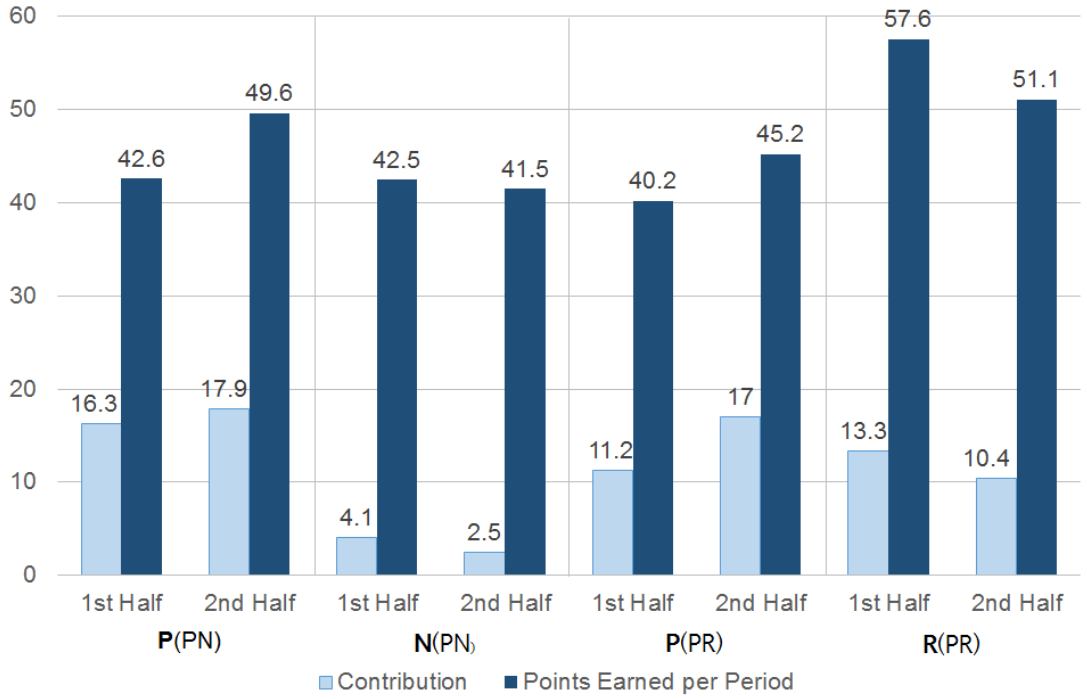


FIGURE 2.2: Average contribution and period scores for periods 1-15 (1st half) and 16-30 (2nd half)

community if given the option between punishment and no sanction. The players achieve high levels of cooperation by using punishment in the initial periods.

The right panel of the graph, containing the PR treatment shows more variance over time: in the beginning, almost all the players choose the reward community, which gradually reaches the highest popularity around period 15 with about 90% of the population selecting it. The data support our hypothesis H_2: people initially prefer the reward community. However, the trend reverses in line with our hypothesis H_3 in the second half of the experiment; players start to migrate to the punishment community in increasing numbers. Thus our results demonstrate that there is no time-persistent preference for rewards. Subjects choose indeed the reward community initially, but as time passes they migrate to the punishment institutions.

Possible explanations for these migration patterns are the change in the rates of contribution and profits of different communities. Although a Mann-Whitney U-test shows that the average contribution is not significantly different in the

treatments (PN 10.75 and PR 13.5 p=0.504)[11], a closer look at single communities reveals that the level of contribution differs significantly between them and that communities with P in the PN treatment condition[12] performs best, with an average contribution of 17.4. It thus reconfirms that punishment seems to be better suited for sustaining a higher level of contribution. The reward community in our experiment secures a relatively stable level of contribution at ca. 60% of the full endowment level (11.9), however, with a decreasing trend.

In the sanctioning regimes it is not only the level of contributions but also the punishing and respectively rewarding behavior of players which affect the payoffs in any given period. Figure 2.2 displays average contributions and average period scores[13] separated for the first and the second half of the experiment since the migration from the reward community begins approximately in period 15. The profits of the R(PR) community and P(PR) are significantly different averaged for the 30 periods and for the first half of the experiment (Wilcoxon signed-rank test p=0.018 and p=0.03 respectively). Despite the decrease in the second half, the profits remain higher than those of the P(PR) but the difference is not significant anymore (p=0.13).

The distributed points in Figure 2.1 show that the difference in profits between the groups stem from the active use of sanctioning. The use of reward and punishment points is not symmetric: punishment points in the P(PR) community are barely used, even when the group expands and attracts new members, while rewards are distributed frequently, although their use declines in the course of the game. Overall, profits in the reward community stay higher throughout the game, but this does not stop migration toward the punishment regime.

Turning to the sanctioning behavior in punishment communities, let us point out that the establishment of an effective punishment community in the PN treatment requires players to discipline each other in the first periods of the game only. A high level of contributions in the later periods relies on the threat of punishment,

[11]For all the further analysis only groups of 2 and more players were considered. The p-values can be found in Table A1 and Table A2 in the Appendix. All non-parametric statistical tests reported are two-tailed and take communities as units of observations.

[12]Hereafter denoted as P(PN), other groups are denoted accordingly.

[13]An average period score refers to the total number of points a player received in the current round. That is, it is the sum from the outcomes of the public good stage and sanctioning stage.

but for it to be credible the society has to undergo welfare losses in the earlier periods. In the punishment regime in the PR environment, however, the threat of punishment alone appears to be sufficient to promote cooperation from the very beginning. This observation is remarkable since there is gradual migration into P communities both in PN and PR. Migration between the two sanctioning institutions leads to less inefficiencies than migration into the punishment regime from a sanction free environment. The former type of migration promotes better adjustment of contribution norms. Migrants already experienced an alternative institutional setting (i.e., R) in which some community members used the (costly) sanctioning mechanism. Thus it seems less likely that new members of P(PR) question whether other group members will actually use punishment. As a consequence, they adjust their contributions accordingly making the actual use of punishment redundant. The existence of the reward community postpones the establishment of a flourishing punishment community, yet significantly increases the efficiency in the latter setting.

While in the PR treatment the mere threat of punishment is sufficient to ensure a high level of cooperation, in the reward regime actual rewarding of contributors is required in every period of the game to ensure pro-social behavior in future periods. Reward points are indeed frequently assigned in the beginning of the game but the level is decreasing over time.

For the evolution of a regime, the question of *who* uses a sanctioning mechanism is as important as whether it is used at all. Comparing Figures 2.3 and 2.4, we see that in the R(PR) in the first half of the game free-riders or those who do not contribute fully generously reward contributors[14]. In the second half the total use of points decreases and now it is the contributors themselves, who have to bear the costs of rewarding while not getting any compensation from the free-riders. This development may lead to an unequal distribution of income, which in turn may cause migration towards the punishment community. A first indication of this is the coincidence of (almost) fully contributing subjects who distribute rewards and migrate subsequently (see Figure 2.4). In the next subsection we will analyze the possible reasons that drive migration between communities.

[14]A similar pattern of free-riders rewarding full contributors was found in the related work of Sutter et al., (2010)

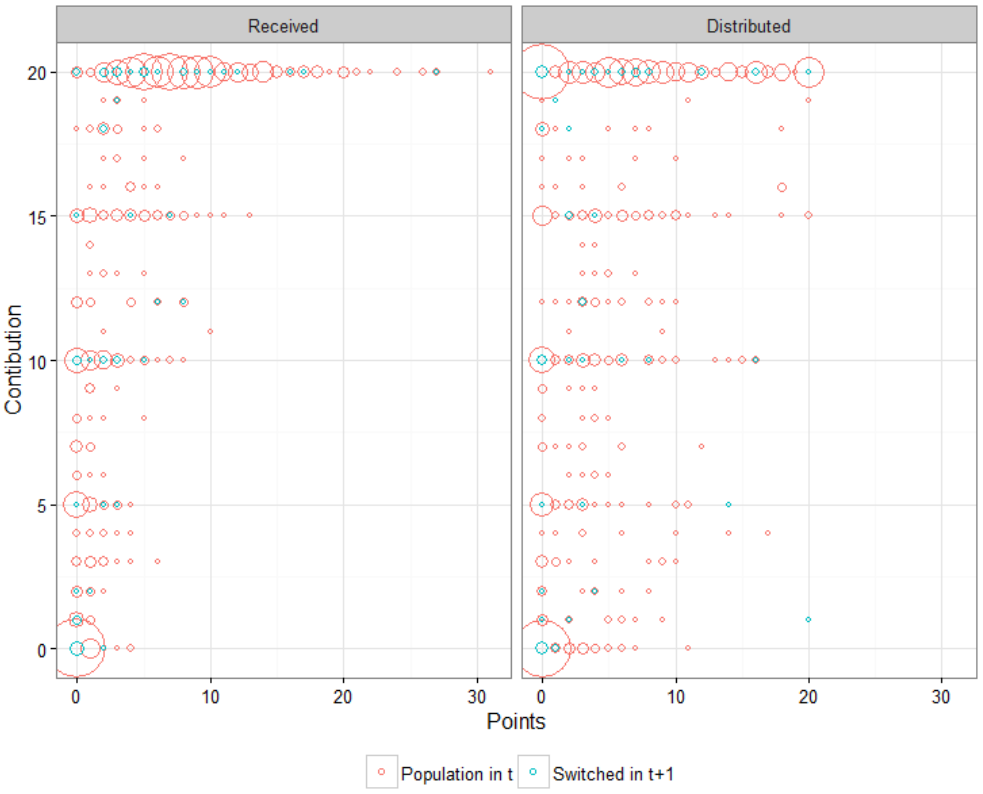


FIGURE 2.3: R(PR): received and distributed points depending on the contribution for the 1st half of the game

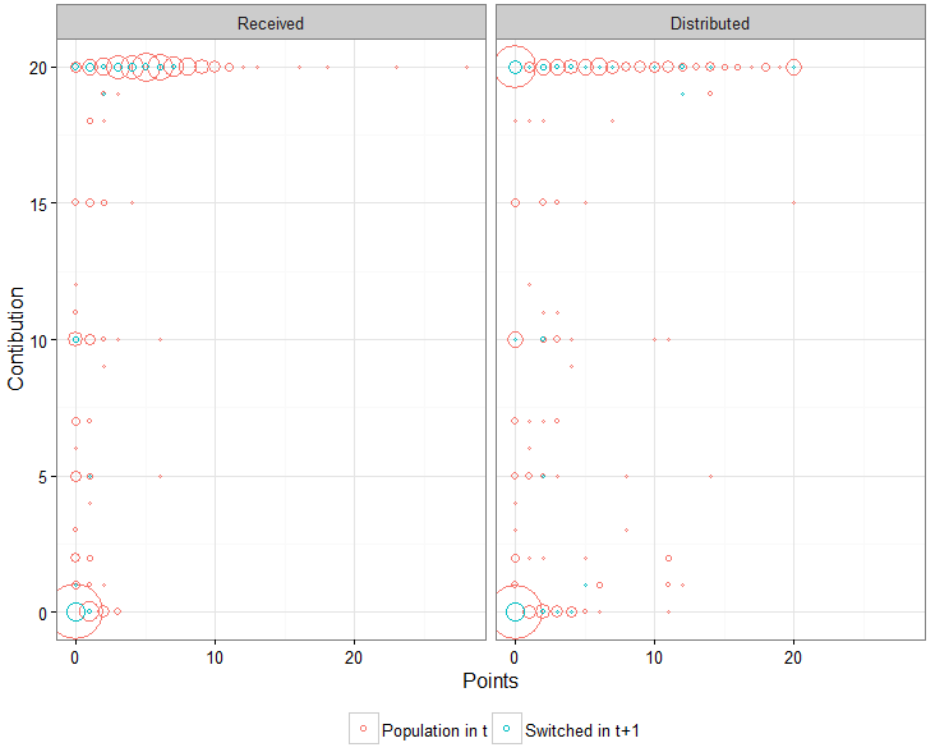


FIGURE 2.4: R(PR): received and distributed points depending on the contribution for the 2nd half of the game

2.3.2 Migration between communities

One intuitive reason for subjects to switch regimes is pursuing higher profits. Players receive comprehensive feedback which includes individual profit of each player in both communities, contribution, received and distributed points (if applicable) for the current period as well as average score for all the periods already played. Therefore, players have well-informed expectations concerning the profits in the alternative community, even without joining it. However, higher profits cannot be the main driver of migration, since earnings remain higher in R(PR) throughout the experiment. Nonetheless, we still observe migration towards P(PR). Therefore, we will analyze other factors beyond mere monetary selfinterest, which potentially triggers migration from the profitable reward community.

There is a number of other potential factors which might explain the migration pattern (e.g., low minimum contribution in the own as well as high average contribution in the alternative community). We test those factors in a series of random effect panel probit regressions with a binary dependent variable taking the value of 1 for each subject, who migrated to the alternative community in the next period (t+1), and zero otherwise. We estimate separate regressions for each community which players potentially leave (e.g., P to N, or R to P). We add several control variables, which allow us to better capture underlying differences between positive and negative sanctions in the PR treatment: antisocial behavior (in addition to received points), difference in income within the community, difference in income between two communities, percentage of free-riders in the community, and the size of the community.

We include antisocial behavior as a dummy variable which takes a value of 1 for a full cooperator, who receives any number of punishment points in P or receives no reward points at all in R. Although withholding rewards may not necessarily be seen as an act of punishment, intuitively these situations are very similar: full cooperators in P invest maximum effort to increase the community's payoff, clearly at the cost of reducing their own payoff. Decreasing their payoffs even further by punishment is likely to decrease their willingness to contribute in the future, and, as such violates the interest of the community. Along the same line of argument, not rewarding full contributors in R violates group interest as well:

while increasing the overall payoff of the community, the full contributors' payoffs remain the lowest in the community if they are not compensated via rewards.[15] Therefore, the variable antisocial behavior provides a proxy for similar effects in the punishment and the reward community.

Difference of income between two communities captures the profit maximizing motive of migration and measures the difference between the average payoff per period between communities.[16]

Since rewards decline in later periods of the experiment, income inequalities may arise within the group between contributors and free-riders. As this inequality could be an important motivation for migration, we included the difference in income within the community in the regressions[17]. We use the total period payoff, that is, the sum of payoffs obtained in the public goods and in the sanctioning stage. For this, for each player the payoff inequality is calculated in the spirit of Bolton and Ockenfels, (2000) as *Difference of income within community= ln(Actual payoff/ Equal payoff)*. As in the Bolton-Ockenfels model, we measure inequality as the difference between the actual individual share and one-nth share of the sum of payoffs assuming that players prefer small differences (precisely, we use the logarithm of the shares as discussed in Jasso, (2007)).[18] From this approach is follows that if one earns more than the equal share, the logarithm has positive value. If subject earns less than the equal share, the value turns negative, and zero reflects perfect equality.

We control for the share of free-riders relative to contributors in the community (measured as percentage of free-riders in the community), since it is possible that

[15]On a related issue Gangadharan et al., (2016) show that rewards are unable to establish high levels of cooperation if they do not achieve equal income distributions.

[16]Difference of income between two communities is defined as follows $DifferenceR(PR) = (AveragePeriodScoreR(PR) - AveragePeriodScoreP(PR)$ and respectively $DifferenceP(PR) = (AveragePeriodIncomeScoreP(PR) - AveragePeriodScoreR(PR)$; therefore, we expect to have same sign of the coefficient for migration in both directions.

[17]Previous research suggests that fairness considerations may be important both for contribution decisions (Fehr and Gächter, 2000) and for formation of stable communities (Robbett, 2014).

[18]Jasso, (2007) originally includes "fair outcome" in the equation instead of the "equal payoff." Since the current paper does not deal with different fairness concepts, for our purposes the principle of fair share appears to be feasible.

the contributors will self-select into the communities with the smaller share of free-riders. Size of the community is controlled for to see if the size of the community plays a role for the migration decision.

Table 2.1 shows the post-estimation of the probit regressions. Our random effect panel probit regressions have standard errors clustered on the experimental subgroup level.[19] The average marginal effects allow to estimate not only the direction of the correlation, but also provide some interpretation for the coefficients. Specifications 1 and 2 describe migration within the PN treatment, while 3 and 4 display the coefficients for the PR treatment. The regression confirms that there are different motivations for migrating between rewards and punishment within the PR treatment. Moreover, results show that different mechanisms drive migration from the punishment community in the two treatment conditions. That is, the punishment community in itself functions differently depending on the alternative institution.

Column 1 shows that apart from minimum contributions in their own community and average contribution in the alternative community, it is the income differences within a community and between competing communities which drive the participants away from the non-sanctioning community. As for migrating from P(PN) to N(PN), being punished increases the probability of migration in the next period but there is no significant effect of antisocial punishment. The level of contributions in own as well as in the alternative community influences the migration decision. It is also the only direction of migration where the size of the group has significant effect on the decision to change the group: players are less likely to leave bigger punishment community.

Columns 3 and 4 contains the estimations for the migration within the PR treatment. Antisocial behavior (i.e., not assigning any points to full contributors) clearly has the strongest impact on the decision to migrate from the reward to the punishment community (column 3). Further of significant impact are minimum and maximum contribution in the own community, the difference of income within the community and between the communities as well as the share of free-riders. The migration from the punishment to the reward community (column 4) again

[19]The group of 12 participants, which remained unchanged throughout the experiment.

Table 2.1: Marginal effects of the determinants for migration

Switched in t+1	(1) N to P	(2) P to N	(3) R to P	(4) P to R
Received Points		0.00781***	-0.000441	-0.00267
		(0.00205)	(0.00169)	(0.0165)
Antisocial Behavior		0.00400	0.0828***	0.272***
		(0.0190)	(0.0243)	(0.0235)
Minimum of Others' Contribution in Own Community	-0.0108*	-0.00214**	-0.00239***	-0.00490
	(0.00581)	(0.000960)	(0.000908)	(0.00548)
Maximum of Others' ontribution in Own Community	-0.00574	-0.00282***	-0.00579***	-0.0154
	(0.00551)	(0.000507)	(0.00147)	(0.0137)
Average Contribution in Other Community	0.0122***	0.00536***	0.000867	0.0277***
	(0.00446)	(0.000911)	(0.00128)	(0.00666)
Difference of Income within Community	-0.630***	0.000316	-0.0404**	-0.0489
	(0.181)	(0.0172)	(0.0195)	(0.173)
Difference of Income between Communities	-0.00569***	0.000393	-0.000943**	-0.00237
	(0.00168)	(0.000379)	(0.000417)	(0.00304)
% of Free-Riders in Community	-0.000980	-0.000497	0.000983**	0.00364*
	(0.000777)	(0.000507)	(0.000468)	(0.00203)
Size of the Community	0.00851	-0.00489***	-0.00256	-0.00944
	(0.00974)	(0.00181)	(0.00628)	(0.0103)
Observations	609	1,440	2,039	346
R^2	0.44	0.56	0.35	0.62

Note: Radom-effects Probit estimation; standard errors in parentheses,
*** p<0.01, ** p<0.05, * p<0.1;
R^2 is calculated for the respective panel probit specification and is mentioned here for completeness. R^2 is calculated after McKelvey & Zavoina.

is mainly determined by the antisocial behavior (i.e., punishing full contributors). For this decision the average contribution of the alternative community and the share of free-riders are significant driving factors as well. Neither the level of contributions in one's own community (minimum and maximum contribution) nor the size of the communities influence the migration decision from P to R.

Notably, the number of sanctioning points itself does not affect the likelihood to migrate in neither R to P nor P to R case. The coefficient for the antisocial punishment, however, is significant, positive and by far of the highest magnitude for both directions of migration. Thus, not the size of sanctioning drive players away, but the instance of antisocial punishment dramatically increases the likelihood of switching. In other words, if sanctions are used to hinder and not to promote cooperation, the alternative community is seen as more attractive.

Earlier, we speculated that migration – particularly of contributors, who suffer the most from the free-riding – might be triggered by the share of free-riders in the group. Yet, increasing numbers of non-contributors have a significant, but rather small effect. The significant and comparatively large effect of unequal distribution of income within the community allows the following interpretation: the share of free-riders in itself seems to have little influence on the decision to leave the community. Only if free-riding is the primary source of inequality within the community and non-contributors do not generously assign reward points to contributors and do not reduce inequality, migration is triggered (recall Figures 2.3 and 2.4). Therefore, the influence of the share of non-contributors is of minor importance only.

Along the same line of arguments, the difference of income within the community seems to be a major driver for the migration from R to P, and even more so from N to P (see specifications 1 & 3). In turn, the more equal are outcomes within the community, the less likely players are to migrate. Thus, for the rewarding regime to function, it is important to ensure that outcomes are evenly distributed. In fact, it is of pivotal importance not only how many rewarding points do contributors receive in total,[20] it is also crucial who assigns them. If free-riders fail to coordinate the equalization of outcomes within the community, contributors migrate to the

[20]Recall that there are enormous efficiency gains in R even for mutual rewards of two players.

alternative institution. Subjects seem to prefer a punishment community with more evenly distributed payoffs over more profitable reward institution.

Overall, our regression results show that internal, within-community factors (such as minimum and maximum contribution and the distribution of incomes) which are not dependent on the performance of the alternative community are of major importance for the decision to migrate from R to P. Migration from P to R on the contrary, depends to a greater extent the performance of the alternative institution. Hence, the P(PR) community grows with players unsatisfied with the distribution of incomes in the R(PR) community. As the antisocial behavior and inequality play the major role for migrating in the upcoming period, it is likely that full contributors who are not rewarded at all (thus suffering from both factors) are the first switch to the alternative regime. Active rewarding of contributors, therefore, appears to be necessary not only to ensure a high level of contribution in the next periods, but also to keep contributors in the community and prevent migration to the punishment community in the first place. As the use of reward points decreases steadily over time, more and more contributors remain unrewarded and migrate.

Taking a closer look at who migrates first into non-populated[21] punishment community in the PR treatment (Fig.2.5) confirms the former claim: we see that the number of migrating subjects with a previous contribution of 20 is twice as high as the number of free-riders and those who contributed moderately. While most participants return to the reward community after one period, a large fraction of the migrated full contributors remains in the punishment community. A proportionality test supports that mostly full contributors migrate first and stay in the unpopulated punishment community: the proportion of full contributors migrating into an unpopulated punishment regime and remaining there for more than one period is significantly higher than the share of full contributors within the reward community in general ($p=0.06$). Within those who migrated into an unpopulated punishment regime and remained there for more than one period the proportion of full contributors (14 out of 21) is significantly higher than proportion of free-riders

[21] A community is considered to be unpopulated if its population consists of up to 1 person.

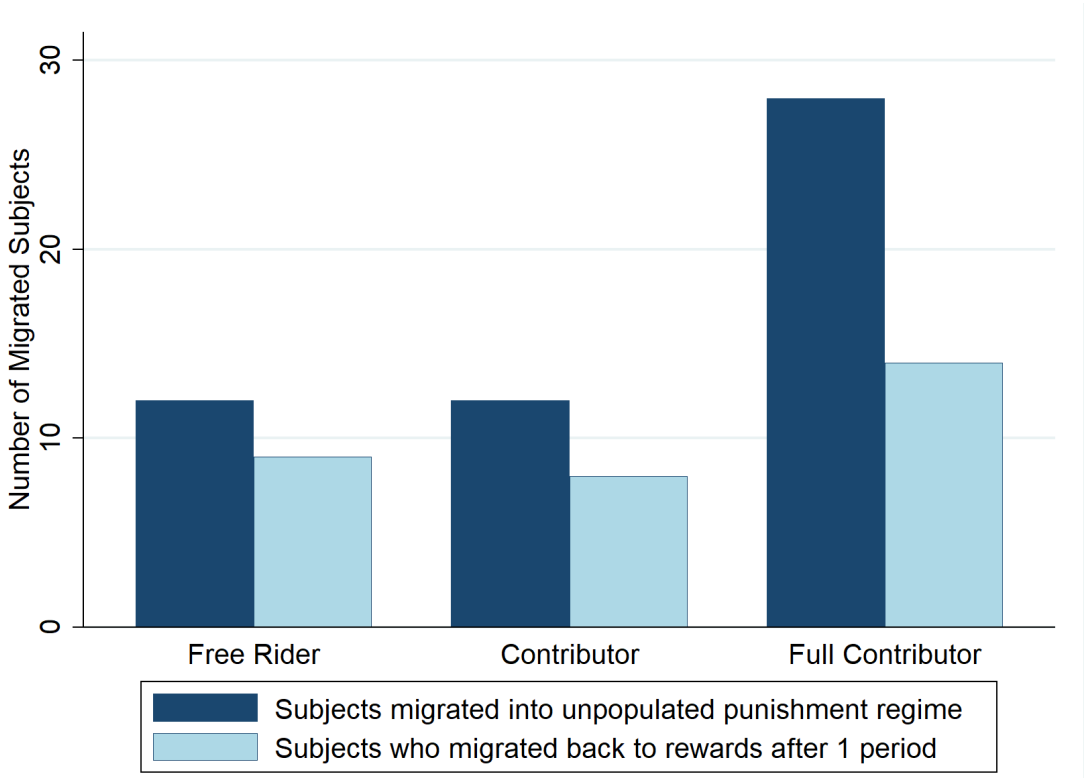


FIGURE 2.5: Subjects, who migrate into a non-populated punishment community classified by their last contribution before migrating (full contributors contributed 20, free-riders – 0, contributors – any amount between 19 and 1). Aggregate over all periods

(3 out of 21, p= 0.000) and of contributors (4 out of 21, p=0.000), as well as of the two latter categories combined (p=0.0005)[22].

2.4 Conclusion

It seems that the conditions triggering the migration to the punishment community are unequal income distribution within the reward community and assigning no reward points to full contributors. While punishment institutions were found to be initially unattractive and avoided if possible by a majority of experimental studies (e.g., Dal Bó et al., 2016; Sutter et al., 2010), our results suggest that

[22]Our findings are in line with the results of Fehr and Williams, (2017) who find also that prosocial individuals are the first to migrate from an institution without punishment and manage to establish a norm of full cooperation

players gradually join the punishment institutions, leaving unequal reward institutions: in line with the findings of Dal Bó et al., (2016), there may be an initial resentment against an institution that potentially inflicts direct negative effects on subjects. Nonetheless, cooperators start "exploring" this rule if income inequalities prevail within the group. In turn, a reward-based society runs the risk of losing the most pro-social members when it fails to acknowledge their contribution. Even highly efficient institutional arrangements lose their attractiveness due to the nonrecognition of individual fairness concerns.

In addition, our experimental results indicate an important avenue for the emergence of sanctioning institutions, which allows to promote cooperation with the help of punishment, but without initial losses of efficiency: if there is a competition between punishment and reward communities, the gradual migration allows the punishment mechanism to sustain high levels of cooperation persistently by threat (rather than actual punishment). To be more precise, migrants already experienced an institutional setting in which some community members defend contribution norms by (costly) sanctioning. There is no need to question whether other group members execute the punishment mechanism, so new members adjust their contributions accordingly. This makes the actual use of punishment redundant, while leading to more homogeneously cooperating communities as envisioned by Tiebout (1956). This result gains additional importance in the light of the recent findings of Gächter, Mengel, et al., (2017): if profits from previous periods are endowments for the next, than the welfare loss from punishment in the early periods grows exponentially and is particularly harmful. Here, the gradual migration into punishment communities from existing competing institutions may promote high levels of cooperation without the initial loss of welfare due to actual punishment.

Our findings highlight the importance of a factor which is easily overlooked in socio-economic research: the institutional frame of open societies reflects an ongoing institutional competition. Institutional designs do not develop in isolation, they rise and fall in competition with alternative mechanisms. That is to say, real world settings are rich on social dilemmas and potential institutional arrangements

for them. The gradual comparison and movement between alternative settings assure the emergence of persistent solutions for dilemmas serving best our needs, both in terms of efficiency and – equally important – in terms of equality.

Bibliography

Andreoni, James, William Harbaugh, and Lise Vesterlund (2003). "The Carrot or the Stick: Rewards, Punishments, and Cooperation". In: *The American Economic Review* 93.3, pp. 893–902.

Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch (2014). "hroot: Hamburg registration and organization online tool". In: *European Economic Review* 71, pp. 117–120.

Bolton, Gary E and Axel Ockenfels (2000). "ERC: A theory of equity, reciprocity, and competition". In: *The American Economic Review*, pp. 166–193.

Charness, Gary and Matthew Rabin (2002). "Understanding social preferences with simple tests". In: *The Quarterly Journal of Economics* 117.3, pp. 817–869.

Dal Bó, Ernesto, Pedro Dal Bó, and Erik Eyster (2016). "The demand for bad policy when voters underappreciate equilibrium effects". In: *The Review of Economic Studies*.

Dickinson, David L, David Masclet, and Marie Claire Villeval (2015). "Norm enforcement in social dilemmas: An experiment with police commissioners". In: *Journal of Public Economics* 126, pp. 74–85.

Fehr, Ernst and Simon Gächter (2000). "Fairness and retaliation: The economics of reciprocity". In: *The Journal of Economic Perspectives* 14.3, pp. 159–181.

– (2002). "Altruistic punishment in humans". In: *Nature* 415.6868, pp. 137–140.

Fehr, Ernst and Klaus M Schmidt (1999). "A theory of fairness, competition, and cooperation". In: *The Quarterly Journal of Economics* 114.3, pp. 817–868.

Fehr, Ernst and Tony Williams (2017). "Creating an Efficient Culture of Cooperation". In: *Working Paper*.

Fischbacher, Urs (2007). "z-Tree: Zurich toolbox for ready-made economic experiments". In: *Experimental Economics* 10.2, pp. 171–178.

Fischbacher, Urs and Simon Gächter (2010). "Social preferences, beliefs, and the dynamics of free riding in public goods experiments". In: *The American Economic Review* 100.1, pp. 541–556.

Gächter, Simon, Friederike Mengel, Elias Tsakas, and Alexander Vostroknutov (2017). "Growth and inequality in public good provision". In: *Journal of Public Economics* 150, pp. 1–13.

Gächter, Simon, Elke Renner, and Martin Sefton (2008). "The long-run benefits of punishment". In: *Science* 322.5907, pp. 1510–1510.

Gangadharan, Lata, Nikos Nikiforakis, and Marie Claire Villeval (2016). "Equality concerns and the limits of self-governance in heterogeneous populations". In: *Working Paper*.

Gardner, Roy, Elinor Ostrom, and James M Walker (1990). "The nature of common-pool resource problems". In: *Rationality and Society* 2.3, pp. 335–358.

Gürerk, Özgür, Bernd Irlenbusch, and Bettina Rockenbach (2006). "The competitive advantage of sanctioning institutions". In: *Science* 312.5770, pp. 108–111.

– (2014). "On cooperation in open communities". In: *Journal of Public Economics* 120, pp. 220–230.

Hauert, Christoph (2010). "Replicator dynamics of reward & reputation in public goods games". In: *Journal of Theoretical Biology* 267.1, pp. 22–28.

Herrmann, Benedikt, Christian Thöni, and Simon Gächter (2008). "Antisocial punishment across societies". In: *Science* 319.5868, pp. 1362–1367.

Hilbe, Christian and Karl Sigmund (2010). "Incentives and opportunism: from the carrot to the stick". In: *Proceedings of the Royal Society of London B: Biological Sciences*, rspb20100065.

Jasso, Guillermina (2007). "Studying justice: Measurement, estimation, and analysis of the actual reward and the just reward". In: *IZA Discussion Paper*.

Milinski, Manfred and Bettina Rockenbach (2012). "On the interaction of the stick and the carrot in social dilemmas". In: *Journal of Theoretical Biology* 299, pp. 139–143.

Nikiforakis, Nikos and Dirk Engelmann (2011). "Altruistic punishment and the threat of feuds". In: *Journal of Economic Behavior & Organization* 78.3, pp. 319–332.

Rand, David G, Anna Dreber, Tore Ellingsen, Drew Fudenberg, and Martin A Nowak (2009). "Positive interactions promote public cooperation". In: *Science* 325.5945, pp. 1272–1275.

Robbett, Andrea (2014). "Local institutions and the dynamics of community sorting". In: *American Economic Journal: Microeconomics* 6.3, pp. 136–156.

Rockenbach, Bettina and Manfred Milinski (2006). "The efficient interaction of indirect reciprocity and costly punishment". In: *Nature* 444.7120, pp. 718–723.

Sefton, Martin, Robert Shupp, and James M Walker (2007). "The effect of rewards and sanctions in provision of public goods". In: *Economic Inquiry* 45.4, pp. 671–690.

Sutter, Matthias, Stefan Haigner, and Martin G Kocher (2010). "Choosing the carrot or the stick? Endogenous institutional choice in social dilemma situations". In: *The Review of Economic Studies* 77.4, pp. 1540–1566.

Tiebout, Charles M (1956). "A pure theory of local expenditures". In: *The Journal of Political Economy*, pp. 416–424.

A.1 Tables

The tables below contain the Mann-Whitney U tests for different subgroups with populations of more than 1 player (see tables A1&A2). Table A3 is an alternative representation of the data displayed on Fig.2.1 in the main body of the paper.

Table A1: Contribution, Mann-Whitney U test

Contibutions	P(PN) 17.4	N(PN) 4.1	PN 10.8
P(PR) 15.0	p=0.0455	p=0.0043	
R(PR) 11.9	p=0.0455	p=0.0027	
PR 13.5			p=0.504

Table A2: Points per period, Mann-Whitney U test

Points per period	P(PN) 47.7	N(PN) 42.5	PN 45.1
P(PR) 44.7	p=0.32	p=0.15	
R(PR) 54.4	p=0.01	p=0.003	
PR 49.6			p=0.03

Table A3: Table representation of Figure 2.1

Periods	Contribution				Distributed points			% of Subjects in the Punishment Group	
	P(PN)	N(PN)	P(PR)	R(PR)	P(PN)	P(PR)	R(PR)	PN	PR
1-3	14.9	6.2	10.6	12.2	2.8	1.2	6.9	38.0	15.5
4-6	18.3	3.1	13.7	15.1	1.1	0.5	5.9	48.6	9.5
7-9	19.1	0.9	13.6	14.2	1.0	1.3	4.0	62.0	8.9
10-12	19.0	0.5	15.1	12.6	1.2	0.2	3.1	70.8	7.1
13-15	19.5	0.3	17.3	13.1	0.5	0.5	4.0	72.2	8.7
16-18	19.3	0.7	20.0	12.5	0.2	0	3.4	76.4	9.5
19-21	18.3	3.0	19.7	10.4	0.2	0.1	2.1	82.4	17.1
22-24	18.4	3.7	17.8	9.5	0.3	1.2	1.9	82.9	24.6
25-27	18.6	3.1	19.2	10.6	0.3	0.6	2.3	83.3	31.7
28-30	18.8	2.9	19.9	8.0	0.4	0.1	1.4	85.2	38.9

A.2 Instructions: PR treatment[23]

Instructions for the experiment

General instructions

At the beginning of the experiment you will be randomly divided into **2 groups of 12 participants** each. During the experiment, you will be interacting only with the members of the same subgroup.

At the beginning of the experiment each participant receives a **starting endowment of 1000 points**. Before the beginning of the experiment there are 5 questions of understanding.

Procedure

Experiment consists of **30 rounds.** Each round consists of 2 stages. The first stage includes the choice of the group and decision about the contribution into the common project. In the second stage participants can affect the outcome of other members of his/her group.

First Stage

(i) Choice of the group

In the first stage, each participant can decide, which group to join.

There are two different groups:

	Affecting the income of the other members of the group
Group	A: Through assignment of negative points
	B: Through assignment of positive points

(ii) Contribution to the common project

Each round in the beginning of the first stage each member of the group receives an **endowment** of **20 Points**.

You decide how many of the 20 Points you want to **contribute** to the common project. The remaining part of the endowment, you **keep to yourself.**

Calculation of your income from the first stage

Your income from the first stage consist of two parts:

- **Points which you kept to yourself** = Endowment – your contribution to the common project
- Your **profit from the common project** = 1.6 x sum of all the contribution to the project of all the community members/ number of members in the community

Therefore, **your income from first stage** is:

20 – your contribution to the common project +
1.6 x Sum of all the contributions to the common project of all the members of the community / Number of members in the group

The profit from the common project will be calculated according this formula individually for each member of the community. **Please note:** Each member of the community receives the same profit

[23]This is a translation of the original instructions in German which can be obtained from the authors upon request.

from the common project. It means that each member of the group profits from **all** the contributions to the project.

Second Stage

Assignment of Points

In the second stage, you see how much each member of the group contributed to the common project. **(Please note: The sequence of players is reshuffled every round.** Therefore, it is not possible to identify a group member over different rounds by their position in the displayed list.)

By assigning points, you can **increase, decrease** or leave the income of the other group members **unchanged**.

In the second stage of every round each participant receives **20 additional points**. You decide how many of these 20 points you want to **assign** to the other group members. You **keep** the remaining points to yourself. To check how many points you have assigned, press the *Calculate Points* button on your screen.

- **Each positive point,** which you assign to a group member, **increases the income of this group member by 3 Points**.
- **Each negative point,** which you assign to a group member, **reduces the income of this group member by 3 points.**
- If you assign **0 points** to a group member, **the income of this group member remains unchanged.**

Calculation of your income from the second stage

Your income from the second stage consist of three parts:

- **Points kept to yourself** = 20 – sum of the points you assigned to the other group members.
- **plus 3 times the number of positive points** you received from the other group members

or

- **minus 3 times the number of negative points** that you received from other group members.

Therefore, **your income from second stage** is:

20 – sum of the points you assigned to the other group members

+3 x (Positive points, which you received from the other group members)

or

– 3 x (Negative points, which you received from the other group members)

Calculation of your round income

Your round income is calculated as follows:

	Your income from the first stage	= 20 – your contribution to the common project + 1.6 x Sum of all the contributions to the common project of all the members of the community / Number of members in the group
+	Your income from the second stage	= 20 – sum of the points you assigned to the other group members +3 x (Positive points, which you received from the other group members) or – 3 x (Negative points, which you received from the other group members)
=	Your round income	

Special case: single group member

Should you be **the only member in your group**, you receive 20 points in the first stage and 20 points in the second stage, that is your round income is **40**. You can act neither during the first nor during the second stage.

Information at the end of the round

At the end of the round, you will receive a detailed overview of the results from all groups. For each member of the group, you will see: their contribution to the project, their income from the first stage, points they assigned (if possible), points they received (if possible), their income from the second stage and their round income.

History

Starting in round two, before the start of a new round you will see an overview of the average results (as above) of all previous rounds.

Total income

The total income from the experiment consists of the starting capital of 1000 points plus the sum of the round incomes of the 30 rounds.

At the end of the experiment, your total income will be paid at an exchange rate of 1 € per 100 points.

Please note:

No communication is allowed during the entire experiment. If you have a question, please raise your hand. All decisions are anonymous, meaning that none of the other participants or the experimenters will know the identity of the person who has made a particular decision. The payout is also anonymous: no participant will know how much other participants have earned.

Good Luck!

Chapter 3

Costs of Institutions and Cooperation in Open Communities

Marina Chugunova[1]

Abstract

On many occasions, people can define which rules will be applied within a certain community (e.g., a team or a firm) and when making such choice they take into account not only how effective the rules are in promoting the desired behavior, but also how costly they are. Yet, players may disproportionally abandon an efficient institution if it involves a fixed provision cost. In the laboratory experiment we study how fixed provision costs for a cooperation-enhancing peer-punishment institution in a public goods game affect the popularity of an institution, rates of cooperation and punishment. We find that in the presence of provision costs the punishment community is overwhelmingly popular with the players choosing it even when it is less profitable than the alternative. By the end of the experiment, players contribute 99% of their endowment in the costly institution community without an increase in the use of punishment points.

Keywords: open communities, voting with feet, cooperation, sanctions, public goods, costly institutions

JEL-Classification: C72, C92, H41.

[1] Financial support of the Graduate School of Business, Economics and Social Science, University of Hamburg, is gratefully acknowledged.

3.1 Introduction

A large strand of research is dedicated to studying how different institutional frameworks perform with regard to promoting pro-social behavior and helping to overcome the problem of free-riding (e.g. Andreoni, Harbaugh, et al., 2003; Fehr and Gächter, 2002). However, not only the performance of an institution matters, but also how likely the people are to implement the framework. In this study we consider how the fixed provision costs of an institution[2] affect its popularity and subsequently cooperation and punishment behavior within community.

A few studies deal with a group choice of institution and consider some form of preference aggregation procedures (see e.g., Andreoni and Gee, 2012; Putterman et al., 2011; Sutter et al., 2010; Tyran and Feld, 2006). In Sutter et al., (2010) participants choose between an institution-free environment (VCM), peer-punishment and peer-rewards systems in exogenously formed groups. Andreoni and Gee, (2012) study the choice between peer-punishment and central punishment mechanisms. In all of these studies more centralized and efficient institutions are provided at no cost if chosen by the players, that is having a more effective institution is free.[3] However, establishing an institution may be seen as a social dilemma by itself, since it requires a costly pro-social action from the members of the society. For instance, at the state level, before governmental institutions can be introduced, citizens have to acquire information on the available alternatives and come to vote. Both acquiring information and voting are pro-social, costly actions, which in a democratic system are also recurring. There is experimental evidence that leads us to believe that these recurring costs of maintaining the institution may play a crucial role in how actively the institution is used and which institutional regime is chosen. As shown in Ramalingam et al., (2016) and Kriss et al., (2016) people's choices of an efficiency-enhancing mechanism are sensitive to its costs. In the latter paper, the authors show in the minimum cooperation game

[2] We differentiate between the costs of having an institution in place (e.g., in the context of law enforcement, having a prison) and costs that are associated with using the institution (e.g., additional costs for every prisoner).

[3] In Andreoni and Gee, (2012) authors consider that hiring the punishing party is costly but then imply that actual use of punishment is free.

that even the low costs of an "institution" (in their case, the price of sending a message) significantly decreases the use of the institution despite high efficiency gains. Markussen et al., (2013) find that the costs of sanctions may be more important than their effectiveness (i.e., how detrimental sanctions are for non-contributors). They consider a mechanism of centralized sanctioning and find that an environment with no sanctions is preferred to a free-riding deterrent and cheap centralized sanctions only by approximately a third of all groups, and that increasing the cost leads to only 15% of all groups choosing a sanctioning institution. In contrast to Fehr and Williams, (2017), their main finding is that peer punishment is widely more popular than a formal sanctioning system. They speculate that it is due to high cost-efficiency of peer-punishment.

In the papers that acknowledge the necessity to finance the establishment of an institution (e.g., Andreoni and Gee, (2012)), it is often modeled as a threshold game. If a certain amount is collected the institution is introduced, but there is no institution otherwise. Such a procedure results in players who would prefer having an institution and those who do not being a part of the same community. Although true in some situations, in many occasions players can vote with their feet and choose the community with the preferred institutional framework, i.e., join the community where all players prefer to have or not to have a certain mechanism in place. That is, one does not face a coordination problem of establishing an institution but rather pays a cost of having access to an already existing one. An intuitive example of such a choice is moving. By moving people can choose between different communities and neighborhoods, which often vary greatly in their overall character, culture and the quality of provided public goods (schools, neighborhood watch, street festivals etc.). Except for some exclusive communities as discussed in Ahn et al., (2008, 2009), there are mostly no formal entry or exit requirements to become a resident of a particular neighborhood, meaning that they are open. However, even open neighborhoods may differ with regard to the price of living there. Communities with more desirable culture or higher levels of public goods provision may in turn be characterized by higher local taxes or higher rent. For instance, Gravel et al., (2006) find that the quality of local public schools in Paris affects the housing prices in the area. Highlighting non-material differences of neighborhoods, Falk and Zehnder, (2013) found that different districts in Zurich

vary significantly in the levels of social capital with high income districts attracting more investments.

The current study builds on these two strands of research. In the environment with open communities and potentially cooperation-enhancing but costly institutions, we consider how the costs of an institution affect migration, cooperation and punishment. We address the question of how fixed recurring provision costs affect the popularity of communities with different institutions, whether they affect the behavior within different communities given that migration is possible and if it promotes sorting of participants.

In our experiment we study the choice between a peer-punishment institution and a voluntary contribution model and assume that monitoring and targeted punishment which lie at the core of peer-punishment are not free but come at a fix provision cost (which can represent, for instance, costs of acquiring information on other's actions). Our experiment tests the robustness of preferences for a peer-punishment institution. A fixed per round cost, which is deducted in addition to the actual use of punishment, renders peer-punishment less effective but allows sending a stronger and more reliable signal regarding a player's cooperation intentions to other players.

Similar to the design of Gürerk et al., (2014) players can choose to join either the free community without a cooperation-enhancing institution or the peer-punishment community, which is associated with a fixed per round cost. Within communities players are offered an opportunity to invest into a group project, which benefits all the players equally. Such a design allows studying both the initial and the over-time preference for institutions. The comparison of an institution with no sanctioning (further referred as VCM) and peer-punishment is of particular interest due to the fact that peer-punishment may, on one hand, lead to higher rates of cooperation, but on the other hand, may be welfare detrimental if it is used too actively (Fehr and Gächter, 2002; Gächter, Mengel, et al., 2017; Herrmann et al., 2008).

While it has been established that if the institution is imposed by a better-informed third party, it can be destructive to cooperation (Bénabou and Tirole, 2011; Kosfeld et al., 2009; Sliwka, 2007; Van der Weele, 2009), an endogenous choice of

institution can trigger the opposite effect (Benabou and Tirole, 2003; Danilov and Sliwka, 2016; Tyran and Feld, 2006). The latter strand of research shows that a choice of an institution allows updating the beliefs on the behavioral norm and coordinating on cooperative equilibrium. If a community choice is a signal, provision costs are likely to make it more credible.

A community choice affects levels of cooperation through signaling and through the institutional framework itself, and our design does not allow to clearly disentangle them.[4] The introduction of the cost may bring suggestive evidence for one of the interpretations. On the one hand, costs make an institution less cost-effective, but a signal more credible (for a broad review see e.g., Connelly et al., (2011)). Therefore if a community choice has primarily signaling function costs should not affect its popularity. Signaling can be the way to explain a "democratic dividend"[5] found in several experiments on voting and the endogenous choice of institutions (Dal Bó et al., 2010; Markussen et al., 2013; Sutter et al., 2010). On the other hand, if the cost-effectiveness of the institution is the main driver of the choice, then making an institution more costly should also make it less popular.

Apart from the effect on the popularity of the institution costs may affect further cooperation behavior in both communities. They can potentially damage rates of cooperation in the punishment community due to sunk costs or alternatively lead to increased cooperation in both groups due to their homogeneity. Gächter and Thöni, (2005) find that groups of "like-minded" people are more likely to cooperate even if they are free-riders.

In contrast to Kriss et al., (2016) and Ramalingam et al., (2016) we find that introduction of provision costs does not affect the popularity of the cooperation-enhancing institution. A punishment institution with per round provision costs grows slower than without provision costs, but it ultimately reaches same levels of popularity. Despite the fact that provision costs make the punishment community less profitable than VCM in the first third of the experiment, ca. 40% of players opt for the punishment community in this period, which suggests that cooperation has, as such, a non-monetary value. The framework of open communities

[4]Signals without the effects of institutions are considered in the research on the role of communication in public goods games. See e.g., Chaudhuri et al., (2006) and Palfrey et al., (2017)

[5]Expressing own institutional preferences leads to increased cooperation among players.

generates high "democratic premium" for cooperative behavior with contributions in the costly punishment community reaching 99% of the endowment. In line with expressive law theories (Sunstein, 1996; Tyran and Feld, 2006), the choice of institution allows players to update expectations about other members of the community and promotes cooperative outcomes. However, as a consequence of updated expectations we would expect to observe a reduction in the use of punishment, which is not the case in our experiment.

The paper proceeds as follows. Section 2 introduces the design of the experiment, Section 3 outlines theoretical predictions and formulates hypotheses. Section 4 brings forward the results, Section 5 concludes.

3.2 Design and Procedures

3.2.1 Design

At the core of the experiment lies a public goods game with a voluntary contribution mechanism, which is commonly used to model social dilemma situations. Players can contribute a share of their endowment to a common project, the sum of all contributions of the group members is multiplied with a certain efficiency factor and the multiplied sum is divided equally among the players. The efficiency factor is such that the group outcome is maximized if players contribute the complete endowment, but the rule of equal split regardless of one's contribution gives rise to free-riding incentives. On top of that, we use the framework of open communities as designed by Gürerk et al., (2006, 2014) to facilitate the comparison of our results to the existing literature. Within this framework players can choose between two communities, which differ in their governing institution. The experiment consists of three stages: *community selection, public goods game* and *sanctioning,* which are repeated for 30 periods. In the experiment players anonymously interact in the groups of 12 players, which remain fixed over periods. We run two treatments that differ with respect to the provision costs of a cooperation-enhancing institution.

Community selection: First, before interacting with any other players, subjects can choose between two communities. The choice set is different between treatments.

We analyze two treatment conditions. In the treatment with costly institution provision (Cost), players can choose between the peer-punishment community, which comes at a fixed per period cost, and the free community without any sanctioning institution. In the Cost treatment players are informed that the fee (τ) for being a member of a peer-punishment community is deducted after the public goods game stage. That is, at the community selection stage, before making a choice players know how many tokens and at which stage of the game will be deducted. We compare the behavior of players in this environment with the Base, where players choose between the same institutional mechanisms (i.e., VCM and peer-punishment), but do not face provision costs for any of them. For the rest of the period, players only interact within their chosen communities.

Public goods game: When the community is chosen, players learn how many other participants joined the same community and each player receives an endowment of 20 tokens. Players are offered an opportunity to invest any number of available tokens into a common project and keep the rest for themselves. The sum of all contributions within one community is multiplied by a factor of 1.6, which is constant regardless of the group size.[6] Profits from the common project are distributed evenly between all members of the community irrespective of the contribution of an individual member.If a community consists of only one member, investment into a common project is not possible, and the player automatically keeps all the points to herself. Earnings from a common project in one community are completely independent from the earnings of the alternative community. Thus, our design models a spatially excludable, congestible public good.

In the Cost treatment the fee (τ) is deducted only at the end of the public goods game stage. Thus, the number of tokens that can be contributed to the common project remains unaffected by the deduction. Provision costs affect the final payoff of the player but do not limit the available action space in any of the phases.

Sanctioning[7]*:* After the public good stage players receive an additional endowment of 20 tokens each and feedback on contributions of every individual in their

[6]Keeping the social return constant over different possible community sizes allows to avoid the scale effect where big groups are more efficient and therefore attract more players.

[7]Instructions were presented in a neutral language. The terms "punishment" or "sanctioning" were not used in the instructions or during the experiment. An English version of the instructions can be found in the Appendix.

community. If the community is equipped with a peer-punishment mechanism, then any player may assign up to 20 punishment points to other community members. Assigning one punishment point reduces one's income by one token and the income of a targeted community member by 3 tokens. A player keeps unassigned points as a part of her endowment. Within the community players are identified only by their ID numbers, which are randomly reshuffled every period. Therefore, our design requires immediate sanctions and makes reputation building and strategic sanctioning impossible. In the community with a voluntary contribution mechanism, participants also receive an additional endowment of 20 tokens, but cannot spend these tokens on sanctions.

At the end of the period, players are provided with a comprehensive feedback on the performance of both communities in the current period. For each individual in both communities, players learn their contributions in a public goods stage, received and distributed points in a sanctioning stage – if applicable – as well as a final payoff for the current round.[8] The first feedback screen is followed by another, which presents the average performance of the two communities for all periods played. It includes an average rate of contributions in the public goods stage, an average number of distributed points – if applicable – and the total payoff of the round. Thus, in our experiment players can make a fully informed decision on which community they want to join in the next period.

Taken together the earnings in each period are calculated as follows. In the community without sanctioning (VCM) in both treatment conditions:

$$\pi_{iVCM} = (20 - c_i + \frac{1.6}{n}S) + 20,$$

where c_i is individual contribution, n is the number of players choosing VCM in the current period ($1 < n \leq 12$) and S is the sum of all contributions in the community.

In the peer-punishment community in the Base treatment, the payoffs are affected by the sanctioning points distributed to other members, s_i, and the sanctions

[8] Although providing full detailed information for both communities appears to be unrealistic, the current study does not focus on the role of information transparency for the institutional choice.

received from other members of the community, (s_{-i}), such that:

$$\pi_{iP(Base)} = (20 - c_i + \frac{1.6}{n}S) + (20 - s_i - 3s_{-i})$$

In the Cost treatment a provision cost (τ) is deducted additionally, which results in:

$$\pi_{iP(Cost)} = (20 - c_i + \frac{1.6}{n}S) - \tau + (20 - s_i - 3s_{-i})$$

The main parametrization is chosen similar to the study by Gürerk et al., (2014). The provision cost $\tau = 2$ is in line with the experimental design of Kosfeld et al., (2009), Markussen et al., (2013), Ramalingam et al., (2016), and Sutter et al., (2010).[9] We opt for a relatively low value of the fixed cost τ for several reasons. First of all, relatively low provision costs allow detecting how minor changes in the efficiency of institutions affect the outcomes. Secondly, previous empirical findings provide evidence that zero may be a special case of the price and that there might be a discontinuity for values close to zero (Shampanier et al., 2007). Thirdly, we consider the choice between no institution and a decentralized peer-punishment institution. Although it is feasible to believe that one may incur some small fixed cost for acquiring information, this cost can not be high, since the institution does not solve the public good problem completely, but rather provides a framework that would facilitate further action. This would not be the case for centralized institutions, which render any free-riding inefficient and resolve a social dilemma situation as such (see, for instance, in Andreoni and Gee, (2012) and Markussen et al., (2013)).

[9] In Sutter et al., (2010) players face a one-time cost of 10 tokens for participating in the initial vote, which determines the institution. Considering the experiment as a whole, the relation between the price and endowment is 1 to 20. In the paper of Kosfeld et al., (2009) the price of an institution was divided by the number of players in the group who opted in and thus is variable: the highest cost of an institution occurs when the club size is 2 and equals 1 token per 20 tokens of the endowment. In Ramalingam et al., (2016) the endowment of players is as well 20 tokens and the right to punish costs 1 token. In Markussen et al., (2013) our parameters would correspond to the cheap informal sanctioning institution. As in our experiment players get 20 points of endowment and additionally 20 points for sanctioning, we keep the provision costs proportional to the existing literature. The pilot of the experiment revealed that players, on the one hand, perceive such costs as negligible and, on the other hand, understand why there is a fixed per round deduction in place.

Participants play this game for 30 periods. To reveal players' changing institutional preferences over time, players repeatedly select themselves at the start of the new period –prior to the contribution stage – into the community with or without a cooperation-enhancing institution. Migration between communities is costless.

3.2.2 Procedure

Experiments were conducted at the WISO Research Laboratory at the University of Hamburg and computerized using z-tree (Fischbacher, 2007). Subjects were mostly students of various majors of the University of Hamburg recruited online via hroot (Bock et al., 2014) and randomly assigned to treatments. No subject participated in the experiment more than once.

We conducted 6 sessions of approximately 2 hours with 144 participants in total. During each session subjects were randomly matched into two groups of 12, which remained unchanged throughout the experiment. Two groups played the game simultaneously, but independently, yielding 6 independent observations for each treatment.

After being seated in cubicles, participants received a copy of the instructions, which were additionally read aloud. A brief comprehension test preceded the experiment. As in the design of Gürerk et al., (2014) subjects were paid no show-up fee but granted a starting endowment of 1000 tokens, equivalent to 10 Euro. The conversion rate (1 token = 10 Eurocent) was common knowledge. After period 30 of the experiment participants filled a short questionnaire and upon completion they were paid privately and in cash. Payments ranged from 19.10 to 25.4 Euro with an average of 23.7 Euro. The average age of the participants was 25, 56% of our subjects were women.

In the Cost treatment the main experiment was preceded by the short version of the social value orientation (SVO) measure as in Murphy et al., (2011).[10] An SVO questionnaire invites people to decide over several distributions of income between oneself and a paired player and thus measures the magnitude of the concern people

[10]The zTree code by Crosetto et al., (2012). The sessions for Base treatment were conducted first and did not include an SVO measure. For the Cost treatment we decided to introduce it to obtain an additional measure of pro-sociality.

have for the welfare of others. Based on this measure players can be classified into the types: individualistic, competitive, cooperative and altruistic. When completing the SVO tasks, players were not informed about the details of the rest of the experiment. In order to not affect the outcomes of the main experiment, the results of the SVO part were revealed only after the completion of the main experiment. Instructions did not suggest any connection between two parts.

3.3 Theoretical Predictions and Hypotheses

In the game each subject faces two (possibly three) decisions: which group to join, how much to contribute and – if applicable– how much to punish.

If we consider players with standard preferences, the equilibrium is straight-forward. In Base treatment, selfish income maximizing players do not punish since punishment is costly and contribute nothing to the common project, making equilibrium strategies (and thus payoffs) in both groups identical. Therefore, selfish income maximizing players are indifferent between joining the VCM community and the punishment community.

In the Cost treatment, the Nash equilibrium behavior is the same, however, players face a provision cost in the punishment community. Therefore, they strictly prefer the VCM community and no one joins the community with a punishment institution. Formally, the introduction of τ simply renders the institution less efficient.

H_1: In the presence of provision costs the punishment community remains unpopulated. Nobody punishes or contributes in the VCM.

If we consider players with some form of other-regarding preferences, the choice of an institution as well as behavior within the group may be different.

When looking at the models of inequality aversion we assume that players only consider their own community (i.e., a group who chose the same institution) and do not derive disutility from inequality with regard to the members of the other community. As payoffs of the communities are completely independent and can

not be affected by members of the alternative community, the performance of one community is unlikely to affect the contribution and punishment behavior of the members of the other group. Thus, disutility from inequality between two communities may play a role only for the community choice in the next period. In this case, disutility from being in the less profitable community reinforces monetary incentives, making players more likely to choose a better (in monetary terms) performing community.

As shown by Fehr and Schmidt, (1999), the theory of inequality-averse preferences is well suited for describing behavior in a social dilemma situation with punishment and is consistent with the various experimental results. If a subject's aversion against inequality is strong enough, she may be ready to punish non-cooperators and enforce a cooperative equilibrium. The predictions of the Fehr-Schmidt model depend on the distribution of parameters of an inequality aversion. According to the empirical estimations in Fehr and Schmidt, (ibid.) and Blanco et al., (2011), there is a considerably high probability of cooperation for small communities but an extremely low probability for large communities (larger than 4 players)[11]. Since in our design the provision cost τ is paid by all the members of the punishment community, it does not affect the inequality within the group. Thus, the introduction of provision costs does not change behavioral predictions for the punishment community. The result of very unlikely cooperation in larger groups is derived under the assumption that the distribution of inequality parameters is common knowledge. If we relax this assumption, then it is possible that by choosing the community players signal their pro-sociality and affect the beliefs of others.[12] As the probability to achieve a cooperative equilibrium under peer-punishment is higher than under the VCM, choosing the punishment community signals pro-sociality. Players can update their expectations about the prevalence of social preferences. If players believe that others are enforcers or conditional cooperators, then contributing is rational.

[11]The game considered in the current experiment is formally analyzed in great detail by Gürerk, (2007). As the game has multiple equilibria, the analysis does not produce clear behavioral predictions.

[12]Relaxing the common knowledge assumption allows including "dividend of democracy" and "expressive law" results of Sunstein, (1996) and Tyran and Feld, (2006) and others in the framework of inequality aversion.

Similar predictions can be derived from the model of Charness and Rabin, (2002) with reciprocity[13]. Beliefs, strength of reciprocity preferences and a standard of a pro-social behavior affect the likelihood of a cooperative equilibrium (see Markussen et al., (2013) for detailed analysis). In line with the intuition above, choosing the punishment community may indicate high valuations of social welfare and affect the contributions of the other members.

Both models rely on punishment or a credible threat of punishment to promote a cooperative equilibrium. Punishment is costly and a player sacrifices some of her profits to decrease the profits of others. As all the players paying τ to join the community forgo parts of their profits, they may be seen as more likely to be ready to do so in the punishment stage as well. Anticipating that other members are ready to enforce full cooperation, all the members of punishment community contribute fully. Therefore the introduction of provision costs while reducing overall efficiency of the institution, can potentially lead to better welfare outcomes.

H_2: The introduction of provision costs allows to better signal pro-sociality and readiness to punish. All players join the costly peer-punishment community and contribute fully. In equilibrium there is no punishment.

Two biases – mental accounting and sunk costs – can play a role in punishment and contribution behavior in the environment with provision costs which may affect Hypothesis 2. If due to mental accounting (Thaler, 1985) a fixed cost is considered to be a part of the punishment budget, then fewer players would be willing to punish, which may undermine the credibility of the punishment threat and harm cooperation. If it is considered to be a part of the contribution budget, then players will be as willing to punish but will reduce contributions by the amount of provision costs. The former case may lead to a collapse of contributions in the punishment community and to players migrating to the VCM community. In the latter case contribution levels are affected but are still higher than in VCM, and therefore players stay in the punishment community.

The sunk cost fallacy may interfere with punishment decisions in the following way. As players already incurred a fixed cost of entering the punishment community

[13]According to Charness and Rabin, (2002) without reciprocity punishment is never implemented as it reduces social welfare. As a result, non-cooperative equilibrium prevails.

they may be more likely to use punishment points afterwards. As players are expected to cooperate fully, punishment is likely to be targeted at full contributors, i.e. to be anti-social. As shown in Chugunova et al., (2017), anti-social punishment is one of the key factors which affects a migration decision and drives players away from a more profitable community. Therefore, if the introduction of provision costs triggers anti-social punishment, it can also foster migration away from the punishment community.

Given that provision costs are smaller than the benefits from the full cooperation, entering a punishment community is not a perfect signal of pro-sociality in the sense that it does not ensure the separation of players by type. Under a threat of punishment which is more credible in the Cost treatment, players who otherwise would not contribute mimic the strategy of full contributors.

The separating equilibrium with pro-social types joining the peer-punishment community and egoistic types entering the VCM can be potentially constructed if we assume that players always act according to their own type and that players derive disutility from free-riding in a cooperative community. High disutility from free-riding could be due to, for instance, conformity (Bernheim, 1994), deviating from the norm (Kessler and Leider, 2012), guilt aversion (Dufwenberg, 2008) or other factors. However, these assumptions appear to be unrealistic as either of approaches would require extreme values of disutility from free-riding combined with aversion towards cooperation to achieve the separating equilibrium. Therefore, we do not consider this approach further. The general intuition is that, since the punishment community is less efficient due to provision costs, a wider range of parameter values would lead to a separating equilibrium. We use different measures of pro-sociality to explore this intuition (social value orientation measure, first period contribution behavior etc.).

H_3: If the peer-punishment community is costly, more pro-social players join it.

Players with inequality averse preferences can be sensitive to the introduction of τ as it reduces "behavioral cooperation premium" in the punishment community as compared to the VCM. Therefore in line with the empirical findings of Kriss et al., (2016), Markussen et al., (2013), and Ramalingam et al., (2016), the introduction

of additional costs (however negligible) should lead to a decrease in the use of a cooperation enhancing mechanism.

H_4: There exists a negative relationship between the costs of an institution and its popularity.

3.4 Results

In this section we consider the dynamics of a community choice, contribution and punishment behavior and explore if the introduction of provision costs leads to sorting.

3.4.1 Migration

Figure 3.1 displays the dynamics of the community choice over time. The bars depict the percentage of the population in the punishment community (values are marked on the left-hand axis), lines stand for the aggregate levels of contribution in the communities and the average per capita use of punishment points with values marked on the right-hand axis. Results on the left panel show that we successfully replicated the results of Gürerk et al., (2014), who also documented slow growth of the punishment community, active punishment behavior in the first periods of the experiment and high levels of cooperation further on. On the right panel we see that the introduction of negligible provision costs of the institution does not dramatically change the dynamics of the community choice and we observe a very similar pattern.[14]

The punishment community in the Cost treatment[15] grows slower than in Base. In order to be able to better detect the changes in the community dynamics we separate the experiment into thirds. A proportionality test confirms that the share of players opting for punishment institution in Base is significantly higher in the

[14]See group averages and the results of the statistical tests in Appendix. Since contribution and punishment is not possible in groups with a single member, only groups of two and more players are considered in a further analysis of punishment and cooperation.

[15]Further on denoted as P(Cost), no institution community denoted as VCM(Cost), communities in the Base treatment is denoted accordingly.

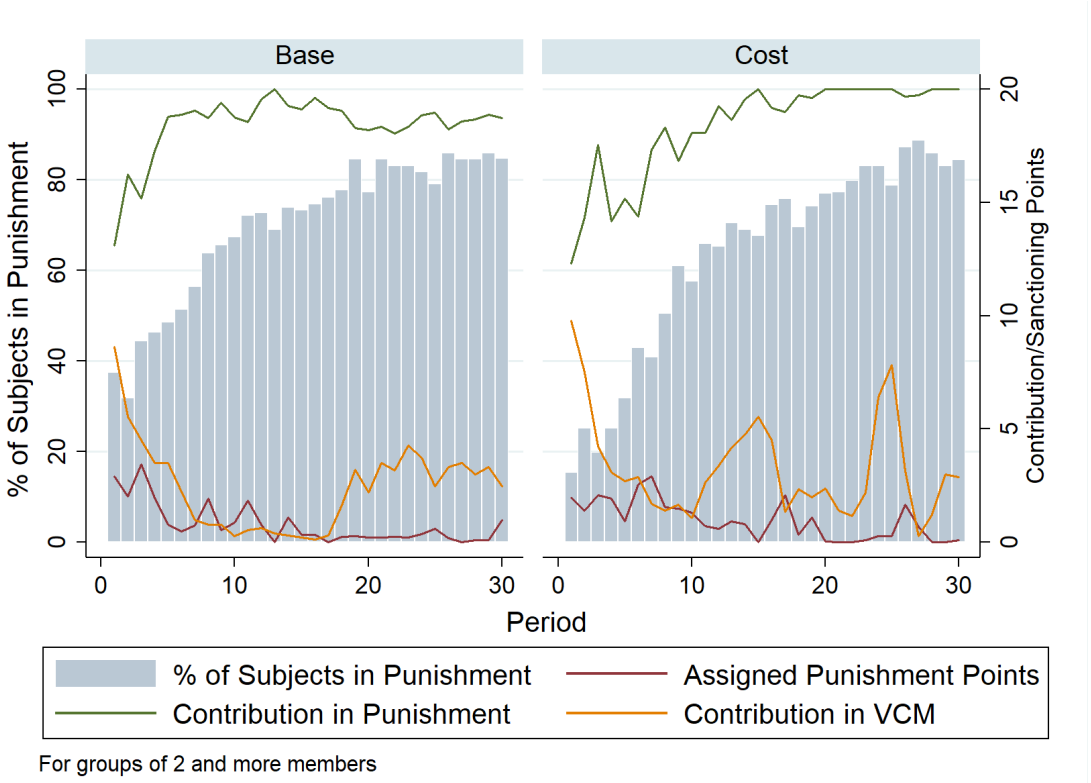


FIGURE 3.1: Community choice, contribution and the use of points over periods

first and second thirds of the experiment (periods 1 till 10 51.4% in P(Base) and 36.9% in P(Cost) p=0.00, periods 11-20 75.3% in P(Base) and 71.3% in P(Cost) p=0.08). In the last third of the experiment the share of players in the punishment community is non-different (83.8% in P(Base) and 83.2% in P(Cost) p=0.72) in both treatments.[16] That goes in sharp contrast with the predictions of the model of selfish preferences and we therefore reject Hypothesis 1. Despite slower growth, P(Cost) is as popular as P(Base) community by the end of the experiment. This does not allow us to make a clear conclusion regarding the link between provision costs and the popularity of an institution. While the punishment community is less popular in the beginning as suggested by Hypothesis 4, by the end of the experiment it is as popular as in the Base, which is in line with Hypothesis 2.

[16]For testing the community choice of players we include single member groups into the analysis.

3.4.2 Contributions

Apart from the speed of community growth we do not observe significant differences in the contribution or punishment behavior, which would be triggered by the introduction of provision costs. Contribution levels between two punishment communities are non-different whether we consider all the periods together or separate the experiment into thirds. That is, members of P(Cost) do not withhold their contributions due to the fact that they faced a fee to be a part of the community. It suggests that mental accounting does not affect the contribution decisions of the players.[17]

The voting with feet framework allows reaching very high levels of cooperation within the punishment community which is in line with arguments on a "dividend of democracy" (Sutter et al., 2010; Tyran and Feld, 2006). In the seminal experiment with peer-punishment by Fehr and Gächter, (2000) as well as in multiple others the authors observe contribution rates of ca. 60% of an available endowment. In the current experiment we do not explore how endogenous and exogenous choice of institution affects contributions, however, the positive effect of endogenous choice was established in the study by Gürerk et al., (2014). In their experiment, the exogenously imposed punishment institution, which mirrored the group size of the matched endogenous community choice session, led to contribution rates of ca. 55% (ca.90% in the treatments with voting with feet). We therefore further consider whether the introduction of fixed per round provision costs affect cooperation. In our framework players opt for even higher levels of contribution. In the third part of the experiment an average contribution in the P(Base) community is 90% or 17.9 points and in P(Cost) cooperation almost reaches its possible maximum with 99%. i.e. 19.9 points. Although the difference in contribution rates between two punishment communities in the last third is

[17]A contribution decision in the P(Cost) treatment can be also subject to moral licensing if joining the punishment group and paying the provision cost is considered to be a moral decision (Seçilmiş, 2018). Since contribution levels in two treatments are non-different, players do not see community choice as a moral decision and do not engage in a moral licensing. However, the share of free-riders in the punishment community is significantly higher in the Cost treatment (1.7% in P(Cost) and 0.6% in P(Base), Mann-Whitney U test p=0.005), which may suggest that although there is no difference in means, some players may withhold contributions due to the imposed fee.

not significant (p=0.3)[18] it is notable that the maximum contribution has been achieved.

Regarding the contribution behavior in the group without an institution we do not observe any additional spillover effects that come from the introduction of provision costs. As such we do not observe benefits of sorting as, for instance, in Gächter and Thöni, (2005), who exogenously sort players into groups according to their previous behavior and find that even in the groups of free-riders there is a high level of cooperation. In our experiment in the VCM players contribute on average about 20% of their endowment, which is in line with the results obtained in the experiments without migration (e.g., Fehr and Gächter, (2000)). That means that an endogenous sorting is not restrictive enough to induce benefits of homogeneity.

3.4.3 Use of punishment points

We had two competing hypotheses regarding the effect of the provision costs on punishment behavior. On the one hand, provision costs may be seen as sunk costs and people therefore would be using sanctions more. On the other hand, if provision costs are considered to be a part of the"punishment budget", players may be less likely to spend additional points on further punishment. The use of punishment points is mainly triggered by the contributions of other members of the group, i.e., it is sanctioning an undesirable behavior. To check whether there are differences in the use of punishment points we run a random-effects panel regression (see Table 3.1, panel 1), in which we control for the effects of a treatment, the average contribution of other players in the community and the level of contribution of an individual on her distributed punishment points as a dependent variable.[19] While the regression shows that fewer punishment points are indeed assigned if the average contribution of other players is high and that players who contribute more punish more, the dummy for treatment is not significant.

[18] All statistical tests reported are non-parametric, two-tailed and take communities as units of observations.

[19] We checked several other feasible specifications to test the robustness of the coefficients. The treatment dummy is not significant for most of them, although as in case for panel 2 it reaches 5% significance level.

N of distributed points	(1)	(2)
Treatment	0.0931	-2.638**
	(0.377)	(1.314)
Avg Contribution Others	-0.477***	-0.402***
	(0.0726)	(0.0903)
Free-Rider		-4.502**
		(1.926)
Free-Rider × Treatment		2.851**
		(1.242)
Contribution	0.104***	
	(0.0374)	
Constant	7.649***	12.46***
	(2.210)	(3.453)
Observations	2,875	2,875
Number of ID	141	141
R-squared	0.12	0.12

Robust standard errors in parentheses, clustered at the group level
*** p<0.01, ** p<0.05, * p<0.1
Note: Treatment dummy – 0 for Base and 1 for Cost,
Free-Rider dummy – 0 for Free-Rider and 1 else.

TABLE 3.1: Random-effects panel regression: Use of punishment points

Therefore we can conclude that there are no differences in the use of punishment triggered by the treatment manipulation. However, the effect of provision costs on punishment may be heterogeneous with different types of players reacting differently. To deeper explore how punishment is used by different types of players we classify them in two following ways.

First, we look into the use of punishment points depending on the player's own contribution and classify players as free-riders if their own contribution in the current period is zero, as full contributors if they contributed 20, and as contributors if they fall between the other two categories. We find that in the P(Base) free-riders distribute significantly more punishment points than both full contributors in the same treatment and free-riders in Cost (5.53 and 0.65 p=0.1 and 0.73 p=0.07 respectively)[20]. The fact that there is the difference in the punishment behavior of free-riders between treatments raises the question of who receives these points

[20]This result should be treated with caution as there are less free-riders in the Base treatment than full contributors and the result for free-riders is likely to be heavily influenced by outliers.

and if the introduction of provision costs leads to less anti-social punishment, i.e., punishment of fully cooperative members of the group.

Data suggest that rates of anti-social punishment are non-different between treatments (Base 6% of all distributed points and Cost 4% p=0.5), which means that in both treatments punishment points are directed predominantly at those who do not contribute fully. This result is in strong contrast with the results of Ramalingam et al., (2016). The core difference lies in the design feature of the latter experiment that not all players bear the provision costs of punishment but only those who volunteered. As a consequence volunteers punish those who contributed fully but did not incur provision costs. In our design provision costs do not contribute to the inequality within the community and therefore do not trigger anti-social behavior.

Although the difference in punishment between treatments is not significant, the fact that in the Cost treatment free-riders engage less often in punishing behavior suggests that at least for this group of players provision costs belong to the "punishment budget". Punishment in both treatments is targeted towards the players who do not contribute fully and therefore promotes higher rates of cooperation.

The second classification approach is based on the social value orientation questionnaire Murphy et al., (2011) and can provide only limited insight into the data as it was a part of the Cost treatment only.

According to the social value orientation questionnaire about 40% of the subjects in the Cost treatment are pro-social and the other 60% are individualistic.[21] There is a weak but significant correlation between the type and its contribution behavior in the VCM group (correlation 0.15 p=0.00), but not in the punishment group (correlation 0.03 p=0.3). That is, in the absence of institutions players behave in line with their type, however, the introduction of an institution can effectively change their behavior. Under the same institution the types do not differ in the level of contributions, but they appear to differ in the punishment behavior with individualist players punishing more (1.05 points) than pro-social (0.79 points p=0.00). This result may be due to the fact that pro-social players care about

We check if the result replicates in the regression (see Table 3.1, panel 2). The direction and significance of coefficients is consistent with the results of the non-parametric tests.

[21] None of the subjects were classified as altruistic or competitive.

the total welfare of the group and therefore may be reluctant to use punishment points. In turn, individualist players do not take the groups welfare into account in their punishment decision and therefore are more likely to punish non-contributors to insure higher contribution levels and as a consequence higher own profits in the next periods.

3.4.4 Use of migration opportunities and welfare impact of punishment

As discussed above the opportunity to "vote with the feet" leads to higher levels of cooperation, however, migration opportunities are not used extensively. In both treatments a large share of subjects did not migrate at all (i.e., remained in the community they chose in period one) or migrated once. Players migrate more actively in the Cost treatment (2.3 times in both directions in Base and 3.8 times in Cost p=0.04). On average in Base players migrate once from P to VCM community and 1.4 times from VCM to Punishment. The introduction of provision costs prolongs the period of active migration. There players migrate on average 1.6 times from Punishment to VCM and 2.3 times in the opposite direction. Consider in addition Fig.3.2 which displays the share of players who newly migrated[22] and share of those who remained in a certain community. The sum of shares of newly migrated players shows how actively migration opportunities are used in a certain period.[23] In both treatments players migrate more actively in the beginning of the experiment but even at its highest point in Base migration barely reaches 30% of subjects changing communities (if we disregard the direction of migration). Therefore linking our results back to the results of Gürerk, (2007), we can conclude that it is rather the opportunity to change the community than its actual use that promotes higher levels of cooperation.

As expected P(Cost) treatment is less profitable due to the deduction of provision costs. Members of P(Base) receive on average 47.66 points per round, whereas in P(Cost) 45.32 points (p=0.078). That is, player in P(Cost) earn on average 2.34

[22]Players who changed the community as compared to the previous period.

[23]In the first period all players are coded as "remained" since players made an initial choice of the community.

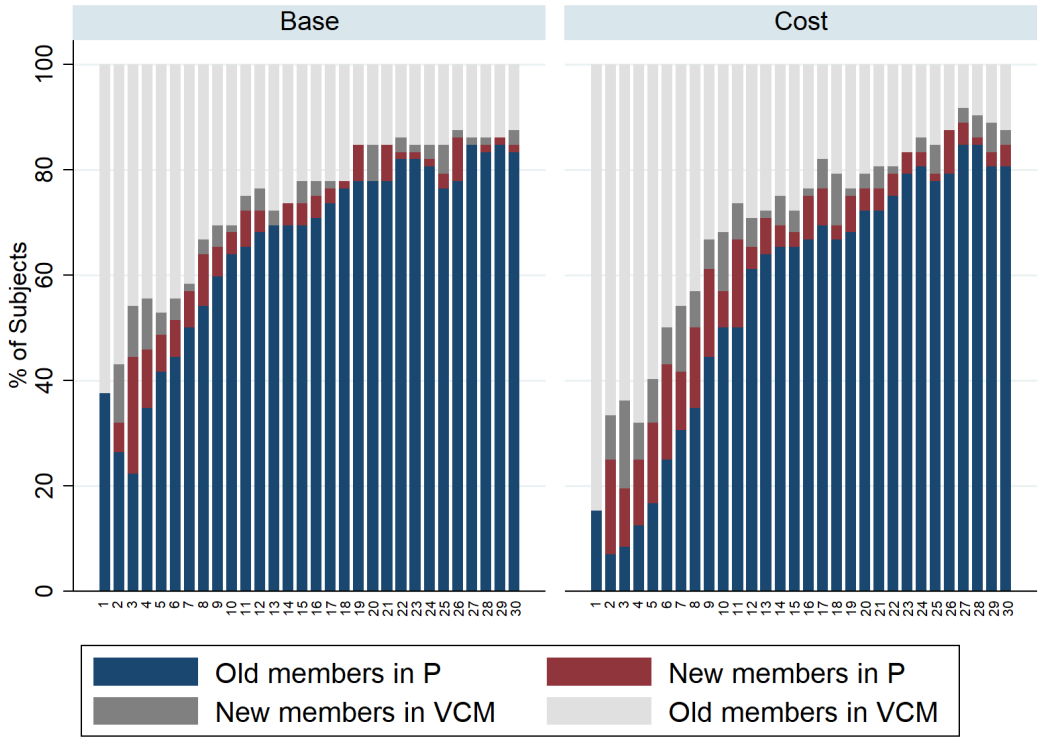


FIGURE 3.2: Share of newly migrated and remained members in the communities. Use of migration opportunities

points fewer than in P(Base), which almost perfectly corresponds to the provision cost parameter.

Reconfirming previous findings of Fehr and Gächter, (2000), Gächter, Mengel, et al., (2017), and Gächter, Renner, et al., (2008) we find that peer-punishment is welfare detrimental in the short run. In the first third in the Base treatment round outcomes are non-different between communities (VCM 42.7, P 41.7 p=0.42), in the Cost treatment VCM is more profitable than peer-punishment and the difference almost reaches conventional significance levels (VCM 42.5, P 36.8 p=0.11). Nevertheless, in the first third on average 37% of players join the punishment group and it is also a period when the punishment group grows the fastest. In the second third of the experiment P(Cost) (43.7) performs as good as VCM(Cost) (42.2 p=0.20) and attracts as much as 71% of the players and only in the last third it becomes more profitable (48.3 and 42.2 p=0.02 with 83% of the population). P(Base) compares better to the VCM(Base) becoming significantly more profitable already in the 2nd third of the experiment (p=0.05).

It is important to consider how many players remain in the punishment community during its unprofitable period (i.e., first third) or whether in every period players who experienced that the punishment community is unprofitable leave and get replaced by the unexperienced players from VCM.

Fig. 3.2 displays shares of population who migrated in a certain period as well as shares of members who remained. We establish that in the first third of the experiment significantly more new players migrate towards the punishment community than those who leave for VCM (16.5% and 10% respectively, p=0.006,). About 25% of the community members remain in the peer-punishment community in every period while it is unprofitable. For the rest of the experiment the difference in the shares of newly migrated players in the communities are non-different. That is, although the punishment community is less profitable than the VCM in the beginning not all the players migrate away after experiencing that it is less profitable. The punishment community attracts more new members than it loses to the VCM.

The fact that people join the P(Cost) community before it becomes profitable suggests the following. First of all, there is no "natural resentment to punishment" (Fehr and Rockenbach, 2003). On the contrary, if players believe that a punishment institution is potentially welfare enhancing they would remain there even if it has not revealed its potential yet. The other conclusion, which could be drawn from this observation is that players prefer to cooperate and earn less than not to cooperate but earn more, which means that there is some non-monetary utility from cooperation. That interpretation is in line with the results of Abbink et al., (2017). They find that even if cooperation is not welfare enhancing, but welfare detrimental, players enforce a cooperative norm.

In the last third of the experiment we observe a stable share of players being members in the punishment community, full cooperation, almost no use of punishment and very low levels of migration. Nevertheless ca. 20% of players remain in the VCM community both in the Cost and the Base treatments. That may be preliminarily interpreted in line with our sorting hypothesis, which we consider in greater detail in the following section.

To sum up, our results suggest that it is not the purely monetary benefits of a peer-punishment institution that attract players. Even if a peer punishment community is modified to be less profitable (or cost-efficient in terms of Markussen et al., (2013)) due to the deduction of provision costs, the absolute majority chooses it and contributes fully. Players join the punishment group before it becomes profitable, suggesting that they opt for an institutional framework and not simply maximize their income.

3.4.5 Sorting

Further we explore if the introduction of provision costs helps players to sort into communities by type. Admitting that potentially there can be separation along a very wide range of criteria, we look only into a few possible classifications. We check if there is a separation depending on a first period cooperation type, on political preferences of players and on their social value orientation type.

First, we classify them according to the contribution behavior in the period one as a revelation of the initial intention to cooperate (similar approach is used, for instance, in Soest et al., (2016)). As before, we apply a rather strict criteria: players who contributed 20 are full contributors, who contributed 0 are free-riders and all others fall in the broad "contributor" category. According to this classification in Base 22% of the sample are full contributors, 73% contributors and 4% free-riders. In the Cost treatment the distribution is 29%, 59% and 11% respectively (Kruskal-Wallis equality-of-populations test p=0.87). In Base in the first half of the experiment, all the initial free-riders remain in the VCM community and migrate to the punishment community afterwards.[24] There is almost no change in the share of full-contributors as almost all of them join the punishment community from the very beginning and remain there throughout the game. In the Cost treatment we observe more variation. There free-riders also initially join the VCM, but as soon as in period 6, they start to join the costly punishment community. Contrary to expectations, the provision costs deter some of the full contributors from joining the punishment group. Initially, full contributors are split almost evenly

[24]Fig. A1 displays the composition of communities over time based on the first period classification of the contribution behavior.

between two communities, afterwards the share in the punishment group grows to ca. 80% of all full contributors choosing the peer-punishment group, however, the punishment community never attracts all of them. Therefore, provision costs do not foster separation in line with this classification of players. The fact that all the "initial" free-riders join the punishment community does not prevent the P(Cost) from reaching 99% of cooperation in the last third. As mentioned before, players do not behave according to the initial type in the presence of the cooperation promoting institution.

As a part of socio-demographic questionnaire participants reported their political attitudes. Players were asked to report their political preferences on a 5 point scale where the very left point of the scale was marked "left" and very right point "right" leaving it for players themselves to interpret the meaning of the intermediate points. The obtained distribution of political attitudes is skewed to the left, which is to be expected from the predominantly student subject pool in Hamburg.[25] Our hypothesis is that players on the right side of the spectrum would be more likely to avoid the punishment community or join it later due to an avoidance of fixed provision costs (i.e. tax). Indeed, while in the Cost treatment in the first periods of the experiment all players with right political allegiances are in the VCM group and join punishment community only later, in Base they join the punishment community from the very beginning.[26] However, this result should be treated with caution as our sample features a very small proportion of players to the right of the political spectrum.

As political attitudes were self-reported, in the Cost sessions[27] we applied the anchoring technique to control for the bias in the reported values. Additionally to reporting own political attitudes, players were asked to locate a major German political party (CDU) on the same scale. For a true position of the party on the political spectrum we resort to the values reported in the Manifesto database[28] which are largely consistent with other sources. We use the true value and the players'

[25]Kruskal-Wallis test does not detect differences in the distribution of political preferences between treatments p=0.27

[26]For the composition of communities depending on the political attitudes of its members over time see Fig. A2.

[27] The data for the Cost treatment were collected after the Base treatment, therefore some minor changes and checks were introduced in the Cost treatment only.

[28]https://manifesto-project.wzb.eu/information/documents/visualizations [Accessed 20.06.2018]

reported value to measure the magnitude and direction of an individual bias and to harmonize self-reported values across individuals. Such a de-biasing procedure leads to even more players landing to the left of the spectrum, but does not lead to substantial changes in the self-reported values or affects the observations reported above.

Finally, we consider a classification based on the social value orientations of players. As mentioned above, since the social value orientation data is available only for the Cost treatment, we can not compare the dynamics of community composition over time. Moreover, within the Cost treatment according to the SVO measure there are only two types of players: pro-social and individualist. Taking into account these restrictions, our insights from this measure are limited. We observe that in the period one 80% of all the pro-social subjects are members of the VCM group as well as 88% of all the individualistic subjects. By the last period we observe similar composition of types but already in the punishment community (92% of all the pro-social players and 79% of all the individualistic players). Looking at the proportion of different types within the community, we can see that initially the punishment community is equally populated by pro-social and individualistic players, while in the VCM community, 64% of players are individualistic. In the period 30, 78% of VCM members are individualistic, while in the punishment community the share of individualistic players is slightly lower (57%). That is, by the end of the game the majority of pro-social and individualistic players belong to the punishment group, the smaller VCM group is predominantly populated by the individualists.

The classifications considered here are far from being an exhaustive list of criteria which could potentially drive the separation. Therefore we abstain from making a conclusion that there is no separation by type of some kind in the framework of open communities. However, as the share of players remaining in the VCM community in both treatments is non-different, it appears that provision costs do not contribute to this process.

3.5 Conclusion

Studies of how effective institutions are in promoting desired behavior are undoubtedly important. However, growing empirical research acknowledges that the institutions are opted for, i.e., voted for, by its members. This means that even the most efficient institution will not be of use if for whatever reason nobody is willing to join it. In the current paper we addressed the question of how fixed per period provision costs affect the popularity of a cooperation-enhancing institution.

We found that the introduction of provision costs does not dramatically change the pattern of migration. One may argue that provision costs parameter might be too low to trigger a behavioral change. In our design, provision costs are low, first of all, because the peer-punishment mechanism as such does not resolve the social dilemma but merely provides a framework, where cooperation is more likely. Secondly, if players are prone to a fixed cost aversion, then we would expect discontinuity to appear close to zero in line with the findings of Shampanier et al., (2007). In fact the introduction of the small provision costs is already enough to render the punishment community less profitable than the alternative in the first third of the experiment. This means that, if the initial profits of the community are of major importance in a community choice, players should not have joined this community already with our low cost parameter.

In the Cost treatment, the peer-punishment community with provision costs grows slower, but is at the end of day as popular as the one without provision costs. That is, players are ready to bear provision costs to be a member of a potentially better environment. The introduction of provision costs renders the community less profitable, thus aggravating the problem of the welfare detrimental effect of peer-punishment in the short run. Notably, almost 40% of players join the peer-punishment community in the early periods when it is less profitable than its free alternative. Players do not change communities very often, but the opportunity of migration itself appears to generate a high democratic premium for cooperation. The introduction of provision costs seems to amplify this effect even further, since cooperation levels in the costly peer-punishment community by the end of the game reach their maximum. Although the increase in contributions is not significant

as compared to the peer-punishment without provision costs, it is notable that extra costs which at first sight are supposed to harm the community, promote cooperative behavior. Taken together results of the experiment prove that the majority of players when choosing between peer-punishment and no institution in a social dilemma situation prefer peer-punishment and that these preferences are robust towards changes in the effectiveness of the punishment institution.

The world is rich in social dilemmas of different scale and magnitude and people employ different institutional arrangement to overcome them. The gradual comparison and movement between alternative settings assure the emergence of persistent solutions for such dilemmas. This study is one of a few, which considers performance of institutions not in isolation, but in relation and competition to each other.

The framework of open communities allows for multiple avenues for further research. By comparing multiple institutional designs we can establish what is the cornerstone of a popularity of institution. As recent findings have demonstrated, costs and effectiveness of institutions may be of lesser importance than assumed. Potentially, it is important to consider what is the role of information transparency and migration costs in generating the cooperation premium of voting with feet.

Bibliography

Abbink, Klaus, Lata Gangadharan, Toby Handfield, and John Thrasher (2017). "Peer punishment promotes enforcement of bad social norms". In: *Nature Communications* 8.1, p. 609.

Ahn, Toh-Kyeong, R Isaac, and Timothy C Salmon (2008). "Endogenous group formation". In: *Journal of Public Economic Theory* 10.2, pp. 171–194.

– (2009). "Coming and going: Experiments on endogenous group sizes for excludable public goods". In: *Journal of Public Economics* 93.1, pp. 336–351.

Andreoni, James and Laura K Gee (2012). "Gun for hire: delegated enforcement and peer punishment in public goods provision". In: *Journal of Public Economics* 96.11, pp. 1036–1046.

Andreoni, James, William Harbaugh, and Lise Vesterlund (2003). "The Carrot or the Stick: Rewards, Punishments, and Cooperation". In: *The American Economic Review* 93.3, pp. 893–902.

Benabou, Roland and Jean Tirole (2003). "Intrinsic and extrinsic motivation". In: *The Review of Economic Studies* 70.3, pp. 489–520.

Bénabou, Roland and Jean Tirole (2011). "Identity, morals, and taboos: Beliefs as assets". In: *The Quarterly Journal of Economics* 126.2, pp. 805–855.

Bernheim, B Douglas (1994). "A theory of conformity". In: *Journal of Political Economy* 102.5, pp. 841–877.

Blanco, Mariana, Dirk Engelmann, and Hans Theo Normann (2011). "A within-subject analysis of other-regarding preferences". In: *Games and Economic Behavior* 72.2, pp. 321–338.

Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch (2014). "hroot: Hamburg registration and organization online tool". In: *European Economic Review* 71, pp. 117–120.

Charness, Gary and Matthew Rabin (2002). "Understanding social preferences with simple tests". In: *The Quarterly Journal of Economics* 117.3, pp. 817–869.

Chaudhuri, Ananish, Pushkar Maitra, Susan Skeath, et al. (2006). *Communication, Advice and Beliefs in an Experimental Public Goods Game*. Monash University.

Chugunova, Marina, Wolfgang Luhan, and Andreas Nicklisch (2017). "When to Leave Carrots for Sticks: On the Evolution of Sanctioning Institutions in Open Communities". In: URL: http://bedarfsgerechtigkeit.hsu-hh.de/dropbox/wp/2017-09.pdf.

Connelly, Brian L, S Trevis Certo, R Duane Ireland, and Christopher R Reutzel (2011). "Signaling theory: A review and assessment". In: *Journal of Management* 37.1, pp. 39–67.

Crosetto, Paolo, Ori Weisel, and Fabian Winter (2012). "A Flexible z-Tree Implementation of the Social Value Orientation Slider Measure (Murphy et al. 2011)–Manual". In: *Jena Economic Research Paper 2012 - 062*.

Dal Bó, Pedro, Andrew Foster, and Louis Putterman (2010). "Institutions and behavior: Experimental evidence on the effects of democracy". In: *American Economic Review* 100.5, pp. 2205–29.

Danilov, Anastasia and Dirk Sliwka (2016). "Can contracts signal social norms? Experimental evidence". In: *Management Science* 63.2, pp. 459–476.

Dufwenberg, Martin (2008). "Psychological games". In: *The New Palgrave Dictionary of Economics* 6, pp. 714–717.

Falk, Armin and Christian Zehnder (2013). "A city-wide experiment on trust discrimination". In: *Journal of Public Economics* 100, pp. 15–27.

Fehr, Ernst and Simon Gächter (2000). "Cooperation and punishment in public goods experiments". In: *American Economic Review* 90.4, pp. 980–994.

– (2002). "Altruistic punishment in humans". In: *Nature* 415.6868, pp. 137–140.

Fehr, Ernst and Bettina Rockenbach (2003). "Detrimental effects of sanctions on human altruism". In: *Nature* 422.6928, p. 137.

Fehr, Ernst and Klaus M Schmidt (1999). "A theory of fairness, competition, and cooperation". In: *The Quarterly Journal of Economics* 114.3, pp. 817–868.

Fehr, Ernst and Tony Williams (2017). "Creating an Efficient Culture of Cooperation". In: *Working Paper*.

Fischbacher, Urs (2007). "z-Tree: Zurich toolbox for ready-made economic experiments". In: *Experimental Economics* 10.2, pp. 171–178.

Gächter, Simon, Friederike Mengel, Elias Tsakas, and Alexander Vostroknutov (2017). "Growth and inequality in public good provision". In: *Journal of Public Economics* 150, pp. 1–13.

Gächter, Simon, Elke Renner, and Martin Sefton (2008). "The long-run benefits of punishment". In: *Science* 322.5907, pp. 1510–1510.

Gächter, Simon and Christian Thöni (2005). "Social learning and voluntary cooperation among like-minded people". In: *Journal of the European Economic Association* 3.2-3, pp. 303–314.

Gravel, Nicolas, Alessandra Michelangeli, and Alain Trannoy (2006). "Measuring the social value of local public goods: an empirical analysis within Paris metropolitan area". In: *Applied Economics* 38.16, pp. 1945–1961.

Gürerk, Özgür (2007). "Endogenous Institution Choice in Social Dilemmas". PhD thesis. Dissertation, University of Erfurt.

Gürerk, Özgür, Bernd Irlenbusch, and Bettina Rockenbach (2006). "The competitive advantage of sanctioning institutions". In: *Science* 312.5770, pp. 108–111.

– (2014). "On cooperation in open communities". In: *Journal of Public Economics* 120, pp. 220–230.

Herrmann, Benedikt, Christian Thöni, and Simon Gächter (2008). "Antisocial punishment across societies". In: *Science* 319.5868, pp. 1362–1367.

Kessler, Judd B and Stephen Leider (2012). "Norms and contracting". In: *Management Science* 58.1, pp. 62–77.

Kosfeld, Michael, Akira Okada, and Arno Riedl (2009). "Institution formation in public goods games". In: *American Economic Review* 99.4, pp. 1335–55.

Kriss, Peter H, Andreas Blume, and Roberto A Weber (2016). "Coordination with decentralized costly communication". In: *Journal of Economic Behavior & Organization* 130, pp. 225–241.

Markussen, Thomas, Louis Putterman, and Jean-Robert Tyran (2013). "Self-organization for collective action: An experimental study of voting on sanction regimes". In: *Review of Economic Studies* 81.1, pp. 301–324.

Murphy, Ryan O, Kurt A Ackermann, and Michel JJ Handgraaf (2011). "Measuring Social Value Orientation". In: *Judgment and Decision Making* 6.8, pp. 771–781.

Palfrey, Thomas, Howard Rosenthal, and Nilanjan Roy (2017). "How cheap talk enhances efficiency in threshold public goods games". In: *Games and Economic Behavior* 101, pp. 234–259.

Putterman, Louis, Jean-Robert Tyran, and Kenju Kamei (2011). "Public goods and voting on formal sanction schemes". In: *Journal of Public Economics* 95.9-10, pp. 1213–1222.

Ramalingam, Abhijit, Sara Godoy, Antonio J Morales, and James M Walker (2016). "An individualistic approach to institution formation in public good games". In: *Journal of Economic Behavior & Organization* 129, pp. 18–36.

Seçilmiş, Erdem (2018). "An experimental analysis of moral self-regulation". In: *Applied Economics Letters* 25.12, pp. 857–861.

Shampanier, Kristina, Nina Mazar, and Dan Ariely (2007). "Zero as a special price: The true value of free products". In: *Marketing Science* 26.6, pp. 742–757.

Sliwka, Dirk (2007). "Trust as a signal of a social norm and the hidden costs of incentive schemes". In: *American Economic Review* 97.3, pp. 999–1012.

Soest, Daan van, Jan Stoop, and Jana Vyrastekova (2016). "Toward a delineation of the circumstances in which cooperation can be sustained in environmental and resource problems". In: *Journal of Environmental Economics and Management* 77, pp. 1–13.

Sunstein, Cass R (1996). "On the expressive function of law". In: *University of Pennsylvania Law Review* 144.5, pp. 2021–2053.

Sutter, Matthias, Stefan Haigner, and Martin G Kocher (2010). "Choosing the carrot or the stick? Endogenous institutional choice in social dilemma situations". In: *The Review of Economic Studies* 77.4, pp. 1540–1566.

Thaler, Richard (1985). "Mental accounting and consumer choice". In: *Marketing Science* 4.3, pp. 199–214.

Tyran, Jean-Robert and Lars P Feld (2006). "Achieving compliance when legal sanctions are non-deterrent". In: *The Scandinavian Journal of Economics* 108.1, pp. 135–156.

Van der Weele, Joel (2009). "The signaling power of sanctions in social dilemmas". In: *The Journal of Law, Economics, & Organization* 28.1, pp. 103–126.

A.1 Tables and Figures

Table A1: Mann-Whitney U test: Overall

Contributions	VCM(Base) 4.1	P(Base) 17.4	10.75
VCM(Cost) 4.1	p=0.87	p=0.004	
P(Cost) 18.5	p=0.004	p=0.63	
11.32			

Points per round	VCM(Base) 42.5	P(Base) 47.7	45.1
VCM(Cost) 42.5	p=0.87	p=0.04	
P(Cost) 45.3	p=0.055	p=0.078	
43.9			

Table A2: Mann-Whitney U test: Periods 1-10

Contributions	VCM(Base) 4.4	P(Base) 16.0	10.2
VCM(Cost) 4.3	p= 0.87	p=0.007	
P(Cost) 14.4	p=0.004	p=0.42	
9.3			

Points per round	VCM(Base) 42.7	P(Base) 41.7	42.2
VCM(Cost) 42.5	p= 0.87	p= 0.42	
P(Cost) 36.8	p= 0.15	p=0.15	
39.7			

Table A3: Mann-Whitney U test: Periods 11-20

Contributions	VCM(Base) 0.9	P(Base) 17.4	10.8
VCM(Cost) 3.6	p=0.46	p=0.01	
P(Cost) 17.5	p=0.01	p=0.87	
11.2			

Points per round	VCM(Base) 40.6	P(Base) 48.4	45.3
VCM(Cost) 42.2	p=0.46	p= 0.05	
P(Cost) 43.7	p=0.13	p=0.11	
43			

TABLE A4: Mann-Whitney U test: Periods 21-30

Contributions	VCM(Base) 3.1	P(Base) 17.9	12
VCM(Cost) 3.7	p=0.39	p=0.02	
P(Cost) 19.9	p=0.01	p=0.31	
13.4			

Points per round	VCM(Base) 41.8	P(Base) 49.6	46.5
VCM(Cost) 42.2	p=0.38	p=0.02	
P(Cost) 48.3	p=0.01	p=0.31	
45.8			

Table A5: Table representation of Figure 3.1

Periods	Contribution				Distributed points		% of Subjects in the Punishment Group	
	Base		Cost					
	P	VCM	P	VCM	P(Base)	P(Cost)	Base	Cost
1-3	14.9	6.2	14.7	7.8	2.8	1.8	38.0	20.2
4-6	18.3	3.1	14.6	2.9	1.1	1.8	48.6	33.4
7-9	19.1	0.9	17.5	1.6	1.0	2	62.0	50.9
10-12	19.0	0.5	18.5	2.4	1.2	0.9	70.8	63.0
13-15	19.5	0.3	19.4	4.8	0.5	0.6	72.2	69.1
16-18	19.3	0.7	19.3	2.7	0.2	1.1	76.4	73.4
19-21	18.3	3.0	19.9	1.9	0.2	0.4	82.4	76.3
22-24	18.4	3.7	20	3.3	0.3	0.1	82.9	82.0
25-27	18.6	3.1	19.8	3.8	0.3	0.9	83.3	85.0
28-30	18.8	2.9	20	2.4	0.4	0.03	85.2	84.6

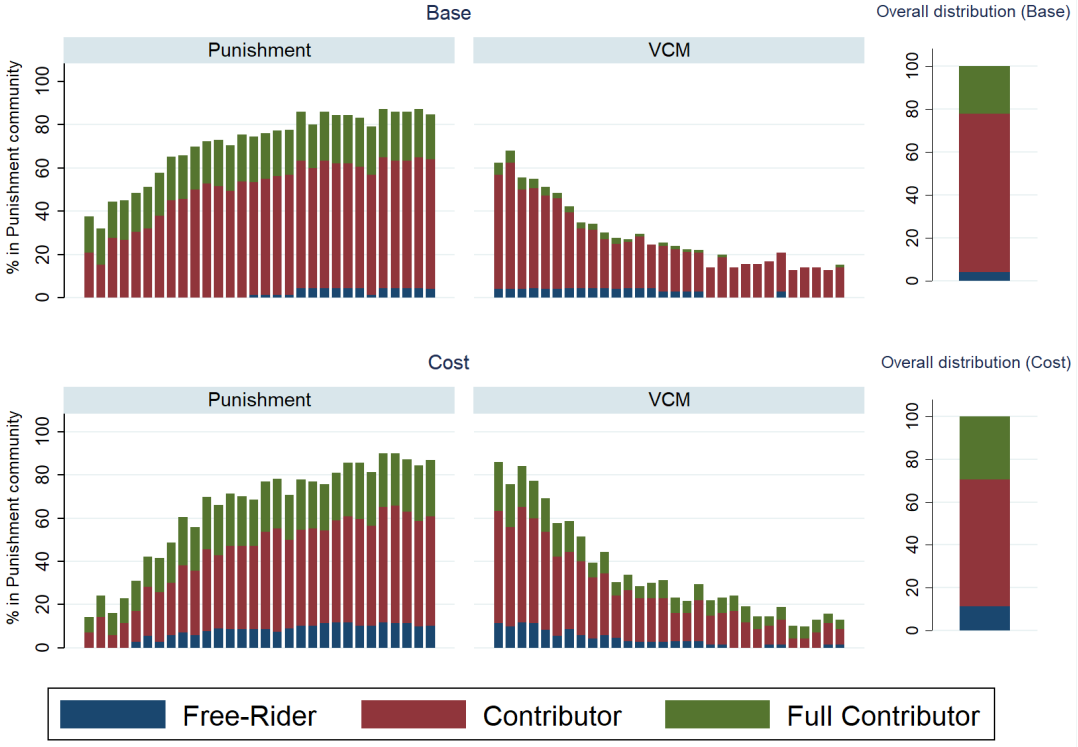


FIGURE A1: Community composition based on cooperation behavior in period one over time

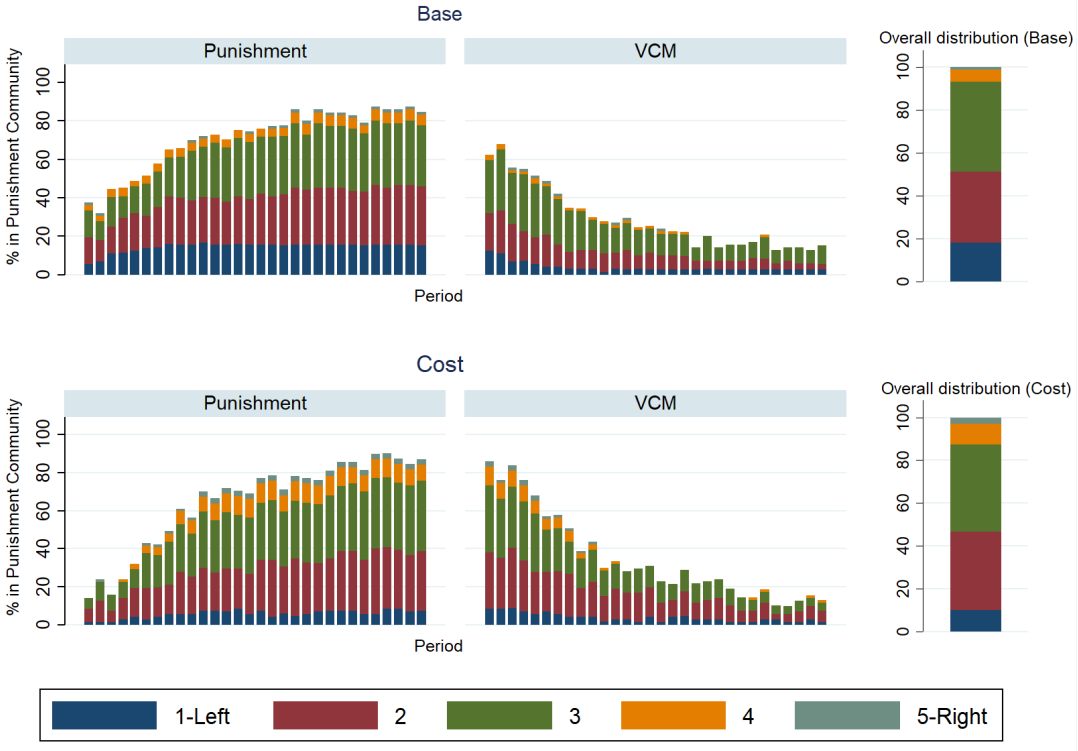


FIGURE A2: Community composition based on political views over time

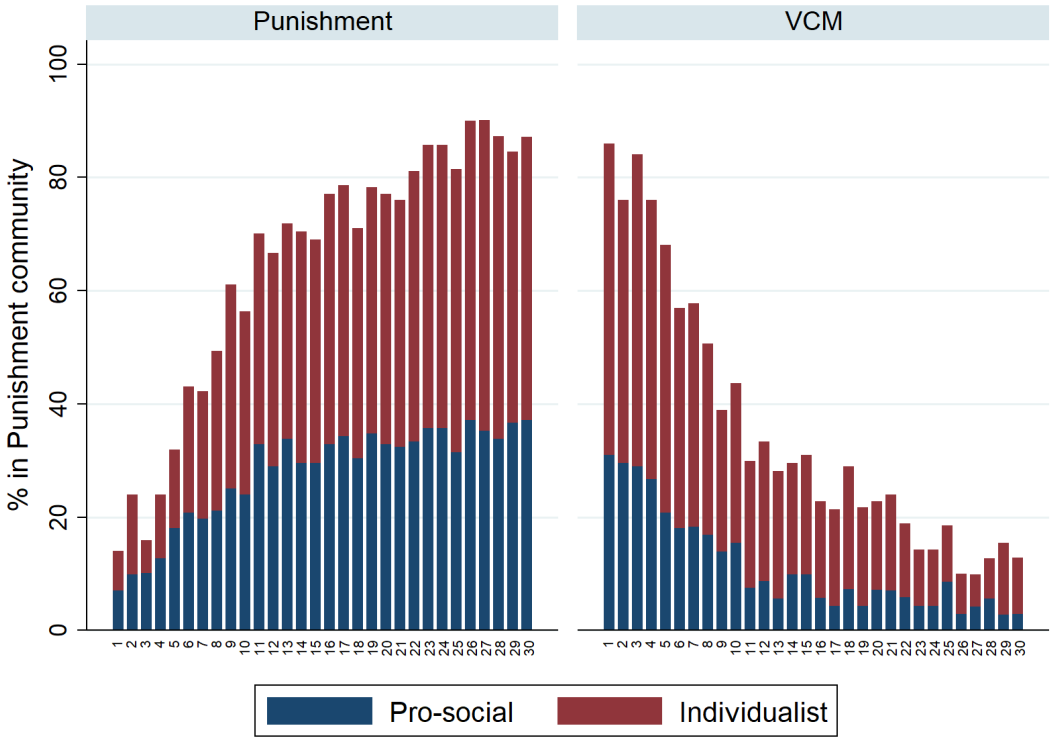


FIGURE A3: Community composition based on social value orientation type over time

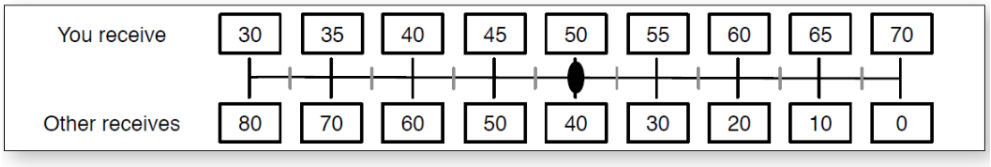
A.2 Instructions: Social value orientation scale[29]

Instructions

Experiment consists of 2 parts and a questionnaire. First, you receive instructions for part one of the experiment. You will receive instructions for part two at the end of part one.

In this task you have been randomly paired with another person, whom we will refer to as the other. This other person is someone you do not know and will remain mutually anonymous. All of your choices are completely confidential. You will be making a series of decisions about allocating resources between you and this other person. For each of the following questions, please indicate the distribution you prefer most by marking the respective position along the midline. You can only make one mark for each question. Your decisions will yield money for both yourself and the other person. In the example below, a person has chosen to distribute money so that he/she receives 50 cents, while the anonymous other person receives 40 cents.

Example:



There are no right or wrong answers, this is all about personal preferences. After you have made your decision, please click OK. You can choose only one distribution per task. As you can see, your choices will influence both the amount of money you receive as well as the amount of money the other receives. After you and the other made all 6 decisions, one decision in each group will be randomly chosen. You will learn which decision was chosen at the end of the part two. After the completion of part two, your payment for both parts will be summarized and shown on the screen in Euro.

Please note:

No communication is allowed during the entire experiment. If you have a question, please raise your hand. All decisions are anonymous, meaning that none of the other participants or the experimenters will know the identity of the person who has made a particular decision. The payout is also anonymous: no participant will know how much other participants have earned.

[29] This is an adaptation of the instructions provided at Murphy et al., (2011).

A.3 Instructions: the Cost treatment[30]

Instructions for the experiment

General instructions

At the beginning of the experiment you will be randomly divided into **2 groups of 12 participants** each. During the experiment, you will be interacting only with the members of the same subgroup.

At the beginning of the experiment each participant receives a **starting endowment of 1000 points**. Before the beginning of the experiment there are 4 questions of understanding.

Procedure

Experiment consists of **30 rounds.** Each round consists of 2 stages. The first stage includes the choice of the group and decision about the contribution into the common project. In the second stage participants can affect the outcome of other members of his/her group.

First Stage

(i) Choice of the group

In the first stage, each participant can decide, which group to join.

There are two different groups:

	Affecting the income of the other members of the group
Group	A: Through assignment of negative points
	B: Through assignment of positive points

If you choose **Group A**, 2 points will be deducted **from your income of the first stage**

(ii) Contribution to the common project

Each round in the beginning of the first stage each member of the group receives an **endowment** of **20 Points**.

You decide how many of the 20 points you want to **contribute** to the common project. The remaining part of the endowment, you **keep to yourself.**

Calculation of your income from the first stage

Your income from the first stage consist of two (three) parts:

- **Points which you kept to yourself** = Endowment – your contribution to the common project
- Your **profit from the common project** = 1.6 x sum of all the contribution to the project of all the community members/ number of members in the community
- Deduction of 2 points if one is a member of Group A

Therefore, **your income from first stage in Group A** is:

20 – your contribution to the common project +
1.6 x (Sum of all the contributions to the common project of all the members of the community / Number of members in the group) – 2 points

[30]This is a translation of the original instructions in German which can be obtained from the author upon request.

Therefore, **your income from first stage in Group B** is: 20 – your contribution to the common project + 1.6 x Sum of all the contributions to the common project of all the members of the community / Number of members in the group

The profit from the common project will be calculated according this formula individually for each member of the community. **Please note:** Each member of the community receives the same profit from the common project. It means that each member of the group profits from **all** the contributions to the project.

Second Stage

Assignment of Points

In the second stage, you see how much each member of the group contributed to the common project. **(Please note: The sequence of players is reshuffled every round.** Therefore, it is not possible to identify a group member over different rounds by their position in the displayed list.)

By assigning points, you can **decrease** or leave the income of the other group members **unchanged**.

In the second stage of every round each participant receives **20 additional points**. You decide how many of these 20 points you want to **assign** to the other group members. You **keep** the remaining points to yourself. To check how many points you have assigned, press the *Calculate Points* button on your screen.

- **Each negative point,** which you assign to a group member, **reduces the income of this group member by 3 points.**
- If you assign **0 points** to a group member, **the income of this group member remains unchanged.**

Calculation of your income from the second stage

Your income from the second stage consist of two parts:

- **Points kept to yourself** = 20 – sum of the points you assigned to the other group members.
- **minus 3 times the number of negative points** that you received from other group members.

Therefore, **your income from second stage in Group A** is: 20 – sum of the points you assigned to the other group members – 3 x (Negative points, which you received from the other group members)
Therefore, **your income from second stage in Group B** is: 20

Calculation of your round income in Group A

Your round income is calculated as follows:

	Your income from the first stage	= 20 – your contribution to the common project + 1.6 x (Sum of all the contributions to the common project of all the members of the community / Number of members in the group) – 2 points
+	Your income from the second stage	= 20 – sum of the points you assigned to the other group members – 3 x (Negative points, which you received from the other group members)
=	Your round income	

Calculation of your round income in Group B

Your round income is calculated as follows:

	Your income from the first stage	= 20 – your contribution to the common project + 1.6 x (Sum of all the contributions to the common project of all the members of the community / Number of members in the group) – 2 points
+	Your income from the second stage	= 20
=	Your round income	

Special case: single group member

Should you be **the only member in your group**, you receive 20 points in the first stage and 20 points in the second stage, that is your round income is **40**. You can act neither during the first nor during the second stage.

Information at the end of the round

At the end of the round, you will receive a detailed overview of the results from all groups. For each member of the group, you will see: their contribution to the project, their income from the first stage, points they assigned (if possible), points they received (if possible), their income from the second stage and their round income.

History

Starting in round two, before the start of a new round you will see an overview of the average results (as above) of all previous rounds.

Total income

The total income from the experiment consists of the starting capital of 1000 points plus the sum of the round incomes of the 30 rounds.

At the end of the experiment, your total income will be paid at an exchange rate of 1 € per 100 points.

Good Luck!

Chapter 4

Redistribution and Production with the Subsistence Income Constraint: a Real-Effort Experiment

Marina Chugunova, Andreas Nicklisch and Kai-Uwe Schnapp[1]

Abstract

A large body of literature demonstrates that redistribution leads to inefficiencies due to distorted work incentives. Yet, this result is typically obtained in environments that leave people absolutely free in their labor-leisure allocation decisions. We challenge this assumption and study labor supply decisions in a framework with a subsistence income constraint and a redistribution system which supports disadvantaged players. Our real-effort experiment offers two sets of results. First, the introduction of a moderate subsistence income threshold leads to a decrease in performance. Second, varying the cause of taxation in this environment does not affect labor responses of taxpayers, but enhances the productivity of welfare recipients if they are the beneficiaries of the redistribution.

Keywords: redistribution, subsistence income, implicit costs of taxation, needy players, meaningfulness of taxation

JEL-Classification: C91, H21, H24

[1]Financial support by the German Research Foundation (DFG) for Nicklisch (Ni 1610/1-1) and Schnapp (Schn 1282/4-1) is gratefully acknowledged.

4.1 Introduction

Economic inequality is on the rise today even in the richest as well as the most equality-oriented economies. There is a broad consensus among researchers that increasing inequality is not harmless and may cause severe social and political tensions within a society (e.g., Piketty, 2014; Staab and Nachtwey, 2016), and potentially lead to a decay of it (Chugunova et al., 2017). According to Joseph Stiglitz, a former chief economist of the World Bank, inequality may lead to increasing societal division, distributive struggle, and even violent conflict (Stiglitz, 2012). Thus, curbing inequality is an important goal for a society. Attempts to do so range from radical proposals like the introduction of a universal basic income, to more moderate and widespread policies such as unemployment benefits and family allowances. To many economists, however, redistribution is a cure that is potentially worse than the disease it is supposed to heal. They fear that redistribution may harm economic growth, as it distorts labor incentives.[2] Hence, redistribution may come with high implicit costs since it reduces overall production within societies (e.g., see the seminal article by Mirrlees, (1971)).

There is a large variety of implicit costs caused by redistribution: income taxation may bias career plans, transfers may influence the decision to investment in human capital or to save for later periods (e.g., Blundell and MaCurdy, (1999) and Blundell and Walker, (1986)). The consequences may be complicated, multilayered, and contradictory. But even in the absence of complex scenarios and long-lasting effects of redistribution, predicting an individual response to redistribution in terms of working time is far from easy. The reason for this is that the individually chosen mix of work and leisure in response to taxation or transfer payments reflects two effects: first, with income taxation and transfers work becomes less and leisure becomes more attractive, implying that individuals work less (substitution effect). Second, taxation makes it harder for taxpayers or easier for transfer recipients to reach a certain income target that individuals need for their daily lives (income effect).

[2]Therefore, there is a broad economic tradition analyzing the interference between redistribution and growth (e.g., Feldstein, 1999; Harberger, 1964; Keane, 2011).

In the presence of a subsistence income constraint, one can think of situations in which the marginal increase of an income tax may lead to an increase in labor time. In other words, tax variations may cause an income effect that outweighs the substitution effect between labor and leisure time. In turn, the substitution effect may dominate the income effect and therefore a marginal increase of an income tax may lead to a decrease in labor time. Farkas et al., (1996) find that 65% of Americans (participants of the Public Agenda survey) claim that the most "upsetting" thing about transfers and redistribution is that "it encourages people to adopt the wrong lifestyle and values" (ibid., p.9).

While in theory the link between redistribution and labor supply is clear. The empirical literature is far from being conclusive in this regard (see Keane, (2011) and Keane and Rogerson, (2015)).[3] Moreover, there is little evidence that targeted transfer programs for poor citizens discourage work (Banerjee et al., 2017). This unsettled picture regarding the effect of redistribution is further reflected in the results of field studies on wage increases. Here, a number of field experiments fail to document labor increases associated with wage increases (Camerer et al., (1997) and Farber, (2005, 2008)). Camerer et al., (1997), for instance, find in a field experiment with cab drivers that the wage elasticity of drivers is substantially negative. This means that a wage increase even leads to less labor and more leisure time. Along the same line of argument, Fehr and Goette, (2007) conduct a controlled field experiment with bicycle messengers. By assigning a wage premium of 25% per ride for a certain time, they provide evidence for both substitution and income effect. While bicycle messengers take more shifts (meaning that they substitute leisure time with labor time as standard theory would predict), they also take a smaller number of rides per shift. This means by implication that they work less hard as the income target is reached nonetheless.

Despite this evidence from the field, experimental economists in their micro-analysis of tax-responsiveness focused primarily on the substitution effect. For

[3]Interestingly, Keane, (2011) reports on important gender differences with respect to the responsiveness of labor supply to changes in wages and taxes: "At least for males, it is fair to say that most economists believe labor supply elasticities are small. [...] For women, in contrast, it is fair to say that most studies find large labor supply elasticities,..." (ibid., p.1). We will come back to this point in the analysis of our results.

instance, in a laboratory experiment without explicit minimum income restriction, Agranov and Palfrey, (2015) show that higher tax rates are associated with lower (aggregated) labor supply.[4] Other studies suggest that there are psychological biases that substantially affect the implicit costs of taxation: Kessler and Norton, (2016) and Sussman and Olivola, (2011) provide evidence that people are tax averse, meaning that they react stronger to monetary losses, if the loss is due to a "tax" than if it is due to some other reasons, for instance, a wage cut.

However, the crucial question is how taxation affects the labor supply in an environment with salient substitution and income effects (i.e., in a setting with an explicit subsistence income restriction)?[5] To address this issue, we run a series of real-effort laboratory experiments. Each experiment consists of three production phases and is played by a pair of players. In each phase, participants solve math tasks and are paid for each correctly solved task. The number of tasks is not limited. Participants are therefore free to decide how fast (i.e., how hard) to work. In the first phase, that we use as a measure for personal ability, all participants are offered the same tasks. In the second phase, some randomly chosen players are exogenously disadvantaged which results in their lower productivity. In the third phase, exogenous productivity differences are lifted again and both players solve the same tasks for twice the previous piece rate.

A Base treatment condition (BA) exactly follows this outline. In the treatment Income Threshold (IT) to study the effect of the subsistence threshold, we exogenously impose an income threshold in phase two. Only players who meet this threshold are allowed to participate in the high income third phase. Due to the

[4] Agranov and Palfrey, (2015) follow with their setting the seminal taxation model introduced by Meltzer and Richard, (1981). Meltzer and Richard assume a non-linear relationship between taxation and productivity in democratic societies and hence base their work on the assumption that there is an internally optimal level of taxation in a democracy. This optimum balances the interaction between productivity and the amount of redistribution.

[5] Fulfilling basic material needs as well as ascertaining a certain minimum degree of inclusion in society's "normal" activities in societal and political life is known in philosophy as the sufficientarian position (Frankfurt, 2000). This position claims that that societies should secure a decent minimum standard of living for all its members, above which there is room to improve individually and freely. An attempt to identify such a minimum standard as a set of general needs has been developed by Sen with the "capabilities approach" (Sen, 1973, 2000). In the same vein, Nussbaum, (2000, 2011) has introduced the concept of thresholds for a set of human "functionings" that are required for a dignified life: "(A)ll should get above a certain threshold level of combined capability, in the sense not of coerced functioning but of substantial freedom to choose and act" (Nussbaum, 2011, p. 24).

imposed productivity differences, the subsistence income threshold is easy to reach for "high productivity" players, but is extremely challenging for the disadvantaged "low productivity" players. In further treatments we introduce taxation varying the way tax revenues are spent between treatments. In a treatment condition Transfer (TR), tax revenues are transferred to the low productivity player in a pair at the end of the second phase. This transfer helps the disadvantaged player to reach the income threshold allowing her to qualify for the high income third phase.

In TR we introduce redistribution that is "meaningful" in the sense that transfers help recipients to meet the minimum income restriction, while labor income for a taxpayer is "meaningful" as well since it allows the taxpayer herself to meet the subsistence threshold. As a consequence, redistribution is legitimate and not a mere act of monetary transfers. Several experimental studies have shown that the perceived legitimacy of redistribution and taxation can heavily influence its implicit cost. For example, the acceptance of redistribution decreases when those who benefit – the welfare recipients – can themselves be held responsible for their dire straits (e.g., Cappelen, Hole, et al., 2007; Cappelen, Sørensen, et al., 2010; Fong, 2001). Therefore, if the tax is collected for recipients who are perceived as fully responsible for their neediness, or if the collected taxes are "wasted" in a way that nobody benefits altogether, this is likely to trigger higher implicit costs of taxation (i.e., a decrease in production in the society). Those results are in line with empirical findings that pro-sociality is a major factor that mitigates the implicit costs of taxation (e.g., Buch and Engel, 2014).[6] In other words, the more taxation helps those who need help, the less inefficient it becomes.

To challenge those results in a context with minimum income requirements, we test taxation and redistribution in a set of treatments varying the purpose for which the tax revenues are spent. As mentioned above, the transfer in TR supports a player who is in a disadvantaged position and is not responsible for her lower productivity. In the treatment Outsider (OS), and Waste (WA) we gradually decrease the meaningfulness of taxation in the sense that there is no disadvantaged "low productivity player" in both treatments. In OS the tax is transferred to an

[6] Aarøe and Petersen, (2014) show in addition that cross-national differences in support for redistribution are due cultural stereotypes about welfare recipients laziness or bad luck.

uninvolved third party, in WA the tax revenues are "wasted" such that nobody (including an experimenter) benefits from it. We use productivity differences across treatments to see whether redistribution to needy members of the society leads to different labor responses than wasteful redistribution or redistribution to less deserving individuals. That is, between TR and OS we vary the social distance between the taxpayers and recipients.[7] Although in both treatments the parties remain anonymous, in TR taxpayers are likely to feel closer to the recipients due to the fact that both of them take part in the whole length of the experiment.

With respect to the introduction of a non-binding subsistence income constraint we find that it lowers the productivity of the highly productive players and bifurcates the productivity of the disadvantaged players. The performance increases among the more talented transfer recipients and decreases among the less talented ones.[8]

The introduction of taxation, however, does not cause negative labor responses. On the contrary, players increase their performance to compensate for the "loss" of income. In addition, the way tax revenues are spent does not affect the labor response of taxpayers significantly. The results suggest that the cause of taxation leads to high variation in labor supply of low talent taxpayers, but essentially does not affect the performance of highly productive ones, who work equally hard in all treatments with taxation. In case of disadvantaged players, redistribution boosts productivity, remedies the division caused by the presence of the subsistence income threshold and leads to increased performance especially of low talent disadvantaged individuals. That is, players who need help the most increase their effort if they can expect a transfer of an uncertain amount that helps them reach the income threshold.

The remainder of the paper proceeds as follows. Section 2 introduces the game and our treatment variations. Section 3 sketches our theoretical analysis and develops behavioral hypotheses. In Section 4 we analyze the results of our experiment. Section 5 concludes.

[7] Ample evidence shows that other-regarding preferences are the most pronounced if social distance between players is small. Thus, increasing social distance weakens altruistic motivations (Bohnet and Frey, 1999).

[8] We use *talent* throughout the text when we talk about the different ability to do the required computations that subjects bring to the lab. We use *productivity* when we talk about the differences in productivity that we externally impose by giving subjects different tasks.

4.2 Design and Treatments

First, we describe the basic setup, that we use in both sets of treatments. Further we explain treatment variations that allow us to measure the implicit costs of redistribution with a subsistence income restriction and the effect of different causes of redistribution.

Our real-effort experiment consists of three phases of 30 minutes each, during which participants can perform simple calculation tasks. The game is played in random pairs, which remain unchanged throughout the experiment. In phase one and three both players compute the sum or the difference of two three-digit numbers. If the task is solved correctly, a new task appears on the screen; otherwise, the same task is presented again. We deliberately opt for a tedious and repetitive task that requires a lot of concentration and offer an extended time span to perform this task, so that we minimize the joy and intrinsic motivation of work, and increase the importance of monetary incentives. As the number of tasks is not limited, participants are free to decide how fast, that is, how hard, to work. Unlike other tax experiments, participants are not offered an alternative leisure activity to avoid an experimenter demand effect.[9]

The piece rate payment φ differs between the first and third phases: in the first (third) phase participants earn 0.05 Euro (0.10 Euro) per solved task. During the first phase, we elicit the motivation and initial ability of the players to perform the task. Note that instructions for phases two and three of the experiment are distributed only after the end of phase one. The second phase differs in five treatment conditions.

4.2.1 The effect of a subsistence income threshold

To study the effect of the introduction of the subsistence threshold in our experiments, we compare the Base treatment condition (BA) with the Income Threshold

[9]Research on productivity shows that not performing a task, that is tedious and requires a lot of concentration, allows to save cognitive bandwidth and therefore can be considered as leisure (e.g, Kool et al., 2010). In addition to reducing their effort, subjects were allowed to use any materials (except for calculators and mobile phones) they brought with them into the laboratory.

treatment condition (IT). In both treatments we randomly assign one player within each pair to be player R (recipient) and one to be player T (taxpayer).[10] The types of players differ only with respect to the second phase: T solves tasks with two three-digit numbers as she did in phase one, while R solves tasks with two five-digit numbers. Irrespective of the type of a player, the piece rate is 0.05 Euro per task.

While in BA there is no subsistence threshold, in IT both T and R face a subsistence income threshold: if a player can not earn the subsistence income (in our case modeled as an exogenously imposed threshold of $S = 3.20$ Euro), she can not take part in the subsequent phase three (but waits in the laboratory until the end of the experiment without additional payment). The subsistence threshold is calibrated such that it is on average just too high to be reached by R on her own, but easy to achieve for T.

4.2.2 The effect of meaningful taxation

With the other set of treatments we analyze the effect of the cause of redistribution in the environment with a subsistence income requirement. For this purpose, we consider three treatments: Transfer (TR), Outsider (OS) and Waste (WA). In this set of treatments we introduce mild taxation on the labor income.[11] That is, T's income in the second phase is taxed with an income tax (τ) of 30%. Between treatments we vary how meaningful redistribution is.

The Transfer treatment models the system with justified redistribution. Within the pair of players tax revenues are transferred to the disadvantaged player R and added to her income from the second phase.[12] Notice that player R receives the transfers unconditionally.[13] Player T (player R) qualifies for the third phase

[10] Players were denoted as type A and B during the experiment to avoid any associations. We introduced types prior to the start of the first phase to avoid players assuming a relation between types and first phase performance.

[11] Deliberately, we introduce a moderate taxation rate that equals roughly the average tax rate across the OECD countries.

[12] We model only net-positions. That is, we do not tax player R and reimburse her afterwards.

[13] Such a design reflects the majority of real world redistribution mechanisms: Gentilini et al., (2014) find that 119 developing countries have implemented at least one type of unconditional cash assistance program, while 52 countries have conditional cash transfer programs for poor households.

if the income minus the tax (plus the transfer) satisfies the subsistence income requirement.

In OS and WA there is no disadvantaged player who benefits from the transfer. That is, player T performs the task and is nevertheless taxed with a 30% income tax. In OS tax revenue is transferred to an "outsider" (labeled as player O). Players O do not take part in any of the working phases of the experiment. They are invited to join the session at the end of the second phase, randomly matched one-to-one to players T, receive tax revenue as a transfer and leave the laboratory. Even if their received transfer is above the threshold they cannot take part in the third phase.

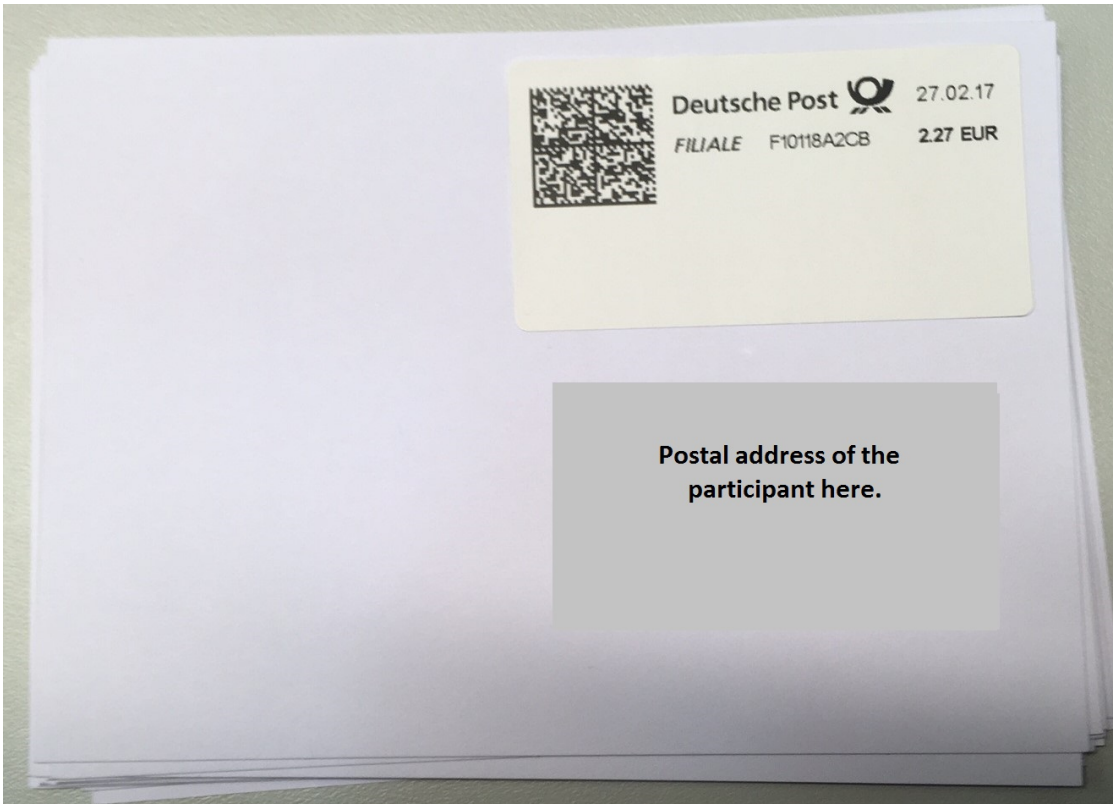


FIGURE 4.1: Example of the postcard from the Stamp treatment

In WA we decrease the meaningfulness of redistribution even further. Before the start of the experiment player T writes her postal address on a blank white postcard (see Fig.4.1). Tax revenue is used to buy postal stickers to send the blank postcard to player T immediately after the experiment.[14] The value of each sticker

[14]A postal sticker is a stamp with an individualized price encrypted in a QR code which is used to pay for postal services (see Fig.4.1).

equals exactly the tax revenue of the respective player.[15] Thus, tax revenue is wasted (neither taxpayers themselves nor experimenters profit from the way taxes are spent), but taxation cannot be interpreted as a mere wage cut.

Table 5.1 summarizes the payoffs of the second phase in all treatment conditions, Y_i denotes i's correctly solved tasks in the second phase.

TABLE 4.1: Payoffs in the second phase by treatment and player type

Treatments	**Player T**	**Player R**	**Player O**	**Threshold**	**Obs.**[16]
BA	φY_T	φY_R	–	–	30
IT	φY_T	φY_R	–	S	32
TR	$(1-\tau)\varphi Y_T$	$\varphi Y_R + \tau(\varphi Y_T)$	–	S	30
OS	$(1-\tau)\varphi Y_T$	–	$\tau(\varphi Y_T)$	S	30
WA	$(1-\tau)\varphi Y_T$	–	–	S	30

Our design incorporates three important features to focus on our research question. First of all, we introduce a one-to-one correspondence between taxpayers and transfer recipients. Player T is individually responsible for determining the size of the transfer to R and therefore there is no coordination problem among several players T, i.e., no so-called bystander effect.

Secondly, due to the predetermined direction of transfers, subjects know in advance whether they are taxpayers or transfer recipients. The choice of a systematically disadvantaged player within the pair is random and therefore transfer recipients both deserve support without any doubt and can not be held accountable for their lower productivity.

Lastly, we create the least favorable conditions for redistribution by reinforcing the entitlement towards one's earnings. As previous studies on donation show (e.g., Cherry et al., 2002; Ogawa et al., 2012), if donors earn incomes (i.e., they work for their incomes), they are less likely to redistribute. In our experiment, we use a real-effort task making taxpayers feel entitled to their earnings and thus less willing to transfer.

[15] All actual tax revenues in the experiment exceeded the price of sending a postcard. Thus, it is correct to believe that all the postcards reached players T within several days after the experiment.

[16] Number of independent observations, that is, the data for a pair of players in BA, IT, and TR, as well as for individual players in WA and OS.

4.2.3 Procedure

The experiment was conducted at the WISO Research Lab at the University of Hamburg using z-tree (Fischbacher, 2007). Subjects were mostly students of various majors of the University of Hamburg, recruited online via hroot (Bock et al., 2014) and randomly assigned to treatments. No subject participated in the experiment more than once.

Before the start of the experiment, participants placed their cellphones into the provided envelopes and sealed them. They had access to the envelopes only after the payment. In addition, participants were asked to keep their bags and personal belongings outside of their cubicles. Thus, participants had no phones or calculators to help them solving the tasks. Pen and paper were provided to be used for calculations. After that, participants were seated in cubicles and received a copy of the instructions for the first phase. Additionally, the instructions were read aloud. Subjects were randomly matched in groups of two in BA, IT and TR, which remained unchanged throughout the experiment and played simultaneously, but independently from other pairs. In WA and OS the game was played individually. Before any action took place a type of the player within the pair was displayed on her screen. Although types were revealed, the actual difference between types was made clear only after the end of phase one. After the end of phase one, instructions for the rest of the experiment (i.e., the second and third phases) were distributed and read aloud. After each phase players had a short break of three minutes.

We conducted 15 sessions with 244 participants in total. Sessions ran with 16-20 subjects per session for BA, IT, and TR and with 15 subjects per session for WA, OS yielding ca. 30 independent observations per treatment (see Table 5.1). At the end of the experiment, subjects were paid privately and in cash. Payments ranged from 3.40 to 57.2 Euro with an average of 26.0 Euro for approximately 140 minutes.

4.3 Hypothesis

4.3.1 Effort provision with and without a subsistence threshold

In the following, we assume that players hold a twice differentiable, additive utility function with constant positive benefits and convex increasing working costs. Both, benefits and costs depend on effort. Effort subsumes in our experiment a combination of endeavor and skills. In the following, we assume players to differ with respect to their costs (i.e., their skills). That is, for two players i and j we label player i as more talented than player j if for every given work speed $\bar{v}$ it holds $c_j(\bar{v}) > c_i(\bar{v})$. In turn, players gain constant marginal benefits from working. That is, their benefits increase constantly by φ (the piece rate payment) in the work speed (per minute) v_i such that their income per phase equals $30v_i\varphi$. In sum, this yields

$$U(v_i) = \varphi v_i - c_i(v_i)$$

such that the individual costs c_i follow $\dot{c}_i(v_i) > 0$ and $\ddot{c}_i(v_i) > 0$.

Without the subsistence income threshold in all phases of the experiment, players maximize their utility by choosing v_i such that $\varphi = \dot{c}_i(v_i)$ (first order condition). Fig. 4.2 shows examples for the optimal work speed in BA: the marginal costs for players with lower math skills grow faster ($c_i'(v_i)$) than those for players with higher math skills ($c_i''(v_i)$). Consequently, former players choose a lower work speed (v_i') than the latter (v_i'').

The introduction of the threshold in the second phase has two effects: first, it creates a reference point at the income equaling the subsistence income S. Let us denote with v_i' the work speed necessary to meet the subsistence income requirement (i.e., $30v_i'\varphi = S$), so that i qualifies for the third phase. In line with the majority of models on reference dependent utility (Kőszegi and Rabin, 2006), we assume that marginal utilities differ to the left and right of the reference point S. Particularly, we consider the case when the marginal utility to the left of S (i.e., every additional Euro that closes the gap to S) is φ, while the marginal utility

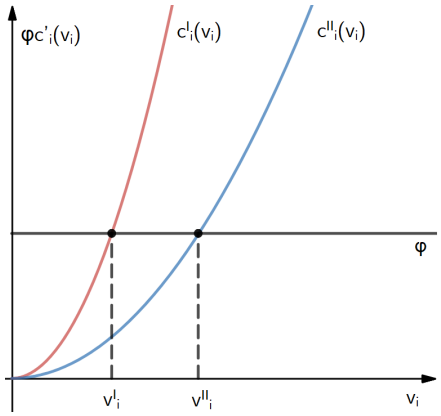


FIGURE 4.2: Optimal work speed in BA

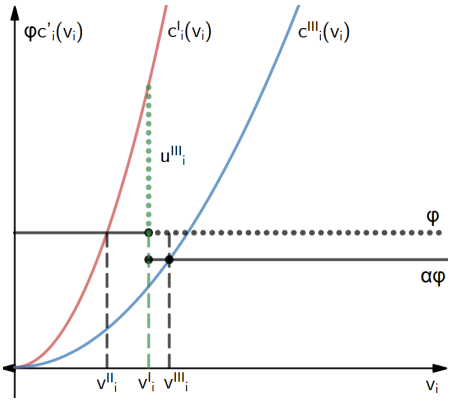


FIGURE 4.3: Optimal work speed in IT

to the right of S (i.e., every additional Euro beyond S) is considerably lower: $\alpha\varphi$ with $1 < \alpha < 0$. Thus, we assume that the marginal utility of work speed is φ to the left of v_i' and $\alpha\varphi$ to the right of v_i'.

Second, i has to consider that v_i may or may not be high enough to qualify i for the third phase. In other words, choosing the work speed of v_i' qualifies i for the third phase of the experiment implying that i gains additionally u_i^{III} denoting the expected utility gained in the third phase. Thus, i's utility function changes to

$$U(v_i) = \begin{cases} \varphi v_i - c_i(v_i) & \text{if } v_i < v_i', \\ \varphi v_i + u_i^{III} - c_i(v_i) & \text{if } v_i = v_i', \\ \alpha\varphi v_i + u_i^{III} - c_i(v_i) & \text{otherwise.} \end{cases}$$

Consequently, the marginal utility of v_i' is not simply φ but $\varphi + u_i^{III}$ and the marginal utility at work speed v_i' exceeds the marginal utility of $v_i' - \varepsilon$ by u_i^{III}. Notice that this discontinuity of the marginal utility of income creates the incentives for players in a medium cost range to "speed up" and to choose a work speed which allows them to just reach the subsistence income. Fig. 4.3 indicates the borders for such an effect: players who would choose an optimal work speed between v_i'' and v_i' in BA (i.e., players with a marginal cost function below $c_i'(v_i)$) speed up and choose v_i' in IT since their marginal benefit exceeds their marginal costs at work speed v_i'.

In turn, there are a number of players who slow down as a consequence of passing the reference point. That is, players who determine their optimal work speed $v_i''' > v_i'$ according to first order condition $\alpha\varphi = \dot{c}_i(v_i)$ in IT would have chosen a higher work speed in BA (Fig. 4.3 shows players with marginal cost function $c_i'''(v_i)$ as an example for this effect).[17]

Recall that we calibrated the subsistence income requirement in such a way that S is easy to reach for advantaged players, while it is challenging for disadvantaged players in IT. In other words, for the majority of advantaged players the optimal work speed in BA would be to the right of v_i', whereas for the majority of disadvantaged players the optimal work speed in BA would be to the left of v_i'. Since the introduction of a subsistence income requirement enhances productivity for players "to the left of v_i'", while it decreases productivity for players "to the right of v_i'", we predict

Hypothesis 1: Overall productivity of advantaged players (disadvantaged players) in phase two is lower (higher) in IT *than in* BA.

4.3.2 Effort provision with a subsistence threshold and taxation

When analyzing the choice of a work speed in the treatments with redistribution, one has to consider the two types of players separately. For players of type T, the piece rate per solved task φ decreases to $(1-\tau)\varphi := \varphi'$. For players of type R, the subsistence income requirement S decreases to $S - T := S'$ with T being the expected transfer they receive from their matched taxpayer. Replacing S with $S' < S$ makes it easier for transfer recipients to meet the subsistence income requirement implying that the groups of players affected by the reference point and the qualification for the third phase "move to the left".

Fig.4.4 illustrates the case: with transfers, the work speed of v_i''' is expected to yield the subsistence income, so that players who would have chosen a work speed

[17]Notice there are also players whose marginal costs exceed the value of $\alpha\varphi$, but are below φ at a work speed of v_i'. They simply choose v_i' as their optimal work speed in IT although their optimal work would have been faster than v_i' in BA.

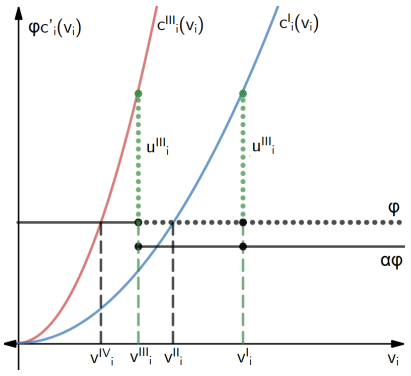


FIGURE 4.4: Optimal work speed for transfer recipients

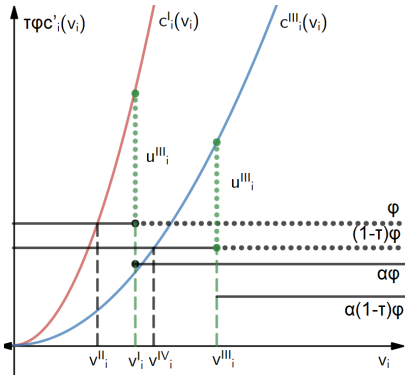


FIGURE 4.5: Optimal work speed for taxpayers

between v_i^{IV} and v_i''' in BA speed up and choose v_i'''. In turn, players who would have speed up without transfers (between v_i'' and v_i' in BA) slow down since they expect to meet the subsistence income requirement easily as a result of the transfers. Since we do not know a priori whether the mass of transfer recipients in TR are affected by the former (speeding up) or the latter effect (slowing down), we cannot hypothesize the overall difference in terms of productivity between IT and TR.

Replacing φ with $\varphi' < \varphi$, it thus becomes "harder" for a taxpayer to meet the subsistence income requirement, as they earn only $(1-\tau)\varphi$ for each task. As a consequence the groups of players affected by the reference point and the qualification for the third phase "move to the right".

Fig. 4.5 illustrates the case: with taxation, the work speed of v_i''' is necessary to yield the subsistence income, so that players who would have chosen a work speed between v_i^{IV} and v_i''' in BA speed up and choose v_i'''. In turn, players who would have speed up without taxation (between v_i'' and v_i' in BA) slow down and choose a work speed according to the first order condition $(1-\tau)\varphi = \dot{c}_i(v_i)$. Finally, players "to the right" of v_i''' lower their work speed considering the first order condition $\alpha(1-\tau)\varphi = \dot{c}_i(v_i)$.

Since we calibrated the income threshold and the tax rate such that mass of taxpayers in TR are located to the right of v_i''', we expect that overall productivity of taxpayers is negatively affected by redistribution. In other words, according to our analysis redistribution with a subsistence income requirement is regarded as a wage cut for taxpayers, similar to theoretical arguments stated in Kessler and Norton, (2016). Therefore, we hypothesize:

Hypothesis 2: Redistribution imposes implicit costs since taxpayers produce less in TR *than in* IT.

Finally, following previous studies showing that the meaningfulness of taxation is crucial for its acceptance and the size of its implicit costs (Schuitema and Steg, 2008), the implicit cost of redistribution differs between TR, OS, and WA.

In TR recipients are disadvantaged and assigned some "neediness" which does not result from their personal responsibility.[18] Therefore, we set up a situation in which redistribution is clearly justified and meaningful: without transfers, it is almost impossible for players R to meet the subsistence income requirement,[19] whereas the meaningfulness of taxation decreases in the treatment conditions OS and WA. While transfers in OS are at least received by another participant, but whose neediness is highly questionable, WA wastes the tax revenue altogether. It seems that taxation is less justified in these treatments implying that inefficiencies increase:[20]

[18] According to Brock and Reader, (2002, p. 433), "An agent has an obligation to help a person in need, if the following conditions all hold, ceteris paribus. For the needy person, severe harm is likely and imminent. He is unable to help himself. He is in his position through causes beyond his control, and it is not the case that he has an informed, voluntary, and enduring desire not to be helped. The agent knows about the needy person's position, knows what is required to help avert the harm, is in a position to help such that the cost of helping is not significant, and her assistance has some good likelihood of being effective. Under such conditions, we can reasonably claim that the agent would be morally required to help the needy person."

[19] We motivate the existence of the threshold and potential income discontinuity by the sufficientarian tradition (as argued in, e.g., "poverty threshold" Benbaji, (2006)): every individual has needs which have to be satisfied at a commonly accepted minimum level to be able to take part in the economic, social, and political life of a given society. For the rather artificial experimental threshold to be perceived as such minimum income level, the participation in the last phase of our experiment is conditioned upon reaching the threshold in the previous phase. Reaching the third phase is made more important by the experimental design through doubling up of payoffs. That is, if a person does not earn enough to ensure her subsistence, she can be considered to be too exhausted to take part in future rounds of production. The threshold is calibrated as required by Brock and Reader, (2002) so that "the cost of helping is not significant" (relative to both earnings in the whole and in the phase of the experiment) and "her assistance has some good likelihood of being effective."

[20] One may argue that WA is the least meaningful of all three treatment conditions, since tax revenues are wasted. Hence one may expect players T to work less in WA than in OS. However, if one assumes taxpayers to be inequality averse, one can think of taxpayer preferring situations in which some of her money is destroyed over situations in which she has to share her money with some undeserving outsider (e.g., in the modified dictator games analyzed by Lazear et al., 2012).

Hypothesis 3: Decreasing the meaningfulness of taxation causes further inefficiencies: taxpayers produce the least in WA*, the most in* TR *and between the two extremes in* OS*.*

4.4 Results

Our data set consists of 29 pairs (i.e., independent observations of 29 players T and 29 players R) for BA, 32 pairs (i.e., independent observations of 32 players T and 32 players R) for IT, 30 (29 players T and 30 players R) for TR, 30 players T for OS and 30 for WA.[21]

As expected, performance of the subjects in all treatments in the first phase is very similar with ca. 125 tasks solved on average in 30 minutes (Kruskal-Wallis equality-of-populations rank test for performance in the first phase p=0.5). Thus the exogenously imposed threshold of 3.20 Euro (equivalent to 64 tasks in BA and IT and 92 tasks in other treatments due to taxation) would imply that players could have solved (on average) 66 and 38 tasks less than in the first phase and still qualify for the third phase.

To measure the performance of the participants controlling for their initial ability we divide the number of correctly solved tasks in the second phase by the number of correctly solved tasks in the first phase. Further, we call this relation between the stages *relative performance*. The relative performance of 1 means that a player solves in the second phase exactly the same number of tasks as in the first one; a relative performance less than 1 means that she reduces her effort and solves less, and the number more than 1 means respectively that she improves her performance. Thus, one can also interpret relative performance as a percentage of change as compared to the first phase.

Before we proceed to the effects of legitimacy of taxation on the labor supply decisions, we analyze how the introduction of the performance threshold affects behavior and compare BA and IT treatments.

[21] We removed three participants (two participants in BA, one in TR) from the sample and did not use their data in the analysis since their ability to perform the task deviated strongly (more than 2.5 standard deviations) from the sample mean. Analysis on the full sample renders qualitatively similar results.

4.4.1 The effect of the introduction of the subsistence income threshold

4.4.1.1 On advantaged players

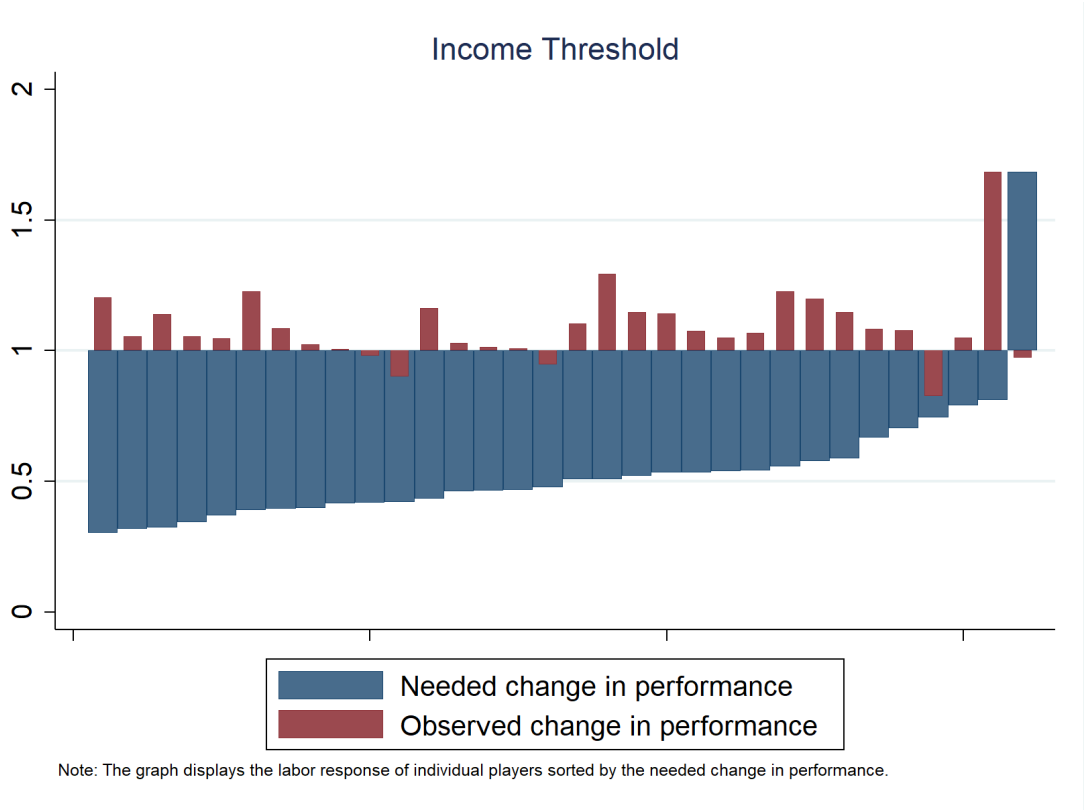


FIGURE 4.6: IT treatment: Individual response to the introduction of the subsistence threshold

Fig.4.6 displays how restrictive the imposed subsistence income threshold is for every individual in the IT treatment. Blue bars show how players T need to adjust their performance in the second phase in order to meet the subsistence income requirement and red bars show the actually observed relative performance. The vast majority of players could lower their performance considerably and still meet the imposed requirement. Therefore, the change in performance in IT treatment as compared to BA should not be driven by the necessity to reach the threshold.

Figure 4.7 depicts the aggregate labor response to the introduction of the income threshold. In line with our Hypothesis 1, we observe that the relative performance of advantaged players is negatively affected. In BA advantaged players increase

their performance by 14% whereas in IT by 9% ($p = 0.07$)[22]. This behavior is consistent with the performance under easily attainable goal as well as with performance under low reference point.

Result 1: Introduction of a subsistence income threshold negatively affects the performance of advantaged players.

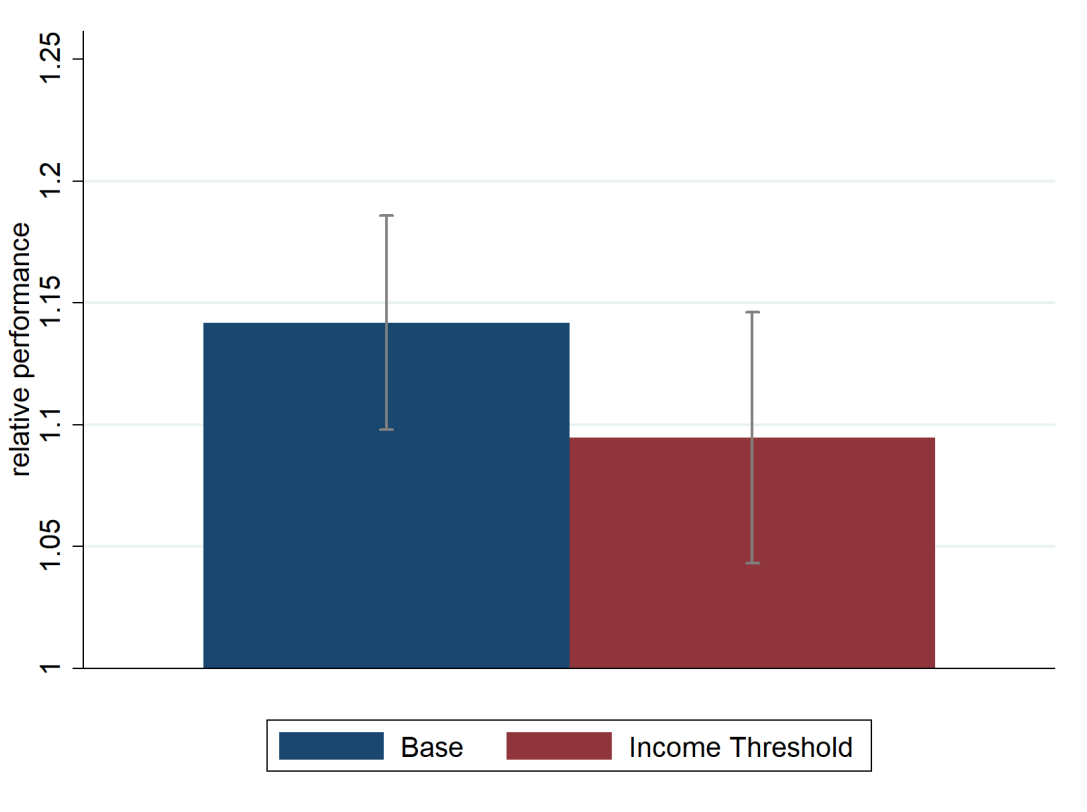


FIGURE 4.7: Relative performance of advantaged players. Effect of the introduction of the income threshold

4.4.1.2 On disadvantaged players

Unlike the advantaged, disadvantaged players face a more difficult task in phase two. Hence, their performance decreases by approximately half. Interestingly, the introduction of the subsistence threshold does not affect their relative performance on the aggregate level (BA (0.49) and IT (0.47), $p = 0.65$).

[22] All reported statistical tests are non-parametric, two-tailed and take individuals as units of observations.

Yet, as theory predicts, behavioral responses differ substantially depending on the talent of players. That is, when we compare the performance of disadvantaged players in BA who would not pass the threshold had it been there with those disadvantaged players in IT who do not make it, we find that the former group performs significantly better than the latter group (0.5 versus 0.41, p=0.05).[23] On the contrary, those disadvantaged players who are sufficiently skilled and would pass the threshold without assistance in BA work less diligent (0.45) compared to those in IT (0.59, p=0.03).[24]

Our experimental results are partly in line with the theoretical predictions: sufficiently skilled disadvantaged players actually speed up in IT in comparison to BA, while less skilled disadvantaged players work significantly less in IT in comparison to BA. The latter finding, that is unexpected taking into account the theoretical predictions, could be interpreted as an act of desperation of those disadvantaged players who are much slower in solving the tasks than it is necessary to meet the subsistence income requirement.

Result 2: Among disadvantaged players, the introduction of a subsistence income threshold negatively affects less skilled players and promotes higher performance among more able players.

4.4.2 The effect of meaningfulness for taxation

In this section, we address the question how the introduction of taxation in the environment with the subsistence income constraint affects performance, and whether the variation of the cause of taxation affects labor supply decisions. For this, we compare the relative performance of IT with the redistribution treatments. In the second part of the analysis, we compare TR, OS and WA treatments between each other. In the following, we first consider the effects of redistribution on taxpayers, then on transfer recipients and afterwards analyze the time dynamics of exerted effort.

[23] 20 disadvantaged players in BA and 21 players in IT fall into this category.

[24] The test is based on 9 observations for BA and 11 for IT and therefore should be treated with caution.

4.4.2.1 On taxpayers

Fig. 4.8 shows that even with the deduction of 30% in TR, OS and WA treatments, the vast majority of taxpayers increase their relative performance in the second phase. Recall that the threshold was easily reachable for taxpayers who could have lowered their effort in the phase two (i.e., consider the blue bars in Fig. 4.8). That is, our results contradict our Hypothesis 2 (predicting a relative performance of less than 1) in all three treatment conditions TR, OS and WA.

Figure 4.9 displays the response to taxation on the aggregate level. The data show that players enhance their relative performance on average by ca. 10% in IT, in three redistribution treatments their performance increases by additional 5%. As our experiment lasts over 140 minutes in total with 90 minutes of active calculations, the relative performance reflects both learning and fatigue. However, there is no reason to believe that those effects are treatment specific and, thus, we see a generally positive effect of taxation.

Mann-Whitney U-tests confirm that relative performance in TR (1.18) and OS (1.16) is higher and significantly different from IT (1.09; $p = 0.005$ for TR and $p = 0.03$ for OS). The difference between IT and WA (1.17; $p = 0.14$) is not significant (see Fig.4.9). Therefore, our results are in contrast to Hypothesis 2: the effect is positive, largest in TR, less pronounced in OS and in WA.

Result 3: Players T produce significantly more in TR and OS than in IT.

Hence, in our experiment, we do not observe the predicted negative effect of meeting the subsistence income threshold on the productivity of taxpayers.[25] It could be the case that taxation limits the salience of the subsistence income threshold and that players are more focused on another income target. In our case the income obtained in the first phase could be an alternative reference point. In other words, taxpayers could speed up in the second phase in general since they want to earn in the second phase as much as they did in the first one.

Analyzing the differences among redistribution treatments, one can see that the relative performance in all three treatments (TR, OS and WA) is not significantly

[25] We will elaborate on this issue in more detail below.

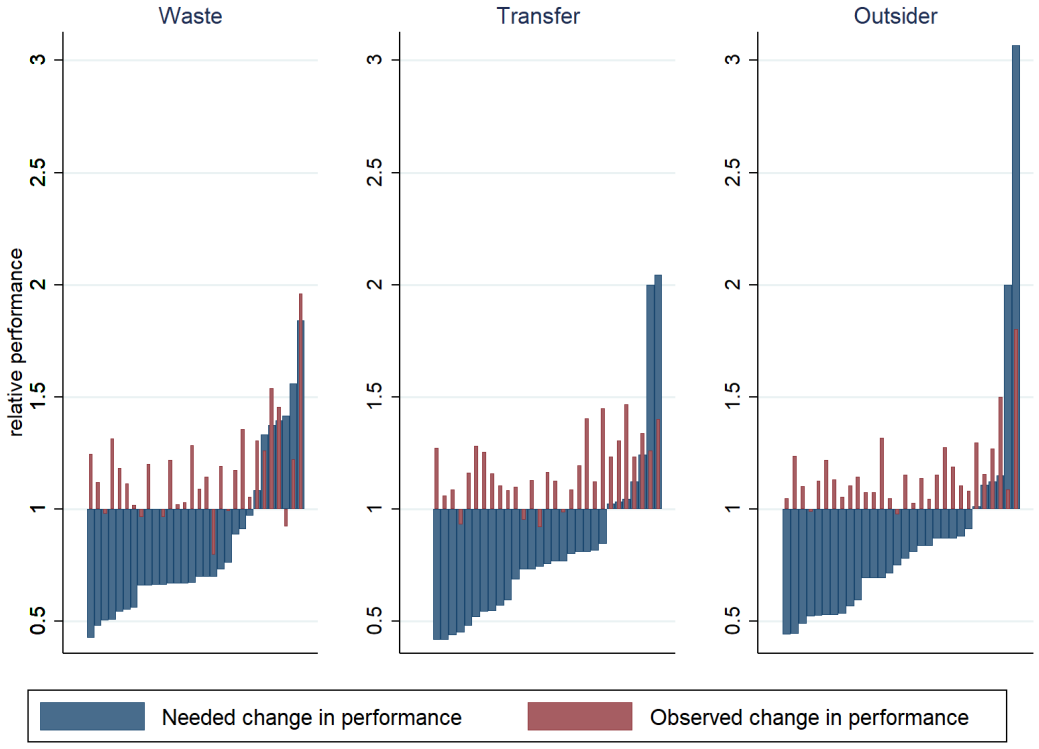


FIGURE 4.8: Relative performance required to reach the subsistence threshold and actual relative performance for each taxpayer

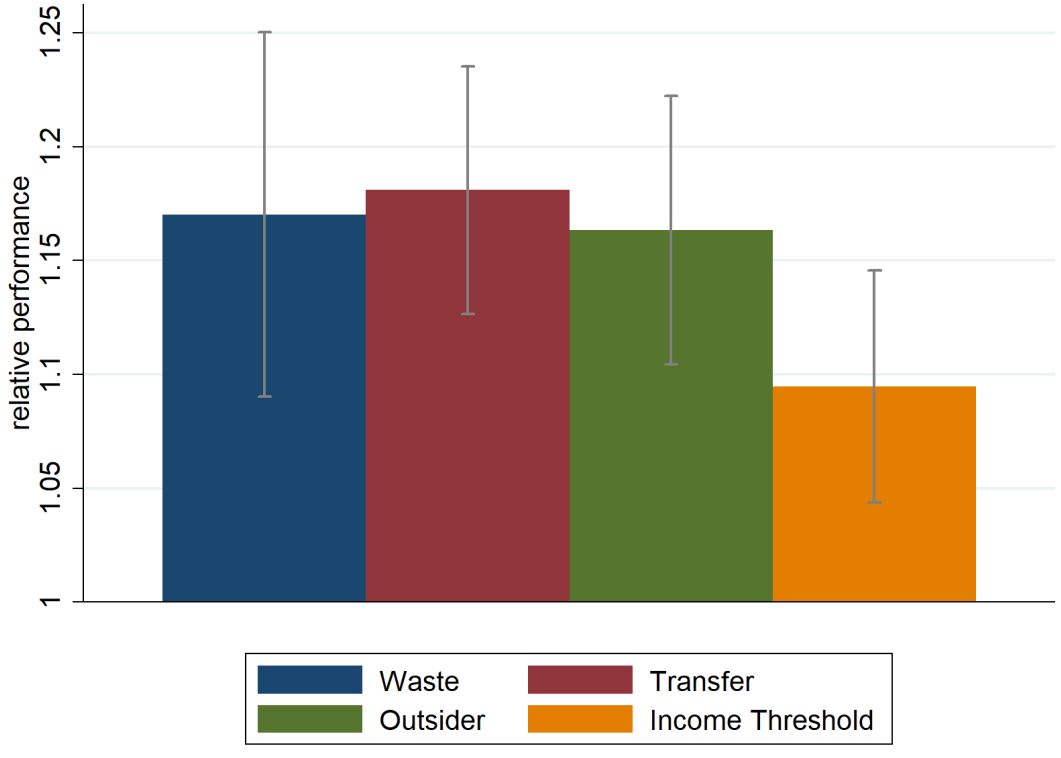


FIGURE 4.9: Relative performance of players T across treatments. 95% confidence interval

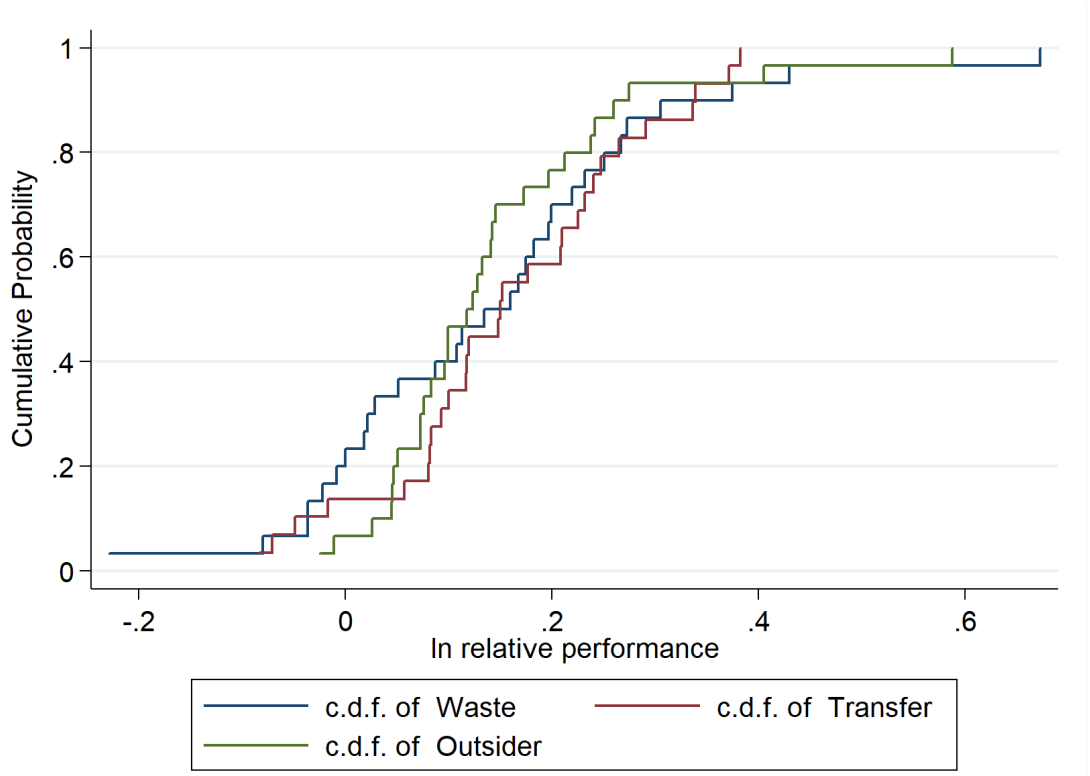


FIGURE 4.10: Players T: Cumulative distribution function of the log of relative performances

different (TR and OS $p = 0.3$, TR and WA $p = 0.55$, OS and WA $p = 0.98$). This finding suggests that in our setting with piece rate and subsistence income requirement players on average do not react to the cause of redistribution.

Notice, however, that there is substantial heterogeneity in response to different redistribution treatments, particularly for WA: Fig.4.10 shows the cumulative distribution function of the log of relative performances for the three treatments. The lower performing half of taxpayers in WA have a substantially lower relative performance than the lower performing half of taxpayers in TR, whereas the relative performance for the higher performing half is rather indistinguishable.

For a more detailed analysis of whether players react heterogeneously to the tax, we classify taxpayers as talented or not talented according to their performance in the first phase: those who are above their treatment's median production are

considered talented and non-talented otherwise.[26] Fig.4.11 shows the relative performance of high and low talent taxpayers along with the 95% confidence interval.

High talent taxpayers are rather insensitive towards the 30% taxation: relative performance in all four treatments is not significantly different from each other. That is, introduction of taxation regardless of its cause does not lead to withdrawal of labor on the side of high performers. Low talent taxpayers are tentatively more sensitive towards treatment manipulations. That is, low talents exert higher effort in TR (1.25) as compared to OS (1.22) and as compared to WA (1.23). Yet, the differences within the redistribution treatments are not significant.[27]

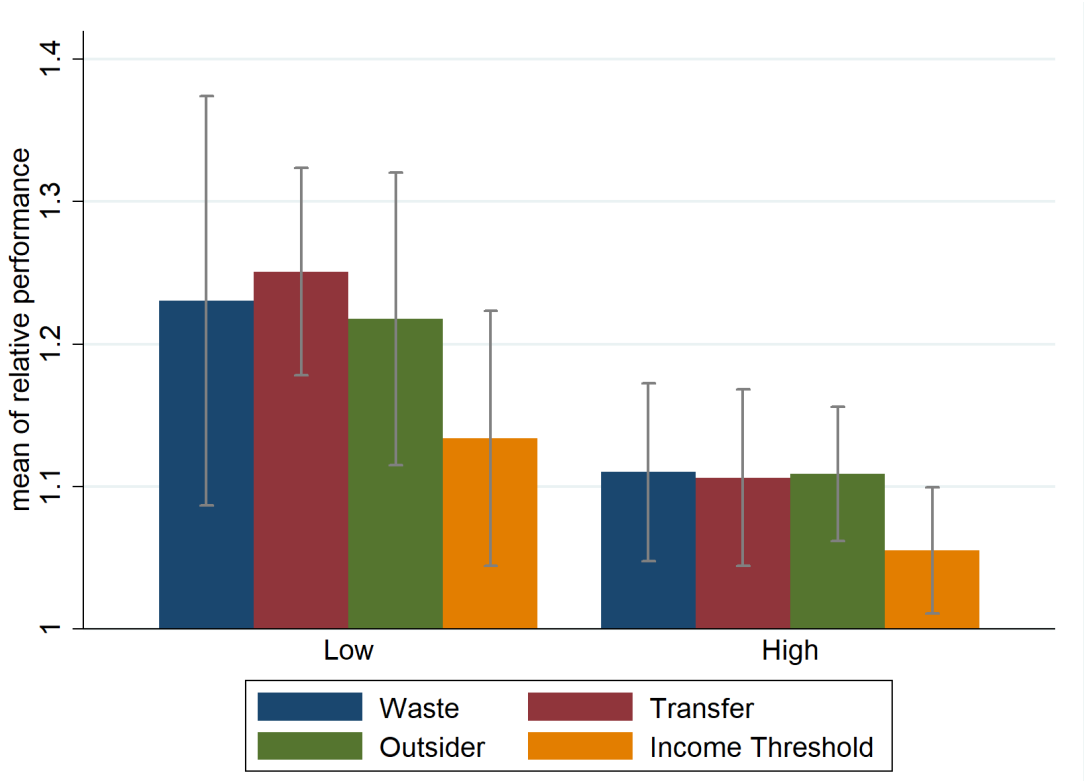


FIGURE 4.11: Relative performance of high and low talent players T. 95% confidence interval

[26]The medians in the first phase are 130 tasks in IT, 122 tasks in TR, 120.5 tasks in OS and 134.5 of correctly solved tasks in WA.

[27]Earlier on, we noticed that there are major gender differences with regards to the acceptance of taxation. Our study does not aim to contribute to this debate. Yet, our experimental data seems to be in line with the previous literature. Women are more sensitive towards the cause of redistribution and increase their performance significantly in TR as compared to their male counterparts. Between treatment comparisons display expected directions but can not be tested reliably.

Another observation is that the variance of the relative performance for low talents differs significantly between the three treatment conditions: the variance is 0.02 in TR, 0.04 in OS, and 0.08 in WA. The difference in variance is not significant for the comparison of OS with other treatments $p = 0.21$ (with TR) and $p = 0.22$ (with WA), however the difference in variance is significant between TR and WA $p = 0.02$. Thus, low talent taxpayers do not respond uniformly to the waste of their tax revenue.

Summarizing our results for taxpayers, we find:

Result 4: In contrast to our Hypothesis 3, the cause of redistribution seems to have little effect on the relative performance of high talent players, but leads tentatively to a positive labor response among low talent participants.

4.4.2.2 On transfer recipients

One could claim that redistribution decreases the level of effort provision among transfer recipients. In our experiment, we fail to find support for this claim. On the contrary, we find that redistribution fosters the productivity of the transfer recipients. On average, we see that transfer recipients enhance their relative performance in the environment with the transfer (0.47 in IT and 0.55 in TR, $p = 0.07$, see Fig.4.12 and the cumulative distribution function of relative performances in Fig.4.13).

Recall that theory predicts a two-fold effect: an increase in productivity for less talented and a decrease in productivity for more talented recipients. To test this prediction, we group transfer recipients into high and low talent according to the median of the treatment performance in the first phase and analyze the differences in relative performance among these groups.

The data offer partial support for our theoretical intuition: indeed, high talent players increase their relative performance in TR (0.5) in comparison to IT (0.43), $p = 0.19$. However, low talent recipients also increase their relative performance in TR (0.6) as compared to IT (0.53), $p = 0.19$. Although the differences are not significant, Fig. 4.13 suggests that the aggregate differences in relative performance

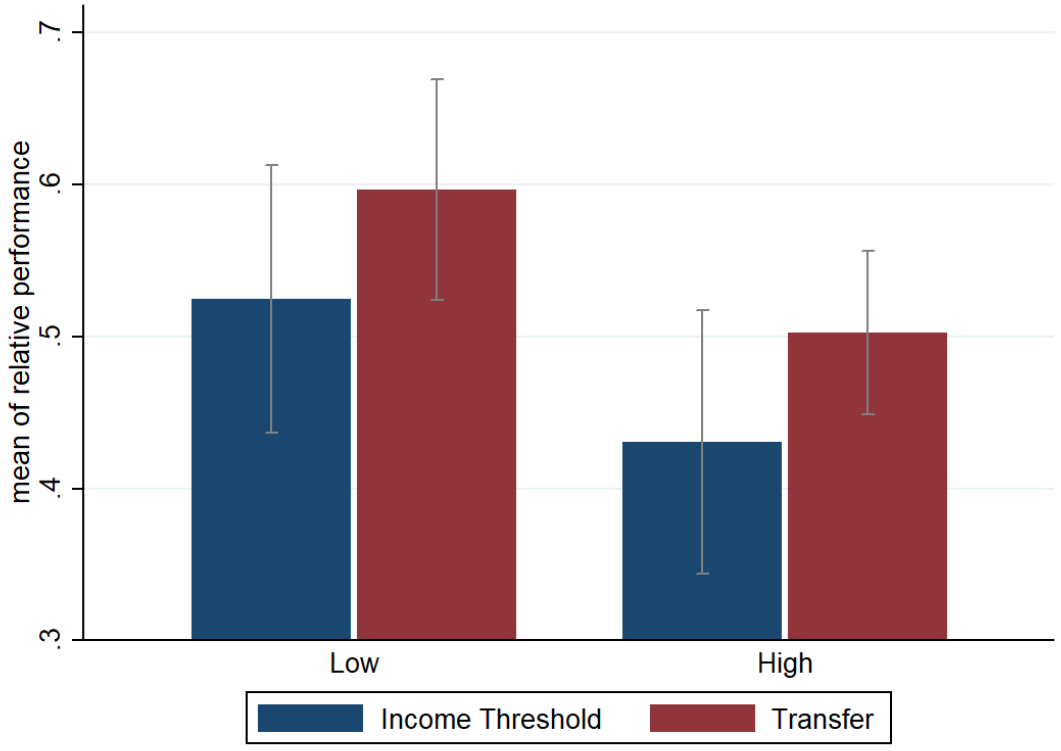


FIGURE 4.12: Relative performance of high and low talent recipients

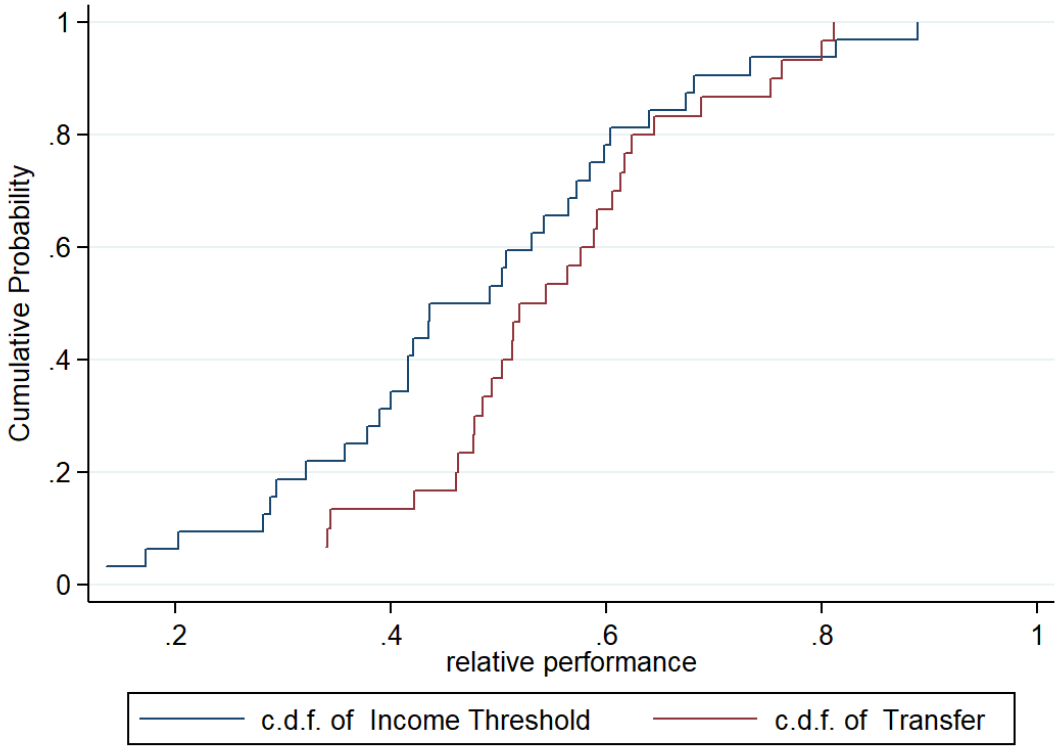


FIGURE 4.13: Players R: Cumulative distribution function of relative performance

are primarily driven by the enhanced performance of the low talent recipients. That is, with redistribution, even low talent players have a chance to take part in the third phase, and so they work harder. As such, redistribution motivates low productivity recipients, those who need support the most, to exert effort.

Result 5: Redistribution leads to higher productivity among the transfer recipients, especially boosting the performance of low talent recipients.

Importantly, redistribution fosters the actual ability of recipients: among them, the correlation between the performance in the first and second phase is higher in TR than in IT (0.71 compared to 0.40 in IT). Introduction of redistribution, even if the size of the transfer is uncertain, mitigates the consequences of the threshold introduction, which worsens the performance of low talent disadvantaged players. In TR it is no longer the case and we observe more consistent relation between true abilities and performance in the second phase.

4.4.3 Does passing the threshold affect the performance of taxpayers?

Our experimental environment with the subsistence income threshold and a tax-free production in the third phase invites taxpayers to either completely stop working or to slow down once they passed the subsistence threshold and secured own participation in the more profitable third phase.

To detect the change in the speed of production due to the threshold we look into the speed of production during the time just before and after the threshold. Fig. 4.14 displays the performance of players three minutes (i.e., 10% of the phase time) before and after the threshold.[28] We observe that in all the treatments the speed of production does not change with passing the threshold. The Wilcoxon signed-ranks test confirms that the speed of production is not significantly different three minutes before and after the threshold (for IT $p = 0.26$, for TR $p = 0.46$, for OS

[28]In all the treatments except for WA all the players produced for at least three minutes after passing the threshold, that is the average is not affected by dropping out of slower players. In WA two players produced only for two minutes after the threshold, thus the displayed average is slightly downward biased. For some sessions time data are missing due to technical issues. The graph is based on the data for 27 taxpayers in IT, 19 taxpayers in TR, 19 in OS, and 27 in WA.

$p = 0.90$, and for WA $p = 0.60$). This result suggests that the threshold is not seen as an important income target.

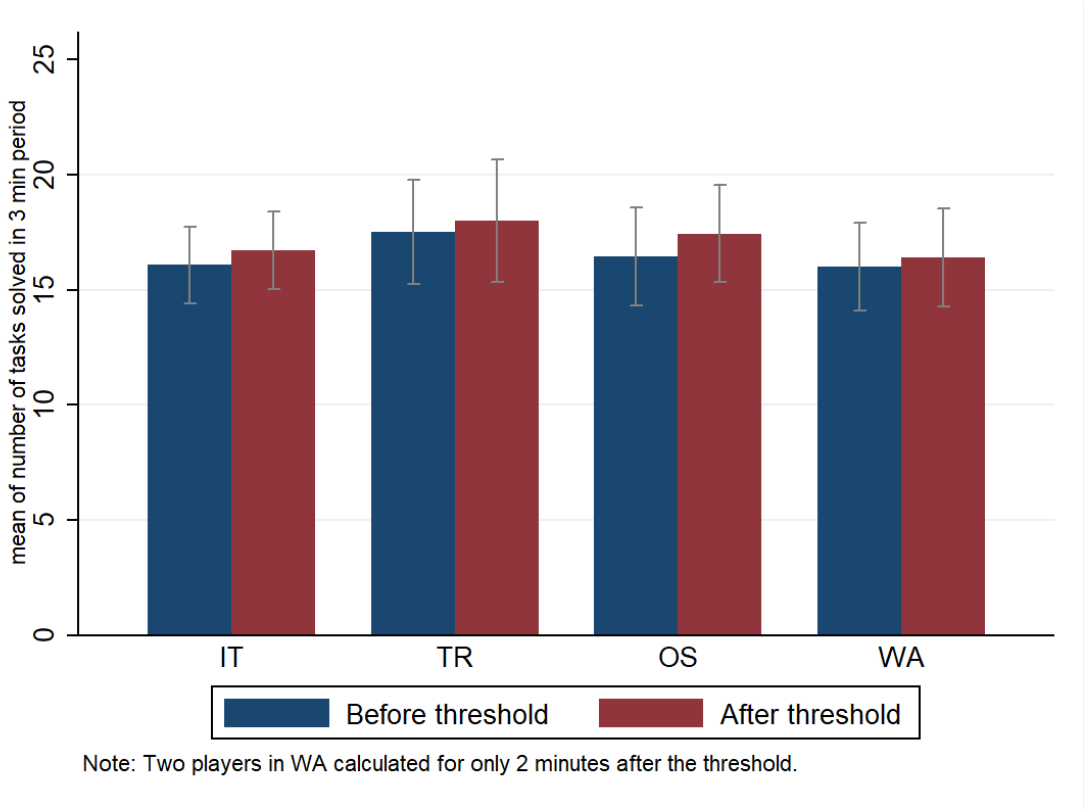


FIGURE 4.14: Mean of the number of correctly solved tasks 3 minutes before and after the threshold. 95% confidence interval

4.5 Discussion and Conclusion

We run a real-effort experiment to measure the effects of redistribution on the labor supply of both taxpayers and transfer recipients. In contrast to previous experiments, we restrict labor-leisure choices by the introduction of a subsistence income threshold which players have to meet in order to continue their participation in the experiment. Additionally, we test whether the meaningfulness of redistribution has an effect on labor choices.

We find that the imposition of the subsistence income threshold leads to a significant decrease in performance. In line with our theoretical prediction, advantaged players work on average 5% less in IT than in BA. We attribute this observation to the reference income effect. In other words, goals that are easily attainable do not

have positive effects for production. As such, our results are in line with previous experiments (Brookins et al., 2017).

In turn, the subsistence income threshold leads to a differential response along the line of initial skills among disadvantaged players. The low talent group of players lowers the performance, while the high talent group of players performs better in the environment with the threshold.

The introduction of a redistribution system in favor of a disadvantaged player mitigates this split and leads to more subjects performing to their true ability. Apart from that, receiving transfers yields the predicted positive effect: the work speed of disadvantaged players increases significantly with the effect being driven by less able players. Thus, within the redistribution system, the expected transfer motivates those who need it the most to enhance performance.

With the regard to the performance of taxpayers, their productivity increases between the first and second phase by 18.1% in TR, by 16.3% in OS, and by 17% in WA (without taxation, the productivity increases "only" by 9.5% in IT). These results complement earlier field evidence on wage variation (e.g., Camerer et al., 1997; Fehr and Goette, 2007). In all these studies, there is no unidirectional adjustment of labor supply when wages vary, highlighting the importance of target incomes for the changes in labor supply associated with varying wages. The fact that we do not observe differences in the speed of production before and after passing the threshold suggests that players focus on another target income, which is likely to be their income in phase one. That is, by working harder taxpayers in our experiment try to compensate for the loss of income due to taxation.

Finally, the comparison of productivity gains among the three redistribution treatments (TR, OS, WA) does not reveal a significant effect of varying causes of taxation on labor supply. Nonetheless, we have tentative evidence that low talent taxpayers are more sensitive towards variations in the cause of taxation: they perform best if they anticipate the transfer within the framework of a needs-based redistribution.

Overall, the results of our experiments imply that the redistribution itself and especially redistribution supporting needy peers does not necessarily lead to inefficiencies. It may lead to higher productivity among taxpayers, who exert more

effort to reach their income target and may support disadvantaged members of the society, thus leading to more people participating in production and to more production within the society as a whole.

Although the results of the laboratory experiments can not be translated directly into economic policy since there are many other factors affecting the implicit costs of redistribution (e.g., transaction costs, bureaucratic inefficiencies, leaky bucket effect, etc.), we can see that the redistribution system has the potential to be a powerful albeit undervalued tool to boost the welfare of the society with the help of meaningful and, therefore, accepted redistribution among its members.

Bibliography

Aarøe, Lene and Michael Bang Petersen (2014). "Crowding out culture: Scandinavians and Americans agree on social welfare in the face of deservingness cues". In: *The Journal of Politics* 76.3, pp. 684–697.

Agranov, Marina and Thomas R Palfrey (2015). "Equilibrium tax rates and income redistribution: A laboratory study". In: *Journal of Public Economics* 130, pp. 45–58.

Banerjee, Abhijit V, Rema Hanna, Gabriel E Kreindler, and Benjamin A Olken (2017). "Debunking the stereotype of the lazy welfare recipient: Evidence from cash transfer programs". In: *The World Bank Research Observer* 32.2, pp. 155–184.

Benbaji, Yitzhak (2006). "Sufficiency or priority?" In: *European Journal of Philosophy* 14.3, pp. 327–348.

Blundell, Richard and Thomas MaCurdy (1999). "Labor supply: A review of alternative approaches". In: *Handbook of Labor Economics*. Vol. 3. Elsevier, pp. 1559–1695.

Blundell, Richard and Ian Walker (1986). "A life-cycle consistent empirical model of family labour supply using cross-section data". In: *The Review of Economic Studies* 53.4, pp. 539–558.

Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch (2014). "hroot: Hamburg registration and organization online tool". In: *European Economic Review* 71, pp. 117–120.

Bohnet, Iris and Bruno S Frey (1999). "Social distance and other-regarding behavior in dictator games: Comment". In: *American Economic Review* 89.1, pp. 335–339.

Brock, Gillian and Soran Reader (2002). "Needs-centered ethical theory". In: *The Journal of Value Inquiry* 36.4, pp. 425–434.

Brookins, Philip, Sebastian Goerg, and Sebastian Kube (2017). "Self-chosen goals, incentives, and effort". In: *Working Paper*.

Buch, Claudia M and Christoph Engel (2014). "Effort and redistribution: Better cousins than one might have thought". In: *Working Paper*.

Camerer, Colin, Linda Babcock, George Loewenstein, and Richard Thaler (1997). "Labor supply of New York City cabdrivers: One day at a time". In: *The Quarterly Journal of Economics* 112.2, pp. 407–441.

Cappelen, Alexander W, Astri Drange Hole, Erik Ø Sørensen, and Bertil Tungodden (2007). "The pluralism of fairness ideals: An experimental approach". In: *The American Economic Review* 97.3, pp. 818–827.

Cappelen, Alexander W, Erik Ø Sørensen, and Bertil Tungodden (2010). "Responsibility for what? Fairness and individual responsibility". In: *European Economic Review* 54.3, pp. 429–441.

Cherry, Todd L, Peter Frykblom, and Jason F Shogren (2002). "Hardnose the dictator". In: *The American Economic Review* 92.4, pp. 1218–1221.

Chugunova, Marina, Wolfgang Luhan, and Andreas Nicklisch (2017). "When to Leave Carrots for Sticks: On the Evolution of Sanctioning Institutions in Open Communities". In: URL: `http://bedarfsgerechtigkeit.hsu-hh.de/dropbox/wp/2017-09.pdf`.

Farber, Henry S (2005). "Is tomorrow another day? The labor supply of New York City cabdrivers". In: *Journal of Political Economy* 113.1, pp. 46–82.

– (2008). "Reference-dependent preferences and labor supply: The case of New York City taxi drivers". In: *American Economic Review* 98.3, pp. 1069–82.

Farkas, Steve, Jean Johnson, Will Friedman, and Ali Bers (1996). *The Values We Live by: What Americans Want from Welfare Reform: a Report from Public Agenda*. Public Agenda.

Fehr, Ernst and Lorenz Goette (2007). "Do workers work more if wages are high? Evidence from a randomized field experiment". In: *American Economic Review* 97.1, pp. 298–317.

Feldstein, Martin (1999). "Tax avoidance and the deadweight loss of the income tax". In: *Review of Economics and Statistics* 81.4, pp. 674–680.

Fischbacher, Urs (2007). "z-Tree: Zurich toolbox for ready-made economic experiments". In: *Experimental Economics* 10.2, pp. 171–178.

Fong, Christina (2001). "Social preferences, self-interest, and the demand for redistribution". In: *Journal of Public Economics* 82.2, pp. 225–246.

Frankfurt, Harry (2000). "Gleichheit und Achtung". In: *Gleichheit oder Gerechtigkeit. Texte der neuen Egalitarismuskritik*. Ed. by Angelika Krebs. Frankfurt a.M.: Suhrkamp Verlag, pp. 38–49.

Gentilini, Ugo, Maddalena Honorati, Ruslan Yemtsov, et al. (2014). "The state of social safety nets 2014". In: *The World Bank, Washington, DC*.

Harberger, Arnold (1964). "Taxation, resource allocation, and welfare". In: *The role of direct and indirect taxes in the Federal Reserve System*. Princeton University Press, pp. 25–80.

Keane, Michael (2011). "Labor supply and taxes: A survey". In: *Journal of Economic Literature* 49.4, pp. 961–1075.

Keane, Michael and Richard Rogerson (2015). "Reconciling micro and macro labor supply elasticities: A structural perspective". In: *Annu. Rev. Econ.* 7.1, pp. 89–117.

Kessler, Judd B and Michael I Norton (2016). "Tax aversion in labor supply". In: *Journal of Economic Behavior & Organization* 124, pp. 15–28.

Kool, Wouter, Joseph T McGuire, Zev B Rosen, and Matthew M Botvinick (2010). "Decision making and the avoidance of cognitive demand." In: *Journal of Experimental Psychology: General* 139.4, p. 665.

Kőszegi, Botond and Matthew Rabin (2006). "A model of reference-dependent preferences". In: *The Quarterly Journal of Economics* 121.4, pp. 1133–1165.

Lazear, Edward P, Ulrike Malmendier, and Roberto A Weber (2012). "Sorting in experiments with application to social preferences". In: *American Economic Journal: Applied Economics* 4.1, pp. 136–163.

Meltzer, Allan H and Scott F Richard (1981). "A rational theory of the size of government". In: *Journal of Political Economy* 89.5, pp. 914–927.

Mirrlees, James A (1971). "An exploration in the theory of optimum income taxation". In: *The Review of Economic Studies* 38.2, pp. 175–208.

Nussbaum, Martha C. (2000). *Women and Human Development. The Capabilities Approach*. Cambridge (UK): Cambridge University Press.

– (2011). *Creating Capabilities. The Human Development Approach*. Cambridge MA: Belknap Press.

Ogawa, Kazuhito, Toru Takemoto, Hiromasa Takahashi, and Akihiro Suzuki (2012). "Income earning opportunity and work performance affect donating behavior:

Evidence from dictator game experiments". In: *The Journal of Socio-Economics* 41.6, pp. 816–826.

Piketty, Thomas (2014). *Capital in the 21st Century*. Harvard University Press Cambridge, MA.

Schuitema, Geertje and Linda Steg (2008). "The role of revenue use in the acceptability of transport pricing policies". In: *Transportation Research Part F: Traffic Psychology and Behaviour* 11.3, pp. 221–231.

Sen, Amartya K. (1973). *1973. On Economic Inequality . Clarendon Press.* Oxford: Clarendon Press.

– (2000). "Social Justice and the Distribution of Income". In: *Handbook of Income Distribution*. Ed. by A. B. Atkinson and F. Bourguignon. Amsterdam: Elsevier, pp. 59–85.

Staab, Philipp and Oliver Nachtwey (2016). "Market and Labour Control in Digital Capitalism". In: *tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society* 14.2, pp. 457–474.

Stiglitz, Joseph E (2012). *The price of inequality: How today's divided society endangers our future*. WW Norton & Company.

Sussman, Abigail B and Christopher Y Olivola (2011). "Axe the tax: Taxes are disliked more than equivalent costs". In: *Journal of Marketing Research* 48.SPL, S91–S101.

A.1 Instructions: TR treatment[29]

Instructions

General rules for participants

You are now going to participate in an economic experiment. At the end of the experiment you will receive a payment. How much you will earn depends on your activities and partly on the activities of other participants. Therefore, it is important that you read the following explanations carefully.

Please do not talk to each other after the start of the experiment. Please do not try to communicate in any other way, unless you are directly asked to during the experiment. If you have any questions, please let us know by a hand sign. An experimenter will come to you to answer your questions. Not following these rules will lead to the exclusion from the experiment and from all payments. Your decisions in the experiment and your answers in the following questionnaire are **anonymous.** Your identity is revealed only to the experimenter, but your answers can not be matched to your identity.

The experiment consists of **three** parts. **You will first receive the instructions for part one of the experiment.** The instructions for parts two and three will be distributed after completion of the first part.

For the duration of the experiment you are randomly assigned by the computer to a group of two players. Apart from you, your group has one more person. **In all three parts of the experiment you are together with the same person in a group.** In each group, there is a person A and person B. The computer assigns the roles randomly at the beginning of the experiment. The decisions of the other person in your group may have an effect on how much you earn. The decisions of people who are not in your group definitely have no influence on how much you earn.

After completing all three parts of the experiment, your payment for all three parts will be summed up and displayed on the screen in Euro. It will be paid in cash at the end of the experiment. After completing the experiment, please stay in your cubicle until we start paying off. During the payment procedure, please wait in your cabin until you are called to collect your payment individually. No other participant will see how much you have earned. Please bring along all the materials you have received from us to the payment.

Before we start the experiment, we kindly ask you to seal your mobile phone in the envelope with your cubicle number. We will collect the envelopes and will return them back together with your payment at the end of the experiment.

Part 1

In this part, you and the other person in your group can individually solve math problems. You have **30 minutes**. Your income in this part is completely independent of the other participant in your group and depends on how many tasks you solve correctly. The same applies to the other person in your group: The other person's income depends solely on the number of the tasks she will solve correctly and is independent of the number of tasks that you will solve. The tasks which you can solve are addition and subtraction calculations with two three-digit numbers. In subtraction tasks the result can never be negative. If you solve the task correctly, pressing the “Next” button will bring you to the next task. **Please do not use pocket calculators or similar tools.** You and the other person solve the same calculation tasks in the same order. **For each correctly solved task you get 0.05 Euro.**

If you have any questions, please show it with a hand sign.

[29]This is a translation of the original instructions in German which can be obtained from the authors upon request.

Part 2

In this part of the experiment you can again individually solve computing tasks. You have **30 minutes.** If you solved the task correctly, then pressing the "Next" button will bring you to the next task. **Please do not use pocket calculators or similar tools.**

The tasks which you can solve are addition and subtraction calculations. However, Person A will add or subtract two three-digit numbers and Person B will add or subtract two five-digit numbers. Reminder: At the beginning of the experiment, the computer randomly assigned one participant in the group to be Person A and another participant to be Person B. The number of solved tasks in part 1 had no effect on this. **For each correctly solved task, you receive 0.05 Euro, regardless of whether it is three or five-digit numbers.**

At the end of Part 2, Person A must give away 30% of her earned income. This means that person A's income from Part 2 is 70% of the income from correctly solved tasks. The rest is transferred to Person B and increases the likelihood that she will be able to participate in Part 3 (more details below). That means that as income from Part 2 Person B receives, in addition to her earned income, 30% of the properly solved tasks of Person A.

Part 3

All participants who got more than € 3.20 in Part 2 can participate in Part 3 of the experiment. This means that person A (without the sum transferred to person B) must earn at least € 3.20 in Part 2 in order to participate in Part 3. For person B, this means: if the income from her solved tasks and the transfer payment received from person A together amounts to at least 3.20 Euro, person B can participate in Part 3. If a participant does not participate in Part 3, she must nevertheless stay in the laboratory until the end of Part 3. In Part 3, you can again individually solve calculation tasks. You have **30 minutes**. The tasks you can solve are addition and subtraction calculations of two three-digit numbers (both for Person A and Person B). In subtraction tasks the result can never be negative. If you solve the task correctly, pressing the "Next" button will bring you to the next task. **Please do not use pocket calculators or similar tools.** If both players in the pair participate, they get the same calculation tasks in the same order. **For each correctly solved task, you get 0.10 Euro.**

If you have any questions, please show it with a hand sign.

Chapter 5

Is Time on Our Side? On the Benefits of Committing to Charities

Marina Chugunova, Andreas Nicklisch and Kai-Uwe Schnapp[1]

Abstract

In this paper, we consider how the timing of a donation decision relative to the production of disposable income affects charitable giving. In a real-effort laboratory experiment with a subsistence income constraint, we study the willingness to donate, the amounts donated as well as the change in labor supply of donors triggered by giving. We find that more people donate if the decision to donate is made before production as a binding pledge. Donors boost their performance, that is driven by their pro-sociality and not by the reference dependence of income. Our results suggest that a commitment to giving is a viable strategy for raising charitable funds.

Keywords: charitable giving, subsistence income, redistribution, timing
JEL-Classification: C91, D03,D64,D31.

[1]Financial support by the German Research Foundation (DFG) for Nicklisch (Ni 1610/1-1) and Schnapp (Schn 1282/4-1) is gratefully acknowledged.

5.1 Introduction

Economic inequality is on the rise today even in rich and equality-oriented economies. At the same time, there is a broad consensus among researchers that increasing inequality is not harmless. Joseph Stiglitz, a former chief economist of the World Bank, warns, that inequality may result in increasing societal division, distributive struggle, and even violent conflict (Stiglitz, 2012). While the transfer of income from better-off echelons of a society to its more needy strata is common, there are different ways of raising money for such transfers. The two most important ways of raising transfer funds are the collection of mandatory payments (i.e., taxes) by the state and the collection of voluntary donations by charities.

Along the different nature of the payments (taxes are mandatory, donations are voluntary), there is another important difference between both mechanisms: taxes are generally known before the transfers are produced while donations are usually determined afterwards. While there are good reasons from a strategical point of view to make tax rates known ex-ante (e.g., hold-up problems as in Van Huyck et al., (1995)), it seems "natural" to determine donations ex-post eliminating the uncertainty about the total size of production for the donor (i.e., the donation decision is made after a disposable income is acquired).

However, in the real world, charities often use strategies where people commit to future payments out of future earnings, that is people pledge before the income is earned. For instance, the Founders pledge[2] allows tech entrepreneurs to make a legally binding commitment to donate 2% or more of their potential proceeds to a chosen charity. At "Giving what we can"[3] people pledge to donate 10% of their future income. Recently some banks started to offer similar charitable schemes to their clients. If a customer opts for a charitable feature of a checking account, some percentage of the average balance (monthly or yearly) is transferred to a charity. For instance, Land of Lincoln Credit Union and Cornerstone Community Bank donate a percentage of the average monthly balance of all depositors'

[2]https://founderspledge.com/ [accessed 22.03.2018]
[3]https://www.givingwhatwecan.org/ [accessed 22.03.2018]

checking accounts to customers' chosen charity organization or group. This commitment strategy is also used by more conventional charity organizations like the Red Cross[4], or can be a part of religious systems such as zakat[5] in Islam.

This paper analyzes the role of time in charitable giving. We ask how the timing of a donation decision relative to the production of disposable income affects the willingness to donate, the effort, one is willing to exert and, as a consequence, the amount of resources available for redistribution. As such, we consider unconditional binding donation pledges over the earned income, that is potentially uncertain. Since giving decisions affect working incentives, our study provides insights into the indirect consequences of charitable giving, namely its effect on the labor supply of donors as well as transfer recipients.

Giving to others has been an important topic for public economics for several decades and has been a manifestation of other-regarding preferences. Researchers studied extensively, on the one hand, what motivates people to give (Camerer, 2011; Engel, 2011). For instance, people may be willing to give due to a "warm glow" (Andreoni, 1989, 1990), and an enhanced self or social image (Andreoni and Bernheim, 2009; Bénabou and Tirole, 2006)). On the other hand, what can impede giving (e.g., allowing to disguise or excuse selfish actions Dana, Cain, et al., (2006) and Dana, Weber, et al., (2007)). Level of own income also has effect of donation rate but the evidence is mixed. While Eckel and Grossman, (2003, 2008) and Eckel, Grossman, and Milano, (2007) find a positive relationship between income and giving, other studies find a negative one (Erkal et al., 2011) or no significant relationship at all (Andreoni and Vesterlund, 2001; Buckley and Croson, 2006).

Benefits of pre-commitment are well-established for some realms of human behavior. Pre-commitment is known to mitigate inconsistencies between expected preferences and actual behavior, for example, in decisions on retirement savings (Thaler and Benartzi, 2004) or health-related issues (DellaVigna and Malmendier, 2006). At the same time, the effects of pledging on charitable giving are not clear. Breman, (2011) is the first to analyze the importance of commitments to future transfers for charitable giving. In a field experiment with the pool of existing

[4]https://www.redcrossblood.org/CommitToGive [accessed 22.03.2018]

[5]"Obligatory charity": 2.5% of one's wealth is to be donated to needy and poor. https://www.islamichelp.org.uk/zakat/ [accessed 22.03.2018]

donors, the author finds that people donate more if they can pledge to donate later, instead of donating immediately.

There have been two approaches developed to explain the relations between pledging and time of giving. Both gather empirical evidence in favor of their theories, although they rely on different mechanisms and generate contradicting hypotheses. Dreber et al., (2016) develop and test a model based on Kahnemann's idea of the fast and the slow system of thinking. In their experiment, a "short-run self" is assumed to be intuitive and altruistic, while a "long-run self" is rational and selfish. Observing that delaying payments decreases giving, they find evidence supporting their hypothesis. Andreoni and Serra-Garcia, (2016) depart from standard models of giving by assuming that utility from giving may be generated not only by the act of giving but also by advance pledging. Their model predicts that advance pledging increases giving (however, subsequent reneging if possible). Andreoni and Serra-Garcia, (ibid.) also present evidence in support of their hypothesis. After putting forward their own model the authors also consider how their data relate to the model suggested by Dreber et al., (2016). They find that among dynamically inconsistent subjects the majority of subjects (62%) choose to give when the choice is made in advance and not to give immediately, while the remaining 38% behave the opposite way (i.e., in line with Dreber et al., (ibid.)). Therefore the overall effect of the delay appears to depend on the distribution of subjects' types in the population. The most recent research touches upon the issue of uncertainty about one's own income and the timing of the actual donations relative to the decisions to donate. Reinstein et al., (2018) explore conditional commitments over an uncertain monetary bonus, which results from a lottery. They find that allowing for conditional commitments increases the donation rate.

In our experiment, we focus on the role of time for charitable giving, when disposable income is not granted, but earned. For this, we use the real-effort experiment introduced by Chugunova et al., (2017) which allows us to analyze giving and labor decisions in the environment with an exogenously imposed income threshold. Specifically, we negatively manipulate the productivity of random players and allow for charitable transfers to these disadvantaged low productivity players. The experiment has three production phases. In all phases players earn an income based on their individual performance. The tasks in the first phase are identical

for all players. Performance in this phase serves as a benchmark of ability. The third phase is a high income phase, where all players solve the same kind of tasks for a higher piece rate. In the second phase, players face an income threshold that has to be satisfied in order to qualify for the third phase. However, in the second phase, half of the player is randomly disadvantaged in the sense that they face a much harder task and therefore are less likely to meet the income threshold. Disadvantaged and, respectively, advantaged players are paired and the latter may choose to transfer a share of own earnings to the disadvantaged player in order to help him qualify for phase three.

This setup has two consequences for the advantaged players: it motivates giving to the disadvantaged player who is likely not to meet the income threshold. At the same time, it constrains giving through the necessity to fulfill the threshold.

Between treatments we vary the time when a donation decision is made. In one treatment advantaged players can make a donation pledge prior to production. We call this the ex-ante treatment (AT). Here they face uncertainty regarding their own productivity. If the pledge is too high, they may fail to qualify for the high income phase themselves. In a second treatment, they can donate immediately after production. We call this the ex-post treatment (PT). In this setting, there is no uncertainty about reaching the threshold any longer. However, the perception of being entitled to one's earnings may limit the amount donated or even lead to no donation at all.

Our results show that a larger share of players chooses to donate in AT as compared to PT. If people donate, the chosen donation rate is not different between treatments. Importantly, the average production in the ex-ante treatment is significantly higher than in the ex-post treatment. The effect is driven by an increased performance of the donors in AT and results in transfers which are almost 25% higher in the ex-ante as compared to ex-post condition. Our findings are in line with findings of Reinstein et al., (2018) and extend to the realm of unconditional pledges. The change in the timing of the donation does not affect the performance of the donees.

The remainder of the paper proceeds as follows. Section 2 presents the design in more detail and states predictions regarding the response to the treatment manipulation. Section 3 reports the results, while Section 4 concludes.

5.2 Experimental Design and Procedures

5.2.1 Design

The experiment consists of three phases of 30 minutes each, during which participants can perform simple computations (similar to Niederle and Vesterlund, (2007)). Players compute sums or differences of two three-digit or five-digit numbers. If a task is solved correctly, a new task appears on the screen; otherwise, the same task is presented again. The number of tasks is not limited and, therefore, participants are free to decide how fast, that is, how hard, to work. We deliberately opted for a tedious repetitive task over an extended period of time, in order to minimize joy of work and intrinsic motivation, such that players respond predominantly to pure monetary incentives.

Participants receive a piece rate per correctly solved tasks. This rate ϕ is at 0.05 Euro in phase one and two of the experiment and is then doubled in phase three to 0.10 Euro. The number of correctly solved tasks is always shown in one corner of the screen.

At the very beginning of the experiment, players are randomly matched into groups of two. These groups remain unchanged throughout the experiment. Within a group, each player is randomly assigned a type. The two types are S (supporter, we called this type "advantaged players" in the introduction) and R (recipient, "disadvantaged players" in the introduction).[6]

The type of players only matters in phase two of the experiment. Here S solves tasks with two three-digit numbers as before, while R now has to solve tasks with two five-digit numbers. The piece rate, however, is 0.05 Euro per task irrespective of a type of a player. The difference between the types is explained to the players

[6]In the experiment itself the types of players were denoted A and B to avoid any associations.

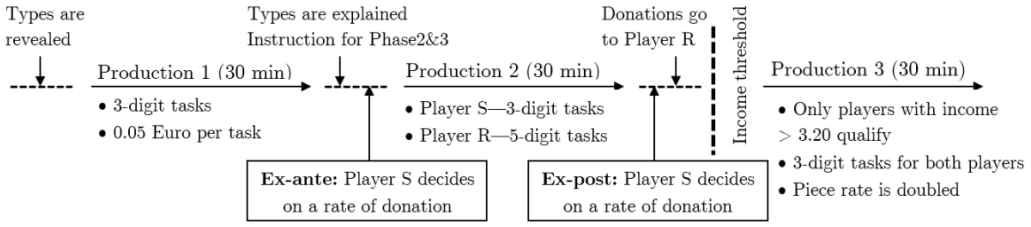


FIGURE 5.1: The timeline of the experiment

between phases one and two. Thus, players receive detailed information about types and their consequences, when it becomes relevant. Furthermore, in this way our reference measurement of individual productivity in the first phase is not affected by the type.

In phase two both players S and R face a subsistence income threshold, which is set exogenously at 3.20 Euro. If a player does not reach this threshold in the second phase, she cannot take part in the high income third phase of the experiment but has to, nevertheless, wait in the laboratory until the end of the experiment. The subsistence threshold is calibrated such that it is on average just too high to be reached by R, but easy to pass for S. This income threshold is supposed to induce and justify redistribution within our experimental society to those who need financial support to stay in the game and earn an appropriate total income from the experiment (i.e. players R).[7] Notice that we inform players about their types at the beginning of phase one, but do not reveal any additional information. Therefore, players cannot form beliefs that typing is related to performance in the first phase. That is, being advantaged is the result of the lottery and not the result of winning a competition (Durante et al., 2014). Figure 5.1 presents the general timeline of the experiment.

Without taking part in the third phase, players receive approximately 7.40 Euro for our 140-minute experiment. This is well below expectations which according to the laboratory guidelines are at about 12 Euro per hour. The average income from our experiment for all three phases is slightly above 26 Euro. This amounts to roughly 12 Euro per hour and is thus in line with expectations.

[7]Donation behavior is often modeled as a public good game in the literature (see Ledyard, (1995)). We, however, believe that a donor should not receive monetary gains from giving and therefore model it as a dictator game following Cherry et al., (2002) and Ogawa et al., (2012).

In this general setup, we run two treatment conditions in which we vary the timing of the donation decisions. In the ex-ante treatment (AT), players S commit to a donation rate b in steps of 5% prior to production. This implies that there is uncertainty about the amount of money that will be earned in phase two. The donation share is deducted from S's income Y_S after phase two and transferred to player R immediately and without a loss of value. In the ex-post (PT) treatment players S decide about the share of income for donation when the production in phase two is over. That means the donor knows exactly how much she earned and may make a donation decision taking into account her own subsistence income constraint. Players S qualify for the third phase if their income minus donation is above the income threshold ($(1-b)Y_S > 3.20$ Euro). Players R qualify if their income plus the transfer satisfies the subsistence income requirement ($Y_R + bY_S > 3.20$ Euro). To explain players S how the donation system works we presented some examples before the donation decision.[8] In both treatments supporters do not know how much the matched player R earned. We, therefore, elicit the maximum willingness to donate. In addition, due to the one-to-one correspondence of supporters and recipients, supporters do not face coordination problems. In this environment, we can observe how the interplay between uncertainty and the perception of entitlement towards one's earnings affects the donation decision and labor supply. Table 5.1 summarizes the payoff schemes of the two treatment conditions in the second phase of our experiment.

Table 5.1: Payoffs in the second phase by treatment and player type

Treatments	**Player S**	**Player R**	b **is set**	**Number of obs.**[9]
AT	$(1-b)Y_S$	$Y_R + bY_S$	individually before production	39
PT	$(1-b)Y_S$	$Y_R + bY_S$	individually after production	30[10]

[8] Identical examples were shown in both treatments and can be found in the Appendix.

[9] The number of independent observations, that is, the data for a pair of players.

[10] We analyze data for 29 pairs of observations in PT. We removed two participants from the sample since individuals heavily deviated in their computational skills from the other participants (i.e., their performance in the first phase was more than 2.5 standard deviations from the mean performance.

5.2.2 Predictions

In the following, we assume that people first choose the preferred work speed and then the donation rate.[11] Following the approach by Andreoni and Serra-Garcia, (2016) we assume that players S perceive a "warm glow" of giving as one element of their additive separable utility function. At the time, i is deciding upon b the size of the expected (egocentric) utility from phase three is fixed in PT, but uncertain in AT. In other words, in PT i, while deciding upon b, knows whether she passed the threshold to phase three; therefore, i's expected utility in phase three is independent of her decision regarding b.[12] In AT, however, the donation share in phase two negatively influences the likelihood that i passes the threshold. Therefore, the optimization of b following the rationale that the marginal egocentric benefit equals the altruistic benefits leads to smaller donation shares in ex-ante than in ex-post, since higher donation shares are associated with smaller expected utility in phase three (and, thus, smaller marginal egocentric benefits).

Hypothesis 1: Players S donate higher shares in ex-post than ex-ante.

Regarding the speed of production, predictions are unclear. Players S may work harder in ex-ante to meet the income threshold: since i keeps less for herself in PT than in AT, the (marginal) costs of production play a more important role in PT than in AT lowering S's effort. As a consequence, it is a priori unclear under which condition the generated transfers are larger.

Hypothesis 2: The size of a transfer in absolute terms can be smaller or larger ex-ante than ex-post.

Finally, another important aspect of this issue is how Players R adjust their labor supply as a reaction to the treatment manipulation. In a related paper, Chugunova et al., (2017) show that players R perform better if they are beneficiaries of a

[11] Due to the nature of a laboratory experiment, in our formal analysis we do not introduce time discounting in any form.

[12] Technically speaking, this is not entirely true if i voluntarily chooses b such that she does not pass the threshold. Since this is a rather unusual observation in ex-post, we will not include this case in our analysis.

redistribution (i.e., if they are sure to receive a positive transfer of an uncertain amount). Taken at the extreme, Hypothesis 1 may drive recipients to expect a zero transfer in AT and a positive transfer of uncertain amount in PT. Therefore, recipients may enhance their performance in PT as compared to AT.

Hypothesis 3: Players R exert more effort ex-post than ex-ante.

5.2.3 Experimental features

Our experiment is a real-effort laboratory experiment with a tedious task that is carried out for a lengthy period of time. This setup generates two important features. First, the tediousness of the task helps us to minimize intrinsic motivation to work and better capture the psychology of effort (Charness et al., 2018). Second, it allows us not only to observe donation decisions under various circumstances but also to test the labor response to donation. Since players donate a share of their income, a donation can affect labor supply and ultimately the total amount of money donated. On the one hand, if donation triggers a negative labor response, it can decrease the total amount of transfer. On the other hand, Imas, (2014) shows that in some contexts people work harder if they work for charities as compared to working for their own benefit. Although in our environment hard work benefits both self and other, a "working for warm glow" mechanism can affect the labor response positively and thus increase the total amount of money donated.

When facing a real-effort task, subjects have to make a decision on a preferred mixture of work and leisure. An objection to our results could, therefore, be, that we did not offer a plausible leisure opportunity, and hence conclusions regarding labor supply must be invalid. However, research on productivity tells us, that not performing a task, that is tedious and requires a lot of concentration, can be considered as leisure (e.g, Kool et al., 2010). In addition, even in the environment we offered, participants who opted not to work, found alternative (i.e. leisure) activities to carry out. This is demonstrated by the fact that we found origami pieces and sheets of paper with drawings in the cubicles of some participants after the sessions.

Another effect possibly tainting our results is, that the transfer is perceived by supporters as a consumption good. We worked against this effect by making the donation instrumental (e.g., Deutsch, 1975). Reaching the income threshold in the second phase has little value as such. However, it allows a player to exert effort and earn income in the high income third phase of the experiment. How much a player benefits from phase three is, again, contingent on her own effort in this phase. Thus, the donation enables a recipient to benefit from phase three but does not ensure a certain payout.

In our design, the generated donations are transferred to a randomly disadvantaged player in the laboratory, and not to a real-world charity. This approach is used in many lab experiments on giving (e.g., Cherry et al., 2002). The purpose of this setup is to isolate the donation decision from a number of other effects: beliefs about the efficiency of suggested charities, beliefs about how beneficent charitable work is as such, beliefs about the worthiness of recipients, and beliefs about the impact individual contributions have. The strongest concern is the worthiness of recipients or how accountable are the recipients in their dire straits. The accountability principle is proven to play a role in redistribution decisions (Cappelen et al., 2013). In our laboratory setting, we ensure that the recipients are disadvantaged at random, that is, they obviously are not accountable for their own disadvantage and thus deserving. Additionally, using recipients in the laboratory allows to control for transaction costs and the timing of the transfer. The full donation amount goes to the recipient immediately. Last, not least, the donation is very likely to be efficiency enhancing, because it increases the recipient's chance to qualify for the high income third round of the experiment and earn additional income, instead of sitting idle.

Our experiment uses a one-to-one correspondence of donors and recipients. Player S is individually responsible for determining the size of the transfer to R and thus there is no coordination problem among several players S.

Last, not least, in PT we create the least favorable condition for redistribution by reinforcing the perception of entitlement towards their earnings among supporters. As previous studies on giving show (e.g., Cherry et al., 2002; Durante et al., 2014; Ogawa et al., 2012), if donors work to obtain their endowments, they are less

likely to redistribute. In PT donating after production makes donors feel entitled to their earnings and thus less willing to transfer money to the other player.

5.2.4 Procedure

The experiment was conducted at the Vienna Center for Experimental Economics (VCEE) using z-tree (Fischbacher, 2007) in January 2017. Subjects were mostly students of various majors of the University of Vienna, recruited online via ORSEE (Greiner, 2015) and randomly assigned to treatments. No subject participated more than once. The average age of the participants was 24 with more female participants (62%) taking part in the experiment.

Before the beginning of the experiment, participants placed their cellphones into provided envelopes and sealed them. They had access to the envelopes only after their payment. In addition, participants were asked to keep their bags and personal belongings outside of their cubicles. Thus, participants had no phones or calculators to help them solve their tasks. They were warned that the experimenter will monitor usage of forbidden devices throughout the experiment, and in case of breach of rules participants will be removed from the experiment without payment. Pen and paper were provided for calculations. Participants were seated in cubicles and received a copy of the instructions for the first phase. Additionally, the instructions were read aloud.

Subjects were randomly matched in groups of two, which remained unchanged throughout the experiment and played simultaneously, but independently from other pairs. Before any action took place the type of a player was displayed on the screen. Although types were revealed, the actual difference between types was made clear only after the end of phase one. At this time, instructions for the second and third phase of the experiment were distributed and read aloud. After each phase players had a short break of three minutes. Throughout the experiment players did not communicate with each other. In both treatments, AT and PT, players S were given identical examples, clarifying the donation procedure directly before making their donation decisions. Supporters could choose any donation

rate as a percent of earned income from 0% (i.e. donate nothing) to 100% (i.e. donate all) with the step of 5%.

We conducted 7 sessions with a total of 138 participants.[13] Sessions ran with 18-20 subjects per session yielding ca. 30 independent observations per treatment (see Table 5.1). At the end of each session, subjects were paid privately and in cash. Payments ranged from 2.60 to 50.40 Euro with an average of 26.20 Euro for approximately 140 minutes.[14]

5.3 Results

5.3.1 Supporters

Our data show that in line with the previous findings more people are willing to donate in the ex-ante condition even if the pre-commitment is binding. We further on refer to those players who chose to donate any amount higher than 0% as donors. First of all, 75% of supporters are donors in AT, while only 55% are donors in PT. Two-sample test of proportions confirms that the proportion of donors in AT is significantly higher than in PT (p=0.04). Although donors in PT treatment choose slightly higher donation rates, the difference between the donation rates between ex-ante and ex-post is not significant. In PT donors on average give 20.6%, while in AT – 18.7% (Mann-Whitney U-test[15] p=0.6). While we find that more people pledge to donate, it does not result in higher donation rates. Therefore we can not confirm the results of Breman, (2011) that pre-commitment leads to higher donation rates. However, our experiment differs in important ways, which could explain this difference. The core one is the sample. Breman, (ibid.) worked with people who are already donors and in our sample, the decision to donate was made

[13] We removed the data from 2 participants from PT since their math ability was more than 2,5 SD away from the mean.

[14] In addition to their earnings from the real-effort task all participants received a 5 Euro flat payment for filling in the questionnaire at the end of the experiment. Participants did not anticipate this extra payment. It therefore cannot have affected their behavior in the production phases. The additional payment was necessary to meet the payoff standards of VCEE. We did not include the flat payment into the range of payments presented above.

[15] All non-parametric statistical tests reported are two-tailed taking individuals as units of observations.

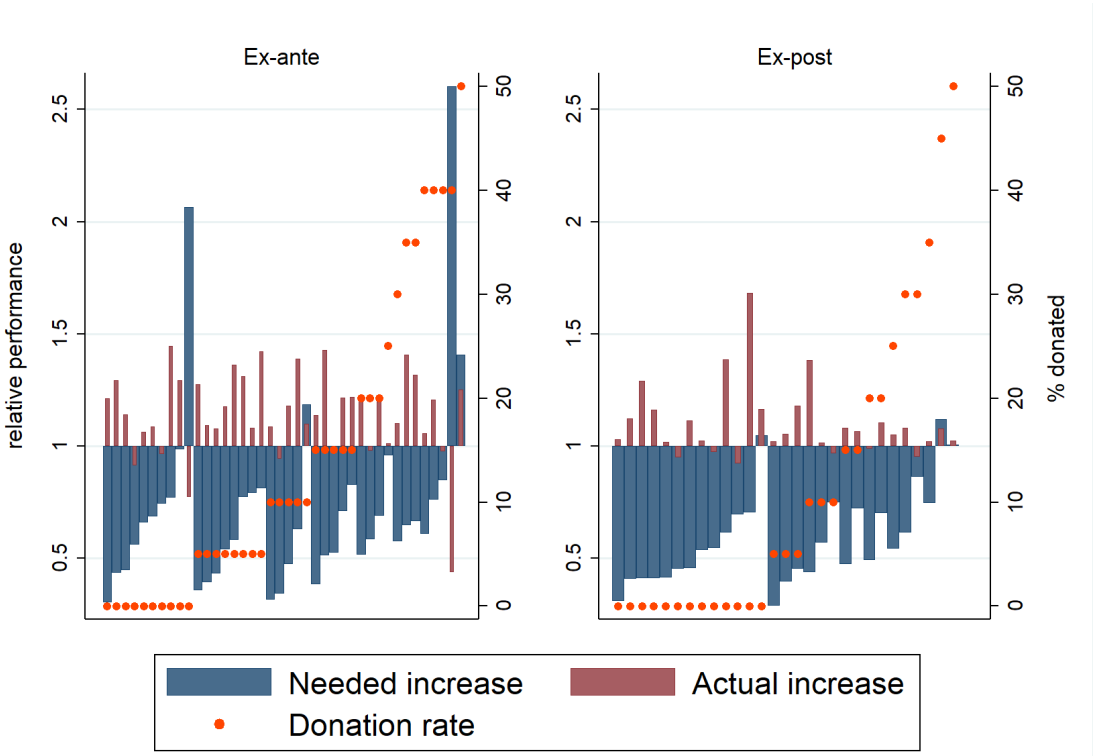


FIGURE 5.2: The individual labor response to a chosen donation rate

for the first time. If we consider the donation rates chosen on average by all the players, who had an opportunity to donate, we find that the per capita donation rate is higher in ex-ante treatment (the difference is however not significant: ex-ante 14.0% and ex-post 11.4%)

Contrary to the predictions of a theory with selfish preferences, donation rates in both donation treatments are significantly different from 0 (one sample t-test AT p=0.00, PT p=0.0001). That is in sharp contrast with the results of Cherry et al. (2002), even when supporters were to decide about the earned endowments and anonymously the majority chose to donate, behaving altruistically.

Therefore based on our data we reject Hypothesis 1:

Result 1: Significantly more people committed to donating in ex-ante than ex-post. Donation rates are non-different in both treatments but significantly different from zero.

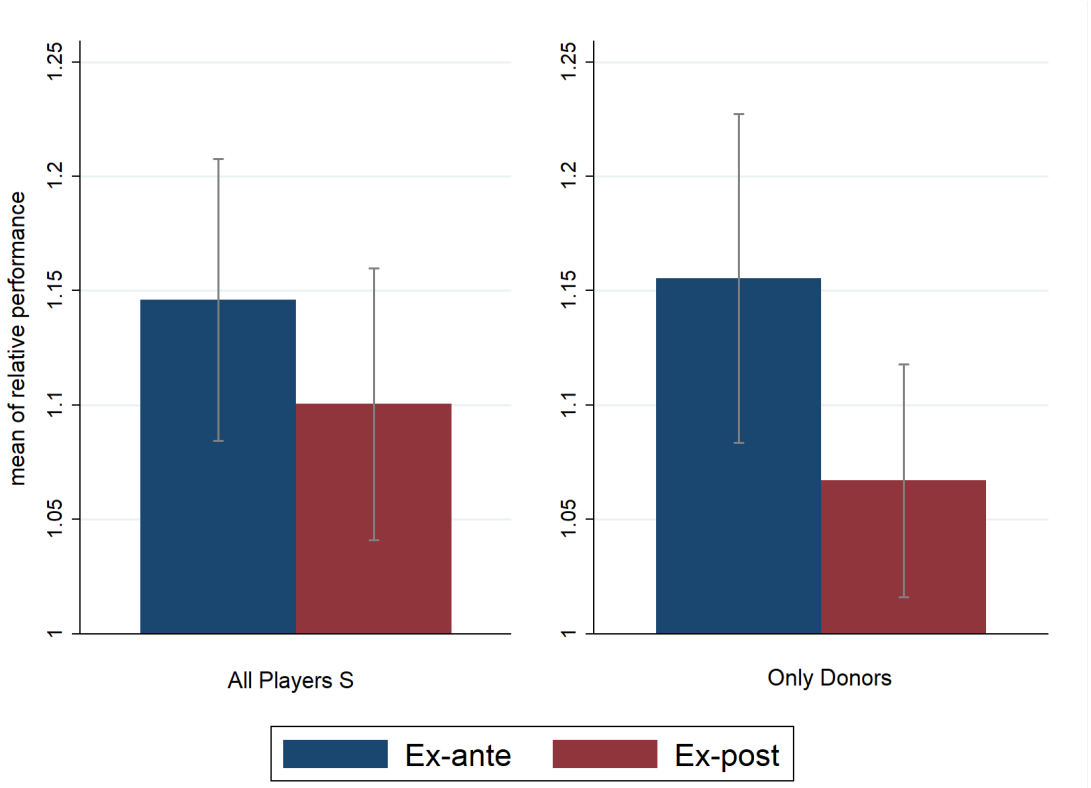


FIGURE 5.3: Aggregate labor response to a chosen donation rate

As hypothesized above, chosen donation rates may affect the labor choices of players. In the first phase players in both treatments perform equally well, meaning that the initial ability of players is non-different (127 tasks in AT and 138 in PT, p=0.12). To measure the performance of the participants controlling for their initial ability we divide the number of correctly solved tasks in the second phase by the number of correctly solved tasks in the first phase. We further refer to this relation between the stages as *relative performance*. The relative performance of 1 means that a player solves in the second phase exactly the same number of tasks as in the first one; a number less than 1 means that she reduces her effort and solves less, and a number more than 1 means respectively that she improves her performance. Thus, one can also interpret relative performance as a percentage of change as compared to the first phase.

Fig. 5.2 shows that the threshold of 3.20 Euro was easily attainable for the majority of S players even if we take into account individually chosen donation rates. Blue bars depict the level of relative performance each player could have exerted to pass the threshold under the chosen rate of donation (marked as orange dots

with the values on the right Y-axis), and as one can see most of the players could have decreased their performance by almost half and would still qualify for production in phase three. Therefore labor responses in phase two are not driven only by the necessity to satisfy it. Red bars correspond to the actual labor response we observed and they lie mostly above the reference line of 1 (no change in production). Thus, supporters in both treatments performed better in phase two as compared with phase one. The figure also gives an overview, which taxation rates were chosen by every single individual in the session.

Figure 5.3 displays the average relative performance of the supporters. Mann-Whitney U-test confirms that relative performance in AT (1.15) is significantly higher than in PT (1.10 p=0.04). Considering the labor performance of donors[16], we find that the relative performance of donors in AT (1.16) treatment is significantly higher than in PT treatment (1.07) (p=0.01, the right panel of Fig.5.3). Thus, the experimental results shed some light on the inconclusive Hypothesis 2:

Result 2: Relative performance in AT is significantly higher than in PT both for the whole population of players S and for donors only.

As the donation rates are non-different in PT and AT, but the relative performance in AT treatment is higher, there may occur differences in the absolute amount transferred to the disadvantaged players. In AT on average players S give away the earnings from 20.8 solved tasks to recipients. In PT it is only 15.7 tasks. Although it accounts for a 25% increase transfer in AT, the difference does not reach levels of statistical significance (p=0.27).

Result 3: Donation rates and exerted effort result in a large however non-significant difference in generated transfers.

In absolute terms, players donate 1.04 Euro on average in AT and 0.79 Euro in PT. That implies, that supporters usually expect recipients to exert some effort to qualify on their own. Only 4 players deviate from this pattern and donate 3.20 Euro or more, that is, the full amount necessary to qualify. We do not observe any focal point in the donated amounts.

[16]Our sample does not contain enough non-donors to make a conclusive statement about their performance.

In the following, we explore the mechanism, which may drive the enhanced performance of donors in AT and consider what factors shape a donation decision.

5.3.1.1 Why do donors work harder in AT?

In PT players S determine how much to donate by choosing a donation rate which will be applied to the known income, whereas in AT the resulting transfer is a product of a chosen donation rate and exerted labor.

Potentially, there are two mechanisms which could motivate donors to work harder in ex-ante treatment. The first mechanism is driven by the reference dependent income. Camerer et al., (1997), Farber, (2005, 2008), and Fehr and Goette, (2007) established that labor decisions are heavily affected by the reference point. A reference dependence may come into play since the decision to donate decreases the piece rate and therefore requires an increased effort to reach the reference point. The second potential mechanism is the result of social preferences. Due to the payoff structure of the experiment if one commits to donating than production profits both the player herself and the recipient. Therefore by increasing labor supply, the supporter can influence how much will be transferred in absolute terms. This mechanism is in line with the findings of Imas, (2014).

Although on the aggregate level both mechanisms produce similar results, they differ in the predictions of what happens after the reference point is reached. The former mechanism implies that after the reference income the speed of production should decrease, while the latter suggests a rather constant speed, i.e., no change of the production speed after the reference income is reached.

We use this difference to distinguish between the mechanisms. In our environment, income from the phase one is a likely reference point. Phase one and phase two both last 30 minutes and players S are offered the same type of tasks and therefore they could use their earnings in phase one as a reference point for phase two. The number of solved tasks is always shown in the corner of the screen and therefore players can easily see if they reached the target. For the analysis, we consider only those players who earned more in the second phase than in the first one, i.e. those who earned at least their reference income. The reference income differs for

each individual and depends both on how much was produced in phase one and on which donation rate was chosen in phase two. For these players, we calculate the average speed of solving one task before the reference point was reached and after it was reached. As we compare changes within subject we do not have to control for their ability when considering the change in speed. Both if we consider only donors or all the players who satisfy the above-described conditions, we do not find a significant difference in the speed of production before and after the reference point (Mann Whitney U test p=0.37 and p=0.4 respectively).

Another consideration, which casts doubt on the reference income hypothesis is the performance of players in phase three. As piece rate is doubled, under a reference income hypothesis players would be expected to solve only about 50% of the number of tasks solved in phase two. However, in phase three players solve about the same number of tasks as in phase two despite the higher piece rate (p=0.73). If anything, their reference point would be the number of tasks solved, but not income, which appears to contradict the findings on the connection between labor and wage changes (Farber, 2005; Fehr and Goette, 2007). We, therefore, conclude that in phase two with a donation it is not the reference income from phase one either, that drives players to perform better.

To explore further if it is indeed pro-sociality that drives the increase in relative performance or pure self-interest we compare the performance of players who donate and those who did not. Although Figure 5.2 shows that the subsistence threshold was not binding for the players, to account for it we look only at the performance after the threshold was satisfied.[17] As we compare the speed of performance between subjects, we control for their initial ability. Table 5.2 displays the results of an OLS regression. For the performance after the threshold and controlling for the individual ability the fact that one is a donor (dummy, i.e., regardless of the chosen share) is associated with 2.5 seconds decrease in the average time per task (panel 1). Additionally, donors who give more tend to speed up more (panel 2). This analysis allows us to conclude that:

[17] 3.20 was an exogenously imposed threshold, but as it was affected by chosen donation rates each player de facto faced an individual threshold.

Table 5.2: Speed of production conditional on donation

Speed of production	(1)	(2)
Donor	-2.507**	
	(1.018)	
N tasks (phase 1)	-0.0713***	-0.0743***
	(0.00870)	(0.00886)
Donation Share		-0.0633**
		(0.0232)
Constant	23.45***	22.78***
	(1.571)	(1.563)
Observations	36	36
R-squared	0.740	0.709

Standard errors clustered at a subject level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Result 4: The positive labor response, observed in AT treatment among donors, is not due to the reference dependent income but is a result of pro-social preferences. That is, pro-social players donate and additionally boost their performance since recipients benefit from it.

5.3.1.2 Why do people choose to donate?

Determinants of the donation decision have applied importance for fundraising. Our experiment generates data, which allows us to take a closer look at what determines the likelihood of donation.

After the end of production phase three, participants were asked to fill in the questionnaire. It included socio-demographic data, big five personality questions (Schupp and Gerlitz, n.d.), self-reported risk and trust measures and political orientation, as well as questions of a justice ideology scale (Stark et al., 2000).[18] The justice ideology scale measures attitudes towards major justice ideologies (egalitarianism, individualism, fatalism, and ascriptivism).

After the factor analysis we acquired 4 reliable measures of personality traits (extroversion, neuroticism, openness, consciousness) and 2 scores from the justice

[18]The questionnaire data for one session in AT was lost due to technical issues. Therefore the number of observations for AT treatment differs from the previous section.

ideology scale, which reflect attitudes towards individualism (i.e., role of merit and effort in the output) and fatalism (i.e., general doubt in justice).

Due to a limited number of observations we refrain from drawing conclusions on what influences the chosen donation rate and focus on the binary decision to donate. To establish what affects the likelihood of becoming a donor we run a probit regression with standard errors clustered at the session level and then estimate the marginal effects, which are displayed in Table 5.3.[19] The dependent variable is binary and takes a value of 1 if a player chose to donate more than 0% of their earnings. In AT treatment we control for the skill of a player (i.e., performance in phase one), the opinion regarding the role of the merit in the outcomes, doubt in justice, distrust towards others, risk attitudes, political orientation, and gender. In PT we additionally control for performance in phase two. With this set of controls, we can capture both attitudes and behavior, which may affect the willingness to donate. From the socio-demographic questionnaire, we include only gender as previous research has found that there are gender differences in the domain of charitable giving (see e.g., Wiepking and Bekkers, (2012) for a comprehensive overview).

We find that in AT the skill of a player S increases the likelihood of becoming a donor, the magnitude of the effect is, however, small. Significant and large in magnitude effect comes from a political orientation of a subject. The variable is coded such that higher values correspond to more right political allegiances. That is, players who report adhering to more right political views tend to become donors more often in AT treatment. Gender, beliefs about justice and merit as well as risk attitudes are never significant, while trust gains significance in some specifications, which is however not robust. That is, players who trust other less tend not to become donors, or to put it simpler, people who trust others tend to become donors more often.

In PT neither initial ability nor performance in phase two affect the likelihood of becoming a donor. The regression shows that people who believe that it is in

[19]Following the finding that relation between income and being a donor is non-linear, but U shaped, we additionally tested a specification (2) with the square of performance in phase 2, regression yields qualitatively similar results.

general possible to achieve a just distribution of resources and those who are more trusting are more likely to become donors.

Table 5.3: Marginal Effects: Determinants of becoming a donor

	(1)	(2)
Becoming a donor	AT	PT
Skill	0.00613***	-0.000243
	(0.000928)	(0.00838)
Performance Phase 2		0.00106
		(0.00578)
Merit	-0.0271	-0.144
	(0.106)	(0.0932)
Belief in Justice	-0.0490	0.421**
	(0.124)	(0.182)
Distrust	-0.131	-0.239***
	(0.0982)	(0.0720)
Male	-0.000218	-0.00126
	(0.154)	(0.0544)
Risk	0.0124	0.00595
	(0.0194)	(0.0547)
Political views	0.161**	0.101
	(0.0728)	(0.116)
Pseudo R2	0.29	0.44
Observations	30	29

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

5.3.2 Labor supply of transfer recipients

Following the findings of Chugunova et al., (2017), a procedure which generates transfer may affect the labor supply of transfer recipients. However, it is not the case for the timing of the donation decision. The average relative performance of transfer recipients in AT and PT is not significantly different between treatments. More difficult task slows the welfare recipients down by ca. half (relative performance in AT 0.51 and in PT 0.54 p=0.47). In phase two recipients worked to their true ability: the correlation between the number of tasks solved in phase one and two is significant in both treatments but is higher in PT (0.81) than in AT (0.61).

Result 5: The timing of the donation decision of the counterparts does not influence labor supply of welfare recipients.

5.4 Conclusion

In this paper we studied how the timing of donation decisions relative to the production of disposable income affects the willingness to donate and labor supply. Our experiment provides an insight in the labor response both on the supply and on the demand side of redistribution. We find that more people commit to donating in the ex-ante condition and that they choose to donate a share of their income that is not different as compared to donors in the ex-post treatment. Controlling for the initial ability, donors in the ex-ante treatment also boost their performance in phase two more than their counterparts in the ex-post treatment. Regarding the mechanism which drives this enhancement, our data suggest that it is triggered by pro-sociality, not by a reference income. A higher share of donors, non-different donation rates and increased labor supply in AT result in a large, however, insignificant difference in generated monetary transfers to disadvantaged players.

We find that political views to the right of the spectrum as well as initial skill level increase the likelihood of becoming a donor in the ex-ante condition. In the ex-post treatment people who believe that a just distribution of resources can be achieved and those who trust others are more likely to donate non-zero amounts.

Our experiment introduces several features of the working environment that allow capturing the nature of labor supply decisions better. We employ a repetitive tedious real-effort task with long production phases to reinforce the importance of the monetary incentives to the players. The novel feature of our experiment is the introduction of a subsistence income constraint. This constraint mimics the fact that labor-leisure choices are restricted by subsistence, that is, by consumption that cannot be undercut by anyone.

Several concerns shall be mentioned with respect to the results of the current study. Our experimental design was likely to create a strong feeling of entitlement

towards one's earnings in the ex-post treatment. This feeling was amplified by the tediousness of the task requiring a lot of concentration and is essentially meaningless. With a more intrinsically valuable task, it may well be, that we would observe different results. However, our study can capture the lower bound of a willingness to donate and as such tested a donation in an adverse environment.

Overall, the findings of the current study suggest that committing to donating from future earnings is a viable strategy for fundraising. Several important questions remain, however, for future investigation. First of all, it is important, which arrangement (i.e., ex-ante or ex-post) will be chosen by players if they can opt for either of the mechanisms. Secondly, the fact that commitment is binding and does not allow for reneging may play a role for donation and labor decisions. Therefore, we consider this study to be a promising invitation for future research to shed more light on this exciting aspect of human life.

Bibliography

Andreoni, James (1989). "Giving with impure altruism: Applications to charity and Ricardian equivalence". In: *Journal of Political Economy* 97.6, pp. 1447–1458.

– (1990). "Impure altruism and donations to public goods: A theory of warm-glow giving". In: *The Economic Journal* 100.401, pp. 464–477.

Andreoni, James and B Douglas Bernheim (2009). "Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects". In: *Econometrica* 77.5, pp. 1607–1636.

Andreoni, James and Marta Serra-Garcia (2016). *Time-inconsistent charitable giving*. Tech. rep. National Bureau of Economic Research.

Andreoni, James and Lise Vesterlund (2001). "Which is the fair sex? Gender differences in altruism". In: *The Quarterly Journal of Economics* 116.1, pp. 293–312.

Bénabou, Roland and Jean Tirole (2006). "Incentives and prosocial behavior". In: *American Economic Review* 96.5, pp. 1652–1678.

Breman, Anna (2011). "Give more tomorrow: Two field experiments on altruism and intertemporal choice". In: *Journal of Public Economics* 95.11-12, pp. 1349–1357.

Buckley, Edward and Rachel Croson (2006). "Income and wealth heterogeneity in the voluntary provision of linear public goods". In: *Journal of Public Economics* 90.4-5, pp. 935–955.

Camerer, Colin (2011). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

Camerer, Colin, Linda Babcock, George Loewenstein, and Richard Thaler (1997). "Labor supply of New York City cabdrivers: One day at a time". In: *The Quarterly Journal of Economics* 112.2, pp. 407–441.

Cappelen, Alexander W, James Konow, Erik Ø Sørensen, and Bertil Tungodden (2013). "Just luck: An experimental study of risk-taking and fairness". In: *The American Economic Review* 103.4, pp. 1398–1413.

Charness, Gary, Uri Gneezy, and Austin Henderson (2018). "Experimental methods: Measuring effort in economics experiments". In: *Journal of Economic Behavior & Organization* 149, pp. 74–87.

Cherry, Todd L, Peter Frykblom, and Jason F Shogren (2002). "Hardnose the dictator". In: *The American Economic Review* 92.4, pp. 1218–1221.

Chugunova, Marina, Andreas Nicklisch, and Kai-Uwe Schnapp (2017). "Redistribution and Production with the Subsistence Income Constraint: a Real-Effort Experiment." In: URL: `http://bedarfsgerechtigkeit.hsu-hh.de/dropbox/wp/2017-18.pdf`.

Dana, Jason, Daylian M Cain, and Robyn M Dawes (2006). "What you don't know won't hurt me: Costly (but quiet) exit in dictator games". In: *Organizational Behavior and Human Decision Processes* 100.2, pp. 193–201.

Dana, Jason, Roberto A Weber, and Jason Xi Kuang (2007). "Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness". In: *Economic Theory* 33.1, pp. 67–80.

DellaVigna, Stefano and Ulrike Malmendier (2006). "Paying not to go to the gym". In: *American Economic Review* 96.3, pp. 694–719.

Deutsch, Morton (1975). "Equity, equality, and need: What determines which value will be used as the basis of distributive justice?" In: *Journal of Social Issues* 31.3, pp. 137–149.

Dreber, Anna, Drew Fudenberg, David Levine, and David Rand (2016). "Self-control, social preferences and the effect of delayed payments". In: *Working Paper*.

Durante, Ruben, Louis Putterman, and Joël Weele (2014). "Preferences for redistribution and perception of fairness: An experimental study". In: *Journal of the European Economic Association* 12.4, pp. 1059–1086.

Eckel, Catherine and Philip J Grossman (2003). "Rebate versus matching: does how we subsidize charitable contributions matter?" In: *Journal of Public Economics* 87.3-4, pp. 681–701.

– (2008). "Men, women and risk aversion: Experimental evidence". In: *Handbook of Experimental Economics Results* 1, pp. 1061–1073.

Eckel, Catherine, Philip J Grossman, and Angela Milano (2007). "Is more information always better? An experimental study of charitable giving and Hurricane Katrina". In: *Southern Economic Journal*, pp. 388–411.

Engel, Christoph (2011). "Dictator games: A meta study". In: *Experimental Economics* 14.4, pp. 583–610.

Erkal, Nisvan, Lata Gangadharan, and Nikos Nikiforakis (2011). "Relative earnings and giving in a real-effort experiment". In: *American Economic Review* 101.7, pp. 3330–48.

Farber, Henry S (2005). "Is tomorrow another day? The labor supply of New York City cabdrivers". In: *Journal of Political Economy* 113.1, pp. 46–82.

– (2008). "Reference-dependent preferences and labor supply: The case of New York City taxi drivers". In: *American Economic Review* 98.3, pp. 1069–82.

Fehr, Ernst and Lorenz Goette (2007). "Do workers work more if wages are high? Evidence from a randomized field experiment". In: *American Economic Review* 97.1, pp. 298–317.

Fischbacher, Urs (2007). "z-Tree: Zurich toolbox for ready-made economic experiments". In: *Experimental Economics* 10.2, pp. 171–178.

Greiner, Ben (2015). "Subject pool recruitment procedures: organizing experiments with ORSEE". In: *Journal of the Economic Science Association* 1.1, pp. 114–125.

Imas, Alex (2014). "Working for the "warm glow": On the benefits and limits of prosocial incentives". In: *Journal of Public Economics* 114, pp. 14–18.

Kool, Wouter, Joseph T McGuire, Zev B Rosen, and Matthew M Botvinick (2010). "Decision making and the avoidance of cognitive demand." In: *Journal of Experimental Psychology: General* 139.4, p. 665.

Ledyard, John O (1995). *Public Goods: A Survey of Experimental Research. S. 111–194 in: J. Kagel/A. Roth (Hrsg.), Handbook of Experimental Economics.*

Niederle, Muriel and Lise Vesterlund (2007). "Do women shy away from competition? Do men compete too much?" In: *The Quarterly Journal of Economics* 122.3, pp. 1067–1101.

Ogawa, Kazuhito, Toru Takemoto, Hiromasa Takahashi, and Akihiro Suzuki (2012). "Income earning opportunity and work performance affect donating behavior: Evidence from dictator game experiments". In: *The Journal of Socio-Economics* 41.6, pp. 816–826.

Reinstein, D, G Riener, and C Kellner (2018). "Commitments to Give-if-you-Win Exceed Donations After a Win". In: *Working Paper*.

Schupp, Juergen and J.-Y. Gerlitz. "Big Five Inventory - SOEP (BFI - S)". In: *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. Ed. by D. Danner and A. Glöckner-Rist. Mannheim: GESIS. URL: http://zis.gesis.org/pdf/Dokumentation/Schupp+%20Big%20Five%20Inventors%20SOEP%20%28BFI-S%29.pdf.

Stark, Gunnar, Stefan Liebig, and Bernd Wegener (2000). "Gerechtigkeitsideologien: Theoretische Grundlagen, Operationalisierung und Ergebnisse für die Erhebung von Einstellungen zur sozialen Gerechtigkeit". In: *ZUMA-Informations-System Sozialwissenschaftlicher Skalen (ZIS) Version* 4.

Stiglitz, Joseph E (2012). *The price of inequality: How today's divided society endangers our future*. WW Norton & Company.

Thaler, Richard and Shlomo Benartzi (2004). "Save more tomorrow^{TM}: Using behavioral economics to increase employee saving". In: *Journal of Political Economy* 112.S1, S164–S187.

Van Huyck, John B, Raymond C Battalio, and Mary F Walters (1995). "Commitment versus discretion in the peasant-dictator game". In: *Games and Economic Behavior* 10.1, pp. 143–170.

Wiepking, Pamala and René Bekkers (2012). "Who gives? A literature review of predictors of charitable giving. Part Two: Gender, family composition and income". In: *Voluntary Sector Review* 3.2, pp. 217–245.

A.1 Instructions: Ex-ante treatment[20]

Instructions

General rules for participants

You are now going to participate in an economic experiment. At the end of the experiment you will receive a payment. How much you will earn depends on your activities and partly on the activities of other participants. Therefore, it is important that you read the following explanations carefully.

Please do not talk to each other after the start of the experiment. Please do not try to communicate in any other way, unless you are directly asked to during the experiment. If you have any questions, please let us know by a hand sign. An experimenter will come to you to answer your questions. Not following these rules will lead to the exclusion from the experiment and from all payments. Your decisions in the experiment and your answers in the following questionnaire are **anonymous.** Your identity is revealed only to the experimenter, but your answers can not be matched to your identity.

The experiment consists of **three** parts. **You will first receive the instructions for part one of the experiment.** The instructions for parts two and three will be distributed after completion of the first part.

For the duration of the experiment you are randomly assigned by the computer to a group of two players. Apart from you, your group has one more person. **In all three parts of the experiment you are together with the same person in a group.** In each group, there is a person A and person B. The computer assigns the roles randomly at the beginning of the experiment. The decisions of the other person in your group may have an effect on how much you earn. The decisions of people who are not in your group definitely have no influence on how much you earn.

After completing all three parts of the experiment, your payment for all three parts will be summed up and displayed on the screen in Euro. It will be paid in cash at the end of the experiment. After completing the experiment, please stay in your cubicle until we start paying off. During the payment procedure, please wait in your cabin until you are called to collect your payment individually. No other participant will see how much you have earned. Please bring along all the materials you have received from us to the payment.

Before we start the experiment, we kindly ask you to seal your mobile phone in the envelope with your cubicle number. We will collect the envelopes and will return them back together with your payment at the end of the experiment.

Part 1

In this part, you and the other person in your group can individually solve math problems. You have **30 minutes**. Your income in this part is completely independent of the other participant in your group and depends on how many tasks you solve correctly. The same applies to the other person in your group: The other person's income depends solely on the number of the tasks she will solve correctly and is independent of the number of tasks that you will solve. The tasks which you can solve are addition and subtraction calculations with two three-digit numbers. In subtraction tasks the result can never be negative. If you solve the task correctly, pressing the "Next" button will bring you to the next task. **Please do not use pocket calculators or similar tools.** You and the other person solve the same calculation tasks in the same order. **For each correctly solved task you get 0.05 Euro.**

If you have any questions, please show it with a hand sign.

[20]This is a translation of the original instructions in German which can be obtained from the authors upon request.

Part 2

In this part of the experiment you can again individually solve computing tasks. You have **30 minutes.** If you solved the task correctly, then pressing the "Next" button will bring you to the next task. **Please do not use pocket calculators or similar tools.**

The tasks which you can solve are addition and subtraction calculations. However, Person A will add or subtract two three-digit numbers and Person B will add or subtract two five-digit numbers. Reminder: At the beginning of the experiment, the computer randomly assigned one participant in the group to be Person A and another participant to be Person B. The number of solved tasks in part 1 had no effect on this. **For each correctly solved task, you receive 0.05 Euro, regardless of whether it is three or five-digit numbers.**

At the beginning of Part two, Person A can choose a percentage. The income from the chosen percent of the tasks correctly solved in Part two will be deducted from Person A and transferred to Person B. This means that Person A can decide how much she is willing to give away to another person. All values between 0 and 100 are possible (with the step of 5%). Since it is a percentage, "100" means giving away everything, while "0" means giving away nothing. For example, if Person A chooses "50", that means that Person A transfers to Person B a half of the income from the tasks which she correctly solved in Part 2. Person B learns how much Person A transferred to her only at the end of Part 2. Person B does not learn how many tasks Person A solved or which percent rate she chose. The transfer to another person in the group increases the likelihood that she will be able to participate in Part 3 (more details below).

Part 3

All participants who got more than € 3.20 in Part 2 can participate in Part 3 of the experiment. This means that person A (without the sum transferred to person B) must earn at least € 3.20 in Part 2 in order to participate in Part 3. For person B, this means: if the income from her solved tasks and the transfer payment received from person A together amounts to at least 3.20 Euro, person B can participate in Part 3. If a participant does not participate in Part 3, she must nevertheless stay in the laboratory until the end of Part 3. In Part 3, you can again individually solve calculation tasks. You have **30 minutes**. The tasks you can solve are addition and subtraction calculations of two three-digit numbers (both for Person A and Person B). In subtraction tasks the result can never be negative. If you solve the task correctly, pressing the "Next" button will bring you to the next task. **Please do not use pocket calculators or similar tools.** If both players in the pair participate, they get the same calculation tasks in the same order. **For each correctly solved task, you get 0.10 Euro.**

If you have any questions, please show it with a hand sign.

Examples (displayed on the screen before the donation decision).

For instance, on the next screen you will choose that you want to give 90% of your income to another participant. That means, e.g., if you solve 10 tasks (5 cents per task), then you would keep 5 cents and transfer 45 cents to another participant.

For instance, on the next screen you will choose that you want to give 20% of your income to another participant. That means, e.g., if you solve 10 tasks (5 cents per task), then you would keep 40 cents and transfer 10 cents to another participant.

Appendix A

Summary

Chapter 2: *When to Leave Carrots for Sticks: On the Evolution of Sanctioning Institutions in Open Communities.*

There is substantial evidence that punishment is more effective than rewards to maintain cooperation in social dilemmas. Yet, previous findings suggest that people generally dislike negative sanctioning institutions and avoid them when possible. We take a new perspective by directly comparing the migration between punishment and reward regimes in a social dilemma. In our laboratory experiments participants continuously "vote with their feet" by migrating between punishment and reward communities. In line with previous research, the vast majority of subjects in our experiment opts initially for the reward institution. Over time, however, more subjects start to join the less profitable punishment community. Analyzing the conditions which trigger migration, we show that full contributors are the first to migrate to the non-populated punishment communities. Followers cooperate almost fully without the requirement of actual punishment, thus reducing the welfare costs of the institution. Individual data suggest that income differences within communities and missing compensations for cooperators in the reward community are key factors for the decision to migrate.

Viele Studien zeigen, dass Kooperation in sozialen Dilemmata effektiver durch Bestrafung als durch Belohnungen aufrechterhalten werden kann. Dennoch zeigen frühere Ergebnisse auch, dass Individuen negative Sanktionsmechanismen ablehnen und diese nach Möglichkeit vermeiden. Indem wir in dieser Studie die direkte Migration zwischen Bestrafungs- und Belohnungssystemen in einem sozialen Dilemma vergleichen, können wir eine neue Perspektive in den Diskurs einbringen. In unserem Laborexperiment haben Teilnehmer die Möglichkeit, in einer fortwährenden "Abstimmung mit den Füßen" zwischen Bestrafungs- oder Belohnungssystemen

zu migrieren. Die große Mehrheit der Teilnehmer wählt anfänglich das Belohnungssystem, was sich mit den Ergebnissen vorangegangener Studien deckt. Im Laufe der Zeit wechseln die Teilnehmer unseres Experimentes jedoch zunehmend in das weniger profitable Bestrafungssystem. Bei der Analyse, welche Bedingungen zur Migration führen, zeigt sich, dass vollständig kooperative Spieler die ersten sind, die in die noch „unbesiedelten“ Bestrafungssysteme migrieren. Die nachfolgenden Spieler kooperieren ebenfalls nahezu vollständig, sodass die tatsächliche Anwendung der Bestrafung nicht notwendig ist und die Kosten der Bestrafungsinstitution sinken. Die individuellen Entscheidungen lassen darauf schließen, dass die Migrationsentscheidung hauptsächlich auf Einkommensunterschieden und fehlender Kompensation für kooperative Spieler innerhalb des Belohnungssystems basiert.

Chapter 3: *Costs of Institutions and Cooperation in Open Communities*

On many occasions, people can define which rules will be applied within a certain community (e.g., a team or a firm) and when making such choice they take into account not only how effective the rules are in promoting the desired behavior, but also how costly they are. Yet, players may disproportionally abandon an efficient institution if it involves a fixed provision cost. In the laboratory experiment we study how fixed provision costs for a cooperation-enhancing peer-punishment institution in a public goods game affect the popularity of an institution, rates of cooperation and punishment. We find that in the presence of provision costs the punishment community is overwhelmingly popular with the players choosing it even when it is less profitable than the alternative. By the end of the experiment, players contribute 99% of their endowment in the costly institution community without an increase in the use of punishment points.

Oftmals können Individuen bestimmen, welche Regeln innerhalb eines bestimmten Systems (z.B. eine Arbeitsgruppe oder ein Unternehmen) gelten. Beim Treffen dieser Entscheidung berücksichtigen sie nicht nur, wie effektiv die Regeln erwünschte Verhaltensweisen fördern, sondern auch deren Kosten. Daher könnten Spieler eine effiziente Institution überproportional ablehnen, wenn diese mit fixen Bereitstellungskosten einhergeht. In diesem Laborexperiment wird untersucht, wie fixe Bereitstellungskosten einer kooperationsfördernden Bestrafungsinstitution in einem Öffentlichen-Gut-Spiel die Beliebtheit der Institution sowie Kooperations-

und Bestrafungsraten beeinflusst. Es zeigt sich, dass das Vorhandensein von Bereitstellungskosten zu großer Beliebtheit des Bestrafungssystems bei den Spielern, die es gewählt haben, führt, selbst wenn es weniger profitabel als das Alternativsystem ist. Am Ende des Experimentes tragen Spieler in dem teuren Bestrafungssystem 99% ihrer Ausstattung bei, ohne dass die Anwendung von Bestrafung erhöht wird.

Chapter 4: *Redistribution and Production with the Subsistence Income Constraint: a Real-Effort Experiment.*

A large body of literature demonstrates that redistribution leads to inefficiencies due to distorted work incentives. Yet, this result is typically obtained in environments that leave people absolutely free in their labor-leisure allocation decisions. We challenge this assumption and study labor supply decisions in a framework with a subsistence income constraint and a redistribution system which supports disadvantaged players. Our real-effort experiment offers two sets of results. First, the introduction of a moderate subsistence income threshold leads to a decrease in performance. Second, varying the cause of taxation in this environment does not affect labor responses of taxpayers, but enhances the productivity of welfare recipients if they are the beneficiaries of the redistribution.

Ein Großteil der Literatur zeigt, dass Umverteilung aufgrund verzerrter Arbeitsanreize zu Ineffizienzen führt. Dieses Ergebnis bezieht sich allerdings typischerweise auf Rahmenbedingungen, in denen Individuen vollständig frei über die Allokation von Arbeit und Freizeit entscheiden können. Wir hinterfragen diese Annahme und untersuchen stattdessen Entscheidungen zum Arbeitsangebot in einer Umgebung, die durch ein Existenzminimum und ein Umverteilungssystem zur Unterstützung benachteiligter Spieler charakterisiert ist. Unser Experiment, in dem reale Leistung erbracht werden muss, führt zu zwei Ergebnissen. Zum einen sorgt die Einführung eines moderaten Schwellwertes als Existenzminimum zu einer Verringerung der erbrachten Leistung. Zum anderen bleibt das Arbeitsangebot der Steuerzahler unbeeinflusst, wenn die Verwendung der Steuern variiert wird. In dem Fall, dass die Steuern benachteiligten Spielern zugute kommen, erhöht sich jedoch deren Produktivität.

Chapter 5: *Is Time on Our Side? On the Benefits of Committing to Charities*

In this paper, we consider how the timing of a donation decision relative to the production of disposable income affects charitable giving. In a real effort laboratory experiment with a subsistence income constraint, we study the willingness to donate, the amounts donated as well as the change in labor supply of donors triggered by giving. We find that more people donate if the decision to donate is made before production as a binding pledge. Donors boost their performance, that is driven by their pro-sociality and not by the reference dependence of income. Our results suggest that a commitment to giving is a viable strategy for raising charitable funds.

In dieser Studie untersuchen wir, wie die Spendenbereitschaft von dem zeitlichen Ablauf zwischen der Spendenentscheidung und der Produktion des verfügbaren Einkommens beeinflusst wird. In einem Laborexperiment, in dem ein Existenzminimum modelliert ist und reale Leistung erbracht werden muss, untersuchen wir die Spendenbereitschaft, die Höhe der Spendenbeträge sowie spendeninduzierte Änderungen des Arbeitsangebots. Unsere Ergebnisse zeigen, dass mehr Spieler spenden, wenn die verbindliche Spendenentscheidung vor der Einkommensproduktion getroffen wird. Spender erhöhen ihre Leistung, was durch Prosozialität und nicht durch referenzabhängige Einkommensziele getrieben zu sein scheint. Somit könnten verbindliche Spendenzusagen eine sinnvolle Strategie darstellen, um das Spendenaufkommen für wohltätige Zwecke zu erhöhen.

Appendix B

Backmatter

B.1 Eidesstattliche Versicherung

Ich, Marina Chugunova, versichere an Eides statt, dass ich die Dissertation mit dem Titel: "Institutional Consequences of Justice on Cooperation and Redistribution Systems" selbst und bei einer Zusammenarbeit mit anderen Wissenschaftlerinnen oder Wissenschaftlern gemäß den beigefüugten Darlegungen nach §6 Abs. 3 der Promotionsordnung der Fakultäat Wirtschafts- und Sozialwissenschaften vom 24. August 2010 verfasst habe. Andere als die angegebenen Hilfsmittel habe ich nicht benutzt. Zudem erkläre ich, dass keine kommerzielle Promotionsberatung in Anspruch genommen wurde und dass die vorgelegte Arbeit nicht schon einmal ineinem früheren Promotionsverfahren angenommen oder beurteilt worden ist.

Ort/Datum Unterschrift Doktorandin

B.2 Selbstdeklaration bei kumulativen Promotionen

Für die Kapitel 2 dieser Dissertation liegt die Eigenleistung für

- das Konzept / die Planung bei 50%
- die Durchführung bei 60%
- der Manuskripterstellung bei 60%

Die vorliegende Einschätzung in Prozent über die von mir erbrachte Eigenleistung wurde mit dem am Artikel beteiligten Koautoren, Andreas Nicklisch und Wolfgang Luhan, einvernehmlich abgestimmt.

Für das Kapitel 3 der vorliegenden Dissertation liegt meine Eigenleistung in allen drei Bereichen (Konzept/Planung, Durchführung und Manuskripterstellung) bei 100%.

Für die Kapitel 4 dieser Dissertation liegt die Eigenleistung für

- das Konzept / die Planung bei 50%
- die Durchführung bei 60%
- der Manuskripterstellung bei 60%

Die vorliegende Einschätzung in Prozent über die von mir erbrachte Eigenleistung wurde mit dem am Artikel beteiligten Koautoren, Andreas Nicklisch und Kai-Uwe Schnapp, einvernehmlich abgestimmt.

Für die Kapitel 5 dieser Dissertation liegt die Eigenleistung für

- das Konzept / die Planung bei 50%
- die Durchführung bei 60%
- der Manuskripterstellung bei 60%

Die vorliegende Einschätzung in Prozent über die von mir erbrachte Eigenleistung wurde mit dem am Artikel beteiligten Koautoren, Andreas Nicklisch und Kai-Uwe Schnapp, einvernehmlich abgestimmt.

Ort/Datum Unterschrift Doktorandin