Abuse,

Competition,

Cooperation:

Essays in Behavioral Economics

Elina Khachatryan

Abuse, Competition, Cooperation: Essays in Behavioral Economics

Dissertation zur Erlangung des akademischen Grades
Doktor der Wirtschafts- und Sozialwissenschaften (Dr. rer. pol.)

vorgelegt im Fachbereich Wirtschaftswissenschaften der Universität Kassel
von
Elina Khachatryan

Erstgutachter
Professor Dr. Björn Frank, Universität Kassel

Zweitgutachter
Professor Dr. Astrid Dannenberg, Universität Kassel

Tag der Disputation: 24. Oktober 2016

Acknowledgments

Coming to Kassel to do my PhD in behavioral economics has been the best decision of my life. And of course, first and foremost I have my supervisor Prof. Björn Frank to thank for this! It has been a tremendous honor and pleasure to work with him! His creativity and genuine love for economics have been contagious. Prof. Frank has given me the wings to explore and to experiment. (Pun intended).

I would also like to thank Dr. Christoph Bühren. I could not imagine having a more delightful and supportive colleague. Besides being a wonderful colleague, Christoph is also a mentor and a great friend.

I am very thankful to both, as well as Ms. Heike Krönung, for making the 3rd floor of Nora-Platiel-Straße 4 a very special place for me. A place where I always felt motivated, confident and at ease. A working environment where my creative potential could flourish.

I believe that academic research is a journey! Working towards my PhD for the past four years has been very satisfying and I very much look forward to continuing this fascinating journey. I am thankful to Prof. Astrid Dannenberg for giving me the opportunity to stay directly on the road.

The past four years in Kassel have been enjoyable also because of my wonderful friends, some living here, and some visiting me no matter how busy they get. A special thank you to Rebekka Hüfner and Hripsime Grigoryan for making so many evenings and week-ends so fun.

I would also like to thank my family: my parents Aelita and Stefan Siewert and brother Leon, for loving me and supporting me in all my pursuits. I would like to dedicate this thesis to my mother, who has always been my role model of determination. Thank you for instilling in me this key character from childhood on.

Of course, I must also mention my loyal companion and personal “cheerleader” Keshu, who is supporting me throughout this journey by putting a smile on my face day in and day out.

Zusammenfassung in Deutsch

Diese Arbeit besteht aus vier separaten Essays. In Kapitel 2 wird ein erpresserisches Szenario in die experimentelle Korruptionsliteratur eingeführt. Kapitel 3 betrachtet *displaced aggression* mittels experimentalökonomischer Methoden. In Kapitel 4 wird psychologische Fundierung in die wirtschaftswissenschaftliche Literatur über Geschlechter und Wettbewerbsfähigkeit eingeführt. Eine Lösung für ein bekanntes Problem aus der Literatur zur Kooperation wird in Kapitel 5 vorgeschlagen.

In Kapitel 2, welches den Titel "*Mitigating Extortive Corruption? Experimental Evidence*" trägt, wird eine neue experimentelle Versuchsanordnung eingeführt, in welcher Bottom-Up Ansätze für die Vermeidung eines geringfügig-erpresserischen Szenarios rekonstruiert und erforscht werden. Empfehlungen scheinen besser in einem Umfeld mit persönlichen und regelmäßigen Kontakten abzuschneiden, wohingegen Anzeigen möglicherweise zu Unzufriedenheit und mehr benachteiligter Behandlung seitens der Staatsbeamten führen kann. Im Gegensatz dazu halten Anzeigen und die potentiell daraus resultierenden Sanktionen einen Staatsbeamten eher von einem erpresserischen Verhalten ab, wenn wiederholte Interaktionen nicht vorgesehen sind. Bezüglich der Einbindung von Bürgern in den Überwachungsprozess konnte eine starke Präferenz für Empfehlungen gegenüber Anzeigen festgestellt werden.

Kapitel 3, mit dem Titel "*Letting off Steam! Experimental Evidence on Irrational Punishment*", behandelt ein ökonomisches Experiment, welches *displaced aggression* untersucht. Aggressionen gelten als *displaced* (verlagert), wenn Provokationen nicht direkt geahndet werden und Aggressionen auf ein anderes Ziel, welches frei von jeglichem Fehlverhalten ist, umgeleitet werden. Obwohl dieses Phänomen weit verbreitet ist, wurde es bis dato noch nicht in der experimentellen Ökonomie untersucht. Wir füllen diese Lücke und stellen fest, dass ein beträchtlicher Anteil der Probanden (37%), wenn sie ungerecht behandelt wurden, ihre Mitspieler bestrafen, obwohl diese nicht für die Ungerechtigkeit verantwortlich sind.

Kapitel 4 mit dem Titel "*Does "Negative Nancy" Compete Like "Positive Pete"? An Experiment*" zeigt, wie sich Unterschiede in der Selbstzuschreibungstendenz und des Geschlechts auf das Konkurrenzdenken bei neutralen Aufgaben auswirken. Im Vergleich zu Männern neigen Frauen eher dazu, Verlust ihren mangelnden Fähigkeiten zuzuschreiben, als Pech oder eine schlechte gewählte Strategie als Grund anzugeben. Dennoch scheuen Frauen bei neutralen Aufgaben nicht vor einem Konkurrenzkampf zurück. Selbst sehr selbstkritische Frauen nehmen den Konkurrenzkampf an.

In Kapitel 5 mit dem Titel "*Institution Formation and Cooperation with Heterogeneous Agents*" wird gezeigt, dass Menschen nur dann erfolgreich zusammenarbeiten können, wenn ihre Erwartungen bezüglich Fairness an die Regeln der Institution, die sie festlegen können, angepasst werden. Viele vorangegangene Studien belegen, dass Heterogenität zwischen den Agenten der Zusammenarbeit in der Gruppe schadet. Dieses Experiment bietet eine endogene Institution, die auf die Wahrnehmung der Fairness der Agenten abgestimmt ist und daher trotz Asymmetrien im Nutzen eine hohe Kooperation in der Gruppe wiederherstellen kann.

Table of Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

1. Behavioral and Experimental Economics

Behavioral economics has revolutionized the discipline of economics over the last several decades. It has done so by putting the human back into economics, by recognizing that people sometimes make mistakes in a systematic, predictable way, are reactive to actions of others and are not necessarily as rational as classical economists have traditionally assumed.

Interestingly, before the neoclassical revolution, at the turn of the 20th century, *humans* were still part of the economics discipline. Actually, Camerer and Loewenstein (2003, p.5) clarify that when psychology did not yet exist as a distinct discipline, "[M]any economists moonlighted as the psychologist of their times." The most notable example of course is Adam Smith. His *The Theory of Moral Sentiment* and to a lesser extent his *Wealth of Nations* have insightful commentary on several psychological principles of human behavior. For example:

- On loss-aversion: "Pain…is, in almost all cases, a more pungent sensation than the opposite and correspondent pleasure" (Smith, 1759, p. 176-177).
- On self-control ("present bias"): "The pleasure which we are to enjoy ten years hence, interests us so little in comparison with that which we may enjoy today" (Smith, 1759, p. 273).

- On overconfidence: "The overweening conceit which the greater part of men have of their own abilities" leads them to overestimate their chance of success. (Smith, 1776, p. 1).[1]

Why was then in the ensuing economic literature the *human* substituted by the *homo economicus*, a perfectly rational and self-interested agent? One reason lies in the history of psychology. At the turn of the 20th century, the discipline of psychology was still in its infancy. Prominent economists of the time might not have wanted to associate themselves with this semi-scientific field. Since then, economics has taken up a development trajectory divergent from psychology and the *humans* that come with it. It was first in 1980-s that behavioral economics emerged as a field. Behavioral economics of today manages to combine the best of the two worlds: it is a homage to Adam Smith's "human-oriented" (and realistic) economics, combined with theoretical rigor of the neoclassical tradition.

Laboratory experimentation is the most common method used in behavioral economics.[2] All the four following Chapters of this manuscript are based on laboratory experiments I have conducted, some in collaboration with co-authors.

There are two key differences between experiments in economics and experiments in psychology: 1) monetary and performance-based incentives and 2) no deception of subjects. Early critics of performance-based incentives were concerned that subjects might act in a certain way in the experiment, because they might feel bad using up the experimenter's money. Fortunately, already in 1998, Björn Frank has shown that subjects are indifferent to the fate of experimenter's funds.[3]

In comparison to traditional empirical studies, based purely on real-life observations, in experimental research counterfactual situations can be created for comparison. Laboratory experiments allow a new level of freedom in research, as experimenters can vary one factor at a time between treatments to distinguish between the effects that these different factors may have

[1]These quotes are taken from Richard Thales's (2016) article "Behavioral Economics: Past, Present, and Future," which is based on his presidential address at the annual meeting of the American Economic Association in January 2016.

[2] In the past two decades, behavioral economics has additionally embraced a wider arsenal of methods: from field experiments (e.g. Gneezy and Rustichini, 2000), to computer simulations (e.g. Angeletos et al., 2001), to brain scans (fMRI-s, e.g. McCabe et al., 2001).

[3] Baseline treatment of this experiment was a standard ultimatum game, where cash was substituted with postal stamps. In the experimental treatment, the stamps were burned for the offers that were rejected in the ultimatum. Behavior was very similar in these two treatments.

on behavior. Moreover, experiments seem especially useful for the study of some specific questions, which are particularly difficult to investigate with other methods.

1. Issues where policy proposals are needed.

Experiments are a cheap and easy method for pre-testing (alternative) policies. For example, Chapter 2 seeks to find the most suitable monitoring mechanism for mitigating extortive corruption in different environments. The policy implication of this study is that negative feedback is appropriate for mitigating extortive corruption in urban environments, yet positive feedback is more appropriate for rural environments.

Sometimes the same experimental game can apply to many different real-life situations, where policy advice is needed. For example, the public good game is a workhorse for analyzing a wide range of situations with possible "free-riding". It is also the basis of Chapter 5, which seeks to find an endogenous institution that could restore cooperation in a group, despite the polarizing effects of heterogeneity among the group-mates.

2. Topics that involve clandestine (or even criminal) behavior.

Becker (1968) was first to approach crime as an economic phenomenon. His *Crime and Punishment* defines crime as an optimization problem, where agents decide to commit or not to commit crime, depending on the costs and benefits of this action. Experimental evidence since then has, however, shown that decision-making in regards to deviant behavior is a lot more nuanced. For example, one of the first papers that investigates corruption in the lab shows that surveillance crowds out intrinsic motivation (Schulze and Frank, 2003, p.159): "It is optimal to either monitor with a high frequency or not to monitor at all." Corruption research (see Chapter 2) benefits greatly from economic experiments.

Experimental research on another clandestine behavior – lying, is currently booming. Rosenbaum et al. (2014) provide a meta-study of 63 experiments on lying; 24 of those

(38%) have been published since 2013. Moreover, in 2012 best-selling researcher Dan Ariely published *The (Honest) Truth about Dishonesty,* a 336-page-book that is completely dedicated to experiments on lying. Most of the studies discussed in the book are variations of the Mazar et al.

(2008) experimental game, where subjects can exaggerate the number of quiz questions they have answered correctly so as to get higher payments.

An alternative experimental design for exploring undetectable cheating has been established by Fischbacher and Föllmi-Heusi (2013). In this game participants roll a die in private. Their payoff depends on the number they report that they have rolled, and participants can easily misreport their die roll for monetary gain. Because a fair six-sided die is equally likely to land on any of the six numbers, this experimental design can detect the proportion of liars in the game, without pinpointing the exact individuals, who have lied.

3. Decisions made in the "heat of the moment."

An interesting direction of behavioral research is study of emotion. It might be ethically questionable to put subjects in emotionally charged situations in real life. For example, investigation of displaced aggression (see Chapter 3), if done in real life, might cause somebody actual harm. (Partially, because guarantee of anonymity is very difficult in real life).

Possibly the most curious among studies that investigate how visceral states affect our decision-making is the paper by Dan Ariely and George Loewenstein (2006) that looks at sexual arousal. They find that sexually aroused men are not only more willing to engage in morally questionable behavior to obtain sexual gratification and are more risk-loving in general, but are also unable to predict in advance that their own behavior could be influenced by sexual arousal.

4. Issues that are so complex that it is difficult (maybe even impossible) to study closely just one aspect of the issue with other research methods.

A prime example of this is women's inferior position in the labor force and the gender wage gap. It is difficult to differentiate between the actual differences in women's behavior or preferences and the component of discrimination. Seminal work by Niederle and Vesterlund (2007) has contributed greatly to our understanding of the first component and has launched a new branch of research in experimental economics that looks at gender differences in competitiveness. Chapter 4 of this manuscript adds to this literature by looking at the effect of self-attributions on competitiveness on a gender-neutral task.

5. Topics where social desirability is an issue.

A prime example of this is research on discrimination. There are only few people that think of themselves as racists or sexists, yet racism and sexism are undeniable parts of life. One must study this issue and possible mechanisms for its deterrence without people knowing what is being studied. Otherwise participants might alter their behavior so as to leave a good impression on the researcher. Anderson and Haupert (1999), for example, study how costs of interviewing affect statistical discrimination.[4] When interviewing costs are lower, employers are more likely to make use of interviewing. Once they are better-informed about the candidates, they are less likely to discriminate against individuals that belong to on average less productive group. This study implies that statistical discrimination could be lessened by more easily available individuating information.

Due to availability and convenience, subjects in most laboratory experiments are university students. But for some questions experiments with a specific subject pool can provide an additional insight. A well-known example of this is a field study by Fehr and Leibbrandt (2011). They conduct the public good game with actual fishermen in Brazil. They find that fishermen who contribute more in the public good game experiment are also more likely to have larger holes in their fishing nets. And larger holes are better for the sustainability of the lake as they allow more fish fry to escape the net.

Soldiers have specific professional challenges, and it is particularly interesting to look at how they behave when their desire to retaliate remains unfulfilled. Therefore, the experimental game in Chapter 3 is played with regular students, and additionally also with students who are affiliated with the military.

[4] Subjects in the role of the workers can either have a yellow or a green card. Productivity of the worker is written on the back of the card. Subjects in the role of the employers know the average productivity of the workers with yellow vs. green cards. Before making the hiring decision, employers are given an opportunity for a costly "interview," which in this context means seeing the productivity written on the back of the specific card.

2. A Thesis of "Firsts"

In this section I will discuss the innovative contributions of this manuscript to the economic experimental literature and describe how these ideas came about.

My work has been the "first" in several contexts: 1) First one to introduce an extortive scenario into the experimental corruption literature; 2) First one to look at displaced aggression, using economic experimental methods; 3) First one to introduce psychological underpinnings into the experimental economic literature on gender and competitiveness; 4) First one to offer a solution to a well-known problem in the cooperation literature.

Listed below are co-authorships and the publication status (when applicable) for each of these four projects:

1) The paper in Chapter 2 has been written in collaboration with Sebastian Kube and Björn Vollan. It has been published in 2015 in the ***Journal of Economics and Statistics*** (Special Issue on *Corruption at the Grassroots-level*), Volume 235(2), pages 228-241.
2) The paper in Chapter 3 has been written in collaboration with Christoph Bühren.
3) The paper in Chapter 4 is single-authored.
4) The paper in Chapter 5 has been written in collaboration with Sebastian Kube, Sebastian Schaube and Hannah Schildberg-Hörisch. It has been published in 2015 in the ***European Economic Review***, Volume 78, pages 248-268.

- Chapter 2 ~ First to "Extort"

I have always been interested in understanding the phenomenon of corruption. As I started studying the economic experimental literature on this topic, I was surprised that all of the existing papers (at the time) were about collusive corruption. Possibly collusions are the focus of corruption experiments, because the rule of law is strong in the developed countries, and most researchers exploring this field of research reside in developed countries.

Public officials, in places of strong rule of law, usually do what they are supposed to do; private citizens then usually pay bribes for getting things they are not supposed to get (for some reason, maybe because that hurts someone else). In the developing world, citizens additionally very often pay bribes for things they are supposed to get and are entitled to by law or purchase. But public officials abuse their power and demand additional benefits for the provision of those services.

To me exclusive research focus on collusive corruption seems somewhat privileged. Extortions are a reality for many people in the world. And I personally have had "the honor" of observing "casual" extortions many a time. This is usually done in a polite fashion, with comments such as "I have no time / I am in a hurry!" or "This is a very difficult issue." These comments carry across the message that the public official could do something, if the necessary incentives were in place. But I even have experienced a case of outrageous extortion. Certainly anyone who has a plane ticket, all the necessary paperwork and has arrived to the airport in a timely manner should be able to fly to their destination. But at the Sochi airport (in 2008) this was not the case when one travels with a "rich-looking" American colleague. We were eventually allowed to fly out, but of course all the cash from our wallets had to stay behind.

These experiences served as an inspiration for this chapter. My co-authors and I implemented a novel experimental design that mimics petty extortive corruption and explores bottom-up approaches for its mitigation. In different setups we examine how monitoring by citizens affects public official's tendency to demand bribes and whether citizens are more willing to engage in monitoring if they can recommend rather than report. Our results are mixed. Recommendations seem to perform better in environments with personal and repeated interactions, where reports might cause discontent and further disadvantaged treatment by public officials. In contrast, reports and the sanctions that they potentially cause are more likely to deter public officials from extortive behavior in settings where repeated interaction is not foreseen. Regarding citizens' monitoring involvement, we find a strong preference for recommendations over reports. Moreover, independent of the matching protocol and the endogenous monitoring mechanism, we find that agents in both roles are sensitive to monitoring and detection rate variations: public officials in their decision to demand a bribe and citizens in their decision to monitor.

- Chapter 3 ~ First "Unfulfilled," when Punishing

Economic experimental literature is rich with studies that explore how punishment, or possibility of punishment, can affect decision-making, and whether possibility of punishment is efficient in reaching better outcomes.[5]

To the best of our knowledge, however, there has been no other economic experiment so far where the potentially frustration-provoking agent is impune to punishment. The question then is: do other people pay for his/her "sins"? I think any of us can think of an example where we have observed an outburst against someone who was just "at the wrong place at the wrong time." Therefore, it is worthwhile to investigate not only "rational" punishment, in the sense that punishment could discipline a free-rider or threaten one into cooperation, but it is also important to study the purely irrational punishment.

Chapter 3 is an economic experiment dedicated to the study of displaced aggression. Aggression is displaced, when provocations cannot be directly retaliated against, and aggression is redirected towards a target innocent of any wrongdoing. While this phenomenon is widespread, as yet it had not been explored in experimental economics. We fill this gap and find that a sizeable proportion of subjects (37%), when treated unfairly, punish co-players who are not at all responsible for the unfairness. When in a disadvantaged position, inequity-aversion seems to be the driving force of punishment, yet when treated fairly, some subjects (17%) exhibit status-seeking behavior. Moreover, students affiliated with an armed forces university are much more likely than regular students to engage in displaced aggression.

- Chapter 4 ~ First to "Self-Attribute," when Competing

I often read articles from social psychology for pleasure. One day I came upon the paper "He's Skilled, She's Lucky" by Swim and Sanna (1996).

They show that observed professional success achieved by men is attributed to skill and talent, whereas equivalent success achieved by women is attributed to luck. These results are anything, but uplifting, and clearly, this issue demands further research: Do women internalize prejudiced

[5] Actually Chapter 2 of this manuscript does exactly that. We check how the possibility for punishment fares in comparison to the possibility for recommendation in extortion mitigation.

attributions of the society, and could this possibly explain the difference in competitiveness observed in the experimental economic literature on gender and competitiveness (as in Niederle and Vesterlund, 2007).

To investigate these questions, I adopted tools from social psychology. The aim of the study in Chapter 4 is to bridge the gap between "gender and competitiveness" literature in experimental economics and "gender and self-attribution" literature in social psychology. Chapter 4 reports how differences in self-attribution tendencies and gender affect competitiveness on a neutral task. Compared to men, women tend to ascribe losses to deficient skills rather than bad luck or a poorly chosen strategy. Nevertheless, women do not shy away from competition on a neutral task. Even highly self-critical women embrace competition.

Moreover, in the economic literature on the topic, it is established that women "shy away" from winner-takes-it-all tournaments when the alternative payment method is a piece rate (i.e. getting a certain amount for sure per correct answer). My paper instead investigates preferences for a winner-takes-all tournaments, when the alternative payment method is a lottery.

- Chapter 5 ~ First to Offer a Resolution to the Polarizing Effects of Heterogeneity, when Cooperating.

Imagine working on a project in a team with two other people. But these two group-mates benefit much more from the commonly produced good than you do. Will you

keep working on this project? Probably not. Given that you are shirking, will your group-mates continue to put in effort? Probably not, as well. "Free-riding" by one player, even if justified by unfairness in benefits, is discouraging. Moreover, lower average contribution level makes continued investment less worthwhile even for agents that have relatively higher benefits from the public good.

Now imagine that this group of people has been given an opportunity to establish a set of rules that they commit to. Rules that could help them improve the cooperation level in the group and rip larger benefits from this team project. What set of rules should they agree on? Chapter 5 answers precisely that question.

We observe social preferences to be limiting the scope for institution formation: This experiment shows that people can cooperate successfully only when their expectations of fairness are aligned with the rules of the institution that they can establish.

Many previous experimental studies have shown that heterogeneity among agents distorts cooperation in the group. This experiment offers an endogenous institution that matches people's perception of fairness and therefore, despite asymmetries in benefits, can restore high levels of cooperation in the group. Inequality-averse players frequently object to institutions that fail to address differences in players' benefits from cooperation - even if rejecting the institution causes monetary losses to all players. Relating our findings to previous studies on institution formation, we discuss potential advantages and drawbacks of stipulating unanimous support for implementing institutions that foster cooperation.

References

Anderson, Donna M.; Haupert, Michael J. (1999): Employment and Statistical Discrimination: A Hands-on Experiment. In *The Journal of Economics* XXV (1), pp. 85–103.

Angeletos, George-Marios; Laibson, David; Repetto, Andrea; Tobacman, Jeremy; Weinberg, Stephen (2001): The Hyperbolic Consumption Model. Calibration, Simulation, and Empirical Evaluation. In *Journal of Economic Perspectives* 15 (3), pp. 47–68. DOI: 10.1257/jep.15.3.47.

Ariely, Dan (2012): The (honest) truth about dishonesty. How we lie to everyone--especially ourselves. First edition. New York: Harper.

Ariely, Dan; Loewenstein, George (2006): The heat of the moment. The effect of sexual arousal on sexual decision making. In *J. Behav. Decis. Making* 19 (2), pp. 87–98. DOI: 10.1002/bdm.501.

Becker, Gary S. (1968): Crime and Punishment. An Economic Approach. In *Journal of Political Economy* 76 (2), pp. 169–217. DOI: 10.1086/259394.

Camerer, Colin; Loewenstein, George (2003): Behavioral Economics: Past, Present, Future. In Colin Camerer, George Loewenstein, Matthew Rabin (Eds.): Advances in behavioral economics. Princeton, N.J., Woodstock: Princeton University Press (The Roundtable series in behavioral economics).

Fehr, Ernst; Leibbrandt, Andreas (2011): A field study on cooperativeness and impatience in the Tragedy of the Commons. In *Journal of Public Economics* 95 (9-10), pp. 1144–1155. DOI: 10.1016/j.jpubeco.2011.05.013.

Fischbacher, Urs; Föllmi-Heusi, Franziska (2013): Lies in Disguise - An Experimental Study on Cheating. In *Journal of the European Economic Association* 11 (3), pp. 525–547. DOI: 10.1111/jeea.12014.

Frank, Björn (1998): Good news for experimenters. Subjects do not care about your welfare. In *Economics Letters* 61 (2), pp. 171–174. DOI: 10.1016/S0165-1765(98)00162-1.

Gneezy, Uri; Rustichini, Aldo (2000): A Fine is a Price. In *The Journal of Legal Studies* 29 (1), pp. 1–17. DOI: 10.1086/468061.

Kube, Sebastian; Schaube, Sebastian; Schildberg-Hörisch, Hannah; Khachatryan, Elina (2015): Institution formation and cooperation with heterogeneous agents. In *European Economic Review* 78, pp. 248–268. DOI: 10.1016/j.euroecorev.2015.06.004.

Mazar, Nina; Amir, On; Ariely, Dan (2008): The Dishonesty of Honest People. A Theory of Self-Concept Maintenance. In *Journal of Marketing Research* 45 (6), pp. 633–644. DOI: 10.1509/jmkr.45.6.633.

McCabe, K.; Houser, D.; Ryan, L.; Smith, V.; Trouard, T. (2001): A functional imaging study of cooperation in two-person reciprocal exchange. In *Proceedings of the National Academy of Sciences of the United States of America* 98 (20), pp. 11832–11835. DOI: 10.1073/pnas.211415698.

Niederle, M.; Vesterlund, L. (2007): Do Women Shy Away from Competition? Do Men Compete Too Much? In *The Quarterly Journal of Economics* 122 (3), pp. 1067–1101. DOI: 10.1162/qjec.122.3.1067.

Rosenbaum, Stephen Mark; Billinger, Stephan; Stieglitz, Nils (2014): Let's be honest. A review of experimental evidence of honesty and truth-telling. In *Journal of Economic Psychology* 45, pp. 181–196. DOI: 10.1016/j.joep.2014.10.002.

Schulze, Günther; Frank, Björn (2003): Deterrence Versus Intrinsic Motivation: Experimental Evidence on the Determinants of Corruptibility. In *Economics of Governance* 4 (2), pp. 143–160.

Smith, Adam (1759): The Theory of Moral Sentiments.

Smith, Adam (1776): An Inquiry into the Nature and Causes of the Wealth of Nations.

Swim, J. K.; Sanna, L. J. (1996): He's Skilled, She's Lucky. A Meta-Analysis of Observers' Attributions for Women's and Men's Successes and Failures. In *Personality and Social Psychology Bulletin* 22 (5), pp. 507–519. DOI: 10.1177/0146167296225008.

Thaler, Richard H. (2016): Behavioral Economics. Past, Present, and Future. In *American Economic Review* 106 (7), pp. 1577–1600. DOI: 10.1257/aer.106.7.1577.

CHAPTER 2

Mitigating Extortive Corruption? Experimental Evidence

Abstract

Extortive petty corruption takes place when a public official elicits small bribes from citizens for providing public services that the citizens are legally entitled to receive. We implement a novel experimental design that mimics this phenomenon and explores bottom-up approaches for its mitigation. In different setups we examine how monitoring by citizens affects public official's tendency to demand bribes and whether citizens are more willing to engage in monitoring if they can recommend rather than report. Our results are mixed. Recommendations seem to perform better in environments with personal and repeated interactions, where reports might cause discontent and further disadvantaged treatment by public officials. In contrast, reports and the sanctions these potentially cause are more likely to deter public officials from extortive behavior in settings similar to the stranger matching protocol. Regarding citizen's monitoring involvement, we find a strong preference for recommendations over reports, even among stranger-matching treatments. Moreover, independent of the matching protocol and the endogenous monitoring mechanism, we find that agents in both roles are sensitive to monitoring and detection rate variations: public officials in their decision to demand a bribe and citizens in their decision to monitor.

1. Introduction

Corruption is a widespread phenomenon in many countries. Its detrimental effects have been documented in a number of empirical studies.[6] During recent years, corresponding evidence also stems from laboratory experiments. The convenience of a controlled lab environment not only allows to identify causal relationships, but at the same time provides a testbed for mechanisms that are meant to fight corruption. Most of the existing (experimental) literature on corruption concentrates on collusive corruption, i.e., corrupt interactions where public officials are paid to implement actions that involve negative externalities on third parties.[7]Our paper complements this literature by exploring extortive corruption rather than collusive corruption. Using laboratory experiments, we try to shed light on the effectiveness of potential mechanisms for mitigating extortive corruption.

There seem to be two main features that differentiate extortive corruption from collusive corruption. Firstly, in extortive corruption, the public official is the one who initiates the transaction by refusing to provide a service and demanding additional pay from the private citizen for providing this service in a proper and timely manner. Secondly, extortive corruption generally entails extortion of the private citizen, where the latter is forced to enter the corrupt transaction due to threats from the public official, whereas in collusive bribery both sides tend to willingly choose to engage in a corrupt transaction. According to Transparency International, in 2011 at least 40% of all bribes paid tended to be of extortive nature.[8] In particular in developing countries, the general public faces extortive petty corruption in their everyday life. It majorly affects their access to public services, including such basics as health care, education, police and jurisprudence. At the same time, mitigating extortive corruption turns out to be difficult for at least two reasons. Classical solutions for mitigating (not only) the problems of corruption are based on law and order

[6]For example, corruption in developing countries accumulates into impediments to economic growth and development (Fisman/Svensson 2007; Mauro 1995; Méon/Sekkat 2005; Knack/Keefer 1995). Moreover, there seems to be an unfavorable correlation between corruption and inequality and poverty (Gupta et al. 2002; Olken 2006), as well as health, education and public infrastructure (Reinikka/Svensson 2004; Davoodi/Tansi 1997).
[7] For a comprehensive review of literature on corruption, see Bobkova and Egbert (2012), Abbink and Serra (2012), Abbink (2006), Dusek et al. (2005), Andvig (2005), Renner (2004).
[8] When asked about their last bribe, 40% of citizens report having paid it for either "receiving a service entitled to" or "avoiding a problem with the authorities"; 30% report having paid their last bribe to "speed things up," and another 30% are unclear, where citizens state either "don't know" or "don't remember" (Transparency International, 2011 Global Corruption Barometer).

approaches: deterrence through exogenous monitoring and fines. For the case of extortive corruption, however, detecting and verifying offenses is often difficult. If citizens pay the bribe, the services provided by the public official are legal such that one cannot infer infringements from this action. If citizens do not pay the bribe, it is again difficult to infer corruptive behavior from the public official's action since harm to the citizen is rather subtle. Usually, speed and quality of the services provided are at the public official's discretion so that it is not straightforward how to prove that low quality and delays in providing the service are intentionally meant to increase the pressure on the citizen to pay the demanded bribe.

One might think that a bottom-up approach could help, i.e., relying on endogenous monitoring by the citizens themselves.[9] If they encounter a corrupt official, they report him to initiate an investigation which might lead to a fine if misconduct in office is proved (e.g., Serra 2012). While this could increase initial detection rates, the difficulty of actually verifying the misconduct remains. We use the term endogenous to signify private citizen's involvement into the monitoring mechanism, by providing monitoring authorities with information regarding bribe-demanding tendencies of the public officials the private citizen has interacted with. In contrast, exogenous monitoring (defined similarly in Serra 2012) would mean randomly generated monitoring by the authorities, independent of information from private citizens. Furthermore, in corruption-entrenched regions anonymity of the reporting citizens is questionable due to weaknesses of infrastructure and general lawlessness (see also Abbink et al. 2012). But without guarantees of anonymity or protection - and because provided services are at the discretion of public officials and citizen's access to the services is limited by his relationship with public officials - endogenous reporting might not take place due to a fear of retaliation. This is particularly true if citizens expect to approach the same official (e.g., nurse or policeman) on multiple occasions.

Instead of having endogenous reporting that potentially leads to a fine for the public official, it might be an option to instead use endogenous recommendations for non-corrupt behavior that potentially lead to a monetary bonus for the public official.[10] This might be appealing for two

[9] For example, Rose-Ackerman and Truex (2012) advocate the need and potential benefits of endogenous monitoring by private citizens: "Citizens have an interest in fighting corruption, and if given a voice, they can be a potent force for its reduction."

[10] As Batory (2012) concludes in her study of currently failing anti-corruption laws in Central Eastern Europe, "anti-corruption interventions should pay more attention to raising awareness among target groups, take existing social norms into account, and rely on positive incentives as well as, or rather than, increasing penalties."

reasons. First it might foster citizens' participation in the endogenous mechanism since i) they clearly do not face the threat of retaliation after a recommendation, and ii) they might expect no (or less) retaliation if they refrain from recommending a public official (simply because ‘‘being reported’’ is clearly a negative signal of intentions while the reasons behind ‘‘not being recommended’’ might be ambiguous and thus perceived differently by the public official). Second, if officials actually react to the size of the monetary bonus, the increase in citizens' endogenous participation in the monitoring process increases public official's marginal incentives to being fair, law-abiding, hard-working and non-corrupt.

To test for the relative effectiveness of endogenous reporting and recommendation to mitigate extortive corruption, we conducted stylized laboratory experiments. Subjects were either in the role of the public official (PO) or the citizen. Officials could ask for a bribe. The citizen could then accept or refuse to pay the bribe. Afterwards, the official had to choose between providing a regular service (high payoff to the citizen) or reduced service (low payoff to the citizen). Depending on the treatment, citizens could then report a bribe-asking official or recommend a law-abiding official. An official who was reported (recommended) then had to pay a fine (could receive a bonus) with a certain probability. Within treatment, we varied the size of this probability. Between treatments, we varied the endogenous mechanism (recommend or report) as well as the matching scheme (partner or stranger). Reporting and recommending mechanisms are each played under two different matching protocols, either using a partner-matching or a stranger-matching protocol; i.e., repetitive interaction within the same pair throughout the ten periods (partner matching) or multiple one-time interactions where subjects are randomly re-matched to a different PO or citizen every period (stranger matching). Under both matching protocols, roles are kept constant throughout the experiment.

We observe that in both matching protocols, more bribes are asked under a recommendation mechanism than under a reporting mechanism. If a bribe is asked for, citizens are more likely to pay the bribe under a reporting rather than a recommendation mechanism only in the partner matching. In the stranger protocol, the picture is reversed, meaning that citizens are less likely to pay the demanded bribe under a reporting rather than a recommendation mechanism. Independent of the matching protocol, citizens' willingness to make use of the endogenous mechanism is higher in recommendation than in reporting.

To the best of our knowledge, we are the first to experimentally test the effectiveness of an endogenous recommendation mechanism for mitigating extortive corruption. Most closely related to the paper at hand are the lab experiments on collusive bribes by Abbink (2004) and Serra (2012). Abbink (2004) shows that rotation of interacting pairs significantly reduces the levels of bribes as well as the frequency of inefficient decisions due to bribery. In contrast to our study, Abbink (2004) features a collusive bribery design whereas we are considering an extortive bribery design. Moreover, Abbink (2004) is not studying endogenous reporting mechanisms. This is done in Serra (2012). She analyzes endogenous as well as exogenous reporting in a one-shot experimental setting. Yet again, her design considers collusive corruption where bribes and collusive behavior cause a negative externality on bystanding passive agents. She finds that in the endogenous monitoring treatment less and lower bribes are asked than in the exogenous monitoring treatment, although the cumulative detection rate is significantly lower in the endogenous monitoring treatment.

The outline of the paper is as follows. Sections 2.1 and 2.2 describe the experimental design and procedures. Section 3 presents the empirical findings and Section 4 concludes.

2. Experiment

2.1. Experimental Design

The experimental design is kept very simple to allow for clear and causal identification of behavioral patterns. It is meant to mimic basic features of extortive (petty) corruption (without claiming to be exhaustive): i) the presence of a public official (PO in the following) and a citizen, ii) the possibility to demand a bribe, and iii) PO's discretion over the service that is provided to the citizen. Given that we are interested in endogenous monitoring, we add iv) the option to report a corrupt PO, respectively to recommend a non-corrupt PO. The sequential move game is depicted in the game trees below and essentially proceeds as follows. The PO is the first mover. He decides whether or not to demand a bribe. The size of the bribe is fixed to 10. In case a bribe is requested, the citizen can decide whether or not to actually pay the bribe.

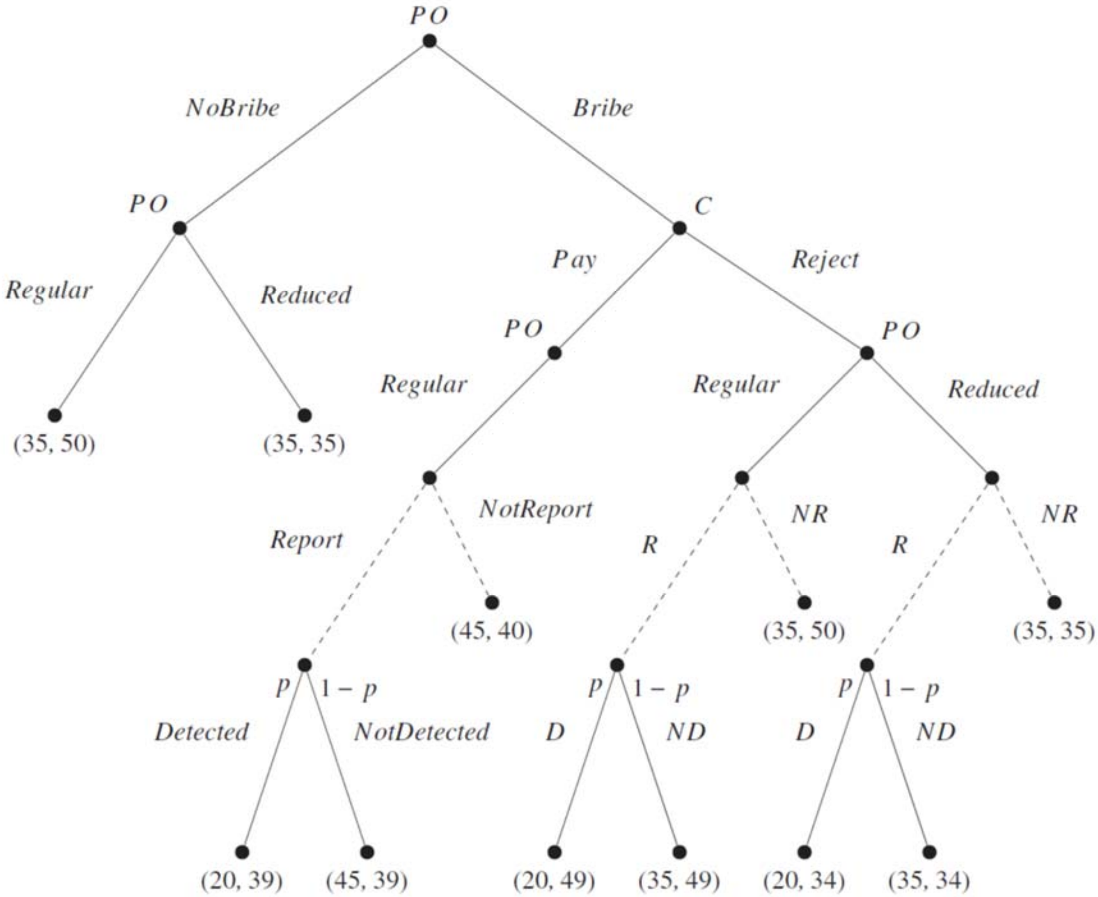
Afterwards, the PO can either provide a reduced service (low payoff of 35 to the citizen) or a regular service (high payoff of 50 to the citizen) at no additional costs for him.[11] Finally, the citizen takes his endogenous monitoring decision.

Depending on the treatment, endogenous monitoring either takes the form of reporting corrupt behavior or recommending non-corrupt behavior. In treatment Report, the citizen is provided with a choice to report a PO if the latter has asked for a bribe, no matter whether or not the bribe has been paid. Reporting costs 1 for the citizen (meant to mimic transaction costs) and initiates an investigation of PO's actions. With a certain detection probability p, PO's corrupt behavior is proved, he loses the bribe that he has received and additionally must pay a punishment fee of 15. As a result, PO's payoff is 20 for that round, no matter whether or not the bribe has been paid. With $1 - p$, the PO is not found guilty of corruption and receives a period payoff of 35 if he did not receive a bribe, respectively 45 if he received a bribe.

In treatment Recommend, citizens are provided with a choice to recommend a PO if the latter has not asked for a bribe. Recommendation also implies costs of 1 for the citizen. A recommended PO has the chance to receive a bonus of 15 with probability p in which case he earns a total of 50 in that round. To keep both feedback mechanisms as comparable as possible, $p \in [.05, .1, .2]$ is the same for both treatments. The dashed lines in Figure 1 show the additionally available move of reporting a corrupt PO, whereas dashed lines in Figure 2 show the additionally available move of recommending a non-corrupt PO. Afterwards, a period ends and the next period starts. At the end of each period, players learn their payoff, and public officials learn if they have been reported or recommended by the citizen, and if so, respectively if they have been fined or rewarded with a bonus. The game is repeated for ten periods in each part.

[11] The game is simplified such that in case a bribe was requested and paid, the PO has no choice but to provide the regular service. This was meant to mimic naturally occurring situations where the citizen requests to receive the service he desires immediately after paying the bribe.

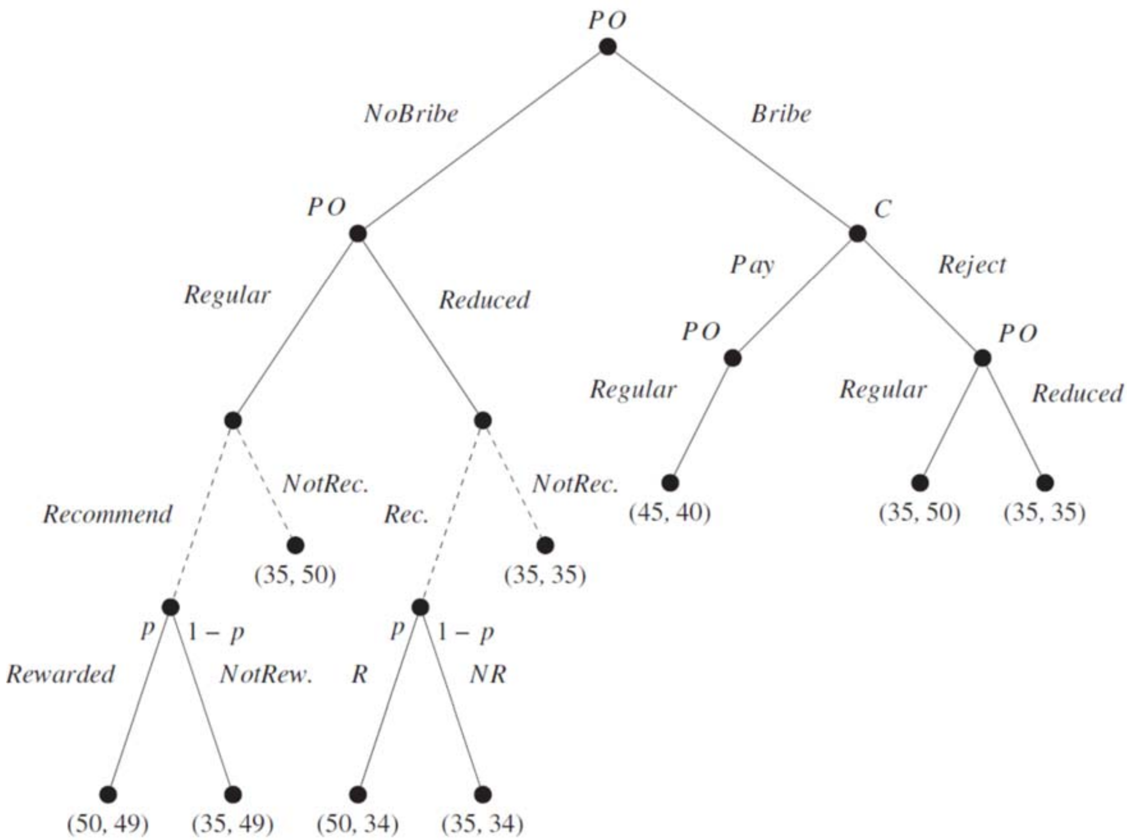
Figure 1: Game Tree for Report Treatments



The 2x2-design leaves us with a total of four treatments. Through cross-treatment comparisons, we can thus examine the effectiveness of the two endogenous monitoring mechanisms in deterring corruption (by comparing how often POs ask for bribes) and engaging public into civil affairs (by comparing how often citizens report, respectively recommend, the PO). Moreover, we can do so under two potentially relevant interaction patterns. Stranger resembles cases where public administration is rather large (e.g., in a large city) or where staff rotation is in place. Partner is more representative of situations where citizens interact continuously with the same officials (e.g., in a small village) which might accentuate reciprocity and fear of retaliation due to the repeated interaction. To keep both monitoring mechanisms comparable, the order of variations in detection rate *p* (from 10% to 5% to 20%) is held constant across all treatments. We vary *p* between the three different parts of the game, where each part consists of 10 periods. Hence, first

10 periods of the game are played with the medium 10% detection rate, second 10 periods with the low 5% detection rate, and last 10 periods of the game with relatively high 20% detection rate. Subjects learn about the change in detection rate only at the beginning of each part. We change the rate of p in different parts of the game to additionally test if behavior of the agents reacts to changes in the detection rate in the Report treatments, respectively changes in the probability of receiving a reward in the Recommend treatments.

Figure 2: Game Tree for Recommend Treatments



2.2. Experimental Procedure

The experiments were run in August 2012 at the BonnEconLab at the University of Bonn, Germany. Subjects were randomly recruited through ORSEE from a pool of about 6000 people, consisting mostly of undergraduate students from the University of Bonn. In total, we had 96 inexperienced subjects, who, according to their ORSEE (Greiner, 2004) record on previous participation at the BonnEconLab, have not taken part in another economic experiment exploring corruption or crime in general. 53% of our subjects were women, subjects had different backgrounds and majors (e.g., Economics, Law, Politics, Physics, Math, Computer Science, Biology, Psychology, Nutrition Science, etc). Subjects were randomly allocated into one of four treatments (24 subjects per treatment).

The experiment was programmed and conducted using z-Tree (Fischbacher, 2007). Interaction occurred anonymously over computers and consisted of playing 30 identically designed periods. In each session we had 24 subjects. Subjects were randomly assigned to a role (PO or citizen) that was kept throughout the experiment. We used loaded instructions featuring expressions such as bribe and public official to ease understanding of the game and payoff structure and to make the normative implications of actions more salient.

Before the start of the experiment, subjects received written instructions. Instructions were read out loud to the subjects. Afterwards, clarifying questions could be posed and control questions had to be solved. Subjects had no problems solving the control questions. Subjects were informed that the experiment consists of three parts at 10 periods each, but only learned at the beginning of each part about the actual parameters being used in that part. Initially, the probability p of being fined, respectively rewarded, was 10% (periods 1-10). At the beginning of the second part, subjects were then informed that p is lowered to 5% in the second part of the game (periods 11-20). At the beginning of the third part, subjects were then informed that p is increased to 20% (periods 21-30). In the two treatments with partner matching, subjects remain in the same pair throughout the first part of the game with no role reversal, then get re-matched to different partners in the second and third parts of the game.

For each part, one period was randomly chosen to be paid out anonymously in cash to the subjects immediately after the session. During the experiment, all payments were in tokens (Taler); with 6.5 Taler = 1 Euro. Total earnings ranged between 15.40 Euros and 21.40 Euros, with an average payment of 18.70 Euros. Sessions lasted for less than two hours.

2.3. Behavioral Predictions

Under standard assumptions of fully rational, money-maximizing behavior, equilibrium predictions do not differ between treatments. In the subgame-perfect Nash equilibria, citizens will neither report nor recommend POs since it is costly to them. Demanded bribes will only be paid by the citizen under the assumption that POs offer reduced services whenever they are indifferent between offering reduced and regular services. In that case, POs will demand a bribe and resulting payoffs will be 45 to the PO and 40 to the citizen. If instead POs are assumed to offer regular service whenever they are indifferent, the payoffs will be 35 to the PO and 50 to the citizen (either a result of no bribe followed by regular service, or asking for a bribe, being rejected and then offering regular service). Given previous evidence (not only) from social punishment in social dilemmas (e.g., Fehr and Gächter, 2000), we know that a significant fraction of people act reciprocally, i.e., are willing to incur costs in order to punish bad behavior or to reward good behavior. If these reciprocity considerations are sufficiently strong, new equilibria might arise. In Recommend treatments, citizens might refuse to pay bribes and/or might recommend regular services; and POs anticipating this might be better off when not asking for a bribe and granting regular service. In Report treatments, citizens might refuse to pay bribes and/or might report corrupt behavior; and POs anticipating this might refrain from bribing and provide regular or reduced service.

Depending on (PO's beliefs about) citizens' actual strength of reciprocity considerations, treatment differences between Recommend and Report might arise. In Recommend, the only way to get back at corrupt POs is by not paying the bribe, which comes at the risk of high costs to the citizen; namely that the PO does only provide reduced service afterwards. POs who expect citizens' reciprocity considerations to be small might thus still ask for a bribe, which would guarantee them at least a payoff of 35. Contrarily in Report, citizens have two options to get back at corrupt POs: not paying the bribe, and reporting the PO. The latter is very cheap so that POs should anticipate to be reported, even when they expect citizens' reciprocity considerations to be small. Thus, POs asking for a bribe run the risk of being fined, in which case they would earn less than the 35 tokens that they can guarantee themselves by not asking for a bribe. Sufficiently loss-averse and/or risk-averse POs might now refrain from bribing.

Reciprocity considerations are likely to interact with the matching protocol, in particular because i) citizens in Partner treatments can "invest" in their reputation while they cannot do so in Stranger treatments, and because consequently ii) POs learn more about citizens' reputation in Partner than in Stranger treatments. So if citizens do invest in their reputation, which implies rejecting bribe demands in Recommend and/or reporting corrupt behavior in Report, we should not expect to see a difference in the frequency of bribes demanded between Report Partner and Recommend Partner (though we might still expect a difference between Report Stranger and Recommend Stranger since reputation building is limited under stranger matching); and for a given type of feedback (report/recommend) we should expect less corrupt behavior under partner than under stranger matching. More- over, since citizens in Partner Recommend can only invest in their reputation by refusing to pay bribes, we might expect to observe more refusals under Partner Recommend than under Partner Report. With respect to the quality provided, however, effects are unclear since POs might want to invest in their reputation, too – which would imply to retaliate by offering reduced services.

3. Results

In this section, we focus on reporting our regression results. All descriptives can be found in Figure 3 at the end of this chapter. Table 1 reports the main results of our experiment with six different regressions separated for partner and stranger matching. We are mainly interested in explaining factors that reduce tendency for demanded bribes but also explore changes in the level of quality provision after the interaction between the public official and the citizen occurs, depending on whether the bribe has been paid or not by the extorted private citizen. The major explanatory variable of interest is Recommend which is indicated by a dummy variable for the sessions where the private citizen was able to recommend a public official. The reference category is formed by those cases where participants could report a public official. We further include variables for the low and

medium detection rates of 5% and 10% in our experiment with the reference category of relatively high detection rate of 20% to see how decrease in detection rate might affect important factors underlying corrupt transactions. The variables period and endgame control for potential learning effects over time.

Table 1: Marginal effects after random effect probit model

	Partner			Stranger		
	Bribe asked	Paid the bribe	Quality offered	Bribe asked	Paid the bribe	Quality offered
recommend	0.228	-0.639**	-0.344	0.440***	0.503***	0.333***
	(0.233)	(0.294)	(0.250)	(0.144)	(0.118)	(0.129)
detection 5%	0.369	0.954**	0.965***	0.919**	0.250	-0.062
	(0.323)	(0.427)	(0.342)	(0.390)	(0.336)	(0.323)
detection 10%	0.049	1.481*	1.385**	0.340	0.318	-0.207
	(0.585)	(0.772)	(0.616)	(0.680)	(0.614)	(0.583)
period	-0.005	0.084**	0.068**	0.012	0.007	-0.007
	(0.028)	(0.037)	(0.029)	(0.032)	(0.029)	(0.028)
endgame	-0.202	0.028	-0.248	-0.236	0.129	0.384
	(0.248)	(0.307)	(0.250)	(0.280)	(0.251)	(0.249)
constant	0.271	-1.349	-1.072	0.289	-0.258	0.589
	(0.757)	(0.989)	(0.792)	(0.867)	(0.778)	(0.741)
observations	576	376	576	576	486	576
ll	-347.3	-206.5	-320.0	-232.5	-311.3	-337.7
chi2	11.73	14.44	11.28	28.43	20.97	9.55

Notes: Standard errors (clustered on matching groups) in parentheses. When analyzing *Paid the bribe*, we restrict our sample to these cases when a bribe was asked. Levels of significances are denoted by ***$p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

In the stranger matching we see that in Recommend more bribes are asked than in Report. In the Stranger treatment, where the private citizen does not feel as deceived by a demanded bribe as in the more intimate Partner treatment, the demanded bribes are also paid out more often which leads to higher quality provision by the public official. While one could argue that this is a second-best situation for the private citizen, it still leads to systematic corruption which is likely to be an undesired outcome for society in the long-run. It also carries emotional costs, since people are likely to dislike being extorted and might find this experience humiliating. On theother hand, better quality provision could be due to less retaliation in the recommend treatments, which diminishes the escalating effects that usually accompany punishment.

Table 2: Marginal effects after random effect probit model for endogenous monitoring: reporting or recommending

	Endogenous Monitoring	
	Partner	Stranger
recommend	0.918***	0.572**
	(0.334)	(0.248)
detection 5%	-0.395	-0.807*
	(0.492)	(0.466)
detection 10%	0.336	-0.886
	(0.889)	(0.838)
period	0.034	-0.035
	(0.042)	(0.040)
endgame	0.201	-0.227
	(0.386)	(0.379)
constant	-0.862	0.691
	(1.143)	(1.056)
observations	268	260
ll	-158.4	-160.6
chi2	19.24	14.92

Notes: Standard errors in parentheses. The dependent variable *Endogenous Monitoring* equals 1 if the citizen reports the PO in treatments Report, respectively recommends the PO in treatments Recommend - and is 0 otherwise. Levels of significances are denoted by *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

In interactions between the same officials and citizens over time (Partner), we do not observe significant differences between Recommend and Report with respect to the number of instances where POs demand a bribe. Also the quality provided does not differ significantly between Recommend and Report. However, we observe that when people can recommend instead of report, they are less likely to pay the demanded bribe. Despite the small costs for recommending, some citizens might have been willing to recommend a fair public official every round, so as to reach the efficient outcome of the game. Thus, when a PO demands a bribe it implies that the efficient outcome cannot be reached, which might induce negative emotions on behalf of the citizen. Moreover, demanding a bribe in Recommend might also be perceived as a signal of PO's distrust that the citizen would continuously recommend the PO. For both reasons, the citizen might be disappointed such that he would refuse to pay a demanded bribe.[12]

[12] In regressions that shed more light on the determinants of citizens' decision to pay the bribe (available upon request), we observe that the willingness to pay the bribe significantly decreases in the number of previous instances of POs demanding a bribe; but only in Recommend Partner. Interpreting this measure as citizen's current amount of

Another key result is that higher rates of being fined after a report, respectively rewarded after a recommendation, lead to less bribes being demanded. This holds in Stranger treatments, whereas more subtle aspects, such as reciprocity and retaliation seem to play a more important role in the Partner treatments. Here public officials might observe in the first periods that the private citizen does not tend to report them and hence can continue asking for bribes with more certainty, in stranger-matching public officials are more sensitive to higher detection rates, since they do not have a chance to learn reporting preferences of the private citizens they work with, and hence must be more careful in their interactions.

As expected, we see in Table 2 that a much larger proportion of subjects makes use of positive endogenous feedback, rather than negative endogenous feedback. This shows that if treated fairly, citizens are willing to pay an extra cost to reward public officials for their good work ethic. Moreover we see that citizens do not treat their own reports and recommendations as pure symbolic signals of disappointment or praise for the PO, but also care about the financial consequences that these carry. So that significantly less endogenous feedback takes place with detection rate of 5% versus 20%.

4. Conclusion

Extortive corruption can often be observed when public officials are given discretion on service provision and can thereby manipulate the private citizens to pay bribes for a service that they are legally entitled to receive. In this paper — and contrary to the existing experimental literature on corruption that focuses on collusive corruption — we made a first step towards a better understanding of extortive corruption using a novel experimental design. We were particularly interested in bottom-up monitoring, i.e., endogenous reports by the citizens. In different setups we examined if endogenous monitoring affects public official's tendency to demand bribes and whether citizens are more willing to engage in monitoring if it constitutes recommendations rather than reports.

disappointment, we take it as supportive evidence for disappointment being of importance in Recommend, but not in Report. Moreover, we do not observe significant effects of the cumulated lag in stranger interactions –suggesting that disappointment seems to be particularly strong in close and intense relationships, i.e., when being exploited repeatedly by the same PO instead of being repeatedly exploited by different POs.

Our results are mixed. Whereas in Abbink (2004) stranger matching leads to less collusive bribery, where both agents benefit from the corrupt transaction, we observe in our experimental game that a public official benefiting from a corrupt transaction is more likely to extort a citizen they most likely will not interact with again. Meanwhile in settings of repeated interaction with the same citizen, prospect of endogenous monitoring is insensitive to the type of monitoring. Moreover, in the stranger treatments, public officials react stronger to report and fines than to recommendations and rewards. They seem to be motivated to abstain from extortive behavior due to fear of reports and sanctions, rather than a prospect of reward for non-corrupt behavior. This might also occur because sanctions tend to be categorized in the loss domain and most subjects tend to be loss-averse. However, receptiveness to recommendations is quite different on the citizen's side. Citizens are vastly willing to engage in endogenous monitoring that is costly to them if it means recommending rather than reporting a public official. This is not only true in the partner matching where there is a stronger bond between the public official and the private citizen; even under stranger matching citizens are frequently rewarding officials that they might not encounter again. Citizens seem to be more upset by public official's unfair behavior, often refuse to pay the demanded bribes and are more willing to engage in endogenous monitoring to reward non-corrupt behavior. Independent of matching protocol and monitoring mechanism, we find that agents in both roles are sensitive to monitoring and detection rate variations: public officials in their decision to demand a bribe and citizens in their decision of involvement in endogenous monitoring.

The latter finding could, for example, explain observations from Transparency International's 2011 Global Corruption Barometer. It reports that 70% of people around the world think that ordinary people can make a difference in the fight against corruption, and 68% could imagine themselves getting involved in the fight against corruption. However, even between otherwise similar countries there are large differences in people's reported willingness to get involved. For example, in Bangladesh only 5% of respondents could imagine themselves getting involved in the fight against corruption, whereas in Liberia, where Nobel Peace Prize-winning President Ellen Johnson Sirleaf has established a strong commitment to anti-corruption, 94% of respondents could imagine themselves getting involved in the fight against corruption. In view of our results, this large disparity possibly demonstrates the significant difference in people's beliefs on how efficient are their reports and how often they would lead to monitoring and detection of corrupt behavior. One could certainly expect very little public involvement in fighting corruption in countries with endemic corruption, such as Bangladesh, where citizens perceive detection rate of misbehavior to be extremely low and therefore do not find it worthwhile getting involved.

If one is willing to draw inferences from the experimental data reported here to natural environments in the field, our observations suggest the following. In settings where people believe that reports or recommendations are likely to be carried through, where this perception might be further improved if independent anti-corruption agencies are involved in the process, general public might be more willing to engage in fighting corruption and public officials might be more receptive to public's complaints. Although more diligent monitoring requires additional infrastructure and spending, changing general public's perception of how thoroughly their feedback is investigated might yield strong long-term benefits.

If we were to compare vis-a-viz the relative efficiency of the two endogenous monitoring systems and were to recommend policy implications, the proper categorization of the real world environment is of great consequence. First, it is likely to make a difference whether citizens face extortive or collusive corruption. For the latter, Serra (2011) demonstrates that endogenous reporting tends to mitigate corruption and increases efficiency, while our results on endogenous reporting in extortive corruption are more pessimistic; the likely reason being the conflict of interest between the PO and the citizen at hand. Second, it seems to make a strong difference whether extortive corruption plays out more in a Partner-matching setting or a Stranger-matching setting. Lack of anonymity and ensuing fear of retaliation speak strongly against the use of negative endogenous monitoring system in the settings of continues interaction between the agents. Moreover, the clear negative statement of intent in case of a report versus a possibility of various vague interpretations for not recommending further aggravates the problems of anonymity, especially since in real life settings, public officials are likely to learn sooner about reports than recommendations.[13] Therefore, citizens in the settings similar to partner-matching are less likely to make use of the negative endogenous feedback mechanism out of fear of angering a public official they might have to face again. Possibly influenced by their own passivity, they are also more likely to give in to the public official's pressure and pay the demanded bribe. In contrast, in environments similar to Stranger-matching, where issues of anonymity are more or less resolved,

[13] It is even possible that public officials would learn about a recommendation only in case of an implemented reward.

private citizens seem to be more likely to make use of the negative endogenous feedback than in environments similar to Partner-matching. Possibly realizing this trend, public officials are less likely to demand bribes if they can be reported and would have to face the consequential threat of a fine. In essence, both endogenous monitoring mechanisms have comparative advantage with public officials or private citizens, but not both at the same time. Possibly a combination of these two mechanisms could lead to the most efficient results in tackling corruption. We leave this question to future research.

Figure 3: Descriptives

This figure shows frequencies of occurrence for the different stages of the game in the first 10 rounds for all of the 4 treatments. It gives an overview of how many bribes have been demanded in each treatment, and depending on that how many were paid, etc.

Treatment	Bribe	Paid	Quality	Report/Recommend
120 REP_P ▷	69 Yes ▷	45 Yes ▷	45 Regular ▷	13 Yes
				32 No
		24 No ▷	6 Regular ▷	0 Yes
				6 No
			18 Reduced ▷	16 Yes
				2 No
	51 No	▷	34 Regular	–
			17 Reduced	–
120 REP_S ▷	92 Yes ▷	48 Yes ▷	48 Regular ▷	8 Yes
				40 No
		44 No ▷	17 Regular ▷	3 Yes
				14 No
			27 Reduced ▷	19 Yes
				8 No
	28 No	▷	9 Regular	–
			19 Reduced	–
120 REC_P ▷	85 Yes ▷	48 Yes ▷	48 Regular	–
		37 No ▷	5 Regular	–
			32 Reduced	–
	35 No	▷	27 Regular ▷	22 Yes
				5 No
			8 Reduced ▷	1 Yes
				6 No
120 REC_S ▷	101 Yes ▷	73 Yes ▷	73 Regular	–
		28 No ▷	4 Regular	–
			24 Reduced	–
	19 No	▷	13 Regular ▷	9 Yes
				4 No
			6 Reduced ▷	1 Yes
				5 No

* Report in treatments REP_P & REP_S; Recommend in treatments REC_P & REC_S.

References

Abbink, K. (2004). Staff rotation as an anti-corruption policy: an experimental study. *European Journal of Political Economy 20* (4), 887–906.

Abbink, K. (2006). Laboratory experiments on corruption. Monash Economics Working Papers archive-38, Monash University, Department of Economics.

Abbink, K., U. Dasgupta, L. Gangadharan, and T. Jain (2012). Letting the briber go free: An experiment on mitigating harassment bribes. *Working paper (SSNR)*.

Abbink, K. and D. Serra (2012). Anticorruption policies: Lessons from the lab. *Research in Experimental Economics 15*, 77–115.

Andvig, J. C. (2005). Experimental research and corruption: A survey of budding research. *Transparency International: Global Coruption Report* , 255–267.

Batory, A. (2012). Why do anti-corruption laws fail in central Eastern Europe? a target compliance perspective. *Regulation & Governance 6*, 66–82.

Bobkova, N. and H. Egbert (2012). Corruption investigated in the lab: a survey of the experimental literature. MPRA Paper 38163, University Library of Munich, Germany.

Davoodi, H. R. and V. Tanzi (1997). Corruption, public investment, and growth. IMF Working Papers 97/139, International Monetary Fund.

Dusek, L., A. Ortman, and L. Lizal (2005). Understanding corruption and corruptibility through experiments. *Prague Economic Papers 2005* (2), 147–162.

Fehr, E. and S. Gächter (2000). Fairness and retaliation: The economics of reciprocity. *The Journal of Economic Perspectives 14* (3), 159–181.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics 10* (2), 171–178.

Fisman, R. and J. Svensson (2007). Are corruption and taxation really harmful to growth? firm level evidence. *Journal of Development Economics 83* (1), 63–75.

Greiner, B. (2004). An online recruitment system for economic experiments. In K. Kremer and V. Macho (Eds.), *Forschung und wissenschaftliches Rechnen 2003. GWDG Bericht 63*, pp. 79–93.

Gupta, S., H. Davoodi, and R. Alonso-Terme (2002). Does corruption affect income inequality and poverty? *Economics of Governance 3* (1), pp. 23–45.

Knack, S. and P. Keefer (1995). Institutions and economic performance: Cross-country tests using alternative institutional measures. *Economics and Politics 7* (3), pp. 207–227.

Mauro, P. (1995). Corruption and growth. *The Quarterly Journal of Economics 110* (3), pp. 681–712.

Meon, P.-G. and K. Sekkat (2005). Does corruption grease or sand the wheels of growth? *Public Choice 122* (1/2), pp. 69–97.

Olken, B. A. (2006). Corruption and the costs of redistribution: Micro evidence from Indonesia. *Journal of Public Economics 90* (4-5), pp 853–870.

Reinikka, R. and J. Svensson (2004). Local capture: Evidence from a central government transfer program in uganda. *The Quarterly Journal of Economics 119* (2), pp. 678–704.

Renner, E. (2004). Wie lässt sich korruption wirksam bekämpfen?: Empirische befunde aus der experimentellen wirtschaftsforschung. *Vierteljahrshefte zur Wirtschaftsforschung / Quarterly Journal of Economic Research 73* (2), pp 292–300.

Rose-Ackerman, S. and R. Truex (2012). Corruption and policy reform. *Yale Law & Economics Research Paper No. 444.*.

Serra, D. (2011). Combining top-down and bottom-up accountability: Evidence from a bribery experiment. *Journal of Law, Economics, and Organization*.

CHAPTER 3

Letting off Steam! Experimental Evidence on Irrational Punishment

Abstract

Aggression is displaced, when provocations cannot be directly retaliated against, and aggression is redirected towards a target innocent of any wrongdoing. While this phenomenon is widespread, as yet it had not been explored in experimental economics. We fill this gap and find that a sizeable proportion of subjects (37%), when treated unfairly, punish co-players who are not at all responsible for the unfairness. When in a disadvantaged position, inequity-aversion seems to be the driving force of punishment, yet when treated fairly, some subjects (17%) exhibit status-seeking behavior. Moreover, students affiliated with an armed forces university are much more likely than regular students to engage in displaced aggression.

1. Motivation

Economic experimental literature has shown us that punishment is sometimes exercised for emotional satisfaction of the punisher, rather than as a disciplining effort of a potential long-term cooperation partner. Fehr and Gächter (2000; and various replications and extensions of this study) demonstrate that subjects use costly punishment against co-players that upset them, even if future interaction is not foreseen.

Our study takes this issue to the next level by looking at the consequences of an unfulfilled desire to punish an anger-provoking co-player. Already in 1939, Dollard et al. coined the term *Displaced Aggression* for situations where an innocent bystander is punished, when the original provocateur of aggression is out of reach.[14] A real-life example of "being at the wrong place at the wrong time" is revealed by Card and Dahl (2011). Using police reports data of domestic violence incidents during the professional football season in United States, they show that "unexpected" losses, (when an objectively better rated team loses), lead to a 10% increase in the rate of at-home violence by men against their wives and girlfriends.

To the best of our knowledge, we are first to implement economic experimental methods to study displaced aggression. Our experiment is an adaptation of the Bartling and Fischbacher (2012) four-person game.[15] One of these four players, the dictator, has the incentive to act selfishly but cannot be penalized for her[16] potentially selfish behavior. The dictator can benefit herself and another player at the cost of two remaining players. Given such an advantaged position in the game, not surprisingly, majority of subjects in the role of the dictator abuse their power.

[14] Strangely, popular culture sometimes seems to encourage displacement of aggression towards inanimate objects. For example, Bushman et al. (1999) report that some self-help books encourage one to punch pillows or break glass to "let off steam." Expression "letting off steam" is not exclusive to the Anglo-Saxon culture. A five-lingual co-author of this paper can think of a colloquialism that expresses the same sentiment in all the five languages she speaks. Scientific evidence, however, points in the completely opposite direction: Outward aggression towards inanimate objects does not help one to "let off steam," but only "fires one more up."

[15] However, we study a completely different question. Bartling and Fischbacher (2012) study whether delegation of the distributive decision to another person shifts also the blame for unfairness to this other person. They find that the decision to delegate is often perceived by the "delegee" as a signal to choose the unfair distribution. And then indeed the "delegees" are more likely to get punished, although the original decision-maker could have directly chosen the fair distribution him/herself.

[16] For simplicity, the dictator (Player A) will always be referred to as "she" and all subjects who can be taken advantage of (Players C) will always be referred to as "he".

The main focus of the paper, however, is the aftermath of such unequal power structure. The potentially disadvantaged subjects are given the possibility to "let off steam," i.e. exercise costly punishment against their non-impune co-players.

We use this simple two-stage game to address four research questions:

1. Do innocent bystanders pay for the "sins" of unfair impune dictators?
2. Do *unkinder* intentions by the dictator cause more displaced aggression towards innocent bystanders?
3. Do bad intentions cause more displaced aggression than *bad luck*?
4. Does association with the military sub-culture affect the level of displaced aggression towards innocent bystanders?

The association with the military in our experiment is defined by subjects' real-life self-selection into the military profession in Germany. Wiens et al. (2015) argue that due to military training, soldiers should be more cooperative and should exert higher social responsibility. They find that in dictator, ultimatum, and trust games, soldiers are more altruistic, more cooperative, more trusting, and more trustworthy than regular students. (See Michailova and Bühren (2015) for similar results.)

The fourth question has been inspired by the knowledge of inter-cultural differences in aggression norms. Through ethnographic studies, Fessler (2006) demonstrates just how wide is the range of cultural differences in aggression norms. Well-being of any Eskimo individual depends heavily on his/her community's common fishing and hunting efforts. Since inter-personal conflict can be detrimental for the whole community, it becomes understandable why in this culture aggressive behavior is considered "childish" and is ridiculed. It is then fascinating that aggression can be considered respectably "manly" and almost admirable in another culture. The whole cinematic industry of "Westerns" attests to the "manliness" of violent "cultures of honor" and chivalry, characteristic to American South and West. (See Sánchez-Jankowski (1991) for a similar analysis on culture and aggression in regards to youth gangs in segregated American cities).

Our study finds that although a large proportion of subjects acts according to standard predictions and does not engage in any costly punishment, a considerable number of subjects engage in displaced aggression and punish their innocent co-players. Subjects associated with the military are more likely to do so. Punishers seem insensitive to the level of *unkindness* by the dictator, and do not react differently to bad intentions vs. bad luck. Additionally, we find that about one fifth of all our subjects exhibit status-seeking behavior, when treated fairly.

Our paper consists of five sections. Section 2 describes our design and particularities of our subject pools. Section 3 introduces two economic behavioral models that compete to predict the behavior of subjects in our experiment. Section 4 describes empirical results and compares the empirics to the theoretical predictions made in Section 3. The last sub-section of Section 4 reports on several additional control treatments we have conducted as a robustness check. Lastly, Section 5 provides a summary of our work and discusses its implications.

2. Experimental Design and Subject Pool Variations

2.1. Design

To study how people react to unfair treatment in situations where the wrongdoer is impune to punishment, we implemented a modified dictator game, inspired by the design of Bartling and Fischbacher (2012).

Similar to Bartling and Fischbacher (2012), our game has four players. Each group consists of a player A, a player B, and two players C. Player A and Players C are active decision-makers in this game.

Our game has two stages. In Stage 1, Player A, the dictator, makes a binary distributive decision. In Stage 2, using strategy-eliciting method, Players C state their punishment decisions for each of the possible distribution scenarios. Thereafter, one of the Players C is randomly selected. This Player's punishment decision is actually implemented and becomes payoff relevant for the whole group. Before they make their punishment decisions, however, Players C do not know whether they have been selected or their co-playing Player C has been selected. They are, however, well-aware that their punishment decision might be payoff-relevant. Punishment is costly for a Player C, who must give up one point of his own to punish a co-player by 5 points.

Player B does not make any payoff-relevant decisions in this game. These passive players, however, are asked to state which distribution they would have chosen had they been randomly assigned to the role of Player A (the dictator).

A very important detail of our design is that Player A, the dictator, cannot be punished. Our design attempts to mimic frustrating conflict situations with strangers, who cannot be retaliated against. For example, frustrating interactions with figures of authority, e.g. a rude police officer. Our research project sets out to explore the aftermath of these situations.

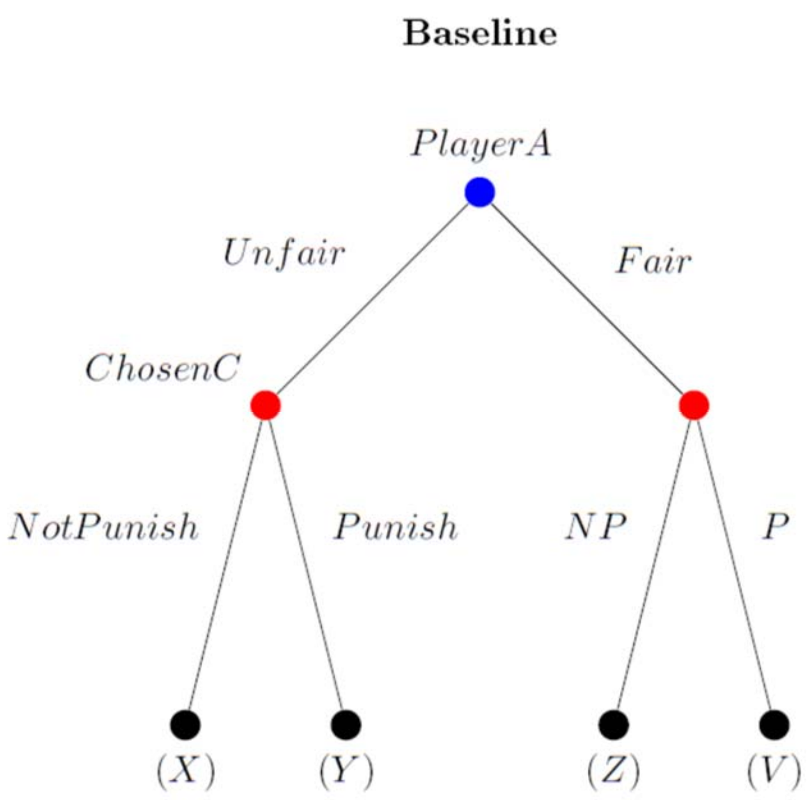
Our experiment has two (main) treatments. The difference between these treatments comes from Stage 1. In *Baseline*, the dictator, Player A, makes a binary choice between an unfair and a fair distribution. Our *Chance* treatment is also a binary dictator game. But in this treatment, player A, must choose between an unfair distribution and delegation to chance. The latter results with equal probabilities in either a fair distribution or an unfair distribution. In both of these treatments, Player A can directly choose an unfair distribution, and only the alternative to this choice is different between the treatments.

Player A begins with 100 points. If Player A directly chooses the unfair distribution, she and Player B in her group receive 45 points each and both Players C in the group receive only 5 points each. In *Baseline*, if Player A chooses the fair distribution, she and all the other 3 players in the group receive 25 points each. In *Chance*, payoffs are the same for an unfair distribution regardless of its source: dictator's direct choice or chance (bad luck). Similarly, a randomly generated fair distribution in *Chance* provides an endowment of 25 points for each of the four players.

It is important to note that in *Chance*, Player A cannot directly implement a fair distribution. The *kindest* choice Player A can make here is to delegate the distributive decision to chance; so to say: "give the fair distribution a chance".

In stage 2, Players C make the punishment decision, before knowing which distribution their group's dictator has chosen. Players C can at most invest 5 points into punishment in any of the scenarios. (This equals his endowment in the unfair distribution scenarios). Moreover, Player C cannot subtract more points from co-players than they already have. (Meaning, that the other non-chosen Player C can be punished at most by one point in the unfair scenarios).

Figure 4: Game Tree *Baseline* treatment



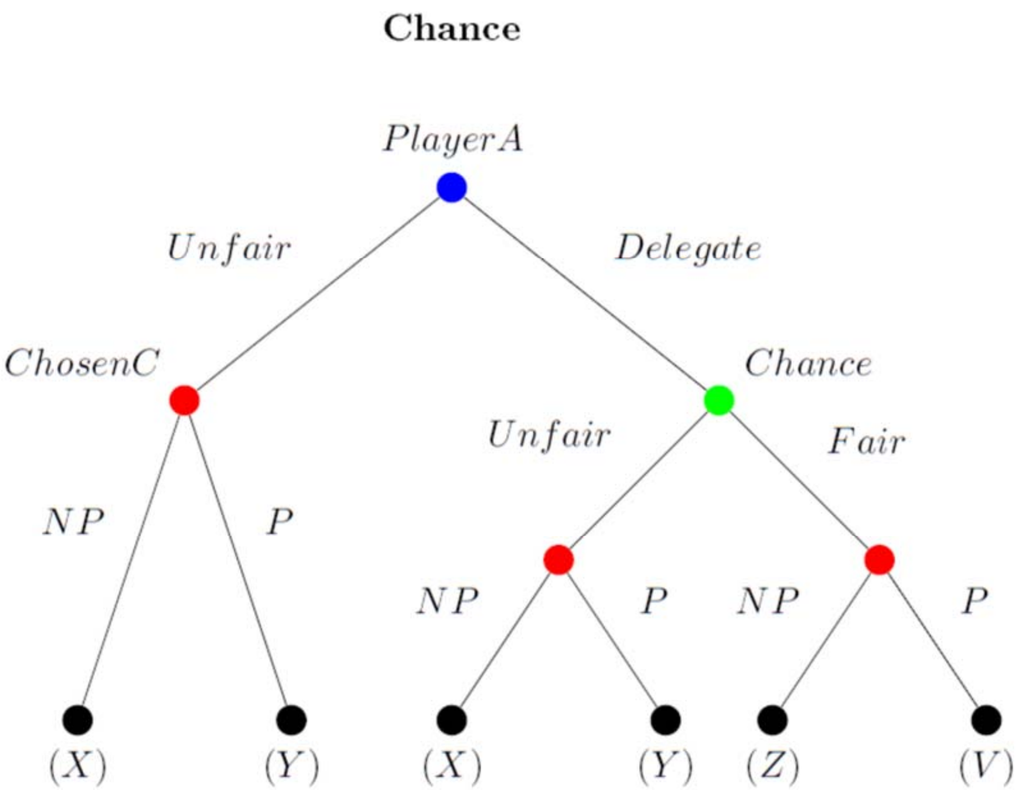
Payoffs:	***X***	***Y***	***Z***	***V***
A	45	45	25	25
B	45	45 - 5b	25	25-5b
Other C	5	5 - 5c	25	25-5
Chosen C	5	5 – b - c	25	25-b-c

There are two scenarios in *Baseline* for which Players C must provide answers: 1) Player A choosing the unfair distribution and 2) Player A choosing the fair distribution.

Since three different scenarios are possible in *Chance*, Players C must state in this case their punishment decisions for each of the three possible scenarios: 1) an unfair distribution directly chosen by Player A, 2) an unfair distribution randomly generated by chance (computer or coin flip, depending on the sessions), and 3) a fair distribution randomly generated by chance (computer or coin flip, depending on the sessions).

Please see the corresponding Figures (4 and 5) for the Game Trees of the *Baseline* and *Chance* treatments, respectively. For simplicity, Player A is always represented by a blue dot, the "chosen" Player C by a red dot, and random chance draw by a green dot. The payoff table below the *Baseline* Game Tree is relevant also to the payoff structure in the *Chance* treatment.

Figure 5: Game Tree *Chance* treatment



Our inter-treatment comparison investigates the second research question of this paper: "Do unkinder intentions by the dictator cause more displaced aggression towards innocent bystanders?" The within-subject comparison of the response to unfairness in *Chance* tests our third research question: "Do bad intentions cause more displaced aggression than bad luck?"

We were interested to see what proportion of subjects would have liked to punish if they could directly retaliate against the unfair dictator. Therefore, after stating their punishment decisions for all the possible payoff-relevant scenarios, Players C were presented with parallel hypothetical scenarios where Player A could also be punished. They were then asked to state their punishment preferences for these hypothetical scenarios.

After the two payoff-relevant stages of the game, all subjects in both of the treatments were asked to estimate what proportion of Players A in their opinion directly chose the unfair distribution (vs. the fair distribution in *Baseline* and vs. delegation to chance in *Chance*).

2.2. Subject Pools

Our fourth research question considers sub-cultural differences in reaction to potential unfairness. A particular subculture comes to mind, when one thinks of potentially frustrating situations. Soldiers must partake in conflict situations because of their professional demands. Moreover, the military system is ridden by strong hierarchy. Soldiers usually cannot talk back, so frustration might stay “bottled up”. Therefore, we are particularly interested to explore how soldiers react to unfair situations, where the potential wrong-doer cannot be retaliated against. In June 2013 and February 2014 we conducted experiments at the Helmut-Schmidt University in Hamburg, which is affiliated with the German military. Subjects were recruited at economics lectures. This subject pool will be mentioned as *Soldiers* from now on.

As a control for this special subject pool, we conducted the same treatments (*Baseline* and *Chance*) with students at the University of Kassel in June - October 2013 and March - December 2014. University of Kassel students were also recruited at economics lectures. This subject pool from now on will be mentioned as *Students*.

Considering the professional challenges of individuals serving in the military, it is a relevant question, whether soldiers are more or less likely than regular students to engage in displaced aggression. On the one hand, discipline seems to be a core value in the military sub-culture. A more evolved sense of discipline might translate into a less emotional approach to unfairness, and therefore less displaced aggression. On the other hand, particularly dominant and aggression-prone individuals possibly self-select into the military profession.

As we compared our *Soldiers* and *Students* subject pools, we felt concerned that besides the difference in career self-selection, which is the variation of interest for our research question, these two subject pools might differ in two more ways:

1) University of Kassel students might have a lower interconnectedness level in comparison to the strong social ties established among the *Soldiers*. Soldiers at the Helmut-Schmidt University in Hamburg spend a lot of time together outside of classroom, including military training and leisure at the shared university dormitories. Michailova and Bühren (2015) argue that due to the trained comradeship of soldiers, students at the Helmut-Schmidt-University belong to a group with stronger social ties than a group of students at any other regular university in Germany. Similarly, Wiens et al. (2015) state that cooperation and social responsibility of soldiers should be higher than that of other students and find evidence for this at the second German university of the military forces in Munich.

2) Experimental sessions at the University of Kassel had a lot lower proportion of male subjects (51%) than the lab sessions at the Helmut-Schmidt University in Hamburg, where the majority of experiment participants (92%) were male. Although, we can certainly control for the gender effects when considering individual subjects' decisions, we were concerned that the gender composition of the co-players in the session that one can observe might also affect one's decisions in the experiment.

To counteract these concerns, we complement our study with another pool of subjects who also have strong social ties and are mostly male. We recruited teams of Ultimate Frisbee players at the German championship in Kassel in September, 2015. Most of the Frisbee players are students of different disciplines. This subject pool will be mentioned as ***Frisbees*** from now on.

Ultimate Frisbee is a very special sport, since it is played without any referee. After each Frisbee match, teams discuss fairness aspects of the specific game and rate their opponent team with spirit points depending on fairness. Frisbee players build a strong community – during tournaments, which are typically played over the weekends, all the players camp together.[17]

Moreover, due to natural composition of the teams, we were able to conduct mostly male sessions (94% male), without calling any attention of the subjects to the gender composition of the sessions.

Our experiment has been computerized using the software z-Tree (Fischbacher, 2007). Most of the experimental sessions were conducted at computer labs at the University of Kassel or the Helmut-Schmidt University in Hamburg. The subject pool of the Frisbee players could not be recruited to

[17] Please see a short video about this sport's philosophy and the Championship event in Kassel under the following link: http://www.rtl-hessen.de/video/9851/deutsche-meisterschaften-im-frisbee.

come to the lab. These players completed a pen and paper version of the game directly at the Frisbee tournament location. We only conducted *Chance* with *Frisbees*.[18]

3. Theory

The behavioral model of frustration by Pierpaolo Battigalli, Martin Dufwenberg and Alec Smith (Battigalli et al., 2015) explicitly considers an example of displaced aggression, and therefore is fitting for our research questions. This model from now on will be referred to as either the B-D-S model or the Frustration model.

To calculate utility using this model, one must first calculate the frustration level caused by a specific scenario. Frustration level depends on the actions of others, and on one's own expectations for those actions. Incidents can either be frustrating (i.e. have a positive frustration level) or non-frustrating (i.e. a frustration level equal to zero).

Intuitively frustration is the difference between what one initially expected and what one can get now, given the choices made by other(s) and one's alternatives in this situation.

Given the potentially frustration-provoking design of our game, where the punishable players (B and other C) cannot be blamed for the provoked frustration, utility according to the B-D-S model depends on one's payoff in a particular scenario, payoff of all the other groupmates in that particular scenario, the frustration level generated by this scenario, and one's own internal irritability level, which is part of one's personality. Given the payoff structure of the game, we will calculate what the parameter for the irritability must be for Player C to choose punishment in a particular scenario.

[18] We conducted two more small (12-persons each) pen-n-paper sessions, so as to control for the potential differences between the computerized and pen-n-paper versions of the *Chance* treatment.

- An all-female-Frisbee-team from the championship held in Kassel in September 2015.
- A mixed-gender mixed-subject students session recruited at the beach volleyball facilities of the University of Kassel, also in September 2015.

Neither of these control pen-n-paper sessions are significantly different in their choices from *Frisbees* or *Students*. Therefore, we could conclude that our pen-n-paper sessions with *Frisbees* are design-wise no different from our computerized sessions with the *Students* and *Soldiers*.

In the most general form, utility of Player C is:

$$U_C(A'sAction, C'sAction) = \pi_C(A'sAction, C'sAction) - \theta_C F_C(A'sAction) \sum_{j \neq C} \pi_j(A'sAction, C'sAction)$$

Where π-s represent material payoffs, θ_C is the parameter for the internal irritability level and $F_C(A'sAction)$ is the frustration level caused by A's action, given Player C's expectations of that action. Clearly, when A's actions do not cause any frustration to Player C, his utility is identical with his monetary payoff.

To check whether Player C would indirectly retaliate against his innocent co-players, first we must calculate the level of frustration for the situation where Player A directly chose an unfair distribution. For example, let's denote with ε and $(1-\varepsilon)$ Player C's expectation of Player A choosing the unfair distribution or choosing the alternative, respectively. (The alternative is a fair distribution in *Baseline*, and is delegation to chance in *Chance*). r denotes, in any given scenario, the probability that Player C punishes. $(1-r)$ is then the probability that Player C does not punish anyone.

If Player A directly chose the unfair distribution, Player C's frustration is:

$$F_C(Unfair) = \max\{0, (1-\varepsilon) * \pi_C("Alternative") + \varepsilon[(1-r) * \pi_C(Unfair, NotPunish) - r * \pi_C(Unfair, Punish)] - \max(\pi_C(Unfair, NotPunish), \pi_C(Unfair, Punish)\}$$

Let's assume player C is standing in front of the choice of punishing or not punishing the passive Player B by one point in the *Baseline*:

$$F_C(Unfair) = \max\{0, 25(1-\varepsilon) + \varepsilon[5(1-r) + 4r] - max(5,4)\} = \max\{0, 20 - 20\varepsilon - \varepsilon r\}$$

If ε (expectation of direct unfairness) is not 100%, which would be very unlikely, Player C is frustrated by dictator's choice: $F_C(Unfair) > 0$.

Therefore, in *Baseline*, when Player A directly chooses the unfair distribution and the chosen Player C chooses not to punish any of his punishable co-players, utility of Player C is:

$$U_C(Unfair, NotPunish) = 5 - \theta_C(20 - 20\varepsilon - \varepsilon r)(45 + 45 + 5)$$
$$= 5 - 95 * \theta_C * F_C(Unfair)$$

If, however, Player C chooses to punish Player B by one point, utility of Player C is:

$$U_C\big(Unfair,\ Punish_{B-1pt}\big) = 4 - \theta_C(20 - 20\varepsilon - \varepsilon r)(45 + 40 + 5)$$
$$= 4 - 90 * \theta_C * F_C(Unfair)$$

In *Baseline*, Player C should choose to Punish Player B by one point, if his personal irritability level is high enough: $\theta_C > {}^{1}/_{5F_C(Unfair)}$.

Alternatively, the *payoff-driven theory of fairness* by Fehr and Schmidt (1999) is a candidate for evaluating our research questions. According to this behavioral model, agents care about their own payoff and their relative standing in comparison to the other players in their group. This model in our paper from now on will be referred to as either the F-S model or the Inequity-aversion model.

In its most general form, utility, according to the Fehr-Schmidt model, is:

$$u_i(x) = x_i - \alpha_i \frac{1}{n-1} \sum_{j \neq i} \max\{x_j - x_i, 0\} - \beta_i \frac{1}{n-1} \sum_{j \neq i} \max\{x_i - x_j, 0\}$$

where x-s represent the material payoffs for the different players and n is the number of players in the group. According to the F-S model, subjects experience disutility from feelings of envy towards co-players that have higher payoffs than them. Subjects also experience disutility from feelings of guilt towards co-players who have lower payoffs than them. The strengths of these feelings for an individual are represented by the parameters α_i and β_i, respectively.

Let's assume Player A has directly chosen the unfair distribution, Player C could then, for example:

i) choose to punish passive Player B by 1 point. In this case Player A would still have 45 points, player B would have 40 points (= 45 - 5), the other Player C would still have 5 points, and the chosen and punishing Player C would have 4 points (= 5 – 1);

ii) choose not to punish anyone. In this case Player A and Player B have 45 points each and the 2 Players C have 5 points each.

The utility equations then for these two scenarios would be:

i) $u_C(45, 40, 5, 4) = 4 - \alpha_C \frac{1}{3}(45 - 4) - \alpha_C \frac{1}{3}(40 - 4) - \alpha_C \frac{1}{3}(5 - 4)$. In this case, chosen Player C has given up one point to punish Player B by 5 points. Therefore, there are now 3 co-players that are better off (in different amounts) than Player C. $u_C(45, 40, 5, 4) = 4 - \frac{78}{3}\alpha_C$.

ii) $u_C(45, 45, 5, 5) = 5 - \alpha_C \frac{2}{3}(45 - 5)$, since there are then 2 players who are better off than the chosen Player C (by 40 points) and no player that is worse off than the chosen player C. $u_C(45, 45, 5, 5) = 5 - \frac{80}{3}\alpha_C$.

Given these calculations, the chosen Player C should choose to punish Player B by one point if $u_C(45, 40, 5, 4) > u_C(45, 45, 5, 5)$. This expression holds true, when α is sufficiently high: $\alpha_C > 3/2$.

For the remainder of this section, all the punishment decisions will first be looked at through the lens of Battigalli-Dufwenberg-Smith (B-D-S) frustration model, and thereafter the inequity-aversion model by Fehr and Schmidt. Moreover, when applicable, additional *intuitive* predictions or predictions based on social psychology literature will also be discussed.

3.1. Inter-Treatment Comparison

When comparing the two (main) treatments, the dictator's direct choice for an unfair distribution signals worse intentions in *Chance* than in *Baseline*, because opportunity costs of not choosing unfair are higher in *Baseline* than in *Chance*.[19] Whereas in *Baseline*, the alternative to the unfair distribution is directly a fair distribution, the alternative in *Chance* still results in an unfair distribution in 50%, which could still benefit Player A and be disadvantageous for Players C. Directly choosing the unfair distribution in *Chance* does not even give the fair distribution a chance. Therefore, possibly Players C are more frustrated by a direct unfair choice in *Chance*

[19] *Baseline*: 45 – 25 = 20 vs. *Chance*: 45 - $\frac{45+25}{2}$ = 10.

compared to *Baseline*. By this logic, they might engage in more displaced aggression in *Chance* than in *Baseline*.

1) B-D-S predictions.

   The B-D-S (frustration) model predicts the opposite: Less punishment in *Chance* than in *Baseline*. Difference in expected payoffs for Players C between the treatments is the cause. In *Baseline*, when Player A is fair, payoff is 25 points for each of the group-mates. In *Chance*, if Player A decides to delegate the distributive decision to chance (computer or coin), expected payoff for Players C is only 15 points ($= \frac{5+25}{2}$). Hence for a Player C, the best case scenario, (when the dictator abstains from the unfair distribution), in the expected payoff terms is worse in *Chance* than in *Baseline*.

   The B-D-S model assumes that subjects are more frustrated by bigger potential losses. Therefore, according to B-D-S, Players C are expected to engage in more displaced aggression in *Baseline* than in *Chance*.

2) F-S predictions.

   Since dictator's direct choice for unfairness results in the exact same distribution in both treatments, the F-S model in both treatments predicts exactly same punishment behavior on the part of Player C.

3.2. Within-Subject Comparison

Our *Chance* treatment allows for <u>within-subject</u> comparison of punishment behavior. The two unfair distribution scenarios can either be caused by bad intentions - dictator's choice, or bad luck - random chance. If intentions matter, subjects should be more frustrated and therefore more aggressive, when a disadvantaged position has been caused by the whim of their co-player, rather than random chance.

1) B-D-S predictions.

   The B-D-S model predicts approximately equal levels of displaced aggression when Player A is directly unfair or when Player A delegates her decision to chance, which results in an unfair distribution.

In *Chance*, when Player A directly chooses the unfair distribution, Player C's frustration level is:

$$F_C(A\_Unfair) = \max\left\{0, (\frac{25+5}{2})(1-\varepsilon) + \varepsilon[5(1-r) + 4r] - max(5,4)\right\}$$
$$= \max\{0, 10 - 10\varepsilon - \varepsilon r\}$$

In *Chance*, when Player A delegates the distributive decision to chance, and the latter generates an unfair distribution, Player C's frustration level is:

$$F_C(Comp\_Unfair) = \max\left\{0, (1-\varepsilon)(\frac{1}{2}[4r + 5(1-r)] + \frac{1}{2}25) + 5\varepsilon - max(5,4)\right\}$$
$$= \max\{0, 10 - 10\varepsilon - 0.5r + 0.5\varepsilon r\}$$

Since parameter r must range between 0 and 1, the difference in the frustration between these two scenarios is marginal ($F_C(A\_nfair) \approx F_C(Comp\_Unfair)$), meaning that approximately equal proportion of subjects should punish their co-players out of frustration in both of the unfair distribution scenarios of *Chance*.

2) F-S predictions.

According to the F-S model, it should not matter if the unfair distribution has been caused by A's bad intentions or the bad luck randomly generated by either the computer or the coin flip (depending on the session). If subjects have inequity-averse preferences, these two scenarios of *Chance* treatment have exactly the same payoff structure, and therefore must result in exactly the same punishment behavior, which depends on individual's personal envy parameter. Subjects with a sufficiently high α (>1.5) should punish Player B in both of these scenarios. For these sufficiently-envious players C, this punishment pattern would bring their own payoff closer to the payoff of their co-players and would thereby boost their utility.

3.3. Who Should Get Punished? (in unfair distributions)

Marcus-Newhall et al. (2000)[20] show that when frustrated, people are more likely to punish an innocent bystander, if the latter is somehow similar to the person who has wronged them. Our experimental design accommodates a renewed test of this theory: each group in our game has one

[20] A meta-study based on 82 papers in social psychology. Interestingly, for many of these studies displaced aggression is not the direct research focus.

punishable co-player that is similar to the dictator (Player B), and another punishable co-player that is similar to the punishing agent (the other Player C). By this logic, frustrated Player C should punish the Player B, but not the other Player C.

1) B-D-S predictions.

   Since utility according to the B-D-S model depends on the sum of co-players' payoffs, there should be no difference in Player C's utility when punishing by one point the passive Player B (who is "rich" like the dictator) or the other Player C (who is "poor"). The condition for one's personal irritability level is the same: $\theta_C > {}^{1}/_{5F_C(Unfair)}$.

2) F-S predictions.

   Player C with F-S preferences should never punish the co-playing Player C. Since unfair distribution leaves the two Players C with equal endowments, punishment of the other Player C could only cause disutility to punishing Player C.

3.4. Should Subjects Punish in Fair Distributions?

There are two scenarios when fair distribution is possible: when Player A directly chooses the fair distribution in *Baseline*, or when in *Chance* Player A delegates her decision to chance, which generates a fair distribution.

1) B-D-S predictions.

   Frustration level equals zero, when distribution is fair.[21] The B-D-S model predicts no displaced aggression, if one is not frustrated.[22]

2) F-S predictions.

In case of fair distribution, all four co-players have an equal endowment. Punishment in this case can only create disutility. Therefore, according to the F-S model, no punishment should be observed in case of fair distribution.

[21] In *Baseline*: $F_C(Fair)_{Baseline} = \max\{0, 5\varepsilon + (1-\varepsilon)[25(1-r) + 24r] - max(25, 24)\} = \max\{0, -20\varepsilon - r - \varepsilon r\} = 0$.
In *Chance*: $F_C(Comp\_Fair)_{Chance} = \max\{0, -10 - 10\varepsilon - 0.5r + 0.5\varepsilon r\} = 0$.

[22] Punishment is costly: $U_C(Fair, NotPunish) = 25 - 0$ vs. $U_C(Fair, Punish) = 24 - 0$.

4. Results

4.1. "Direct Unfairness"

Unfairness rates between Baseline and Chance are not significantly different, for neither Students, nor Soldiers. When aggregating across those two types of treatments, unfairness rate of Soldiers is weakly significantly higher than the unfairness rate of Students (2-sided Fisher's exact test, p = 0.064), and is significantly higher than the unfairness rate of Frisbees (2-sided Fisher's exact test, p = 0.006).

Table 3: Proportion of direct unfairness by subject pool and treatment

Treatment / Subject Pool	**Baseline**	**Chance**	**All**
Students	58% (N=12)	62% (N=13)	60% (N=25)
Soldiers	83% (N=6)	100% (N=6)	92% (N=12)
Frisbees	-	38% (N=13)	38% (N=13)

Average direct unfairness rate for all of the observations is 62%. Average hypothetical "direct unfairness" rate stated by passive Players B is only slightly lower, at 56%.

Moreover, all three types of players (A, B, C) are good at predicting the direct unfairness rate for their own subject pool, although *Frisbees* often overestimate the proportion of unfair dictators. The average estimate of direct unfairness equals 70%. (When isolating just the beliefs of Players C, the average estimate of dictators' direct unfairness is 68%).

Among the *Students* subject pool, which has almost equal number of male and female subjects, female subjects are more generous than their male counterparts. 33% of all female *Students* directly choose the unfair distribution, compared to the 75% of all the male *Students* who do the same (2-sided Fisher's exact test, p = 0.087). Yet, both men and women have similar and close to reality beliefs regarding the proportion of unfair dictators (72% vs. 68%, Mann-Whitney test, p = 0.339).

4.2. Proportion of "Rational" Players

According to standard theory, subjects should not invest any of their own payoff into the punishment of their co-players, if they are only interested in their own payoff. Since punishment is costly, we define subjects as "rational" if they decide not to punish their co-players in any of the presented scenarios.

Table 4: Proportion of "rational" players by subject pool and treatment

Treatment / Subject Pool	**Baseline**	**Chance**	**All**
Students	71% (N=24)	73% (N=26)	72% (N=50)
Soldiers	50% (N=12)	25% (N=12)	37.5% (N=24)
Frisbees	-	69% (N=26)	69% (N=26)

Although the majority of subjects are "rational" (altogether 63%), a considerable proportion of subjects (37%) engage in irrational punishment in at least one of the scenarios they are presented with. *Soldiers* seem more susceptible to irrational punishment than others (63.5% vs. 29% among *Students* and *Frisbees* pooled together, 2-sided Fisher's exact test, $p = 0.007$).

Most of the proceeding results are discussed in terms of binary decision to punish or not to punish. Main results are no different if we instead consider total amount of points invested into punishment.

4.3. Punishment Behavior in Unfair Scenarios

Table 5: Punishment Behavior in Unfair Scenarios

Inter-Treatment Comparsion (Baseline vs. Chance) **Within-Subject Comparison** (Chance: Dictator vs. Computer / Coin flip)

Treatment	**Baseline**	**Chance**	**Chance**
Unfairness By…	Dictator	Dictator	Computer / Coin flip
Students	29% (N=24)	23% (N=26)	27% (N=26)
	p = 0.751		p = 1.000
Soldiers	42% (N=12)	75% (N=12)	50% (N=12)
	p = 0.214		p = 0.250
Frisbees	-	23% (N=26)	19% (N=26)
			p = 1.000

Note: 2nd lines for each subject pool provide 1) p-values of 2-sided Fisher's exact tests for the Inter-Treatment Comparisons, and 2) for Within-Subject Comparisons, the p-values of the McNemar's exact test (commonly used for within-subject observations).

4.3.1. Inter-Treatment Comparison

When comparing *Baseline* and *Chance*, differences in the punishment rate, are not significantly different for neither *Students*, nor *Soldiers* (2-sided Fisher's exact test p = 0.751 and p = 0.214, respectively; p = 0.635 when these two subject pools are combined). This is in line with predictions of the F-S model.

4.3.2. Within-Subject Comparison

When aggregating across all the three subject pools in *Chance*, 18 out of the 64 subjects C (28%) choose to punish their co-players in the scenario where unfair distribution has been caused by chance. In comparison, 21 out of the 64 subjects C (33%) choose to punish their co-player(s), when the dictator in *Chance* directly chose the unfair distribution. Hence, for the majority of subjects, the source of inequality <u>is not important.</u> This empirical result is in line with both F-S and B-D-S models' predictions.

4.3.3. Who Gets Punished?

Among all the available unfair distribution cases, Players C chose to punish in 51 cases (out of the total possible 164 cases). In 63% of these cases the only target of punishment was Player B. These observations are in line with the findings of the Marcus-Newhall et al. (2000) meta-study, since the "rich" passive Player B resembles the "rich" wrongdoing Player A.

Table 6: Who gets punished? (in unfair distributions)

Subject pool	**Treatment**	**Unfairness by…**		**B only**	**C only**	**B & C**	**Punishments N**
Students	Baseline	Dictator		57%	0%	43%	7
	Chance	Dictator		50%	0%	50%	6
		Computer		71%	0%	29%	7
Aggregate of Students				**60%**	**0%**	**40%**	**20**
Soldiers	Baseline	Dictator		40%	0%	60%	5
	Chance	Dictator		56%	0%	44%	9
		Computer		67%	0%	33%	6
Aggregate of Soldiers				**55%**	**0%**	**45%**	**20**
Frisbees	Chance	Dictator		83%	0%	17%	6
		Coin flip		80%	0%	20%	5
Aggregate of Frisbees				**82%**	**0%**	**18%**	**11**
Aggregate of ALL Observations				**63%**	**0%**	**37%**	**51**

Although no subject in the role of Player C punishes just the co-playing Player C, of all the punishments exercised as a response to unfairness, 37% involve not only the passive Player B, but also the co-playing Player C. The latter is not only dissimilar to the potentially wrong-doing Player A, but is the same type of player as the punishing agent himself.

The B-D-S model can partially explain the observed punishment preferences of Players C. According to the predictions of the B-D-S model, it should not matter for the subjects which of their "innocent" co-players they are punishing. Both of these players could be "in the wrong place at the wrong time". (Frustration level is the same). However, by this same logic, the B-D-S model fails to explain absence of Players C who only punish the co-playing Player C, but not Player B.

The F-S model of inequity-aversion, in its original form, cannot explain behavior of subjects C who punish the co-playing Player C. In all the scenarios of unfair distribution, both players C have an endowment of 5 points each. Since they have an equal payoff to begin with, punishment would create inequality, which in turn could only decrease the punishing party's utility.

However, possibly, so many subjects choose to punish Player C along with Player B because they do not wish to be the worst off person in their group.

A modified version of the F-S model that allows for a negative β parameter: "status-seeking" preferences, could explain why the chosen Player C might punish a co-player that has the same payoff as him.[23]

4.3.4. Punishment Behavior in Fair scenarios

Table 7: Punishment in fair scenarios

Treatment / Subject Pool	**Baseline** (A's Decision)	**Chance** (Lucky Chance)	**All**
Students	12.5% (N=24)	19% (N=26)	16% (N=50)
Soldiers	17% (N=12)	25% (N=12)	21% (N=24)
Frisbees	-	15% (N=26)	15% (N=26)

Punishment rates in the fair scenarios are not significantly different across any of the possible comparisons. In summary, 17 out of our 100 subjects C chose to punish their co-player(s) in one of the fair scenarios, caused either by Player A's direct fairness or Player A's delegation combined with good luck.

Interestingly, among cases of punishment, on average more points are invested into punishment when the distribution is fair, rather than unfair: 3.7 points vs. 2.9 points, respectively;[24] which translates into 18.5 points vs. 14.5 points of damage to co-players, respectively. Similarly, when punishing their co-players in the fair scenarios, rather than unfair scenarios, subjects punish their co-players more often by the maximum amount of points possible (5 points): 35% (6 out of 17) vs. 23.5% (12 out of 51) of the time, respectively (but not significantly more according to a two-

[23] The "status-seeking" β must then be inversely correlated to the subject's α (i.e. envy) parameter. Subjects that have a high α parameter, must have an especially low (in the negative domain) β parameter to also punish the other player C: $\beta < -\frac{\alpha+0.75}{2}$.

[24] Subjects who punish in both fair and unfair scenarios, invest significantly more points into punishment in the fair scenarios: Wilcoxon-signed rank test, p=0.007.

sided Fisher exact test, p=0.357). These punishment trends could be due to the income effect in the fair distribution scenarios, meaning that Players C have more resources to spend on punishment. However, they still cannot invest more than 5 points into punishment of co-players.

The B-D-S frustration model cannot explain any of the punishment we see in the fair distribution scenarios. The level of frustration in the fair distribution scenarios always equals zero. When frustration equals zero, no utility can be gained from punishment. Therefore, no subject should engage in punishment in any of the fair distribution scenarios.

In our post-experimental questionnaire, when asked about their general strategies in the game (an open-ended question), quite a few subjects expressed a desire for status. Perhaps most eloquently, one subject wrote: "I do not want that everybody has the same amount of money. It's better when other participants have less than me."[25]

In the fair distribution scenarios, the two non-impune co-players, Player B and the other (non-chosen) Player C, have the same payoff as the punishing party. Therefore, here, a Player C, who wishes to be better off than others, should punish both Player B and Player C. This explains why only 3 Players C (out of the 17 "fair"-punishers) choose to punish only Player B, and no one punishes just the other Player C.

If we look at this phenomenon with the F-S model in mind, this behavioral trend points towards a negative β factor, which is not foreseen by the original F-S model. A modified version of the F-S model that allows a negative β parameter could explain such status-seeking behavior. Moreover, same parameter restrictions would apply as in the previous sub-section (on Punishment of co-playing Player C in unfair distributions): $\beta < -\frac{\alpha+0.75}{2}$.

Hence, individuals that have the corresponding α and β parameters should engage in punishment in both of these directions: punishing (also) Player C in the unfair distribution scenarios and punishing co-players in the fair distribution scenarios. However, these within-subject predictions are not always consistent with our empirical evidence. 17 (= 7 + 10) subjects choose to punish in one of those status-improving scenarios, but not in the other. Additionally, there are 3 subjects (out of all 100 C-s) that punish co-players only in case of a fair distribution.

[25] Translated from German: "Ich möchte nicht, dass alle exakt dieselbe Punktzahl haben. Besser, wenn andere Teilnehmer weniger als ich haben"

		Punish (also) C in Direct Unfairness scenario	
		Yes	*No*
Punish in Fair Scenarios	*Yes*	7	10
	No	7	76

4.4. Discussion of Subject Pool Differences

There are clearly no differences in any of the punishment rate comparisons between *Students* and *Frisbees*. Hence, differences in the interconnectedness level among subjects or gender composition of the sessions do not play a role. For the discussion in this sub-section, we pool together these two subject pools as *Non-Soldiers*.

When aggregating across the two treatments, *Soldiers* punish similarly to *Non-Soldiers* in the fair distribution scenarios, (2-sided Fisher's exact test, p = 0.547), but punish significantly more than *Non-Soldiers* in the direct unfairness scenario (2-sided Fisher's exact test, p = 0.005). In the unfair distribution scenario, generated by computer or a coin flip, punishment behavior between *Soldiers* and *Non-Soldiers* is weakly significantly different (2-sided Fisher's exact test, p = 0.080).

Self-selection into a military career seems to be an important determinant of action when looking at irrational punishment. Our data demonstrates that in the unfair distribution scenarios *Soldiers* are more likely to engage in punishment of their co-players than *Non-Soldiers*. If we assume that subjects have F-S preferences, a higher proportion of *Soldiers* (than *Students* or *Frisbees*) must have an $\alpha_C > 3/2$.

Table 8: Marginal effects of probit regressions for the direct unfairness scenario

	Model (1)	Model (2)	Model (3)	Model (4)
Soldier	0.333***	0.342***	0.370***	0.365***
Chance	-	0.021	0.021	0.021
Male	-	-	-0.111	-0.115
Impatience Measure	-	-	-	0.029
N_C	100	100	100	99

Note: *** corresponds to a 1% significance level. N_C is the number of subjects C.

If we assume that subjects have B-D-S preferences, given that *Students* and *Soldiers* on average expect similar direct unfairness rates from the dictator (among Players C: 71% vs. 77%, respectively; Mann-Whitney test, p = 0.293), the frustration level $F_C(Unfair)$ experienced by *Soldiers* and *Non-Soldiers* is not much different. Therefore, for *Soldiers* to engage in more displaced aggression in the unfair scenarios, they must have on average a higher irritability parameter θ_C than *Non-Soldiers*. To punish a co-player by 1 point *Soldiers* are more likely to fulfil the $\theta_C > 1/5F_C$ condition.

4.5. Control Treatments

In addition to our main two treatments, we conducted another three different control treatments with the *Students* subject pool at the University of Kassel. Subjects were, similarly, recruited at economics lectures.

These control treatments check if punishment behavior is different when: 1) Potentially wrongdoing dictator (Player A) is not impune to punishment – *Retaliate* treatment; 2) Player A is a passive player, and the random chance determines the distribution – *Random* treatment; 3) Players C are informed of Player A's action prior to making their punishment decisions – *Hot* treatment.

Table 9: Retaliate and Random compared to Baseline; Hot compared to Chance

	Baseline	***Retaliate***	***Random***	***Chance***	***Hot***
Proportion Punishing if Dictator Unfair	29%	62.5%	33%	23%	44%
P-value for a 2-sided Fisher's exact test	-	0.017	1.000	-	0.187
N_C	24	32	6	26	16

Note: ***Students*** subject pool only. N_C is the number of subjects C.

1. *Retaliate*: same as *Baseline*, but Player A can also be punished.

Significantly more Players C punish if direct retaliation against potentially wrong-doing Player A is possible. (62.5% vs. 29% among *Students*). Hypothetical punishment rates in *Baseline* and *Chance* are similar to the actual punishment rate of 62.5% in *Retaliate*. When looking at all the observations we have in *Baseline*, when asked to imagine that Player A is not impune to punishment, 67% (24 out of 36) of Subjects C would have liked to punish. Hypothetical punishment (for *Students* and *Soldiers*) is also not much different, at 66%, in *Chance*.

Moreover, among all our subjects in these two main treatments (*Baseline* and *Chance*), 42% of non-punishers (28 out of 67) in the direct unfairness scenarios would have switched to punishment in that same scenario, if they could also punish Player A. (McNemar's (within-subject) exact test, p = 0.000).

2. *Random*: same as *Baseline*, but Player A is a passive player and distribution is randomly generated by the computer. All players are aware that distribution has a 50-50% chance of being fair or unfair.
   Punishment behavior in *Random* is no different from punishment behavior in *Baseline* or punishment behavior in response to randomly generated unfair distribution in the *Chance* treatment (2-sided Fisher's exact test, p = 1.000).
3. *Hot*: same as *Chance*, but Players C know in advance which action Player A has chosen.[26]
   Predictions of both B-D-S and F-S models predict the exact same behavior in *Hot* and *Chance*. When comparing the direct unfairness scenario between these two treatments, we expected more punishment in *Hot*, as unfair treatment here is not a possibility, but a certainty. Yet, although our results go in the direction predicted by our intuition, the proportion of punishers among the *Students* is not significantly different between the treatments (*Hot* vs. *Chance*).

[26] In *Hot*, all 8 dictators opted for the directly unfair distribution. Therefore, we cannot make any comparisons for the two computer-generated scenarios. These scenarios were answered as hypothetical situations. In the hypothetical fair scenario of *Hot*, 25% of Players C would have liked to punish. This is not significantly different from the non-hypothetical punishment rate in the fair scenario of *Chance* (2-sided Fisher's exact test, p = 0.711).

5. Discussion and Conclusion

Table 10: Summary of Theory Comparisons - Predictions and Results

Model	Inter-Treatment Comparison		Within-Subject Comparison		Punishment of Pl. C (any scenario)		Punishment in Fair scenarios	
B-D-S	Baseline > Chance	✗	Baseline ≈ Chance	✓	Yes	✓	No	✗
F-S	Baseline = Chance	✓	Baseline = Chance	✓	No	✗	No	✗
F-S + Status	Same as F-S	✓	Same as F-S	✓	Yes	✓	Yes	✓

We see that a large proportion of our subjects (63% in the main treatments) are rational. For the remaining 37%, a variation of the F-S model, the F-S + Status, seems to best explain our experimental evidence. Punishing subjects are inequity-averse in all the scenarios, where they can be in a disadvantaged position. Punishment that aims to close the payoff gap between themselves and the punishable "rich" Player B is not considerably different between the main two treatments or between scenarios within *Chance*.

Yet, a sub-group of those "disadvantage-averse" subjects choose to create inequality when they are not at a disadvantage. Punishments in the fair scenarios and punishment of the other Player C in the unfair scenarios cannot be explained neither by standard theory, nor by either of the theories we considered. It seems unreasonable to invest resources into punishment of "equals", unless one wants to be better off than them.

Possibly, the unfair power structure of our game activated among subject an updated social norm of unfairness. Maybe, the possibility of being treated unfairly, especially without adequate options for retribution, and correct realization that such treatment is very likely, awakened subject's not even selfish, but spiteful side.

We are not the first study to observe spiteful behavior in the lab. For example, already in 2011, Abbink and Herrmann introduced the mini- JOD (Joy of Destruction) game, where subjects could exercise costly punishment to destroy the endowment of another player, although these subjects had no previous interaction, let alone – any conflict. In their baseline treatment, about 10% of subjects engage in such spiteful punishment. The punishment rate of 17% in our fair scenario seems somewhat similar.

<u>*Summary of results:*</u>

- 1st research question: A sizable proportion of subjects pay for the "sins" of the unfair impune dictator.
- 2nd research question: Dictator's *unkindness* level does not seem to play a role.
- 3rd research question: Punishers seem to be equally reactive to unfavorable circumstances and unfair treatment directed at them: *bad luck* vs. bad intentions, respectively.
- 4th research question: Association with the military sub-culture seems to affect punishment behavior in unfair scenarios.

The motive for irrational punishment as a way of "letting off steam" has been overlooked in the previous literature. It would be interesting to study the role of displaced aggression in future experiments. To differentiate between inequity-averse and "emotional outbursts" type of displaced aggression, it would be interesting to conduct an experiment where passive player's exact endowment is unknown to the potential punisher. For example, a two-stage game, where first stage is a simple two-person dictator game, and in the second stage the recipient in the dictator game can pay to punish a passive player, whose endowment is randomly generated and is unknown to the other players.

References

Abbink, Klaus; Herrmann, Benedikt (2011): The Moral Costs of Nastiness. In *Economic Inquiry* 49 (2), pp. 631–633.

Bartling, B.; Fischbacher, U. (2012): Shifting the Blame. On Delegation and Responsibility. In *The Review of Economic Studies* 79 (1), pp. 67–87. DOI: 10.1093/restud/rdr023.

Battigalli, Pierpaolo; Dufwenberg, Martin; Smith, Alec (2015): Frustration and Anger in Games. In *CESifo Working Paper Series No. 5258*. Available online at Available at SSRN: http://ssrn.com/abstract=2591839.

Bushman, Brad J.; Baumeister, Roy F.; Stack, Angela D. (1999): Catharsis, aggression, and persuasive influence. Self-fulfilling or self-defeating prophecies? In *Journal of Personality and Social Psychology* 76 (3), pp. 367–376. DOI: 10.1037/0022-3514.76.3.367.

Card, D.; Dahl, G. B. (2011): Family Violence and Football. The Effect of Unexpected Emotional Cues on Violent Behavior. In *The Quarterly Journal of Economics* 126 (1), pp. 103–143. DOI: 10.1093/qje/qjr001.

Dollard, John; Miller, Neal E.; Doob, Leonard W.; Mowrer, O. H.; Sears, Robert R. (1939): Frustration and aggression. New Haven, CT, US: Yale University Press.

Fehr, E.; Schmidt, K. M. (1999): A Theory of Fairness, Competition, and Cooperation. In *The Quarterly Journal of Economics* 114 (3), pp. 817–868. DOI: 10.1162/003355399556151.

Fehr, Ernst; Gächter, Simon (2000): Cooperation and Punishment in Public Goods Experiments. In *American Economic Review* 90 (4), pp. 980–994. DOI: 10.1257/aer.90.4.980.

Fessler, Daniel M. T. (2006): The Male Flash of Anger: Violent Response to Transgression as an Example of the Intersection of Evolved Psychology and Culture. In Jerome H. Barkow (Ed.): Missing the Revolution: Oxford University Press, pp. 101–118.

Fischbacher, Urs (2007): z-Tree. Zurich toolbox for ready-made economic experiments. In *Exp Econ* 10 (2), pp. 171–178. DOI: 10.1007/s10683-006-9159-4.

Marcus-Newhall, Amy; Pedersen, William C.; Carlson, Mike; Miller, Norman (2000): Displaced aggression is alive and well. A meta-analytic review. In *Journal of Personality and Social Psychology* 78 (4), pp. 670–689. DOI: 10.1037/0022-3514.78.4.670.

Chapter 3

Michailova, Julia; Bühren, Christoph (2015): Money Priming and Social Behavior of Natural Groups in Simple Bargaining and Dilemma Experiments. In *MAGKS Joint Discussion Paper Series in Economics 30-2015*.

Sánchez-Jankowski, Martín (1991): Islands in the street. Gangs and American urban society. Berkeley: University of California Press.

Wiens, Marcus; Johannemann, Kirsten; Morasch, Karl; Hofmann, Martin (2015): Offizier und Gentleman? Eine experimentelle Untersuchung berufsbezogener Normen am Beispiel des Offiziers. In *Volkswirtschaftliche Diskussionsbeiträge, Universität der Bundeswehr München*.

CHAPTER 4

Does “Negative Nancy” Compete Like “Positive Pete”? An Experiment

Abstract

This paper explores how differences in self-attribution tendencies and gender affect competitiveness on a neutral task. Compared to men, women tend to ascribe losses to deficient skills rather than bad luck or a poorly chosen strategy. Nevertheless, women do not shy away from competition on a neutral task. Even highly self-critical women embrace competition.

1. Introduction

It might not be a coincidence that the common colloquialism contrasts "Positive Pete" to "Negative Nancy" and not negative Nathan or Nick. The current study, firstly, investigates differences in self-attribution tendencies for previous success / failure between genders.

Subjects who attribute their success to potentially fluctuating factors and subjects that attribute their failure to lack of ability[27] are defined as *negative*. In contrast, subjects who attribute their success to their own skill or talent and subjects that attribute their failure to a potentially fluctuating factor are defined as *positive*. Emanating from these definitions, negative subjects should be less likely to believe that they will succeed in the future. Therefore, (secondly), the hypothesis that negative subjects are less competitive than positive subjects is tested.

The present study is based on a trivia quiz shown to be gender-neutral both in performance and performance confidence (by Michailova and Katter, 2014). I chose a neutral task to set apart the potential effects of differences in self-attribution tendencies and gender on one's competitive disposition. However, an interaction of these variables is also worth investigating: possibly only women who ascribe their losses to lack of ability and victories to another source choose to shy away from competition on a gender-neutral task.

The results of the present study show that women are equally competitive on a gender-connotation-free task.[28] The factor most closely correlated with competitiveness is the previous personal record of victory or loss on that task. Explanations for those victories and losses, however, do not affect competitiveness. For example, a large proportion of first-round losers (67.6%), who ascribe their losses to lack of skill or talent, nonetheless choose to compete on the neutral task in the second round, and women are just as likely as men to do so (69.2% vs. 63.6%, respectively).

The current study confirms the claim that "Nancy" is more negative than "Pete," yet finds that even "Negative Nancy" competes like "Positive Pete" on a gender-neutral task.

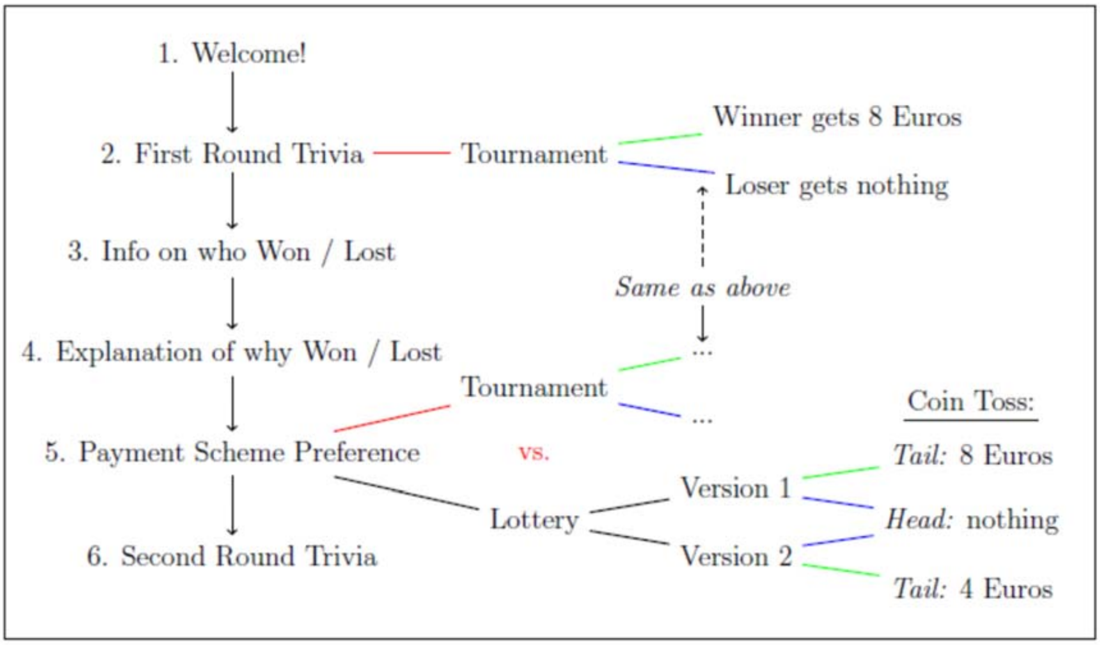
[27] Ability attributions invoke relatively permanent aspects of the self (Platt, 1988). In contrast, factors, such as luck, invested effort, or strategy choice are relatively volatile.

[28] This is in line with Grosse et al. (2014): men are significantly more likely to choose a competitive payment scheme on a typically masculine summation task, whereas no significant differences in competitiveness are observed for a gender-neutral word-order task.

2. Experimental Design

The experiment was conducted in October 2015 at the University of Kassel. 200 first- semester economics students took part in the experiment during their student orientation week.

Figure 6: Relevant Design Steps



Once welcomed into the experiment, subjects were informed that they will take part in a two-person winner-takes-it-all tournament, where the winner is determined by the relative performance on a general knowledge trivia.[29] They were informed that more instructions will follow later.

After the first round of trivia, subjects were informed whether they have won or lost. The number of correct answers was not disclosed, so that subject would not know how well they have performed on the task in absolute terms.

Subjects were then asked to choose from a multiple-choice question the most suitable explanation of why they think they have won or lost in the tournament. They were presented with two ability-related explanations: skill and intelligence. The remaining options on the multiple-choice

[29] If the two competitors answer correctly an equal number of trivia questions, a "winner" is chosen randomly. Winners earn 8 Euros, losers – nothing.

presented the following common, potentially fluctuating explanations: luck, strategy, effort, feeling tired / rested, feeling nervous / relaxed.

Thereafter, subjects were informed that they will take part in another round of trivia and must state their payment scheme preference for this round. They were presented with two payment alternatives: a winner-takes-it-all tournament or a lottery. Hence, competitiveness in the current paper is defined by subject's preference for a tournament over a lottery. A lottery, with a 50% chance of gain and 50% chance of zero payoff mimics the tournament's payoff structure and thereby can better differentiate between non-competitiveness and risk-aversion. This differs from the existing literature, where competitiveness is often defined as a preference for a tournament over a piece rate.

To check whether the preference for a tournament is sensitive to the expected value of its alternative, there were two versions of the lottery. About half of the subjects earn 8 Euros, if they get lucky (which is also the winning amount of the tournament). The other half earn only 4 Euros, if they get lucky.

Subjects were informed that, subsequent to the second trivia quiz and some additional questions, a die roll will determine their payment scheme for the second round, with their preferred payment scheme being implemented 5/6th of the time, if 1, 2, 3, 4 or 5 is rolled. All subjects had to take part in the second round of trivia. The possibility that even lottery-preferring participants would sometimes be paid according to a tournament encouraged all subjects to do their best on the second round of trivia, regardless of their stated payment scheme preference. Winning rates of tournament- vs. lottery-preferring subjects are not significantly different.[30]

Finally, the second round of trivia was played.

Please see Figure 6 for the flow chart of the relevant design steps.

[30] Two-sided Fisher's exact test p=0.476.

3. Results

Figure 7: Descriptive Statistics of Stated Explanations

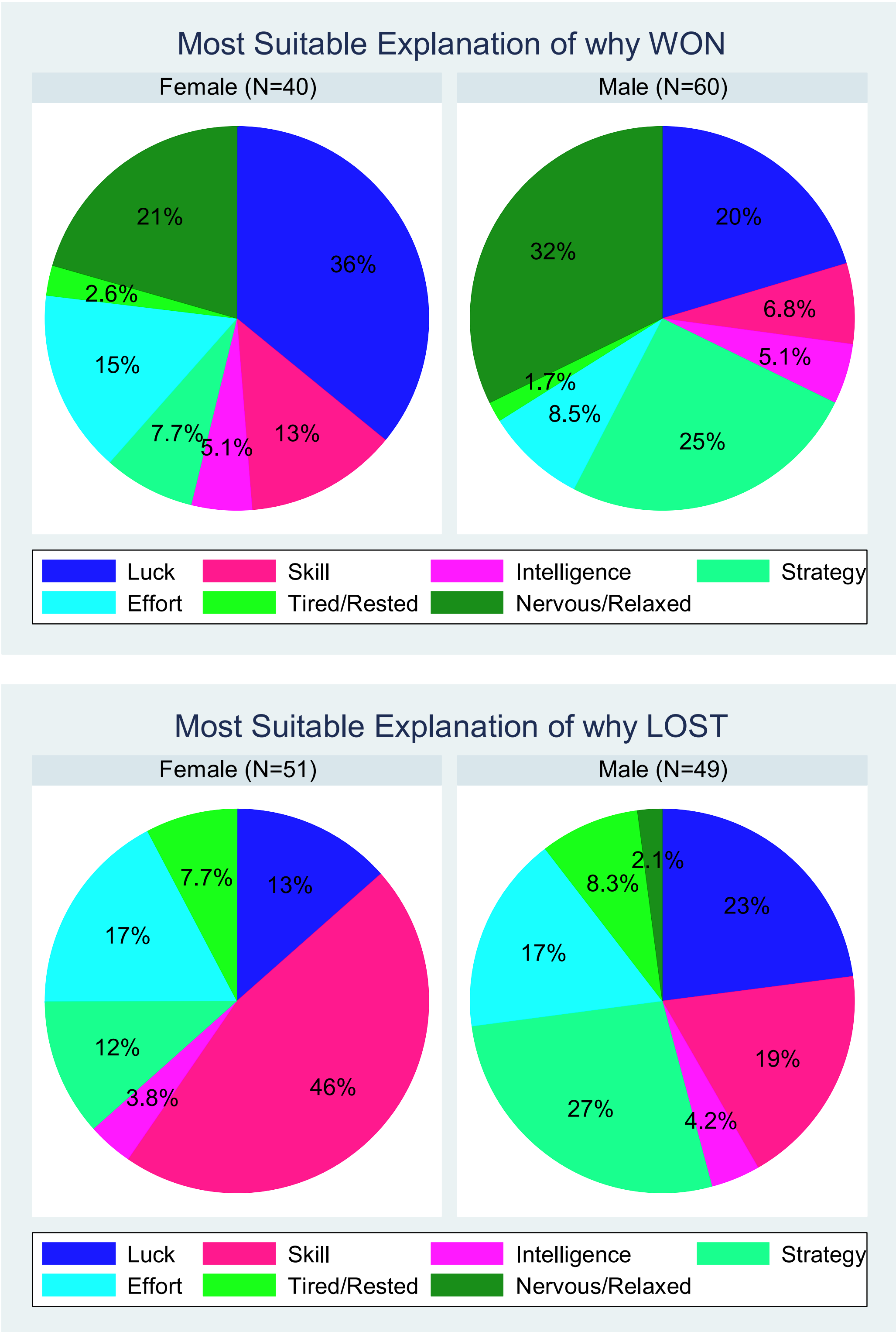


Figure 7 displays the descriptive statistics of the chosen explanations by gender after the first round of trivia. Number of winning and losing female and male participants are written above the pie charts.[31] When contrasting ability-related explanations (in pink- purple) to the other available alternatives (potentially fluctuating explanations), there are no significant gender differences in the victory domain, (Two-sided Fisher's exact test p-value=0.557).[32] In the loss domain, however, women are significantly more likely than men to attribute losses to one's own lack of ability (p=0.007). As is seen in the third pie chart, it is remarkable that after a loss, almost half of all the female participants have a critical skill-deficient self-image.

Table 11 shows the regression results on the dependent variable of interest: competitiveness. For simplicity, a *negativity* dummy is defined for ability attributions among losers and selection of one of the other possible explanations among winners. Neither negativity, nor gender, however, can predict competitiveness on a neutral task (Model 1). Interaction of gender and negativity cannot predict competitiveness on a neutral task, either (p=0.246).[33]

Table 11: Marginal effects after random effect probit model for preferring a Tournament

	Model 1	Model 2
Negativity	0.008 (0.060)	
Ability-attributed Loss		-0.071 (0.105)
Male	0.016 (0.055)	-0.024 (0.100)
Won before	0.205*** (0.060)	
Inferior Lottery	0.202*** (0.060)	0.341*** (0.095)
N	198	100
LL	-84.6	-55.6
Chi2	27.25	12.67

Notes: Standard errors in brackets. Levels of significance: *** p < 0.01, ** p < 0.05, * p < 0.1.

Also resilience on a neutral task, defined as the willingness to compete after a loss, does not seem to depend on gender (Model 2). In contrast, on a quantitative task, Buser (2016) finds that post-

[31] There are no significant gender differences in the winning rate. Two-sided Fisher's exact test p=0.155.
[32] In the rest of the section, p signifies p-values of a two-sided Fisher's exact test.
[33] Chi-square test with 3 degrees of freedom, p=0.267.

loss male subjects select a more challenging and competitive task and female subjects a less challenging and competitive task.

Competitiveness on a neutral task is well-predicted by the personal record of previous victory or loss and the expected value of the alternative payment scheme. Moreover, whereas first-round winners prefer a tournament for the second round regardless of the lottery payoff structure ($p=0.478$), first-round tournament-losers are sensitive to the expected value of the lottery ($p=0.001$). Hence, a risky alternative with a considerably inferior expected value could pressure people to prefer a tournament, even if they have lost before.

4. Conclusion and Outlook

In summary, "Nancy" in my study is undoubtedly more negative than "Pete." This matches a meta-analysis of 266 studies in social psychology, which illustrates that men are more likely than women to demonstrate a self-serving bias (Mezulis et al., 2004).[34] Possibly, women make negative self-attributions due to the common gender expectation of modesty (Daubman et al., 1992).

Yet, it seems that negative self-attributions and competition aversion do not necessarily go hand-in-hand. Neither negative individuals in general, nor negative women in particular, compete less on a neutral task.

In contrast, on a quantitative task women shy away from competition, despite no actual disadvantage in performance (Niederle and Vesterlund, 2007). It is, however, important to note that the stereotype that women have inferior math abilities is deeply rooted in our society[35] and might be responsible for the differences in competitiveness observed in Niederle and Vesterlund (2007) and its numerous replications.

Another potential cause of such divergent results might be the difference in the definition of competitiveness: preferring a tournament over a lottery vs. a piece rate. Hence, possibly the gender

---

[34] A self-servingly biased person attributes positive outcomes to his/her own character and negative outcomes to external factors.
[35] Reuben et al. (2014) verify that this stereotype exists independent of one's own gender or performance and persists despite reliable information on previous performance.

differences observed in the existing literature are due to the differences in risk-aversion, rather than competitiveness.

Further research is needed to check if differences in self-attribution tendencies can explain gender differences in competitiveness[36] on a stereotypically male task.

[36] As it is defined in the present study.

References

Buser, T., 2016. The Impact of Losing in a Competition on the Willingness to Seek Further Challenges. *Management Science* doi:10.1287/mnsc.2015.2321.

Daubman, K.A., Heatherington, L., Ahn, A., 1992. Gender and the self-presentation of academic achievement. *Sex Roles* 27, pp.187–204. doi:10.1007/BF00290017.

Grosse, N., Riener, G., Dertwinkel-Kalt, M., 2014. Explaining Gender Differences in Competitiveness: Testing a Theory on Gender-Task Stereotypes. SSRN Electronic Journal doi: 10.2139/ssrn.2551206.

Mezulis, A.H., Abramson, L.Y., Hyde, J.S., Hankin, B.L., 2004. Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias. *Psychological Bulletin* 130, pp.711–747. doi: 10.1037/0033-2909.130.5.711.

Michailova, J., Katter, J.K., 2014. Quantifying overconfidence in experimental finance. *International Journal of Behavioural Accounting and Finance*. doi:10.1504/IJBAF.2014.067629.

Niederle, M., Vesterlund, L., 2007. Do Women Shy Away From Competition? Do Men Compete Too Much? *The Quarterly Journal of Economics* 122, pp.1067–1101. doi:10.1162/qjec.122.3.1067.

Platt, C.W., 1988. Effects of causal attributions for success on first-term college performance: A covariance structure model. *Journal of Educational Psychology* 80, pp. 569–578. doi:10.1037/0022-0663.80.4.569.

Reuben, E., Sapienza, P., Zingales, L., 2014. How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences of the United States of America* 111, pp.4403–4408. doi:10.1073/pnas.1314788111.

CHAPTER 5

Institution Formation and Cooperation with Heterogeneous Agents

Abstract

Driven by an ever-growing number of studies that explore the effectiveness of institutional mechanisms meant to mitigate cooperation problems, recent years have seen an increasing interest in the endogenous implementation of these institutions. In this paper, we test within a unified framework how the process of institution formation is affected by three key aspects of natural environments: i) heterogeneity among players in the benefits of cooperation, ii) (a)symmetry in players' institutional obligations, and iii) potential trade-offs between efficiency and equality in payoff allocations. We observe social preferences to be limiting the scope for institution formation. Inequality-averse players frequently object to institutions that fail to address differences in players' benefits from cooperation - even if rejecting the institution causes monetary losses to all players. Relating our findings to previous studies on institution formation, we discuss potential advantages and drawbacks of stipulating unanimous support for implementing institutions that foster cooperation.

1. Introduction

"*[...] a set of rules used in one physical environment may have vastly different consequences if used in a different physical environment*"

(Ostrom, 1990, p.22).

Cooperation problems are ubiquitous in many areas in economics, ranging from teamwork or hold-up problems in managerial economics, over community governance or property rights security in development economics, natural resource management or climate protection in environmental economics, trade obstacles or treaty formation in international economics, to tax compliance and the provision of public goods in public economics. Each example certainly has its own distinctive issues, but when it comes to mitigating the underlying cooperation problems, there is usually a common approach: the modification of individuals' incentive-compatibility constraints, such that "free-riding" is no longer the dominant strategy (e.g., Shavell and Polinsky, 2000). These modifications (implicitly or explicitly) impose restrictions on individuals' choice sets, which raises the question whether they will be implemented in the first place (e.g., Gürerk et al., 2006; Tyran and Feld, 2006; Kosfeld et al., 2009; Bierbrauer and Hellwig, 2011; Markussen et al., 2014). In the present paper, we will shed light on this central question – asking in particular to which extent (i) the heterogeneity of the involved players and (ii) the (a)symmetry of the restrictions affects their implementation.

Consider the following example that we use throughout the paper, namely the provision of a public good. If members of a society are perfectly identical and all benefit equally from overcoming this social dilemma, one might expect them to mutually agree on establishing an institution that eliminates the social dilemma.[37] However, controversies might arise when members are heterogeneous and have different stakes in overcoming the social dilemma. In particular, when equality considerations are taken into account, the exact content of the institution is key to successful implementation. Symmetric institutions, in which all members have the same obligations, might be rejected in favor of asymmetric institutions with member-specific obligations – even if this implies monetary losses for all members.

[37] Of course, expected benefits must exceed the costs of implementing the institution. Throughout the paper, we take this for granted by assuming the institution to be costless – notwithstanding that the case of positive costs would be interesting to study (e.g., Kamei et al., 2015).

To causally identify how institution formation is affected by selected aspects of natural environments, we conducted a series of laboratory experiments. The basic underlying game, a public-good game, is a prominent workhorse for studying cooperation problems. Each player receives an endowment and has to decide on its allocation between private consumption and contributions to a public good. Provision of the public good creates benefits for all group members and is socially efficient in terms of the sum of monetary payoffs.[38] However, the individual marginal return from the public good is below the marginal return from private consumption, such that free-riding incentives exist which jeopardize public good provision. To offer players the opportunity to endogenously mitigate the cooperation problem, we add an additional stage that is played prior to the public good game. At this first stage, players decide on implementing an institution using unanimity voting. If all players in the group vote in favor of the institution, they are committed to certain efficiency-enhancing contribution levels in the subsequent public good game.[39] If at least one player votes against the implementation of the institution, the regular public good game is played and each player can freely decide how much to contribute in the second stage.

Players in our setup thus start in the absence of institutions and subsequently decide on the implementation of a joint institution to foster cooperation. In such an initial, lawless state of nature that is characterized by sovereign players facing a social dilemma, it seems natural to use unanimity voting for deciding on the implementation of institutions.[40] In fact, unanimous decision-making is the easiest possible, if not the only, voting procedure that players do not have to explicitly agree upon prior to voting. It does not require players to give up sovereignty, since each player can veto any decision. This is different for non-unanimous voting rules, such as majority voting, where players need to forfeit part of their sovereignty and which therefore typically only emerges after a joint history of cooperation.[41]

---

[38] Throughout the paper, efficiency refers to monetary payoffs.

[39] One could also think of the institution as consisting of two elements: (i) It states a certain obligation for each player, i.e., the exact amount that he is required to contribute in the second stage, and (ii) it installs a deterrent sanctioning technology, i.e., players' contributions are monitored and a player receives harsh punishment when deviating from the required contribution. For reasons of simplicity, the second component is not an explicit part of the experiment. Instead, it is implicitly modeled by restricting a player's choice set in the second stage to the required contribution (see Kosfeld et al., 2009 or Gerber et al., 2013 for similar approaches).

[40] The idea of an initial state of nature that is characterized by sovereign agents in a lawless environment goes back to Rousseau (1762) and Hobbes (1651).

[41] Cooperation in past periods may foster trust and reciprocal behavior among players, which may make them willing to forfeit part of their sovereignty. To give just one example, international organizations, most notably the League of Nations as the precursor of what is now the United Nations, used to apply the unanimity voting rule for

Since our focus is on how institution formation is affected (i) by heterogeneity in players' benefits from cooperation, and (ii) by the (a) symmetry of obligations, we vary these factors in a controlled manner while fixing the decision rule to unanimity voting in all treatments. First, in some treatment conditions (Homogeneous types), all players are of the same type and, thus, receive the same benefits from the public good, while in other conditions (Heterogeneous), there are two types that differ in their marginal benefits. Second, we vary the content of the institution. All players are either obliged to contribute their entire endowment to the public good (Symmetric institution), or obligations differ between the two player types (Asymmetric). While the symmetric institution implies efficient public good provision, but inequality in payoffs for heterogeneous players, obligations in the asymmetric case are chosen such that final payoffs are equalized. This setup allows us to clearly identify the roles of inequality aversion and efficiency concerns in the process of institution formation.

We find that inequality considerations can hamper the formation of efficient institutions meant to foster cooperation. With heterogeneous player types, those with low marginal benefits frequently object to the symmetric institution (about 40% reject it). The same is observed for homogeneous player types with asymmetric institutions (about 45% reject it). On the other hand, support is high when the institution implements equal payoff allocations: the asymmetric institution seems perfectly acceptable for heterogeneous player types, as does the symmetric institution for homogeneous types. In both cases, more than 90% of all votes are in favor of the implementation.

With respect to the sum of monetary payoffs, we observe that efficiency is always lower when institution formation failed than when the institution was implemented. The symmetric institution for homogeneous player types performs best (average efficiency is above 90% of the maximally obtainable sum of payoffs). Compared to this, under heterogeneity both the symmetric and the asymmetric institution lead to lower rates of efficiency, albeit for different reasons. In the former case, average efficiency is lower because the symmetric institution is frequently rejected. In the latter case, heterogeneous player types frequently implement the asymmetric institution, but average efficiency is lower since total obligations and the level of public good provided are lower. The asymmetric institution for homogeneous players performs worst.

voting on matters of substance before World War II. It was only during the post-war growth in international coordination through permanent organizations that non-unanimous voting rules were increasingly applied.

The striking differences in average efficiency and implementation rates between treatments underline at least three important issues. First, our results stress that inequality-aversion can have a strong impact on the process of institution formation. In most of the existing studies on institution formation, introducing social preferences to the theoretical models usually leads to stronger support for the institution; be it because more players want to be part of a coalition than is predicted under standard preferences (e.g. Kosfeld et al., 2009; McEvoy et al., 2015), or because the institution to be implemented allows them to reduce free-riders' payoffs (e.g., Markussen et al., 2014). By contrast, in those cases where inequality-aversion makes a difference in our setup, inequality-averse players are predicted to be less inclined to support the formation of the institution – a phenomenon that has not been discussed so far in the corresponding literature. As can be seen in our data, this easily leads to situations where players forego monetary payoffs by objecting to efficient institutions; in particular given the requirement of unanimous decisions.

However, and this is the second point we would like to stress, the use of unanimity voting for implementing institutions must not always be detrimental to efficiency. On the contrary, it can even help to foster cooperation.[42] Already Wicksell (1964) discusses that institutions based on unanimity or consensus voting can be ideally suited to overcome the canonical problem of free-riding. Unanimity makes individual activism implicitly conditional on the activism of all other parties involved. This mitigates the dilemma of institution formation: those who agree on implementing an institution do not face the subsequent risk of free-riding by non-supporting, and thus non-participating, players (see also Maggi and Morelli, 2006). Consequently, there is no drawback in supporting institutions that are based on unanimous decisions; either all players participate and the institution is formed, or the institution is not created at all. This can be clearly seen when comparing our data to related studies that implicitly allow players to "opt out" of institutions (Kosfeld et al., 2009; Gerber et al., 2013). While efficient and equitable institutions are frequently not implemented in those other studies, we observe that such institutions receive strong support and are implemented most of the times when unanimity is required.

Of course, this is not to say that unanimity will lead to stronger cooperation all the time. The

[42] Apart from this, there is also another desirable feature of unanimity. It is easy to agree on a principle of unanimity, since every party has veto power and freedom of choice is thus granted (at least ex ante, before an institution is implemented). Moreover, recent evidence implicitly suggests that many people value unanimous decisions, and that they have a strong preference for involving all players in the decision-making process (see Decker et al., 2003; Sutter et al., 2010, or Linardi and McConnell, 2011, and the references therein).

unanimity voting rule grants de facto veto rights to every party involved. Therefore, it is crucial that the institution to be voted on addresses idiosyncratic interests amongst the involved parties. We see this in our study, since homogeneous players frequently reject asymmetric institutions, and heterogeneous players regularly reject symmetric institutions. Support for the latter is also found in lab experiments by Banks et al. (1988) and Kesternich et al. (2014), as well as in the survey evidence reported in Reuben and Riedl (2013). The importance of fixing appropriate institutional obligations beforehand is also reflected in the literature that studies homogeneous players' acceptance thresholds on minimum contribution requirements in public good games (Birnberg et al., 1970; Dannenberg et al., 2014; Rauchdobler et al., 2010). Taken together, the evidence strongly suggests that prior to the ultimate voting about the implementation of an institution, great care has to be taken ex ante in designing the institution.

The institution at hand is implicitly built around a centralized authority with a deterrent sanctioning technology, but also other institutional mechanisms could be implemented to foster cooperation (e.g., Apesteguia et al., 2013; Falkinger and Fehr, 2000; Andreoni and Gee, 2012). One could even think about implementing decentralized sanctioning regimes. Of course, the seminal papers by Ostrom et al. (1992), Fehr and Gächter, 2000, and Fehr (2002) started with the basic idea of mutual monitoring and punishment among the members of a group; focusing in particular on the question whether certain behavioral

norms can emerge, even in the absence of formal institutions with a centralized structure. Still, there are some studies where players do vote over the implementation of decentralized sanction regimes (Putterman et al., 2011; Markussen et al., 2014; Kamei et al., 2015). Those studies exclusively focus on majority voting and homogeneous agents. It might be interesting to reconsider their results in our setup with heterogeneous players, or to see how behavior would change when using unanimity voting procedures.

Finally, in particular the results of our two benchmark treatments, where homogenous or heterogeneous players face only a regular public goods game and are not given the option to implement an institution, directly add to a broad strand of literature on "asymmetric" public goods. This literature has different approaches to studying the impact of asymmetries on cooperation; most commonly by varying the ratio of costs to benefits between players of the same group (see, e.g., the seminal paper by Fisher et al., 1995, and the recent work by McGinty and Milam, 2013), or by introducing inequalities in players' endowments, be it explicitly, e.g., Cherry et al. (2005) or implicitly by using different action sets for different agents, e.g., Khadjavi et al. (2014). Just like in our data, where contributions are lower with heterogeneous agents but the difference falls short of being significant, the existing empirical evidence is mixed: with some studies reporting lower contributions in the presence of an asymmetry, some reporting higher contributions, and others finding no effect at all on cooperation levels (see, for example, Anderson et al., 2008, p.1014, for a detailed review of the findings). Maybe most notable for our context are the findings from Riedel and Schildberg-Hörisch (2013). In their setup, asymmetry is implemented by exogenously imposing requirements on minimum contribution levels that differ between the two (otherwise symmetric) players that form a group. These obligations are non-binding and are backed up by non-deterrent sanctions only. Still, the authors observe that non-binding obligations shape contribution behavior, but in contrast to our findings, the effect is only temporary and vanishes over time.

The outline of the paper is as follows. Section 2 describes the experiment design. In Section 3, behavioral predictions for subjects' behavior will be derived, using both standard and social preferences (inequality aversion). Section 4 presents and discusses the empirical results. Section 5 concludes.

2. Experiment

In natural environments, the complexity of the process of institution formation makes it particularly difficult to draw causal conclusions about the conditions under which institutions come into being. As a starting point, we therefore use the controlled environment of laboratory experiments to study central aspects of the endogenous formation of institutions. In this section, we present the design of our experiment and describe the implemented procedures.

2.1. Design

Our design builds on a standard public goods game (VCM game), a frequently used workhorse to study elements of social dilemmas in the lab (e.g., Isaac and Walker (1988)). Each player has a private endowment $E = 20$. Players simultaneously decide on the amount ci that they contribute to a public good, with $0 \leq c_i \leq E$, $i = 1, ..., n$. The benefits from the public good are enjoyed by all players, independent of their individual contribution ci. In some treatments, players are heterogeneous, i.e., not all players benefit from the public good to the same extent. To model heterogeneity, we allow the marginal per capita return (MPCR) γi from the public good to vary across players.[43] Given the contributions of all players (c_1, ..., c_n), player i's material payoff π_i is thus given by

$$\pi_i = E - c_i + \gamma_i \sum_{i=1}^{n} c_i.$$

In all treatments, parameters for γ_i are chosen such that players face a social dilemma. Efficiency, defined as the sum of payoffs of all players, is maximized if all players contribute their entire endowment. Yet, from an individual perspective, each player's material payoff is maximized by not contributing to the public good, regardless of the other players' contributions. Formally, this implies $\sum_{i=1}^{n} \gamma_i > 1$ and $\gamma_i < 1$, $\forall i$.

We form groups of three players (n = 3). Between treatments, we vary two components. First, we vary the composition of players' types γ_i. In some treatments (HOM), players are

homogeneous, i.e., all players are of the same type and thus receive the same benefits from the public good (γ_i = 2⁄3). In other treatments (HET), players are heterogeneous: two players have a high return from the public good (γ_i = 3⁄4) and one player has a low return (γ_i = 1⁄2).[44] The different marginal per capita returns are chosen as to keep total efficiency gains constant between treatments (2 · 3⁄4 + 1 · 1⁄2 = 3 · 2⁄3).

Second, we vary availability and content of the institution. In the benchmark treatments (VCM), there is no institution formation stage and players play a regular public goods game. In the main treatments, there is an institution formation stage first, followed by a contribution stage. In the institution formation stage, a single institution is available and can be implemented via unanimity voting, i.e., the institution is implemented if and only if all players vote in favor of adopting the institution. If the institution is rejected, the regular public goods game without any restrictions on contributions is played. The institution states each player's obligation $\overline{c_i}$, the amount that each

[43] To give just two among many possible examples, nation states differ in their benefit from climate protection or researchers at different stages of their career benefit from joint publications to a different extent.

[44] We choose a single player with a lower return because this setup is sufficient to illustrate the potential weakness of unanimity voting, i.e., already a single player can prevent successful institution formation by vetoing.

player has to contribute to the public good in the second stage if the institution has been implemented. Voting and the implementation of the institution are costless.[45]

The main treatments vary in the type of institution that is available. In general, treatments are designed to reflect a tradeoff between efficiency and equality of payoffs. In treatments with the symmetric institution (SYM), all players are obliged to contribute their entire endowment to the public good if the institution has been implemented. The symmetric institution maximizes the sum of payoffs of all players and, thus, induces the efficient outcome. In treatments with the asymmetric institution (ASYM), one player is required to contribute 8 units, while the two others are obliged to contribute all 20 units to the public good. In treatments with heterogeneous players, the obligation is 20 for the high types, and 8 for the low types. Obligations are chosen such that the asymmetric institution implies equal payoffs for both types of players (36 each), which comes at an efficiency cost. In contrast, with heterogeneous players, the symmetric institution implies inequality in final payoffs (45 for the high types and 30 for the low type). If the asymmetric institution is combined with homogeneous players, one randomly chosen player has to contribute 8 units, while the other two players are obliged to contribute 20 units. The design results in the 2 x 3-treatment matrix shown in Table 12.

Table 12: Treatments

	VCM	SYM	ASYM
HOM	$\gamma_i = (2/3, 2/3, 2/3)$ $\bar{c}_i = (n/a)$	$\gamma_i = (2/3, 2/3, 2/3)$ $\bar{c}_i = (20, 20, 20)$	$\gamma_i = (2/3, 2/3, 2/3)$ $\bar{c}_i = (20, 20, 8)$
HET	$\gamma_i = (3/4, 3/4, 1/2)$ $\bar{c}_i = (n/a)$	$\gamma_i = (3/4, 3/4, 1/2)$ $\bar{c}_i = (20, 20, 20)$	$\gamma_i = (3/4, 3/4, 1/2)$ $\bar{c}_i = (20, 20, 8)$

2.2. Procedures

The computerized experiments (using z-Tree; Fischbacher, 2007) were run at the BonnEconLab of the University of Bonn, Germany in 2012. Student subjects were recruited randomly from all majors (using Orsee; Greiner, 2004) and were randomly assigned to one of the six treatments

[45] These are simplifying assumptions. Qualitatively, the theoretical predictions do not change as long as the gains in individual material payoffs due to implementing the institution outweigh the individual implementation costs.

(between-subject design). For each treatment, we ran two sessions with 24 subjects each. In each session, subjects first received written instructions. To create common knowledge, instructions were read out aloud to the subjects. Afterwards, subjects answered a set of control questions and could pose clarifying questions to ensure understanding of the game's structure and payoffs. Subjects then played the game repeatedly for 20 periods. Interaction took place within the same group of three subjects (partner matching protocol), it was anonymous and decisions were taken in private at the computer. After each voting stage, subjects received feedback on the voting result and the voting behavior of the other two subjects in their matching group. After each contribution stage, subjects were informed about their own payoff and the payoffs and contributions of the other two subjects in their group. After all 20 periods, subjects answered a questionnaire covering socio-demographic characteristics. Each session lasted about 80 min. Accumulated earnings were converted at a rate of 40 tokens = 1 Euro. Total earnings per subject ranged between 10 Euro and 22.5 Euro, with an average of about 16.4 Euro.

Altogether, we had 282 subjects, and observations on 5640 individual decisions. Given the allocation of subjects to the six treatments, repeated interaction in 20 periods and matching groups of 3, we have 16 independent observations per treatment.[46] 39% of our subjects are male, their age ranges from 16 to 42, with an average age of 22 years.

3. Behavioral Predictions

For each treatment, we characterize players' equilibrium behavior under two alternative assumptions concerning the shape of the utility function. First, we assume that each player's utility function coincides with the monetary payoff of the game, π_i, i.e., that players have standard preferences. Second, we assume that (at least some) players have social preferences. From the large set of available approaches to social preferences, including, among many others, Bolton and Ockenfels (2000) or Charness and Rabin, 2002, we use the specification from Fehr and Schmidt (1999): in addition to valuing own monetary payoff, a player suffers from inequality in monetary payoffs, i.e., from others being worse or better off than himself. In our treatment with

[46] Exceptions are treatments HET-VCM and HOM-ASYM, for which we have 15 independent observations since some subjects did not show up.

heterogeneous benefits from the public good, players might vote against implementing an institution that obliges all players to contribute equally to the public good in order to avoid inequality in payoffs. In the remainder of this section, we will provide an intuition for the behavioral predictions for each treatment under the two alternative assumptions on the shape of players' utility functions using the parameters of our design.

Table 13 summarizes the behavioral predictions for players with standard preferences. In basic VCM games, they are predicted not to contribute to the public good at all. Whenever $\gamma_i < 1$, contributing does not pay off from an individual perspective. Condition $\gamma_i < 1$ is met for all players in treatments HOM-VCM ($\gamma = 2/3$) and HET-VCM ($\gamma_l = 1/2$ and $\gamma_h = 3/4$).

Table 13: Behavioral predictions based on standard preferences

		VCM	SYM	ASYM
HOM	voting	-	implement institution	implement institution
	contribution	$c = 0$	$c = 20$	$c(\bar{c} = 20) = 20,\ c(\bar{c} = 8) = 8$
HET	voting	-	implement institution	implement institution
	contribution	$c_h = c_l = 0$	$c_h = c_l = 20$	$c_h = 20,\ c_l = 8$

In all two-stage treatments, predictions are derived using backward induction. Let U^{INST} denote utility when the institution has been implemented, with INST = SYM for the symmetric and INST = ASYM for the asymmetric institution. In the contribution stage, players will compare the utility they receive with the respective institution being in place, U^{INST}, to the utility of the VCM game that is played if the institution has not received unanimous support in the voting stage, U^{VCM}. Unanimity voting ensures that, whenever $U^{INST} \geq U^{VCM}$, it is a best response to the voting behavior of the other players to vote in favor of the institution. If all other players also vote in favor of implementing the institution, the institution will be implemented and the player's preferred outcome is achieved. If, in contrast, at least one other player votes against implementing the institution, the institution will not be implemented and the VCM game will be played. However, the approving player is still equally well off as if he had voted against implementing the institution. Whenever $U^{INST} < U^{VCM}$, a player will vote against installing the institution. In our design, $U^{INST} > U^{VCM} = E = 20$ for all player types in treatments HOM-SYM, HOM-ASYM, HET-SYM and HET-ASYM (see payoffs in Table 12). Consequently, for all treatments, players with standard preferences are predicted to vote in favor of the respective institution. The institution will be implemented and players will contribute according to their individual obligation. To summarize, if players have standard preferences, unanimity voting on the formation of institutions is predicted to help to overcome the social dilemma of public good provision. This result holds irrespective of whether players are homogeneous or heterogeneous and whether a symmetric or an asymmetric institution is voted on.

Table 14: Behavioral predictions based on Fehr-Schmidt preferences

		VCM	SYM	ASYM
HOM	voting	-	implement institution	type $\bar{c}=20$ rejects if α high, type $\bar{c}=8$ rejects if β high
	contribution	(c,c,c), $c \in [0,20]$ if $\beta_i > 1/3 \forall i$; (0,0,0) otherwise	$c=20$	if reject: as in $HOM-VCM$ otherwise: $c(\bar{c}=20)=20$, $c(\bar{c}=8)=8$
HET	voting	-	low type rejects if α_l high	implement institution
	contribution	$(c_h, c_h, c_l = 2/5c_h)$, $c_h \in [0,20]$ if $\beta_h > 2/7$ and $\beta_l > 2/5$; (0,0,0) otherwise	if reject: as in $HET-VCM$ otherwise: $c_h = c_l = 20$	$c_h = 20$, $c_l = 8$

Table 14 displays the behavioral predictions for players with social preferences in terms of inequality aversion (Fehr and Schmidt, 1999). If players have social preferences, there are multiple equilibria in treatment HOM-VCM.[47] The intuition is as follows: If all players are sufficiently averse to advantageous inequality (β sufficiently high),[48] they will exactly match the contribution level $c \in [0, E]$ of the other players to equalize payoffs. If players are not or only mildly averse to advantageous inequality (β low), the only equilibrium that remains is the one with zero contributions of all players. In treatment HET-VCM, the basic mechanism driving the existence of equilibria with positive contributions is the same. If all players are sufficiently averse towards earning more than others, they contribute positive amounts as soon as the other players contribute positive amounts to prevent an unequal payoff distribution. However, to achieve equal payoffs for all three players, the low type contributes less than the two high types.

In treatments HOM-SYM and HET-ASYM, assuming social instead of standard preferences does not change the predictions. In both cases, the proposed institution guarantees equality of payoffs while simultaneously maximizing utility of players who are sufficiently averse to unequal payoffs. Hence again, all players are predicted to vote in favor of the respective institution, it will be implemented, and players will contribute according to their obligation. In treatments HET-SYM and HOM-ASYM, however, predictions based on standard preferences and social preferences differ. In both treatments, players with standard preferences always support the formation of the institution as it offers a higher monetary payoff than the VCM and they do not suffer from unequal payoffs that arise from implementing the institution. In contrast, in treatment HET-SYM, low type players with social preferences who suffer sufficiently from being worse off than the high types (α sufficiently high), object to institution formation. They prefer a lower monetary payoff, but equal payoffs across players in the VCM, to a higher monetary payoff, but

disutility from inequality due to the symmetric institution being in place. Consequently, low type players drive rejections of the proposed symmetric institution. Similarly, in treatment HOM-ASYM, all players potentially have a motive for voting against the asymmetric institution that introduces inequality in payoffs: Players with an obligation of 8 tokens, if they are sufficiently averse to advantageous inequality, and players with an obligation of 20 tokens if they are

[47] The proof is provided in Fehr and Schmidt (1999).
[48] In the model of Fehr and Schmidt (1999), the parameter β captures the intensity of aversion to advantageous inequality, while the parameter α measures the degree of aversion to disadvantageous inequality.

sufficiently averse to disadvantageous inequality.[49]

4. Results

This section is structured along five sets of results concerning differences in voting and contribution behavior across treatments. All results are qualitatively in line with the behavioral predictions presented in Section 3, when assuming that at least some players are inequality averse to an extent that induces their behavior to deviate from the predictions based on standard preferences. We report corresponding results in the text below.

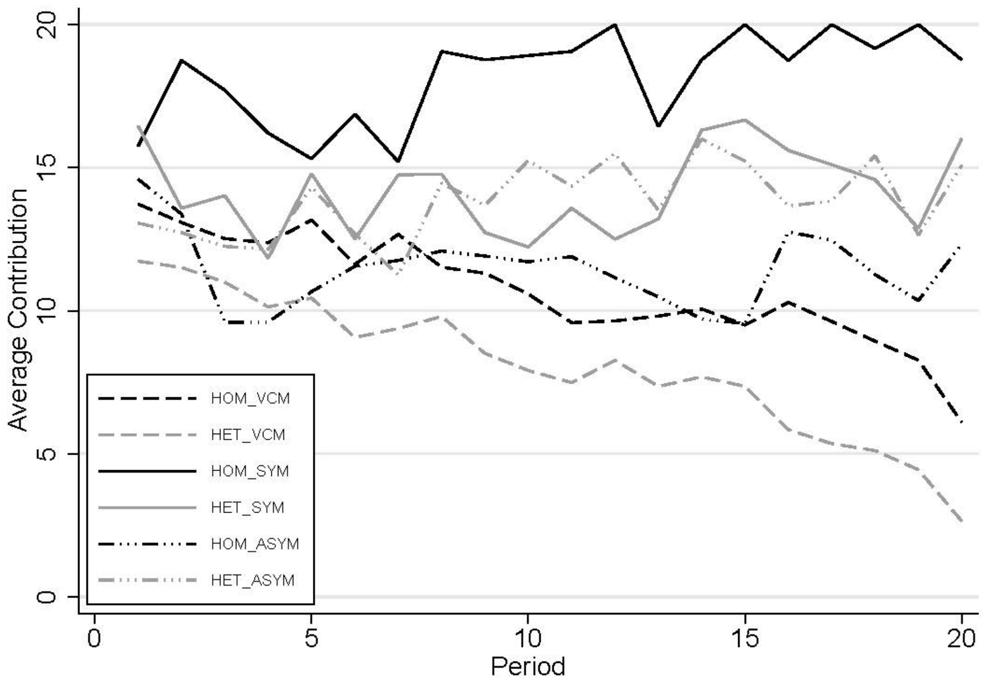
First, we will briefly present results in treatments HOM-VCM and HET-VCM that provide baseline scenarios for comparing whether unanimity voting on institutions increases efficiency. We proceed by discussing under which circumstances unanimity voting on symmetric or asymmetric institutions helps to increase public good provision. We thereby focus on treatment comparisons in which changes in behavior can be attributed to a single change in setup. That means, we either compare treatments with different institutions, while keeping constant the composition of player types (HOM or HET) or we compare treatments with a different composition of player types, while keeping constant the nature of the institution to be voted on (SYM or ASYM).

Tables 15 and 16 contain first descriptive results. Table 15 displays contributions averaged over all periods by treatment. Table 16 shows the share of affirmative votes and

implementation rates averaged over all periods by treatment. Moreover, Figures 8 and 9 display the treatment-specific development of contributions and share of affirmative votes over time.

[49] Preferences for efficiency, see, e.g., Charness and Rabin (2002), are an alternative explanation for rejecting an asymmetric institution. Efficiency seekers should reject institutions that do not induce full contributions in order to contribute more than they were obliged to with the institution being in place (expecting others to contribute more after rejection provides a reason for selfish agents to reject institutions, too). Charness–Rabin preferences predict that rejections of institutions are possible in all treatments. However, all predicted rejections require fairly large amounts of contributions in the subsequently played VCM to be justified. Yet, our data (see Section 4.4) does not indicate higher levels of contributions after an institution is rejected than under the institution, suggesting that efficiency seeking is not a predominant motive for rejections.

Figure 8: Development of average contributions over time



Note: In treatments HOM-VCM and HET-VCM, average contributions decrease over time (HOM-VCM: Spearman's Rho r = - 0:27, p < 0.01 and HET-VCM: r = - 0.47, p < 0:01). In treatments HOM-SYM and HET-ASYM, average contributions increase over time (HOM-SYM: r = + 0.27, p < 0.01, HET-ASYM: r = + 0.21, p < 0:01). In treatments HET-SYM and HOM-ASYM, time trends in contributions are not significant (HET-SYM: r = + 0.08, p = 0.16, HOM-ASYM: r = - 0.03, p = 0.64).

4.1. Baseline Treatments: Homogenous vs. Heterogeneous Players in the VCM

On average, contributions in the standard VCM tend to be lower with heterogeneous than with homogeneous agents. Subjects contribute 10.7 out of 20 units in treatment HOM-VCM and 8.1 units in treatment HET-VCM (Mann–Whitney ranksum test (MWU), p = 0.11).[50]

Result 1: Average contributions in treatment HET-VCM are slightly, but not significantly lower than in treatment HOM- VCM.

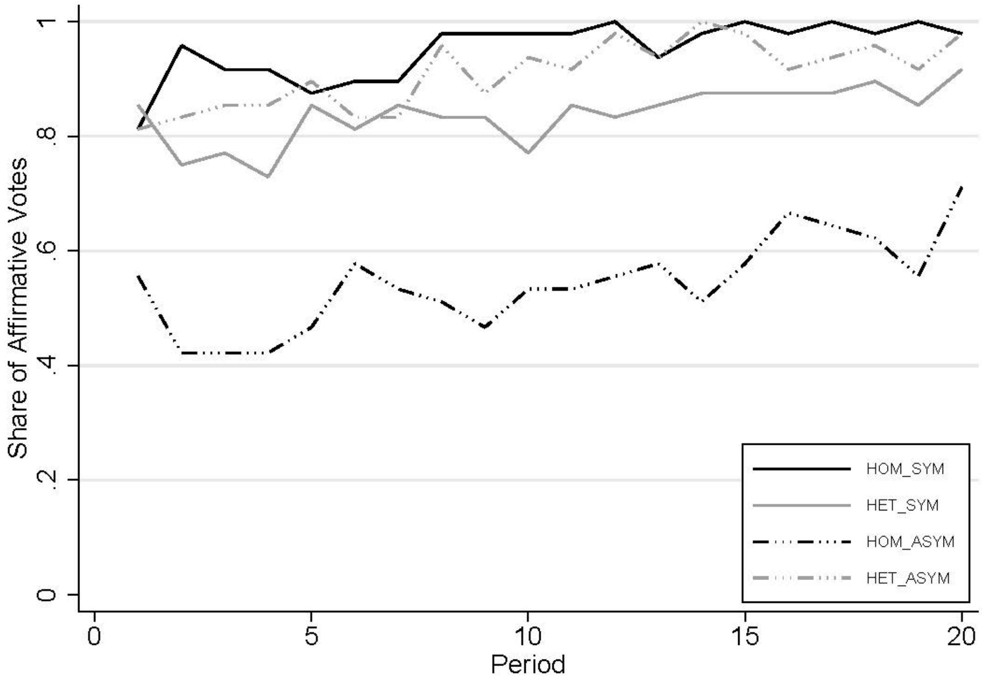
Moreover, we observe that average contributions of low and high types differ in HET-VCM: while low type players contribute only 5.3 units, high type players contribute 9.4 units on average. As a

[50] Throughout the paper, we report two-sided p-values. Each matching group's average contribution is one independent observation.

consequence, average payoffs for the two player types are similar, 26.8 and 28.7 units, respectively. Players of both types seem to intuitively strive for equal payoffs.

4.2. Unanimity Voting on the Symmetric Institution: Homogeneous vs. Heterogeneous Players

Figure 9: Share of affirmative votes over time



Note: In all four two-stage treatments, the share of affirmative votes increases over time (HOM-SYM: Spearman's Rho r = + 0.29, HET-SYM: r = + 0.18, HOM-ASYM: r = + 0.18, HET-ASYM: r = + 0.26, all $p < 0.01$).

Table 15: Average Contributions by Treatment

Institution / Player type		VCM	SYM	ASYM
HOM	overall	10.72	18.18	11.44
		(7.83)	(5.27)	(8.34)
	types $\bar{c} = 20$	–	–	12.12
		–	–	(8.79)
	types $\bar{c} = 8$	–	–	10.09
		–	–	(7.23)
HET	overall	8.05	14.21	13.85
		(6.56)	(7.79)	(7.20)
	high types	9.42	14.77	17.18
		(7.07)	(7.42)	(6.36)
	low types	5.33	13.08	7.21
		(4.33)	(8.38)	(2.90)

Note: Standard deviations are in parentheses.

We first consider the voting behavior of homogeneous players who are confronted with the decision whether to install the symmetric institution that obliges each player to contribute the efficient amount, 20 units. Overall, 95.2% of votes (914 out of 960 votes) are in favor of implementing the symmetric institution. As a result, in 86.6% of all cases, all three players of a group unanimously agree to implement the symmetric institution and it is indeed implemented.

Result 2: In treatment HOM-SYM, average contributions are significantly higher than in treatment HOM-VCM.

On average, subjects contribute 18.2 units in treatment HOM-SYM instead of 10.7 units in treatment HOM-VCM (MWU, $p < 0{:}01$). After some periods of initial learning efficiency is close to 100% (see also Fig. 8). To summarize, in our setup with homogeneous players, unanimity voting on the symmetric institution increases efficiency substantially.

Does unanimity voting on the efficient institution also yield high support if players are heterogeneous, i.e., if the efficient institution introduces unequal payoffs? Again, we start by analyzing behavior in the voting stage. In treatment HET-SYM, the overall share of affirmative votes is lower than in treatment HOM-SYM, 83.9% instead of 95.2% (MWU, $p = 0.01$). Heterogeneous players object the implementation of the efficient symmetric institution more often than homogeneous players. The difference in affirmative votes between treatment HOM-SYM and HET-SYM persists over time (see Fig. 9). Similarly, the overall implementation rate in treatment HET-SYM is 56.3%, substantially lower than in treatment HOM-SYM, 86.6% (MWU, $p = 0.01$). Rejections of the institution are largely due to the voting behavior of low types. In our data, 95.9% of high types vote in favor of implementing the institution in treatment HET-SYM, but only 59.7% of low types do.

As a consequence of the lower implementation rate, average contributions are significantly lower in treatment HET-SYM than HOM-SYM: 14.2 instead of 18.2 (MWU, $p = 0.01$). However, average contributions in treatment HET-SYM are significantly higher than in the VCM with heterogeneous players (MWU, $p < 0:01$). Result 3 summarizes our results for treatment HET-SYM.

Result 3:

1. In treatment HET-SYM, average contributions are significantly higher than in treatment HET-VCM.
2. In treatment HET-SYM, both implementation rate and average contributions are significantly lower than in treatment HOM-SYM.

Overall, if players are heterogeneous rather than homogeneous in their marginal returns from the public good, unanimity voting on the efficient institution does not always result in its successful implementation. Still, compared to the standard public good game in which no institution is available, unanimity voting on the efficient institution increases efficiency substantially – even if players are heterogeneous.[51]

[51] Although the focus of our paper is on distributive fairness, subjects' behavior could be driven by procedural fairness concerns, too. While economists have started studying the latter approach only lately (the first economic experiments are reported in Bolton et al., 2005; see also Krawczyk, 2011, for a theoretical model), the idea of procedural fairness has been prominent in psychology for some decades already (see, for example, Tyler and Lind, 2000 for a summary). Barrett-Howard and Tyler (1986) report that procedural fairness is equal in importance to distributive fairness for subjects who are confronted with allocation decisions. This could provide a potential explanation for the 84% approval rate in HET-SYM, namely if subjects think that equal obligations are procedurally fair in general and/or randomly assigning heterogeneity in the MPCRs to subjects is procedurally fair. Interestingly, if this line of reasoning indeed applies it seems to be done in a self-serving manner, because the support for the institution is much stronger among high types than among low types. We thank an anonymous referee for pointing this out.

Table 16: Share of affirmative votes and implementation rate by treatment

Institution Player type		SYM	ASYM
HOM	affirmative votes		
	overall	.95	.54
	types $\bar{c} = 20$	–	.48
	types $\bar{c} = 8$	–	.68
	implementation rate	.87	.27
HET	affirmative votes		
	overall	.84	.91
	high types	.96	.90
	low types	.60	.94
	implementation rate	.56	.77

Notes: The share of affirmative votes as well as the implementation rate is significantly higher in treatment HOM-SYM than in HET-SYM (MWU, $p = 0.01$ for both), in HOM-SYM than in HOM-ASYM (MWU, $p < 0{:}01$ for both), lower in HET-SYM than in HET-ASYM (MWU, $p = 0.16$ for the share of affirmative votes and $p = 0.08$ for the implementation rate), and significantly higher in HET-ASYM than in HOM-ASYM (MWU, $p < 0{:}01$ for both). The share of low types' affirmative votes in HET-ASYM is significantly higher than in HOM-ASYM (MWU, $p < 0{:}01$). The share of high types' affirmative votes in HET-ASYM is significantly lower than in HOM-ASYM (MWU, $p = 0.04$).

4.3. Unanimity Voting on the Asymmetric Institution: Homogeneous vs. Heterogeneous Players

A potential remedy to the frequent rejections of the symmetric institution by low type players is to design an asymmetric institution that ensures the maximum possible payoffs among the set of all equitable payoff allocations. Obviously, under the asymmetric institution, the low type players' obligation must be lower than under the symmetric institution. As a drawback, the implementation of the asymmetric institution results in a lower level of public good provision than the implementation of the symmetric institution.

Concerning results in treatment HET-ASYM, 91.0% of players vote in favor of implementing the asymmetric institution which results in 77.2% successful implementations. Thus, with heterogeneous players, the asymmetric institution that guarantees equal payoffs for both player types is more than 20% points more likely to be implemented than the symmetric one that induces the efficient outcome, but unequal payoffs across player types (MWU, p = 0.08 for the implementation rate and p = 0.16 for the share of affirmative votes). The higher implementation rate is due to the substantially higher likelihood of low types to vote in favor of the asymmetric institution than the symmetric one: 94.1% instead of 59.7% (MWU, p < 0:01). With 89.5%, the high types' share of affirmative votes for the asymmetric institution is only slightly lower than the 95.9% affirmative votes for the symmetric institution (MWU, p = 0.04).

While implementation rates differ markedly for treatment HET-SYM and HET-ASYM, average contributions do not: 13.9 units in HET-ASYM compared to 14.2 units in HET-SYM (MWU, p = 0.97).[52] There are two opposing effects that cancel each other out: while the higher implementation rate in HET-ASYM increases contributions, implementing the asymmetric institution instead of the symmetric one reduces contributions of the low types from 20 to 8 units. Compared to the benchmark VCM game with heterogeneous players, average contribution levels are significantly higher in treatment HET-ASYM than in treatment HET-VCM (MWU, p < 0.01). We summarize results for treatment HET-ASYM below.

Result 4:

1. In treatment HET-ASYM, the implementation rate is higher than in treatment HET-SYM. In contrast, average contributions in treatments HET-ASYM and HET-SYM are very similar.
2. In treatment HET-ASYM, average contributions are significantly higher than in treatment HET-VCM.

Overall, designing institutions that address players' demand for equal benefits from institution formation seems to be very successful in raising the implementation rate. In many contexts, a higher rate of institution formation could be considered beneficial per se, e.g., due to raising reliability of public good provision or by potentially triggering future institutionalized cooperation. However, increasing the implementation rate by voting on an asymmetric institution will always come at the cost of institutionalizing less than efficient levels of public good provision.

[52] In line with the theoretical predictions, we observe that in HET-ASYM low and high types are equally well off on average. High types' average payoff is 34.0, low types' average payoff is 33.6. In contrast, in treatment HET-SYM, average payoffs of low types are substantially lower than those of high types: 28.2 instead of 37.2. Again, this finding is in line with the theoretical predictions.

To rule out that the high implementation rate in HET-ASYM is due the asymmetry in contributions per se, we now turn to treatment HOM-ASYM. Here, we can explore how the asymmetric institution performs if players are homogeneous, i.e., when it introduces binding rules concerning contributions to potentially increase efficiency, but those rules induce unequal payoffs across players.

Proposing an asymmetric institution to homogeneous players receives relatively low levels of support. The average share of affirmative votes ranges between 40% and 70% over time, resulting in an average implementation rate of only 26.7%. For players with an obligation of 8 units, the share of affirmative votes is 67.7%, while it is 20 percentage points lower for those with an obligation of 20. This might be because both types of players possibly have a motive to vote against the institution, namely aversion to advantageous inequality (for players with an obligation of 8 units) and aversion to disadvantageous inequality (for players with an obligation of 20 units). The share of affirmative votes as well as the implementation rate is significantly higher in treatment HET-ASYM than in HOM-ASYM (MWU, $p < 0.01$ for both).

We have already shown that, with homogeneous players, proposing a symmetric institution helps to overcome the social dilemma of public good provision. This is not the case with an asymmetric institution. The average contributions in treatment HOM-ASYM are not significantly different from average contributions in treatment HOM-VCM (MWU, $p = 0.75$) and significantly lower than in treatment HOM-SYM (MWU, $p < 0.01$).

Finally, the asymmetric institution performs worse for homogeneous than for heterogeneous players, i.e., when it introduces inequality instead of addressing it. With homogeneous players, both the share of affirmative votes and the average contributions are lower (MWU, $p < 0:01$ for affirmative votes and $p = 0.04$ for contributions). This strongly suggests that the success of the asymmetric institution for heterogeneous agents is indeed due to addressing payoff inequalities between agents. Below, we summarize results for treatment HOM-ASYM.

Result 5:

1. Average contributions in treatment HOM-ASYM and HOM-VCM do not differ significantly.
2. In treatment HOM-ASYM, implementation rate and average contributions are significantly lower than in treatment HOM-SYM.

3. In treatment HOM-ASYM, implementation rate and average contributions are significantly lower than in treatment HET- ASYM.

4.4. Contributions by Institution Formation Status

So far, we have analyzed average contributions in a given treatment, averaging over cases of successful institution formation and those of failure to form an institution. We have not studied yet how failure to implement the proposed institution affects contribution levels. If motives for objecting to institution formation differ across treatments, contribution levels in case of failed institution formation could also differ across treatments. For example, inequality aversion could be a plausible motive for voting against institution formation in treatments HET-SYM and HOM-ASYM in which institutions induce unequal payoffs. In treatments HOM-ASYM and HET-ASYM, a preference for efficient levels of public good provision could drive rejections. Rejections of the institution are harder to rationalize in treatment HOM-SYM because implementation of the institution results in maximal and equal payoffs. Consequently, rejections could be due to, e.g., mistakes or pleasure from exerting (destructive) power. These motives could induce negative reciprocity, resulting in contribution levels well below the corresponding VCM. In contrast, efficiency seekers could reject an asymmetric institution aiming at contribution levels that exceed institutional obligations. Players who reject an institution due to inequality aversion have motives to contribute as in the baseline VCM whose equilibria ensure equality of payoffs across players.

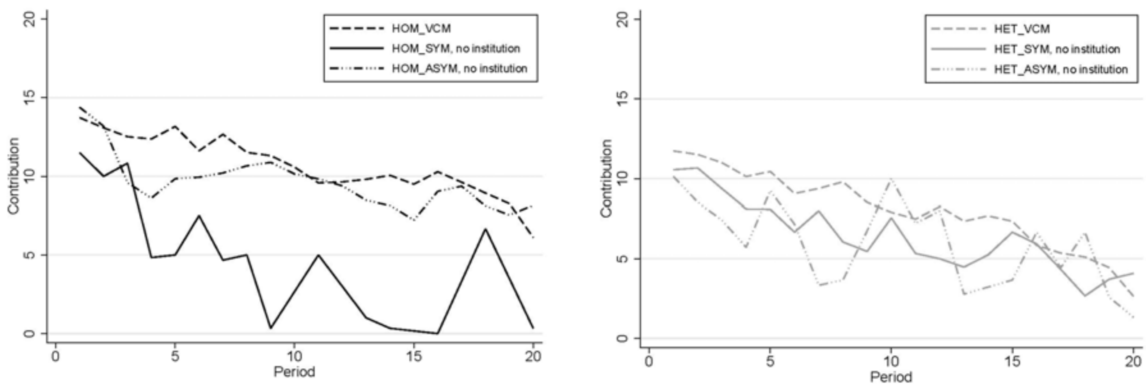
While we did not elicit subjects' individual beliefs about the preferences of players which rejected the institution, presenting results on average contributions in case of failed institution formation is still informative. Table 17 and Figure 10 show that, in the relatively rare case of institution failure (13%), average contribution levels in treatment HOM-SYM are substantially below those of the corresponding VCM (6.4 instead of 10.7 units). In treatments HET-SYM, HET-ASYM, and HOM-ASYM, average contributions are much closer to those of the corresponding baseline VCM. Taken together, our results do not point at a large, "hidden cost" of failed institution formation, namely substantially and frequently reduced contributions in case of failed institution implementation (except for treatment HOM-SYM).

Table 17: Average contributions after failed institution formation

Institution / Player type	VCM	SYM	ASYM
HOM	10.72 (7.83)	6.43 (6.90)	9.78 (8.56)
HET	8.05 (6.56)	6.76 (6.33)	6.59 (7.12)

Note: Standard deviations are in parentheses.

Figure 10: Contributions in case of no institution over time



5. Conclusion

The paper at hand studied the process of institution formation in social dilemmas, in particular the role of heterogeneity among players (i) in their benefits from cooperation and

ii) in their institutional obligations. We found that the potential tension between efficiency and equality in payoffs, originating from these heterogeneities, strongly affected implementation rates of institutions. With heterogeneous players, aggregate implementation rates were significantly lower for institutions featuring equal rather than unequal obligations; and vice versa for homogeneous players – even though failed implementation usually implied severe cutbacks in monetary payoffs. Both with homogeneous and heterogeneous players, failed implementations arose primarily, but not exclusively, from rejections by the disadvantaged players that profited to a lesser degree from the implemented institution. Consequently, institutions which tailored obligations to players' specific heterogeneities were able to gather higher degrees of support. In fact, if benefits from institution formation were evenly distributed across players, we observed strikingly higher implementation and cooperation rates than what has typically been found in related studies that only require non-unanimous support for institutions to be implemented for all members (Kosfeld et al., 2009).

A potential reason for the latter finding is that, in contrast to other decision rules, unanimity voting entails a very strong notion of conditional cooperation. The veto right inherent in unanimity voting makes each player's cooperation decision contingent on the decision of all other players involved. Consequently, the supporting players do not face the risk of being exploited by non-supporting players.[53] On a similar note, no player will ever be governed by an institution that he did not support himself. Both, the notion of conditional cooperation and the retained sovereignty, make unanimity voting an attractive rule to settle on in the first place.

On the other hand, these advantages come at the cost of an increased likelihood of rejecting efficient institutions as well as potentially low levels of cooperation after a rejection has occurred. Already with three players, we saw that these problems exist. With larger groups, one might expect successful institution formation to be even more difficult, in particular if benefits from institution formation are not equally distributed across players. Moreover, our data suggest that voting against the institution is sometimes connected with the implicit costs of making subsequent cooperation more difficult. One might even imagine that rejecting players become

the target of retaliation in other, seemingly unrelated, domains. Both threats might be bigger in large groups, simply because there are more players who might potentially opt against the institution and/or who might retaliate rejections. Yet, for groups deciding on the implementation of an institution that takes care of players' idiosyncrasies, these threats might instead strengthen the power of an unanimity rule. Furthermore, under institutions that lead to inequalities in payoffs, payoff differences might be less salient in large groups because they are harder to recognize – in particular if players do not compare themselves with everyone else in a large population, but rather choose a small reference group consisting of similar others. It would therefore be interesting to check in future studies whether the positive or negative effects dominate when group size is

[53] Parts of this reasoning rely on the strong enforcement mechanism underlying our institutions. It would be interesting to study setups that allow to discriminate between (i) support and (ii) adherence to an institution, and how these factors are affected by the voting mechanism in place. We thank an anonymous referee for pointing this out.

increased.

Follow-up studies might also investigate if aggregate behavioral patterns are affected by changes in other parameters of our design, like the marginal per capita return from cooperation or the exact content of the institution. We observed in our data on heterogeneous agents that, overall, the symmetric and asymmetric institution lead to similar average cooperation rates. This was due to two opposing effects that cancel each other out: while the higher implementation rate for the asymmetric institution generally increases cooperation, total obligations (and thus cooperation rates) are lower than when the efficient symmetric institution is implemented. Although this qualitative finding is not at the heart of our paper, it is still intriguing. Given the quantitative behavioral effects that we observe, one could imagine that average outcomes between symmetric and asymmetric institutions start diverging as the most efficient payoff-equalizing mechanism becomes more inferior to the efficient mechanism.

Along similar lines, natural next steps for future extensions also include more complex institutional arrangements. For example, redistribution might allay disadvantaged member's doubts about the implementation of efficient institutions for heterogeneous agents. The implementation of institutions with hierarchical structures, from simple leader-follower arrangements to multi-layered structures, yield the potential to increase implementation rates and cooperation, too (e.g., Gächter et al., 2010; Hamman et al., 2011; Falk and Kosfeld, 2012). Complementing these variations, one could also shed more light on the performance of different voting rules for implementing given institutions (e.g., Young, 1995; Gillet et al., 2009; Austen- Smith and Feddersen, 2006). More generally,

allowing for richer environments with competing institutions and voting rules opens up the possibility to learn even more about the type of institutions that endogenously arise within a group. Of course, in contrast to our approach, self-selection would make proper causal interpretation more difficult. Still, it would be a nice complement to the current research agenda: understanding what kind of institutions are created by groups, which voting rules are adopted for implementing these institutions, and how these institutions perform under a variety of circumstances.

References

Anderson, L. R., J. M. Mellor, and J. Milyo (2008). Inequality and public good provision: An experimental analysis. *Journal of Socio-Economics 37*, 1010–1028.

Andreoni, J. and L. K. Gee (2012). Gun for hire: delegated enforcement and peer punishment in public goods provision. *Journal of Public Economics 96* (11), 1036–1046.

Apesteguia, J., P. Funk, and N. Iriberri (2013). Promoting rule compliance in daily-life: Evidence from a randomized field experiment in the public libraries of Barcelona. *European Economic Review 64*, 266—284.

Austen-Smith, D. and T. J. Feddersen (2006). Deliberation, preference uncertainty, and voting rules. *American Political Science Review 100* (2), 209–217.

Banks, J. S., C. R. Plott, and D. P. Porter (1988). An experimental analysis of unanimity in public goods provision mechanisms. *The Review of Economic Studies 55* (2), 301–322.

Barrett-Howard, E. and T. R. Tyler (1986). Procedural justice as a criterion in allocation decisions. *Journal of Personality and Social Psychology 50* (2), 296–304.

Bierbrauer, F. J. and M. F. Hellwig (2011). Mechanism design and voting for public-good provision. *MPI Collective Goods Preprint* (2011/31).

Birnberg, J. G., L. R. Pondy, and C. L. Davis (1970). Effect of three voting rules on resource allocation decisions. *Management Science 16* (6), 356–372.

Bolton, G. E., J. Brandts, and A. Ockenfels (2005). Fair procedures: Evidence from games involving lotteries. *Economic Journal 115* (506), 1054–1076.

Bolton, G. E. and A. Ockenfels (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review 90* (1), 166–193.

Charness, G. and M. Rabin (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics 117* (3), 817–869.

Cherry, T. L., S. Kroll, and J. F. Shogren (2005, July). The impact of endowment heterogeneity and origin on public good contributions: evidence from the lab. *Journal of Economic Behavior and Organization 57*(3), 357–365.

Dannenberg, A., A. Lange, and B. Sturm (2014). Participation and commitment in voluntary coalitions to provide public goods. *Economica 81*, 257-275.

Decker, T., A. Stiehler, and M. Strobel (2003). A comparison of punishment rules in repeated public good games an experimental study. *Journal of Conflict Resolution 47*(6), 751–772.

Falk, A. and M. Kosfeld (2012). It's all about connections: Evidence on network formation. *Review of Network Economics 11*(3), Article 2.

Falkinger, J. and E. Fehr (2000). A simple mechanism for the efficient provision of public goods: Experimental evidence. *The American Economic Review 90*(1), 247–264.

Fehr, E. and S. Gächter (2000). Cooperation and punishment in public goods experiments. *The American Economic Review 90*(4), 980–994.

Fehr, E. and S. Gächter (2002). Altruistic punishment in humans. *Nature 415*, 137–140.

Fehr, E. and K. M. Schmidt (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics 114*(3), 817–868.

Fischbacher, U. (2007, February). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics 10*(2), 171–178.

Fisher, J., R. M. Isaac, J. W. Schatzberg, and J. M. Walker (1995). Heterogenous demand for public goods: Behavior in the voluntary contributions mechanism. *Public Choice 85*(3-4), 249–266.

Gächter, S., D. Nosenzo, E. Renner, and M. Sefton (2010). Sequential vs. simultaneous contributions to public goods: Experimental evidence. *Journal of Public Economics 94*(7-8), 515–522.

Gerber, A., J. Neitzel, and P. C. Wichardt (2013). Minimum participation rules for the provision of public goods. *European Economic Review 64*, 209–222.

Gillet, J., A. Schram, and J. Sonnemans (2009). The tragedy of the commons revisited: The importance of group decision-making. *Journal of Public Economics 93* (5-6), 785–797.

Greiner, B. (2004). An online recruitment system for economic experiments. In K. Kremer and V. Macho (Eds.), *Forschung und Wissenschaftliches Rechnen*, Number 63 in GWDG Bericht. Gesellschaft für Wissenschaftliche Datenverarbeitung.

Gürerk, O., B. Irlenbusch, and B. Rockenbach (2006). The competitive advantage of sanctioning institutions.*Science 312* (5770), 108–11.

Hamman, J. R., R. Weber, and J. Woon (2011). An experimental investigation of electoral delegation and the provision of public goods. *American Journal of Political Science 55* (4), 738–752.

Hobbes, T. (1651). Leviathan. Meiner Verlag (1996), Hamburg.

Isaac, R. and J. Walker (1988). Group size effects in public goods provision: The voluntary contributions mechanism. *The Quarterly Journal of Economics 103* (1), 179–199.

Kamei, K., L. Putterman, and J.-R. Tyran (2015). State or nature? Endogenous formal versus informal sanctions in the voluntary provision of public goods. *Experimental Economics 18*, 38–65.

Kesternich, M., A. Lange, and B. Sturm (2014). The impact of burden sharing rules on the voluntary provision of public goods. *Journal of Economic Behavior and Organization 105*, 107–123.

Khadjavi, M., A. Lange, and A. Niklisch (2014). The social value of transparency and accountability: Experimental evidence from asymmetric public good games. *WiSo HH Working Paper Series 12 2014* .

Kosfeld, M., A. Okada, and A. Riedl (2009). Institution formation in public goods games. *American Economic Review 99* (4), 1335–1355.

Krawczyk, M. W. (2011). A model of procedural and distributive fairness. Theory and Decision 70 (1), 111–128. Linardi, S. and M. A. McConnell (2011). No excuses for good behavior: Volunteering and the social environment. *Journal of Public Economics 95* (5), 445–454.

Maggi, G. and M. Morelli (2006). Self-enforcing voting in international organizations. *The American Economic Review 96* (4), 1137–1158.

Markussen, T., L. Putterman, and J.-R. Tyran (2014). Self-organization for collective action: An experimental study of voting on sanction regimes. *The Review of Economic Studies 81* (1), 301–324.

McEvoy, D. M., T. L. Cherry, and J. Stranlund (2014). Endogenous minimum participation in international environmental agreements: An experimental analysis. *forthcoming in Environmental and Resource Economics*.

McGinty, M. and G. Milam (2013). Public goods provision by asymmetric agents: Experimental evidence. *Social Choice and Welfare 40* (4), 1159–1177.

Ostrom, E., J. Walker, and R. Gardner (1992). Covenants with and without a sword: Self-governance is possible. *The American Political Science Review 86* (2), 404–417.

Putterman, L., J.-R. Tyran, and K. Kamei (2011). Public goods and voting on formal sanction schemes. *Journal of Public Economics 95* (9), 1213–1222.

Rauchdobler, J., R. Sausgruber, and J. Tyran (2010). Voting on thresholds for public goods: Experimental evidence. *FinanzArchiv: Public Finance Analysis 66* (1), 34–64.

Reuben, E. and A. Riedl (2013). Enforcement of contribution norms in public good games with heterogeneous populations. *Games and Economic Behavior 77* (1), 122 – 137.

Riedel, N. and H. Schildberg-Hörisch (2013). Asymmetric obligations. Journal of Economic Psychology 35, 67–80. Rousseau, J. J. (1762). *Der Gesellschaftsvertrag*.

Röder-Taschenbuch (1988), Köln.

Shavell, S. and A. M. Polinsky (2000). The economic theory of public enforcement of law. *Journal of Economic Literature 38* (1), 45–76.

Sutter, M., S. Haigner, and M. G. Kocher (2010). Choosing the carrot or the stick? Endogenous institutional choice in social dilemma situations. *Review of Economic Studies 77*(4), 1540–1566.

Tyler, T. and E. Lind (2000). Procedural justice. In J. Sanders and V. Hamilton (Eds.), *Handbook of Justice Research in Law*, pp. 63–91. New York: Kluwer.

Tyran, J.-R. and L. P. Feld (2006, March). Achieving compliance when legal sanctions are non-deterrent. *Scandinavian Journal of Economics 108*(1), 135–156.

Wicksell, K. (1964). A new principle of just taxation. In R. A. Musgrave and A. T. Peacock (Eds.), *Classics in the Theory of Public Finance*. London: Macmillan.

Young, P. (1995). Optimal voting rules. *Journal of Economic Perspectives 9*(1), 5

Appendix

CHAPTER 2

Mitigating Extortive Corruption? Experimental Evidence

German instructions for:

Allgemeine Erklärungen für die Teilnehmer

Sie nehmen nun an einer wirtschaftswissenschaftlichen Studie teil. Wenn Sie die nachfolgenden Erklärungen genau lesen, dann können Sie eine nicht unbeträchtliche Geldsumme verdienen. Es ist daher sehr wichtig, dass Sie diese Erklärungen genau durchlesen und verstehen.

Während der Studie herrscht ein absolutes Kommunikationsverbot. Wenn Sie Fragen haben, dann richten Sie diese bitte ausschließlich an uns: zeigen Sie auf, und wir kommen zu Ihnen. Die Nichtbeachtung dieser Regel führt zum Ausschluss von der Studie und allen Zahlungen.

Die Studie besteht aus genau **drei** Teilen. Sie erhalten nun zunächst die Informationen für den ersten Teil. Informationen zu den anderen beiden Teilen erhalten Sie jeweils vor Beginn des entsprechenden Teils der Studie.

Wie viel Geld Sie nach der Studie erhalten, hängt von Ihren Entscheidungen sowie den Entscheidungen der anderen Teilnehmer ab. Während der Studie sprechen wir nicht von Euro, sondern von Talern, wobei im ersten Teil gilt:

1 Euro = 6,5 Taler
2 Euro = 13 Taler
etc.

Für diesen ersten Teil erhalten Sie am Ende der Studie das Einkommen aus einer zufällig ausgewählten Runde bar ausbezahlt, d.h., jede Runde könnte für Sie potentiell auszahlungsrelevant sein.

Informationen für den ersten Teil

Der erste Teil besteht aus **10 Runden** und jede Runde aus **3 Stufen.** Sie interagieren in jeder Runde mit **demselben** Teilnehmer. Weder während noch nach der Studie erhalten Sie Informationen über die Identität der anderen Teilnehmer.

Zu Beginn werden Sie zufällig einer von zwei Rollen zugeordnet. Ein „Beamter“ interagiert mit einem „Bürger“ in jeder Runde. Jede Runde besteht dabei aus **3 Stufen**:

- In der **1. Stufe** entscheidet der Beamte, ob er ein Bestechungsgeld in Höhe von 10 Talern von dem Bürger einfordern möchte. Wenn ein Bestechungsgeld gefordert wird, muss der Bürger entscheiden, ob er das Bestechungsgeld zahlt. Wenn kein Bestechungsgeld gefordert wurde, muss der Bürger keine Entscheidung treffen.

- In der **2. Stufe** entscheidet der Beamte, ob er dem Bürger entweder eine *reguläre* Leistung oder eine *reduzierte* Leistung zukommen lässt. Falls in Stufe 1 der Beamte ein Bestechungsgeld gefordert hat und der Bürger dieses bezahlt hat, entfällt die Wahl und der Beamte muss die reguläre Leistung erbringen.
  Unabhängig von der Leistung erhält der Beamte eine Auszahlung von 35 Talern. Im Falle der regulären Leistung erhält der Bürger ein Einkommen von 50 Talern und im Falle der reduzierten Leistung ein Einkommen von 35 Talern. Falls der Bürger das Bestechungsgeld bezahlt hat, wird von seinem Einkommen noch das Bestechungsgeld von 10 Talern abgezogen.

- In der **3. Stufe** entscheidet der Bürger, ob er den Beamten meldet. Das Melden des Beamten reduziert das Einkommen des Bürgers um 1 Taler. Wenn der Beamte ein Bestechungsgeld gefordert hat und der Bürger den Beamten gemeldet hat, kann dem Beamten mit einer Wahrscheinlichkeit von 10% (also durchschnittlich in 1 von 10 Fällen) die Forderung des Bestechungsgeldes nachgewiesen werden, woraufhin der Beamte bestraft wird. Durch die Strafe reduziert sich das Einkommen des Beamten um 15 Taler. Zusätzlich verliert er das Bestechungsgeld, wenn der Bürger das Bestechungsgeld bezahlt hat. Danach wird der Beamte über die Entscheidung des Bürgers und über eine eventuelle Strafe informiert. Zudem erfahren beide Teilnehmer die Höhe ihres Einkommens aus der Runde.

Danach beginnt die nächste Runde, wie oben erwähnt, mit **demselben** Teilnehmer.

Abhängig von den Entscheidungen des Beamten und des Bürgers können die folgenden Situationen eintreten:

Beispiel 1: Wenn ein Beamter *kein Bestechungsgeld* einfordert und die *reguläre Leistung* bereit stellt, ist am Ende der Runde das Einkommen des Beamten 35 Taler und das Einkommen des Bürgers 50 Taler.

Beispiel 2: Wenn ein Beamter *kein Bestechungsgeld* einfordert und die *reduzierte Leistung* bereit stellt, ist am Ende der Runde das Einkommen des Beamten 35 Taler und das Einkommen des Bürgers 35 Taler.

Beispiel 3: Wenn ein Beamter *ein Bestechungsgeld einfordert*, der Bürger dieses *nicht bezahlt* und der Beamte die *reguläre Leistung* bereit stellt, ist am Ende der Runde das Einkommen des Beamten 35 Taler und das Einkommen des Bürgers 50 Taler.
Wenn jedoch der Bürger den Beamten gemeldet hat, erhält er stattdessen 34 Taler. Falls der Beamte eine Strafe von 15 Talern erhält, ist sein Einkommen am Ende der Runde 20 Taler. Falls er keine Strafe erhält, ist sein Einkommen am Ende der Runde 35 Taler.

Beispiel 4: Wenn ein Beamter *ein Bestechungsgeld einfordert*, der Bürger dieses *nicht bezahlt* und der Beamte die *reduzierte Leistung* bereit stellt, ist am Ende der Runde das Einkommen des Beamten 35 Taler und das Einkommen des Bürgers 35 Taler.
Wenn jedoch der Bürger den Beamten gemeldet hat, erhält er stattdessen 34 Taler. Falls der Beamte eine Strafe von 15 Talern erhält, ist sein Einkommen am Ende der Runde 20 Taler. Falls er keine Strafe erhält, ist sein Einkommen am Ende der Runde 35 Taler.

Beispiel 5: Wenn ein Beamter *ein Bestechungsgeld einfordert*, der Bürger dieses *bezahlt* und der Beamte somit die *reguläre Leistung* bereit stellen muss, ist am Ende der Runde das Einkommen des Beamten 45 Taler und das Einkommen des Bürgers 40 Taler.
Wenn jedoch der Bürger den Beamten gemeldet hat, erhält er stattdessen 39 Taler. Falls dem Beamten die Forderung von Bestechungsgeld nachgewiesen wurde, erhält er eine Strafe von 15 Talern und verliert das Bestechungsgeld von 10 Talern. Sein Einkommen am Ende der Runde ist damit 20 Taler. Falls dem Beamten die Forderung von Bestechungsgeld nicht nachgewiesen wurde, ist sein Einkommen am Ende der Runde 45 Taler.

Haben Sie noch Fragen? Dann zeigen Sie bitte auf.

Allgemeine Erklärungen für die Teilnehmer

Sie nehmen nun an einer wirtschaftswissenschaftlichen Studie teil. Wenn Sie die nachfolgenden Erklärungen genau lesen, dann können Sie eine nicht unbeträchtliche Geldsumme verdienen. Es ist daher sehr wichtig, dass Sie diese Erklärungen genau durchlesen und verstehen.

Während der Studie herrscht ein absolutes Kommunikationsverbot. Wenn Sie Fragen haben, dann richten Sie diese bitte ausschließlich an uns: zeigen Sie auf, und wir kommen zu Ihnen. Die Nichtbeachtung dieser Regel führt zum Ausschluss von der Studie und allen Zahlungen.

Die Studie besteht aus genau **drei** Teilen. Sie erhalten nun zunächst die Informationen für den ersten Teil. Informationen zu den anderen beiden Teilen erhalten Sie jeweils vor Beginn des entsprechenden Teils der Studie.

Wie viel Geld Sie nach der Studie erhalten, hängt von Ihren Entscheidungen sowie den Entscheidungen der anderen Teilnehmer ab. Während der Studie sprechen wir nicht von Euro, sondern von Talern, wobei im ersten Teil gilt:

1 Euro = 6,5 Taler
2 Euro = 13 Taler
etc.

Für diesen ersten Teil erhalten Sie am Ende der Studie das Einkommen aus einer zufällig ausgewählten Runde bar ausbezahlt, d.h., jede Runde könnte für Sie potentiell auszahlungsrelevant sein.

Informationen für den ersten Teil

Der erste Teil besteht aus **10 Runden** und jede Runde aus **3 Stufen.** Sie interagieren jede Runde mit einem **anderen zufällig bestimmten** Teilnehmer. Weder während noch nach der Studie erhalten Sie Informationen über die Identität der anderen Teilnehmer.

Zu Beginn werden Sie zufällig einer von zwei Rollen zugeordnet. Ein „Beamter" interagiert mit einem „Bürger" in jeder Runde. Jede Runde besteht dabei aus **3 Stufen**:

- In der **1. Stufe** entscheidet der Beamte, ob er ein Bestechungsgeld in Höhe von 10 Talern von dem Bürger einfordern möchte. Wenn ein Bestechungsgeld gefordert wird, muss der Bürger entscheiden, ob er das Bestechungsgeld zahlt. Wenn kein Bestechungsgeld gefordert wurde, muss der Bürger keine Entscheidung treffen.

- In der **2. Stufe** entscheidet der Beamte, ob er dem Bürger entweder eine *reguläre* Leistung oder eine *reduzierte* Leistung zukommen lässt. Falls in Stufe 1 der Beamte ein Bestechungsgeld gefordert hat und der Bürger dieses bezahlt hat, entfällt die Wahl und der Beamte muss die reguläre Leistung erbringen.
  Unabhängig von der Leistung erhält der Beamte eine Auszahlung von 35 Talern. Im Falle der regulären Leistung erhält der Bürger ein Einkommen von 50 Talern und im Falle der reduzierten Leistung ein Einkommen von 35 Talern. Falls der Bürger das Bestechungsgeld bezahlt hat, wird von seinem Einkommen noch das Bestechungsgeld von 10 Talern abgezogen

- In der **3. Stufe** entscheidet der Bürger, ob er den Beamten meldet. Das Melden des Beamten reduziert das Einkommen des Bürgers um 1 Taler. Wenn der Beamte ein Bestechungsgeld gefordert hat und der Bürger den Beamten gemeldet hat, kann dem Beamten mit einer Wahrscheinlichkeit von 10% (also durchschnittlich in 1 von 10 Fällen) die Forderung des Bestechungsgeldes nachgewiesen werden, woraufhin der Beamte bestraft wird. Durch die Strafe reduziert sich das Einkommen des Beamten um 15 Taler. Zusätzlich verliert er das Bestechungsgeld, wenn der Bürger das Bestechungsgeld bezahlt hat. Danach wird der Beamte über die Entscheidung des Bürgers und über eine eventuelle Strafe informiert. Zudem erfahren beide Teilnehmer die Höhe ihres Einkommens aus der Runde.

Danach beginnt die nächste Runde, wie oben erwähnt, mit einem **anderen zufällig bestimmten** Teilnehmer.

Abhängig von den Entscheidungen des Beamten und des Bürgers können die folgenden Situationen eintreten:

Beispiel 1: Wenn ein Beamter *kein Bestechungsgeld* einfordert und die *reguläre Leistung* bereit stellt, ist am Ende der Runde das Einkommen des Beamten 35 Taler und das Einkommen des Bürgers 50 Taler.

Beispiel 2: Wenn ein Beamter *kein Bestechungsgeld* einfordert und die *reduzierte Leistung* bereit stellt, ist am Ende der Runde das Einkommen des Beamten 35 Taler und das Einkommen des Bürgers 35 Taler.

Beispiel 3: Wenn ein Beamter *ein Bestechungsgeld einfordert*, der Bürger dieses *nicht bezahlt* und der Beamte die *reguläre Leistung* bereit stellt, ist am Ende der Runde das Einkommen des Beamten 35 Taler und das Einkommen des Bürgers 50 Taler.
Wenn jedoch der Bürger den Beamten gemeldet hat, erhält er stattdessen 34 Taler. Falls der Beamte eine Strafe von 15 Talern erhält, ist sein Einkommen am Ende der Runde 20 Taler. Falls er keine Strafe erhält, ist sein Einkommen am Ende der Runde 35 Taler.

Beispiel 4: Wenn ein Beamter *ein Bestechungsgeld einfordert*, der Bürger dieses *nicht bezahlt* und der Beamte die *reduzierte Leistung* bereit stellt, ist am Ende der Runde das Einkommen des Beamten 35 Taler und das Einkommen des Bürgers 35 Taler.
Wenn jedoch der Bürger den Beamten gemeldet hat, erhält er stattdessen 34 Taler. Falls der Beamte eine Strafe von 15 Talern erhält, ist sein Einkommen am Ende der Runde 20 Taler. Falls er keine Strafe erhält, ist sein Einkommen am Ende der Runde 35 Taler.

Beispiel 5: Wenn ein Beamter *ein Bestechungsgeld einfordert*, der Bürger dieses *bezahlt* und der Beamte somit die *reguläre Leistung* bereit stellen muss, ist am Ende der Runde das Einkommen des Beamten 45 Taler und das Einkommen des Bürgers 40 Taler.
Wenn jedoch der Bürger den Beamten gemeldet hat, erhält er stattdessen 39 Taler. Falls dem Beamten die Forderung von Bestechungsgeld nachgewiesen wurde, erhält er eine Strafe von 15 Talern und verliert das Bestechungsgeld von 10 Talern. Sein Einkommen am Ende der Runde ist damit 20 Taler. Falls dem Beamten die Forderung von Bestechungsgeld nicht nachgewiesen wurde, ist sein Einkommen am Ende der Runde 45 Taler.

Haben Sie noch Fragen? Dann zeigen Sie bitte auf.

Kontrollfragen

Bitte nehmen Sie sich noch einen Moment Zeit, um die folgenden Fragen zu beantworten bevor die Studie beginnt.

Kontrollfrage 1

Welche der folgenden Aussagen sind richtig:
(Bitte kreisen Sie die richtigen Antworten ein)

A) in der nächsten Runde interagiert der Bürger mit demselben Beamten

B) in der nächsten Runde interagiert der Bürger mit einem anderen Beamten

C) in jeder Runde wechselt die Rolle, die man einnimmt

D) seine Rolle behält man für alle Runden bei

Bitte berechnen Sie das Einkommen des Bürgers und des Beamten in den verschiedenen Szenarien.

Kontrollfrage 2

Der Beamte verlangt ein Bestechungsgeld von 10 Talern.

Der Bürger ist bereit das Bestechungsgeld zu zahlen.

Der Bürger meldet den Beamten.

Dem Beamten kann aber kein Fehlverhalten nachgewiesen werden.

Wie hoch sind die Einkommen der beiden Seiten in dieser Runde?

Einkommen des Bürgers: ______Taler

Einkommen des Beamten: ______Taler

Wie hoch wäre das Einkommen des Beamten, wenn ihm Fehlverhalten nachgewiesen worden wäre? ______Taler

Kontrollfrage 3

Der Beamte verlangt ein Bestechungsgeld von 10 Talern.

Der Bürger ist nicht bereit das Bestechungsgeld zu zahlen.

Der Beamte stellt die reguläre Leistung bereit.

Der Bürger meldet den Beamten nicht.

Wie hoch sind die Einkommen der beiden Seiten in dieser Runde?

Einkommen des Bürgers: ______Taler

Einkommen des Beamten: ______Taler

Wie hoch wäre das Einkommen des Bürgers, wenn die reduzierte Leistung bereitgestellt worden wäre? ______Taler

Kontrollfrage 4

Der Beamte verlangt kein Bestechungsgeld und stellt die reduzierte Leistung bereit.

Wie hoch sind die Einkommen der beiden Seiten in dieser Runde?

Einkommen des Bürgers: ______Taler

Einkommen des Beamten: ______Taler

Allgemeine Erklärungen für die Teilnehmer

Sie nehmen nun an einer wirtschaftswissenschaftlichen Studie teil. Wenn Sie die nachfolgenden Erklärungen genau lesen, dann können Sie eine nicht unbeträchtliche Geldsumme verdienen. Es ist daher sehr wichtig, dass Sie diese Erklärungen genau durchlesen und verstehen.

Während der Studie herrscht ein absolutes Kommunikationsverbot. Wenn Sie Fragen haben, dann richten Sie diese bitte ausschließlich an uns: zeigen Sie auf, und wir kommen zu Ihnen. Die Nichtbeachtung dieser Regel führt zum Ausschluss von der Studie und allen Zahlungen.

Die Studie besteht aus genau **drei** Teilen. Sie erhalten nun zunächst die Informationen für den ersten Teil. Informationen zu den anderen beiden Teilen erhalten Sie jeweils vor Beginn des entsprechenden Teils der Studie.

Wie viel Geld Sie nach der Studie erhalten, hängt von Ihren Entscheidungen sowie den Entscheidungen der anderen Teilnehmer ab. Während der Studie sprechen wir nicht von Euro, sondern von Talern, wobei im ersten Teil gilt:

1 Euro = 6,5 Taler
2 Euro = 13 Taler
etc.

Für diesen ersten Teil erhalten Sie am Ende der Studie das Einkommen aus einer zufällig ausgewählten Runde bar ausbezahlt, d.h., jede Runde könnte für Sie potentiell auszahlungsrelevant sein.

Informationen für den ersten Teil

Der erste Teil besteht aus **10 Runden** und jede Runde aus **3 Stufen.** Sie interagieren in jeder Runde mit **demselben** Teilnehmer. Weder während noch nach der Studie erhalten Sie Informationen über die Identität der anderen Teilnehmer.

Zu Beginn werden Sie zufällig einer von zwei Rollen zugeordnet. Ein „Beamter“ interagiert mit einem „Bürger“ in jeder Runde. Jede Runde besteht dabei aus **3 Stufen**:

- In der **1. Stufe** entscheidet der Beamte, ob er ein Bestechungsgeld in Höhe von 10 Talern von dem Bürger einfordern möchte. Wenn ein Bestechungsgeld gefordert wird, muss der Bürger entscheiden, ob er das Bestechungsgeld zahlt. Wenn kein Bestechungsgeld gefordert wurde, muss der Bürger keine Entscheidung treffen.

- In der **2. Stufe** entscheidet der Beamte, ob er dem Bürger entweder eine *reguläre* Leistung oder eine *reduzierte* Leistung zukommen lässt. Falls in Stufe 1 der Beamte ein Bestechungsgeld gefordert hat und der Bürger dieses bezahlt hat, entfällt die Wahl und der Beamte muss die reguläre Leistung erbringen.
  Unabhängig von der Leistung erhält der Beamte eine Auszahlung von 35 Talern. Im Falle der regulären Leistung erhält der Bürger ein Einkommen von 50 Talern und im Falle der reduzierten Leistung ein Einkommen von 35 Talern. Falls der Bürger das Bestechungsgeld bezahlt hat, wird von seinem Einkommen noch das Bestechungsgeld von 10 Talern abgezogen

- In der **3. Stufe** entscheidet der Bürger, ob er den Beamten meldet. Das Melden des Beamten reduziert das Einkommen des Bürgers um 1 Taler. Wenn der Beamte kein Bestechungsgeld gefordert hat und der Bürger den Beamten gemeldet hat, erhält der Beamte mit einer Wahrscheinlichkeit von 10% (also durchschnittlich in 1 von 10 Fällen) eine Belohnung. Durch die Belohnung erhöht sich das Einkommen des Beamten um 15 Taler. Danach wird der Beamte über die Entscheidung des Bürgers und über eine eventuelle Belohnung informiert. Zudem erfahren beide Teilnehmer die Höhe ihres Einkommens aus der Runde.

Danach beginnt die nächste Runde, wie oben erwähnt, mit **demselben** Teilnehmer.

Abhängig von den Entscheidungen des Beamten und des Bürgers können die folgenden Situationen eintreten:

Beispiel 1: Wenn ein Beamter *kein Bestechungsgeld* einfordert und die *reguläre Leistung* bereit stellt, ist am Ende der Runde das Einkommen des Beamten 35 Taler und das Einkommen des Bürgers 50 Taler.
Wenn der Bürger den Beamten gemeldet hat, erhält er stattdessen 49 Taler. Falls der Beamte einen Bonus von 15 Talern erhält, ist sein Einkommen am Ende der Runde 50 Taler. Falls er keinen Bonus erhält, ist sein Einkommen am Ende der Runde 35 Taler.

Beispiel 2: Wenn ein Beamter *kein Bestechungsgeld* einfordert und die *reduzierte Leistung* bereit stellt, ist am Ende der Runde das Einkommen des Beamten 35 Taler und das Einkommen des Bürgers 35 Taler.
Wenn der Bürger den Beamten gemeldet hat, erhält er stattdessen 34 Taler. Falls der Beamte einen Bonus von 15 Talern erhält, ist sein Einkommen am Ende der Runde 50 Taler. Falls er keinen Bonus erhält, ist sein Einkommen am Ende der Runde 35 Taler.

Beispiel 3: Wenn ein Beamter *ein Bestechungsgeld einfordert*, der Bürger dieses *nicht bezahlt* und der Beamte die *reguläre Leistung* bereit stellt, ist am Ende der Runde das Einkommen des Beamten 35 Taler und das Einkommen des Bürgers 50 Taler.

Beispiel 4: Wenn ein Beamter *ein Bestechungsgeld einfordert*, der Bürger dieses *nicht bezahlt* und der Beamte die *reduzierte Leistung* bereit stellt, ist am Ende der Runde das Einkommen des Beamten 35 Taler und das Einkommen des Bürgers 35 Taler.

Beispiel 5: Wenn ein Beamter *ein Bestechungsgeld einfordert*, der Bürger dieses *bezahlt* und der Beamte somit die *reguläre Leistung* bereit stellen muss, ist am Ende der Runde das Einkommen des Beamten 45 Taler und das Einkommen des Bürgers 40 Taler.

Haben Sie noch Fragen? Dann zeigen Sie bitte auf.

Allgemeine Erklärungen für die Teilnehmer

Sie nehmen nun an einer wirtschaftswissenschaftlichen Studie teil. Wenn Sie die nachfolgenden Erklärungen genau lesen, dann können Sie eine nicht unbeträchtliche Geldsumme verdienen. Es ist daher sehr wichtig, dass Sie diese Erklärungen genau durchlesen und verstehen.

Während der Studie herrscht ein absolutes Kommunikationsverbot. Wenn Sie Fragen haben, dann richten Sie diese bitte ausschließlich an uns: zeigen Sie auf, und wir kommen zu Ihnen. Die Nichtbeachtung dieser Regel führt zum Ausschluss von der Studie und allen Zahlungen.

Die Studie besteht aus genau **drei** Teilen. Sie erhalten nun zunächst die Informationen für den ersten Teil. Informationen zu den anderen beiden Teilen erhalten Sie jeweils vor Beginn des entsprechenden Teils der Studie.

Wie viel Geld Sie nach der Studie erhalten, hängt von Ihren Entscheidungen sowie den Entscheidungen der anderen Teilnehmer ab. Während der Studie sprechen wir nicht von Euro, sondern von Talern, wobei im ersten Teil gilt:

1 Euro = 6,5 Taler
2 Euro = 13 Taler
etc.

Für diesen ersten Teil erhalten Sie am Ende der Studie das Einkommen aus einer zufällig ausgewählten Runde bar ausbezahlt, d.h., jede Runde könnte für Sie potentiell auszahlungsrelevant sein.

Informationen für den ersten Teil

Der erste Teil besteht aus **10 Runden** und jede Runde aus **3 Stufen.** Sie interagieren jede Runde mit einem **anderen zufällig bestimmten** Teilnehmer. Weder während noch nach der Studie erhalten Sie Informationen über die Identität der anderen Teilnehmer.

Zu Beginn werden Sie zufällig einer von zwei Rollen zugeordnet. Ein „Beamter“ interagiert mit einem „Bürger“ in jeder Runde. Jede Runde besteht dabei aus **3 Stufen**:

- In der **1. Stufe** entscheidet der Beamte, ob er ein Bestechungsgeld in Höhe von 10 Talern von dem Bürger einfordern möchte. Wenn ein Bestechungsgeld gefordert wird, muss der Bürger entscheiden, ob er das Bestechungsgeld zahlt. Wenn kein Bestechungsgeld gefordert wurde, muss der Bürger keine Entscheidung treffen.

- In der **2. Stufe** entscheidet der Beamte, ob er dem Bürger entweder eine *reguläre* Leistung oder eine *reduzierte* Leistung zukommen lässt. Falls in Stufe 1 der Beamte ein Bestechungsgeld gefordert hat und der Bürger dieses bezahlt hat, entfällt die Wahl und der Beamte muss die reguläre Leistung erbringen.
  Unabhängig von der Leistung erhält der Beamte eine Auszahlung von 35 Talern. Im Falle der regulären Leistung erhält der Bürger ein Einkommen von 50 Talern und im Falle der reduzierten Leistung ein Einkommen von 35 Talern. Falls der Bürger das Bestechungsgeld bezahlt hat, wird von seinem Einkommen noch das Bestechungsgeld von 10 Talern abgezogen

- In der **3. Stufe** entscheidet der Bürger, ob er den Beamten meldet. Das Melden des Beamten reduziert das Einkommen des Bürgers um 1 Taler. Wenn der Beamte kein Bestechungsgeld gefordert hat und der Bürger den Beamten gemeldet hat, erhält der Beamte mit einer Wahrscheinlichkeit von 10% (also durchschnittlich in 1 von 10 Fällen) eine Belohnung. Durch die Belohnung erhöht sich das Einkommen des Beamten um 15 Taler. Danach wird der Beamte über die Entscheidung des Bürgers und über eine eventuelle Belohnung informiert. Zudem erfahren beide Teilnehmer die Höhe ihres Einkommens aus der Runde.

Danach beginnt die nächste Runde, wie oben erwähnt, mit einem anderen zufällig bestimmten Teilnehmer.

Abhängig von den Entscheidungen des Beamten und des Bürgers können die folgenden Situationen eintreten:

Beispiel 1: Wenn ein Beamter *kein Bestechungsgeld* einfordert und die *reguläre Leistung* bereit stellt, ist am Ende der Runde das Einkommen des Beamten 35 Taler und das Einkommen des Bürgers 50 Taler.
Wenn der Bürger den Beamten gemeldet hat, erhält er stattdessen 49 Taler. Falls der Beamte einen Bonus von 15 Talern erhält, ist sein Einkommen am Ende der Runde 50 Taler. Falls er keinen Bonus erhält, ist sein Einkommen am Ende der Runde 35 Taler.

Beispiel 2: Wenn ein Beamter *kein Bestechungsgeld* einfordert und die *reduzierte Leistung* bereit stellt, ist am Ende der Runde das Einkommen des Beamten 35 Taler und das Einkommen des Bürgers 35 Taler.
Wenn der Bürger den Beamten gemeldet hat, erhält er stattdessen 34 Taler. Falls der Beamte einen Bonus von 15 Talern erhält, ist sein Einkommen am Ende der Runde 50 Taler. Falls er keinen Bonus erhält, ist sein Einkommen am Ende der Runde 35 Taler.

Beispiel 3: Wenn ein Beamter *ein Bestechungsgeld einfordert*, der Bürger dieses *nicht bezahlt* und der Beamte die *reguläre Leistung* bereit stellt, ist am Ende der Runde das Einkommen des Beamten 35 Taler und das Einkommen des Bürgers 50 Taler.

Beispiel 4: Wenn ein Beamter *ein Bestechungsgeld einfordert*, der Bürger dieses *nicht bezahlt* und der Beamte die *reduzierte Leistung* bereit stellt, ist am Ende der Runde das Einkommen des Beamten 35 Taler und das Einkommen des Bürgers 35 Taler.

Beispiel 5: Wenn ein Beamter *ein Bestechungsgeld einfordert,* der Bürger dieses *bezahlt* und der Beamte somit die *reguläre Leistung* bereit stellen muss, ist am Ende der Runde das Einkommen des Beamten 45 Taler und das Einkommen des Bürgers 40 Taler.

Haben Sie noch Fragen? Dann zeigen Sie bitte auf.

Kontrollfragen

Bitte nehmen Sie sich noch einen Moment Zeit, um die folgenden Fragen zu beantworten bevor die Studie beginnt.

Kontrollfrage 1

Welche der folgenden Aussagen sind richtig:
(Bitte kreisen Sie die richtigen Antworten ein)

A) in der nächsten Runde interagiert der Bürger mit demselben Beamten

B) in der nächsten Runde interagiert der Bürger mit einem anderen Beamten

C) in jeder Runde wechselt die Rolle, die man einnimmt

D) seine Rolle behält man für alle Runden bei

Bitte berechnen Sie das Einkommen des Bürgers und des Beamten in den verschiedenen Szenarien.

Kontrollfrage 2

Der Beamte verlangt ein Bestechungsgeld von 10 Talern.

Der Bürger ist bereit das Bestechungsgeld zu zahlen.

Wie hoch sind die Einkommen der beiden Seiten in dieser Runde?

Einkommen des Bürgers: ______Taler

Einkommen des Beamten: ______Taler

Kontrollfrage 3

Der Beamte verlangt ein Bestechungsgeld von 10 Talern.

Der Bürger ist nicht bereit das Bestechungsgeld zu zahlen.

Der Beamte stellt die reguläre Leistung bereit.

Wie hoch sind die Einkommen der beiden Seiten in dieser Runde?

Einkommen des Bürgers: ______Taler

Einkommen des Beamten: ______Taler

Wie hoch wäre das Einkommen des Bürgers, wenn die reduzierte Leistung bereitgestellt worden wäre? ______Taler

Kontrollfrage 4

Der Beamte verlangt kein Bestechungsgeld und stellt die reduzierte Leistung bereit.

Der Bürger meldet den Beamten.

Der Beamte erhält eine Belohnung.

Wie hoch sind die Einkommen der beiden Seiten in dieser Runde?

Einkommen des Bürgers: ______Taler

Einkommen des Beamten: ______Taler

Wie hoch wäre das Einkommen des Beamten, wenn er keine Belohnung erhalten würde? ______Taler

CHAPTER 3

Letting off Steam! Experimental Evidence on Irrational Punishment

English translation of the instructions and control questions for *Chance Treatment* of *Player C*.

General instructions for participant C

We are pleased to welcome you to this economic experiment.

If you read the following instructions carefully, you can earn money, depending on your decisions and those of the other participants. It is thus very important that you read these instructions carefully. If you have questions, please ask us.

Communication with other participants during the experiment is strictly forbidden. Not following this rule will result in exclusion from the experiment and any payments.

During the experiment, we will not speak of Euros, but points. Your entire income will first be calculated in points. The total number of points you earn during the experiment will be converted into Euros at the end of the experiment, where the following conversion rate applies:

5 Points = 1 Euro.

At the end of today's experiment, you will receive the number of points earned during the experiment in **cash**.

We will explain the exact experimental procedure on the next pages.

The Experiment

At the beginning of the experiment, <u>three</u> other participants in the experiment will be randomly assigned to you. You will never learn of the identity of the three persons assigned to you before or after the experiment, nor will the persons assigned to you learn of your identity.

There are three types of participants in this experiment: participants A, B, and C. **You are a participant C.** The three other persons assigned to you are one participant A and one participant B and one participant C. (There are thus one participant A, one participant B, and two participants C).

In this experiment, 100 points will be distributed between the four participants according to Allocation 1 or Allocation 2.

- **Allocation 1:** Participants A and B receive 45 points each and the two participants C receive 5 points each.
- **Allocation 2:** Participant A, participant B and both participants C receive 25 points each.

Participant A can either choose Allocation 1 or delegate the decision to a computer that generates Allocation 1 in 50% of the cases and Allocation 2 in 50 % of the cases.

The table below provides an additional summary of the possible outcomes:

	Participant A's points	Participant B's points	One participant C's points	The other participant C's points
Participant A chooses Allocation 1	45	45	5	5
Participant A delegates and computer generates Allocation 1	45	45	5	5
Participant A delegates and computer generates Allocation 2	25	25	25	25
Participant A delegates and computer generates Allocation 2	25	25	25	25

Either you or the other participant C will be chosen randomly. Only the decision of the randomly chosen participant C will actually determine the payoff outcomes of the game. You will be informed about the random choice of participant C by the computer after you make all your decisions.

As participant C, you have the possibility of giving up your points to deduct points from participant B and/or the other participant C. For each point that you give up, 5 points will be deducted from the corresponding participant(s) that you choose. In case of both allocations 1 or 2, you may give up at most 5 of your points for deducing points from the other participants. However, a participant can never have more points deducted than he or she earned from Allocation 1 or 2.

Participant A is impune, which means that no points can be deducted from participant A.

Participant B cannot make any decisions in this Experiment.

Example 1: Participant A chooses Allocation 1 and the randomly chosen participant C gives up 2 points to deduct 2 x 5 = 10 points from participant B. The following payments then result:

	Participant A	Participant B	The other participant C	The chosen participant C
Allocation	45	45	5	5
Deductions	**-**	**2x5**	**0x5**	2
Payments?	45	45-10=35	5-0=5	5-2=3

Example 2: Participant A delegates the decision and the computer generates Allocation 2 and the randomly chosen participant C gives up 2 points to deduct 1 x 5 = 5 points from participant B and 1 x 5 = 5 points from the other participant C. The following payments then result:

	Participant A	Participant B	The other participant C	The chosen participant C
Allocation	25	25	25	25
Deductions	**-**	**1x5**	**1x5**	2
Payments?	25	25-5=20	25-5=20	25-2=23

Example 3:
The randomly chosen participant C does not give up any points to deduct points from other participants. The points shown on the previous page will then result, depending on the respective allocation.

Your decisions:

Before you as participant C learn which allocation is chosen/generated, and before you know whether you or the other participant C were chosen randomly, we ask you to **make your decision for each of the following three cases:**

- Participant A decides for Allocation 1 (45 45 5 5)

- Participant A delegates the decision and the computer generates Allocation 1 (45 45 5 5)

- Participant A delegates the decision and the computer generates Allocation 2 (25 25 25 25)

In particular, this means that you must indicate for each of the cases whether you would like to deduct points or not, and if yes, how many points would you like to give up to deduct points from the other participants.

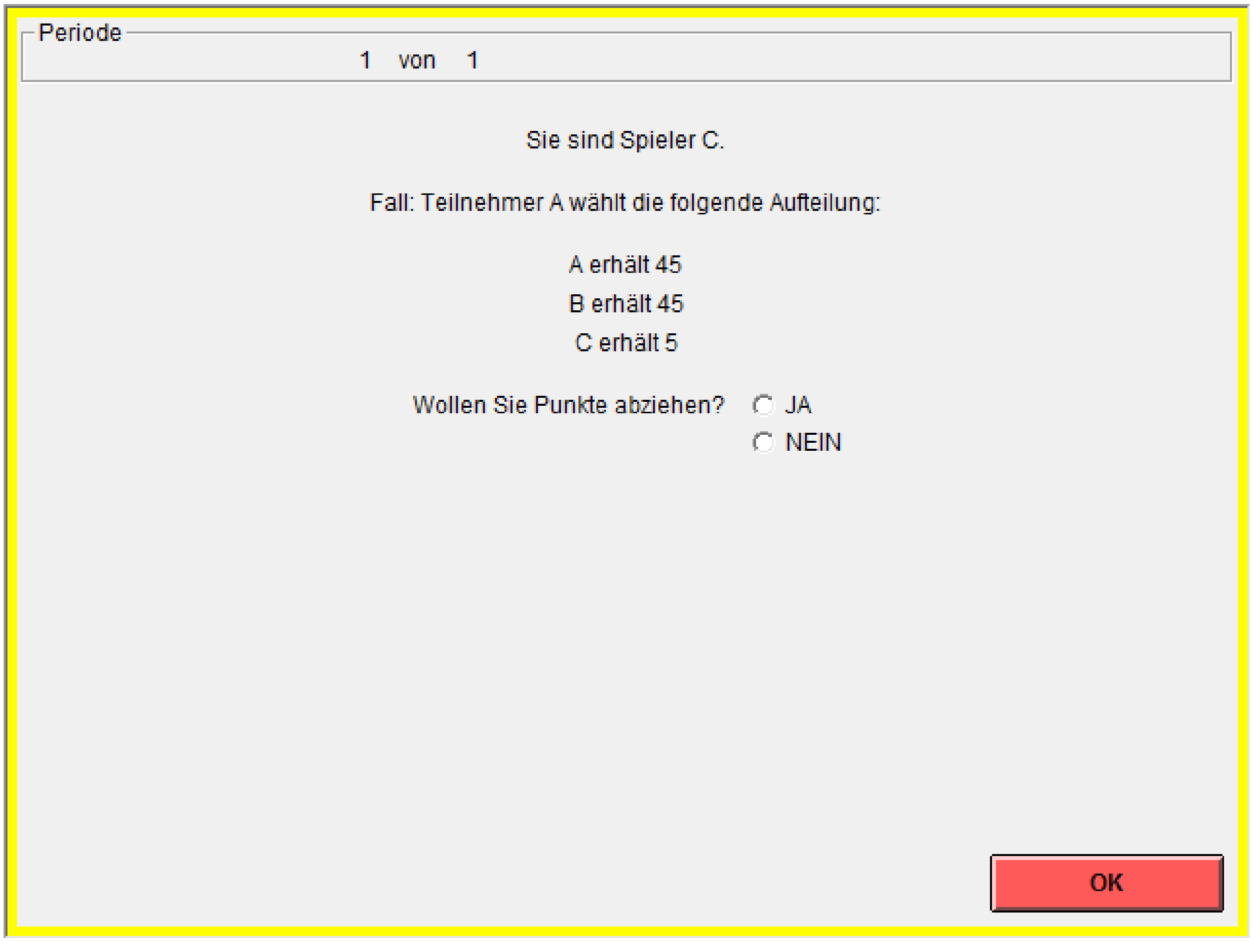
Participant A makes his or her decision without knowing what you or the other participant C would do in the two cases.

If you are randomly chosen, your decision for that case which actually arises from participant A's decision will be implemented.

Your decisions for each of the three cases can therefore be applicable for your payment.

Procedure on the Computer:

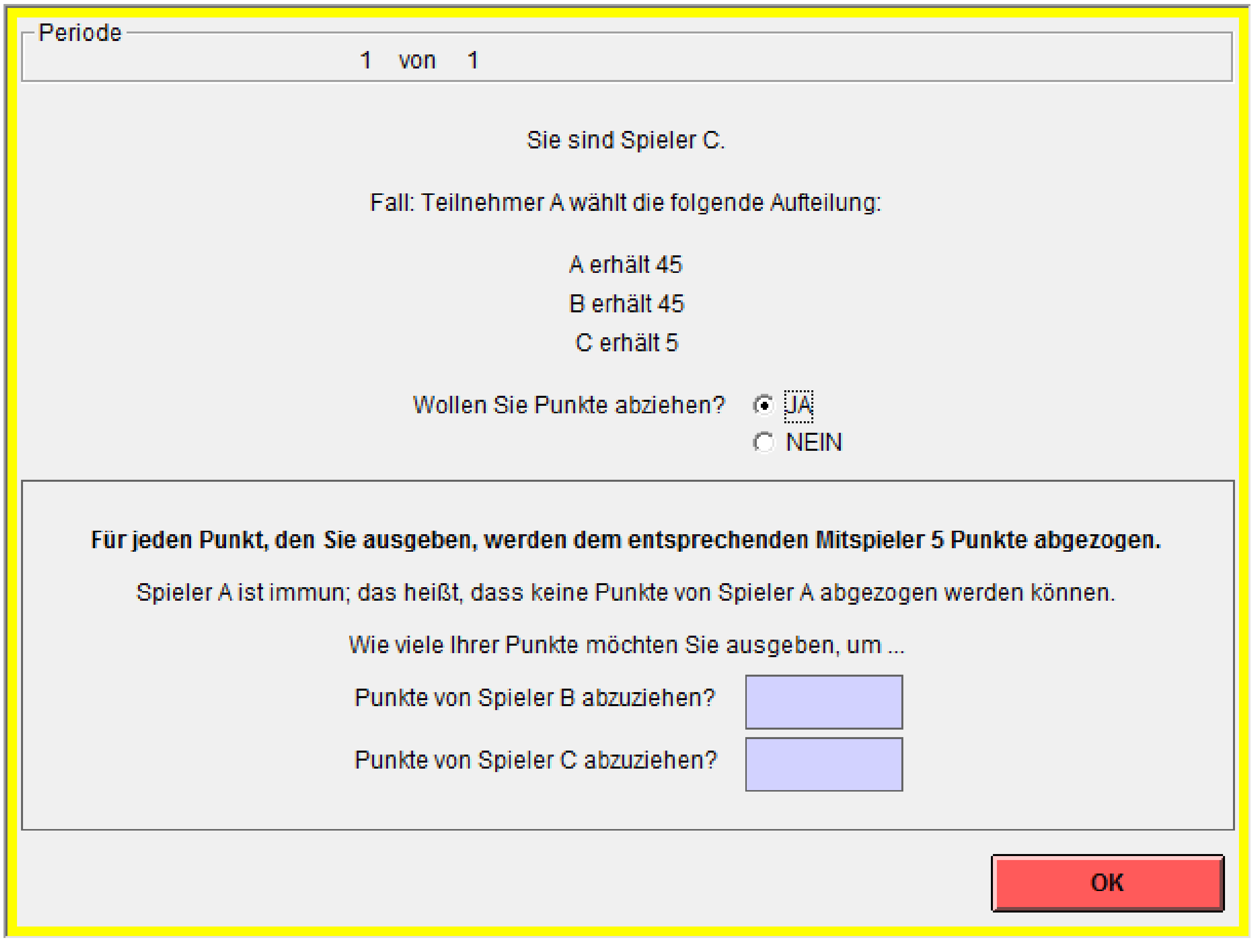
If you are randomly chosen, the decision you make for the case stemming from the decisions of participant A will be implemented. Therefore, each of your three decisions can be relevant for your payment. Please enter your decisions on the following screen:



The example above shows the possible case “**participant A decides for the allocations 45 45 5 5"**.

The screens for the other two outcomes, in which participant A delegates the decision, appear correspondingly. Thus, please take exact note of which case you are making your decisions for!

If you click "YES", the screen shown below then appears:

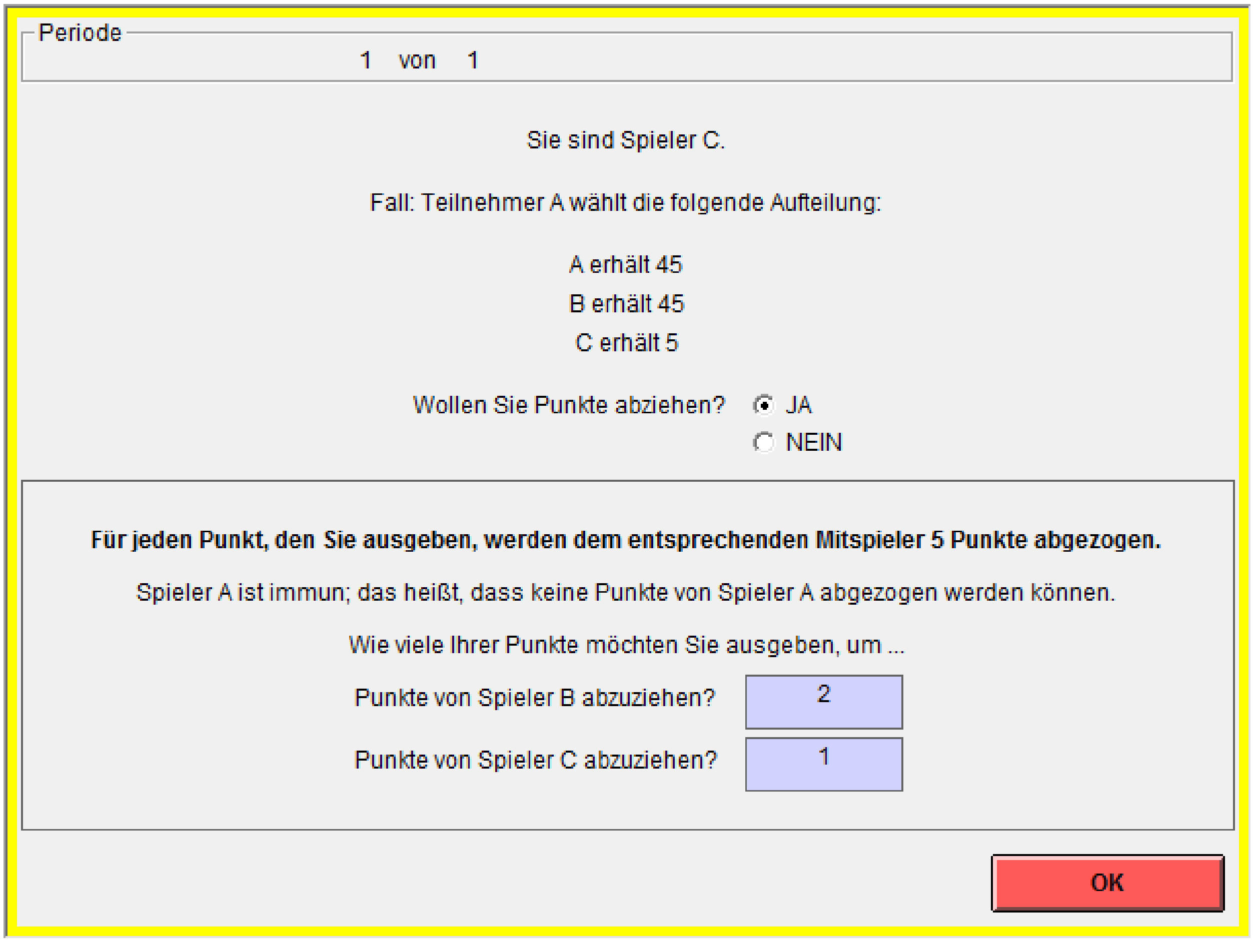


If you click "YES", you can enter the desired amount of point(s) you would like to give up for making five-fold deductions from the other player(s)' earnings in the corresponding two fields. Hence, five points will be deducted from the corresponding participant for every point that you give up.

If you click "NO", the field with the black edge will not appear (or it will disappear again) and you cannot deduct any points.

Once you have made your decision, please click the "OK" button in the lower right corner. You can change your decisions as long as you have not clicked this button.

An example follows on the next page.



In this example, participant C would like to deduct points. He or she has thus clicked on "YES" and the field with the black edge appears. Participant C deducts 2 x 5 = 10 point from participant B and 1 x 5 = 5 points from the other participant C. (This is only an example, and neither a suggestion nor a hint about how you should act.)

After you click the "OK" button, you will get to the next case. You can change your decision as long as you have not yet clicked this button.

Afterwards we will ask you hypothetical questions on how many points you would have liked to give up to deduct five-fold of those points from each of the participants, if participant A would not have been impune. If you click "YES", you would be able to enter the desired amount of point(s) you would like to give up for deductions in three fields that correspond to each of your co-players.

Moreover, we would like to know your expected estimates regarding the decisions made by the other participants.

At the very end we kindly ask you to fill in a questionnaire. All information collected will most certainly be handled with full confidentiality. Do you have any remaining questions?

Practice Questions

Please answer the following practice questions. They only serve to make you more acquainted with the experiment. **The decisions and numerical values in the practice questions are chosen on a purely random basis and are not to be considered as a hint or suggestion as to how you could decide.** Your answers to the practice questions will have no effect on your payment at the end of the experiment.

1. Whose decisions are relevant for the payment at the end of the experiment?

2. Participant A chooses Allocation 1. One participant C is randomly chosen and decides to deduct the point values in bold face type. Please determine the payments which result.

	Participant A	Participant B	The other participant C	The chosen participant C
Allocation	45	45	5	5
Deduction	**-**	**3x5**	**1x5**	
Payments?				

3. Participant A delegates the decision and the computer generates Allocation 1. One participant C is randomly chosen and decides to deduct the point values in bold face type. Please determine the payments which result.

	Participant A	Participant B	The other participant C	The chosen participant C
Allocation	45	45	5	5
Deduction	**-**	**3x5**	**0**	
Payments?				

We would again like to emphasize that the decisions and numerical values in the practice questions are chosen on a purely random basis and are not to be considered as a hint or suggestion as to how you could decide

Please raise your hand after you have solved all the practice questions. We will come to your seat and check your answers.

Once we have checked your answers, it is advisable to seriously consider your decisions in the experiment.

CHAPTER 4

Does "Negative Nancy" Compete Like "Positive Pete"? An Experiment

Relevant Instructions (English Translation).

*Organized according to the design steps described in **Figure 6**.*

Comments written in *Italic*

Welcome!
Welcome to our experiment!

Dear participant, you can earn real money in this experiment. Therefore, it is important that you read these instructions very carefully. If you have any questions please raise your hand and wait for us to come to you.

Communication with other participants during the experiment is strictly forbidden. Not following this rule will result in exclusion from the experiment without payment.

In this experiment, you will be presented with some general knowledge questions. Imagine that you are taking part in a game, like “Trivial Pursuit” or “Who wants to be a Millionaire?”, and you have to choose the correct answer from the three given alternatives.

We will hand you the trivia quiz upside down, with your code number written in large letters on the back. Please make sure that the number written on the back is the same as the code number that you have received on a card in the beginning of the game. Please keep the card with the code number on your desk at all times and keep a good eye on it, since you will need to show your code number card for your cash payment.

For the sake of fairness, all participants must start simultaneously. Please do not turn around the quiz sheet until we tell you to do so.

There are in total four quiz versions. Persons seated next to you will always have a different version from yours. These four versions differ in their difficulty: two of those are comparatively easy, and the other two are comparatively difficult (to answer). You have a 50/50 chance of receiving either an easy or a difficult version.

In each version, there are 15 general knowledge questions and you have exactly 2 minutes to answer as many questions as you can. After those 2 minutes, please turn around your quiz sheets immediately after we ask you to do so and make sure that the back sheet with your code number becomes visible to the experimenters.

You will be assigned a competitor. Your payment in this game will depend on your and your competitor’s relative performance in answering the general knowledge questions. The one of you that answers more questions correctly wins and will receive 8 Euros at the end of the experiment. The loser receives no payment. (If the two competitors correctly answer an equal number of trivia questions, one of the two players will be chosen randomly as a winner and would receive 8 Euros at the end of the experiment and the other player will not receive any payment).

At no point in the game will you know if the trivia quiz you have received is one of the easy or difficult versions. Similarly, you will never know if the trivia quiz your competitor received is one of the easy or difficult versions.

After collecting the quizzes, we will check your answers, (please be patient with us), and will compare your results to the results of your competitor. We will then inform you in written form whether you won or lost the competition. This will not be done orally. (Please remain quite at this time and do not show the news you receive to the subjects sitting around you).

You will not be informed on how many questions you or your competitor have answered correctly, but only whether you won or lost.

More instructions will follow later.

Good Luck!

1. **First Round Trivia**

*Following Michailova's and Katter's (2014) 6 easy, 6 medium-difficulty and 6 difficult trivia questions, gender-neutral both in performance and confidence, 2 comparatively easy and 2 comparatively hard trivia sets were generated with 15 randomly ordered questions each.(*6 easy, 6 medium and 3 hard vs. 3 easy, 6 medium and 6 hard questions). *Ultimately, this is what allowed luck attributions.*

2. **Info on who Won / Lost**

You WON!

Why do you think you won?

OR

You LOST!

Why do you think you lost?

3. Explanation of why Won / Lost

Please cross **one** reason that you think fits best to your case.

You WON, because…

…you think the trivia quiz that you received was relatively easier than your competitor's.

…you are really good at answering this type of questions.

…you are very smart.

…you chose a very good strategy for tackling this task.

…you worked very hard (exerted a lot of effort) to answer as many questions as you could.

…you were well-rested and therefore could perform so well.

…you were very relaxed and therefore could perform so well.

OR

Please cross **one** reason that you think fits best to your case.

You LOST, because…

…you think the trivia quiz you received was relatively more difficult than your competitor's.

…you are not good at answering this type of questions.

…you don’t think you are very smart.

…you chose a bad strategy for tackling this task.

…you did not exert enough effort to answer the questions.

…you were tired and therefore could not perform so well.

…you were nervous and therefore could not perform so well.

4. Payment Scheme Preference

<u>Next Round</u>

You will once more (and last time) receive a quiz. This will consist of 15 general knowledge questions. You have exactly 2 minutes to answer as many questions as you can.

This time there are two quiz versions that differ in difficulty. Participants seating next to you will always have a different version from yours.

You have a 50/50 chance of receiving either an easy or a difficult version. The difficulty level of the version you receive does not depend on the difficulty level of the version you have received in the previous round and it also does not depend on whether you have won or lost in the previous round.

Your competitor in this round is different from your competitor in the 1st round.

In this round, there are two possible payment mechanisms that determine how you would get paid.

Alternative K: You would be paid, as in the previous round, i.e. you receive 8 €, if you win, and 0 € otherwise.

(If the two competitors answer correctly an equal number of trivia questions, one of the two players will be chosen randomly as a winner).

Alternative L: Regardless of your performance on the quiz, a coin will be flipped. If it lands on "Head," you receive 0 €. If it lands on "Tail," you receive 4 €. ***(OR** 8 €, depending on the session).*

You must indicate which one of these two alternative payment mechanisms you prefer. After completing the quiz and answering some additional questions, you will roll a die. If you roll a 1, 2, 3, 4 or 5, you will be paid according to the payment mechanism you prefer. If instead, you roll a 6, you will be paid according to the payment mechanism you did not prefer. In 5/6 of the cases you will be paid according to the payment mechanism you prefer, and in 1/6 of the cases you will be paid according to the payment mechanism you did not prefer. Truthfully indicating the payment mechanism you prefer improves your chances of being paid according to your preferences. In any case, you should do your best answering the trivia questions, since this might determine your payoff.

<u>Examples:</u>

- You prefer Alternative K and roll a 3. If you correctly answer more trivia questions than your assigned competitor, you receive 8 €, and no payment otherwise.
- You prefer Alternative K, and roll a 6. If the coin flip lands on "Tails," you receive 4 € *(**OR** 8 €, depending on the session)*, if on "Head" - 0 €.
- You prefer Alternative L, and roll a 2. If the coin flip lands on "Tails," you receive 4 € *(**OR** 8 €, depending on the session)*, if on "Head" - 0 €.
- You prefer Alternative L, and roll a 6. If you correctly answer more trivia questions than your assigned competitor, you receive 8 €, and no payment otherwise.

Please indicate the payment mechanism you prefer:

□ **Alternative K**

□ **Alternative L**

5. Second Round Trivia

Using www.focus.de "Was Schüler wissen sollten" page (translated as "What pupils should know"), two trivia sets with 15 questions each were generated.

CHAPTER 5

Institution Formation and Cooperation with Heterogeneous Agents

Instructions Translated into English: HET-ASYM Treatment.

Instructions for the other treatments were as similar as possible except for the necessary adjustments concerning the composition of types (in treatments with homogeneous players), the level of obligations (in treatments with the symmetric institution), and the omittance of the first stage in the baseline VCM treatments.

General explanations to the participants

You are now participating in an economic experiment. If you read the following explanations carefully, you will be able to earn a considerable amount of money – depending on your decisions and those of the other participants. Thus it is very important to read these instructions carefully and to understand them.

During the experiment, it is absolutely prohibited to communicate with the other participants. If you have any questions, please ask us: please raise your hand and we will come to your seat. If you violate this rule, you will be dismissed from the experiment and forfeit all payments.

How much money you will receive after the experiment depends on your decisions and those of the other participants. During the experiment, payoffs will be calculated in Taler instead of Euro. Your total income will be calculated in Taler first. The total amount of Taler that you have accumulated during the experiment will be converted into Euro and paid to you in cash at the end of the experiment. The exchange rate from Taler to Euro is as follows:

40 Taler = 1 Euro

The experiment consists of exactly one part. This part is divided into **20 periods**. At the beginning of the experiment you are randomly assigned to a group of three. Thus, there are two other participants in your group. In each group of three, there are two participants of type A and one participant of type B (the difference between type A and type B will be explained in detail shortly). Whether you are of type A or of type B is determined randomly. In all periods your type remains the same, just as the types of the other participants in your group remain the same. You will be interacting with the same two participants in all periods. Neither during, nor after the experiment will you receive any information about the identities of the other participants in your group.

Each period is divided into three stages:

- In the **second stage** you have to decide on how many Taler you contribute to a project and how many Taler you keep for yourself.
- In the **first stage** you can decide if you want to commit yourself and the other participants in your group to certain contributions to the project in stage 2. Only if all participants decide in stage 1 to commit all participants in your group to certain contributions to the project, the contributions will actually be fixed. If not all participants decide to fix the contributions, then you and the other participants in your group will be able to choose any contribution level in the second stage.
- In the **third stage** you get to know the contributions of all participants in your group to the project in stage 2 and the payoffs of all participants in your group in this period.

Detailed information about the course of a period

At the beginning of each period every participant receives 20 Taler. In each period you have to decide on how to use these 20 Taler. You can contribute Taler to a project or put them on a private account. Every Taler that you don't contribute to the project is automatically put on your private account.

Income from your private account:

For each Taler you put on your private account, you earn exactly one Taler. For example, if you put 20 Taler on your private account (thus contributing zero Taler to the project), you would earn 20 Taler from your private account. If, e.g., you would put 2 Taler on your private account (thus contributing 18 Taler to the project), your income from the private account would be 2 Taler. Nobody but you receives Taler from your private account.

Income from the project:

For each Taler that you or another participant in your group contributes to the project, you (and each other participant in your group) earn a certain number of Taler. Each participant's income from the project depends on his or her type and is determined as follows:

$$Type\ \boldsymbol{A's}\ income\ from\ the\ project = \frac{3}{4} * (Sum\ of\ all\ contributions\ to\ the\ project)$$

$$Type\ \boldsymbol{B's}\ income\ from\ the\ project = \frac{1}{2} * (Sum\ of\ all\ contributions\ to\ the\ project)$$

Example 1: the sum of contributions from all participants to the project is 12 Taler (e.g. if you and the two other participants contribute 4 Taler each, or if one of the three participants contributes 12 Taler and the two other participants contribute 0 Taler). Then the two participants in your group who are of type A each receive an income of ¾ * 12 = 9 Taler from the project, and the participant in your group who is of type B receives an income of ½ * 12 = 6 Taler from the project.

Example 2: The sum of contributions from all participants to the project is 36 Taler. Then the two participants in your group who are of type A each receive an income of ¾ * = 27 Taler from the project, and the participant in your group who is of type B receives an income of ½ * 36 = 18 Taler from the project.

<u>Income at the end of a period:</u>

Your income at the end of a period is the sum of your income from your private account and your income from the project:

$Type\ A$:

$Income\ from\ the\ private\ account\ (20 - contribution\ to\ the\ project)$

$+ Income\ from\ the\ project\ (\frac{3}{4} * sum\ of\ contributions\ to\ the\ project)$

$= Income\ at\ the\ end\ of\ the\ period$

$Type\ B$:

$Income\ from\ the\ private\ account\ (20 - contribution\ to\ the\ project)$

$+ Income\ from\ the\ project\ (\frac{1}{2} * sum\ of\ contributions\ to\ the\ project)$

$= Income\ at\ the\ end\ of\ the\ period$

Let us illustrate how your income at the end of a period is calculated using two examples:

Example 1: Assume that you are of type A and contribute 16 Taler to the project, just as the other two participants. The sum of contributions is then 16 + 16 + 16 = 48 Taler. Your income in this example would be:

<u>4 Taler</u> from the private account + <u>¾ * 48 Taler</u> from the project = 4 + 36 = <u>40 Taler</u>.

Example 2: Assume that you are of type A and contribute 0 Taler to the project, while the other two participants contribute 16 Taler each. The sum of contributions is then 16 + 16 + 0 = 32 Taler. Thus, your income would be

<u>20 Taler</u> from the private account + <u>¾ * 32 Taler</u> from the project = 20 + 24 = <u>44 Taler</u>.

The first stage

In the first stage you can decide whether you want to commit yourself and the other participants in your group to a certain contribution to the project in the second stage. All participants decide simultaneously. Only if all participants in your group decide to commit themselves and the other participants to certain contributions, are the contributions in stage 1 actually fixed. In this case contributions will be fixed as follows:

Type A: Contribution of 20 Taler to the project

Type B: Contribution of 8 Taler to the project

If *not all* participants decide to fix the contributions, you and the other participants in your group can freely contribute any number of your 20 Taler to the project in the second stage.

The second stage

At the beginning of the second stage you get to know how each participant in your group decided in the first stage.

If in the first stage all participants decided to fix the contributions in the second stage, then in the second stage you have to contribute the corresponding amount. Thus, if you are of type A you have to enter a contribution of 20 Taler and if you are of type B you have to enter a contribution of 8 Taler. Other inputs are not possible and will automatically be adjusted by the computer program.

In this case the period income of the participants of type A is ¾ * 48 = 36 Taler each and the period income of the participant of type B is 12 + ½*48 = 36 Taler.

If in the first stage not all participants decided to fix the contributions in the second stage, then in the second stage all participants can freely choose any integer contribution between 0 and 20 to the project (0, 1, 2, …, 19, 20).

In this case your period income is computed as indicated above:

Type A: 20 – your contribution to the project + ¾ * (sum of all contributions to the project in your group)

Type A: 20 – your contribution to the project + ½ * (sum of all contributions to the project in your group)

The third stage

In the third stage you get to know the contributions to the project by all participants in your group, as well as their period income. Furthermore, you will again see how each participant in your group decided in the first stage.

Then the current period ends and the next period begins with the same participants. Your type and the types of the other participants remain the same. All participants can then again decide in the first stage whether they want to fix contributions in the second stage. Again, the second stage follows and finally the third stage.

End of the experiment and payment

The experiment ends after 20 periods. Subsequently, we will ask you to answer a few general questions on the computer. Your answers to these questions have no influence on how much money you earn in the experiment. When all participants have filled out the questionnaire, payments will be made. Your total income from the 20 periods will be converted into Euro and paid to you in cash.

Do you have any questions? If so, please raise your hand.

Affidavit

Hiermit versichere ich, dass ich die vorliegende Dissertation selbstständig, ohne unerlaubte Hilfe Dritter angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder unveröffentlichten Schriften entnommen sind, habe ich als solche kenntlich gemacht. Kein Teil dieser Arbeit ist in einem anderen Promotions- oder Habilitationsverfahren verwendet worden.

Elina Khachatryan

Kassel, 13.09.2016