Impartial Decision Makers and other Third Parties in Moral and Social Dilemmas – An Experimental Analysis

Inauguraldissertation

zur

Erlangung des Doktorgrades

der

Wirtschafts- und Sozialwissenschaftlichen Fakultät

der

Universität zu Köln

2014

vorgelegt

von

Dipl.- Volks. Lilia Zhurakhovska

aus

Kiew, Ukraine

Referent: Prof. Dr. Bettina Rockenbach, Wirtschafts- und Sozialwissenschaftliche Fakultät, Universität zu Köln

Korreferent: Prof. Dr. Dirk Sliwka, Seminar für Allgemeine Betriebswirtschaftslehre und Personalwirtschaftslehre, Universität zu Köln

Tag der Promotion: 19.12.2014

Acknowledgments

I would like to express my thanks to Bettina Rockenbach, Christoph Engel, Marco Kleine, and Pascal Langenbach as well as to Armenak Antinyan, On Amir, Sophie Bade, Maria Bigoni, Gary Charness, Angela Dorrough, Susann Fiedler, Uri Gneezy, Sebastian J. Goerg, Kristoffel Grechenig, Olga Gorelkina, Hanjo Hamann, Mark Daniel Heintz, Alexandros Karakostas, Michael Kurschilgen, Janine Meuser, Henrike Moll, Nikos Nikiforakis, Hans-Theo Normann, Serguei Pronkine, Paul Schempp, Paul Smeets, Joel Sobel, J. Forrest Williams, and Erte Xiao for helpful advice, comments, and critique on an earlier version of the thesis. Moreover, I am grateful to my family and friends for their psychological support and their patience.

Content

Introduction

Norms are a cornerstone of social interactions between economic decision makers. More precisely, behavioral norms provide rules that give guidance when people have to take decisions (Bicchieri, 2006). Comprehensible laws are probably the most obvious example of such rules. Everybody knows that they are breaking a rule if they cross the road on a red light or steal. By contrast, many social norms are subjective and imprecise. The answer to the question of what one should do becomes unapparent when it is unknown which behavior others expect and how other people could react to that behavior. Furthermore, many rules are vaguely formulated, i.e., social and distributive norms can often be interpreted individually. The answer to this question becomes even less clear if norms compete, i.e., if one can only obey one rule by breaking another one. Different people can have different reasons to apply different norms (Loewenstein et al., 1993).

People might have difficulties deciding which rule to follow, particularly if their actions have an impact on others. One example of such competing norms is euthanasia. In many countries euthanasia is prohibited by law and yet people often do it because they feel morally obliged to help their relatives or their patients. In the absence of institutions[1] and clearly defined common norms, one reasonable way to decide what one should do might be to form beliefs about the actions of others and to condition own behavior on these beliefs (Fischbacher et al., 2001; Fischbacher and Gächter, 2010). Even knowing the norm, one might decide not to follow it. People might deviate from a behavioral norm if, for example, they perceive the norms as unjust or not legitimate (Tyler, 2006). Imagine a situation in which stealing from someone also means helping another person. In this context a person might face a conflict between several motives: first, people may be driven by selfish motives, second, they may commiserate with the harmed person; third, they may feel sympathy for the potential beneficiary of the action.[2]

The first chapter of the thesis (with Christoph Engel) investigates how subjects behave when facing a normative conflict: Subjects participate in a two-person one-shot prisoner's dilemma experiment. Yet as in a Bertrand-oligopoly situation, in the experiment cooperation with an insider automatically imposes harm on a passive outsider. The treatment manipulations concentrate on the magnitude of the harm. A homo economicus has no reason to cooperate in such a dilemma irrespective of externalities. Yet, if subjects do cooperate in a one-shot prisoner's dilemma in the absence of externalities there are two reasons why, in the presence of negative externality, they might cooperate less. On the one hand, the reason could be that players are less optimistic regarding the other player's cooperativeness. The reasoning for this would be that players do not cooperate because they do not want to be exploited (Rapoport and Chammah, 1965) or that they reciprocate the behavior of the other

[1] The present dissertation uses the term "institution" in the majority of the cases, not to consider some rule or law. Here, the term is rather used for a person or a group of people who enforce a rule. The most prominent example of an institution in this thesis is an authority, i.e., a person who is authorized to take decisions based on someone's actions which influences their outcomes. For a simplification of the writings in the thesis the authority (the player in the role of the authority) is from this point on referred to as "she" and the other participants are referred to as "he."

[2] Selfishness (Smith, 1776) and sympathy for others (Edgeworth, 1881) were introduced in economic theory centuries ago.

active player (Fischbacher et al., 2001; Fischbacher and Gächter, 2010). On the other hand, subjects may cooperate less in such a situation because they are pro-socially motivated. That is, they may care about total welfare and will therefore not want to cause harm to passive bystanders. To disentangle the aforementioned motives of the active players, beliefs about the choices of other active players are elicited. Note that if an active player defects in order to protect the passive outsider he automatically harms the insider. In such a conflict in-group-favoritism (e.g., Chen and Li, 2009) predicts negative externalities will cause only little or no change in the behavior of the active players. In sum, it turns out that in a situation where cooperation harms outsiders, subjects are willing to let down the insider to spare the outsider harm. This even holds when controlling for beliefs. More precisely, relatively small harm is sufficient to isolate this effect. Apparently, when facing a situation of either protecting an innocent outsider or being cooperative with an insider, subjects favor the outsider.

Now, if institutions come into play, normative rules might become more obvious to subordinates, which might help to coordinate their behavior. One way of implementing institutions is by introducing an impartial authority who can punish or reward someone for not behaving according to a particular norm. If an authority is impartial in this context this means that her actions cannot be strategically motived. An impartial authority could be a judge who invests her private time to read her cases carefully, a regulator of an industry appointed by governments, a non-member hired by an association as its discipliner, an officer who receives a fixed income for operating a company owned by a foundation, etc. Now if the authority is impartial but her actions are costly for her it might be difficult for the subordinate to anticipate which behavioral norms the authority will follow. On the one hand, standard economics predicts that people do not engage in costly punishment (or reward) if they do not ever receive a monetary benefit from their action. On the other hand, recent studies argue that people exhibit strong (indirect) reciprocity, which "*means that people are willing to reward friendly actions and to punish hostile actions although the reward or punishment causes a net reduction in the material payoff of those who reward or punish.*" (Camerer and Fehr, 2004, p. 56). Put another way, introducing an impartial authority shifts the challenge from forming beliefs about the behavior of peers to correctly anticipating the norms the authority wishes to implement and the norms she herself will follow. If one anticipates which behavioral norm an impartial authority is willing to reward, one has an incentive to change own behavior according to this norm.

The second chapter (single-authored paper), investigates the introduction of an institution in the form of an impartial authority (third party) into a two-person situation. The impartial authority can reward a stranger for acting according to a desired behavioral norm. The reward is costly and does not lead to higher earnings for her. In the present study the desired norm is the trustworthiness of a second mover toward a first mover. In particular, the study analyzes whether the second mover increases his trustworthiness in anticipation of the reward from the impartial authority as compared to a situation where a reward cannot be expected. Furthermore, it is investigated whether the impartial authority rewards the high trustworthiness of the second mover toward the first mover and if she displays motivational crowding out (Stanca et al., 2009). Put differently, the study investigates whether the authority's reward differs depending on whether the potential recipient is aware of the possibility of receiving a reward or not. To answer these questions, the study applies a trust game followed by a variant of a dictator game.[3] The trustee in the trust game becomes the recipient in the dictator game. Both games are one-shot. The information about the content of the second game is subject to treatment manipulation. In sum, this chapter finds convincing evidence of a positive strong indirect reciprocity and—contrary to Stanca et al. (2009)—no support for the motivational crowding out of positive strong indirect reciprocity. Moreover, the positive strong indirect reciprocity is correctly anticipated by players and leads to higher trustworthiness. To the best of my knowledge, this is the first paper to explicitly study the effect of an anticipated positive strong indirect reciprocity on trustworthy behavior.

The previous paragraph illustrates that subordinates are able to anticipate that complying with desired behavioral norms leads to a reward by an impartial authority. But what if the subjects do not face a simple one-shot sequential-choices situation? In many situations in life one interacts with multiple others and thus more than one person is affected by one's actions. Furthermore, one might face a similar situation involving different people multiple times. If one wants to be a conditional cooperator in such settings one has to form beliefs about the behavioral norm of those people one is currently facing. Obviously, a central impartial authority can help to coordinate (for an overview on organizational economics see Gibbons and Roberts, 2013). Yet, if one is facing different authorities at different points in time, anticipating what norm which authorities are trying to implement becomes an even greater problem. If it is difficult to realize what the desired norm is, you might not even realize if you are deviating from it.

Even when subjects know that they are breaking a rule, they are not always aware of how much damage they can cause before getting punished. Moreover, the extent of the punishment is not necessarily obvious in advance. By far not all laws are as transparent as not crossing the road when the light is red. Numerous laws are subject to interpretation. In a different nexus (e.g., in situations without explicit rules) communication between affected people has been shown to be a helpful coordination instrument (see the meta-analysis by Sally, 1995; the survey by Crawford, 1998; the meta-analysis by Balliet, 2010). Yet communication paired with punishment by an impartial authority has not been studied so far. As mentioned above, a prominent example of impartial authorities trying to enforce rules are judges in court. The decisions judges make often involve punishment. Indeed, in real life punishment is regularly paired with justification. For example, after the court decides whether a criminal is to be sent to prison, a judge usually has to write an explicit verdict stating the reasons for the decision. That is, she has to explain for which crime she chose the punishment and why the punishment is of the respective severity, or why the criminal has not been imprisoned but only put on

[3] The game is a simple dictator game with an efficiency factor, which is called a "helping game" in the literature (e.g., Seinen and Schram, 2006). That is, transfers to a recipient are tripled by the experimenter.

probation or even acquitted. An additional benefit from justifying the punishment is that observers can learn how they should behave. By listening to the verdict against a felon one can learn how to behave in the future. If the reasons for the punishment are made public, even those people who have not been punished so far can adjust their behavior to avoid being punished in the future.

The third chapter (with Christoph Engel) deals with the question of whether justifying punishment can serve as a substitute for monetary punishment, i.e., whether an impartial authority can implement a particular behavioral norm by pairing the punishment with justifications. In the study in a multi-period public goods game subjects can be punished by an impartial authority that does not benefit from the contributions. Punishment is costly for the authority. The group composition changes every period and all subjects remain in their roles throughout the entire experiment. The authority always writes justifications but in the treatments the degree of transparency of the communicated reasons varies gradually. (I.e., either no one receives the justifications, or only the punished person, or all current group members.) In sum, it is found that the more public the reasons are, the less punishment is needed to keep up the same level of cooperation.

As many rules and norms are vague, not only the subordinates but also the authorities do not always know which norm they should try to implement. In a situation in which several competing prominent norms (e.g., distributive rules) are present, communication from a subordinate to an impartial authority can help her to take a reasonable decision. Yet, if an authority is impartial it follows that she will attempt to take an objective decision (cf. Engel and Zhurakhovska, 2012), i.e., she does not have to rely on the consent of the subordinate. Thus, it is possible that introducing communication will not influence her decision. This might explain why so far "voice" (defined as "*participation in decision making by expressing one's own opinion*" (Folger, 1977, p.109)) in procedures involving impartial authorities, has received little attention in economics. Yet having a voice has been widely discussed in other social sciences as an important procedural aspect (cf. Katz, 1960; Tyler et al., 1985; Tyler, 1987; Lind et al., 1990). This concept is of importance in a large set of economic, social, and legal interactions, e.g., in principal-agent relationships (for example, the process of wage setting) or in bargaining situations.

In the last chapter of the thesis (with Marco Kleine and Pascal Langenbach) the interaction between one impartial authority and two subordinates with or without voice is studied. The analysis focuses in particular on the question of whether subjects appreciate having a voice in a decision making process even if their statements do not influence the outcome of the process. In the first part of the experiment the subordinates have to fulfill a real-effort task with an asymmetric workload and an asymmetric piece rate. How the money is generated is common knowledge. Thus, three prominent competing distributive rules evolve: the input-equity rule, the output-equity rule and the equality rule (cf. Nikiforakis et al., 2012). Treatments differ in whether (and if so, how much) one subordinate can communicate with the impartial authority regarding his desired allocation decision. After receiving (or not receiving) this message the authority distributes the money between the two subordinates. In the second part of the experiment, the subject with (or without) voice becomes a sender in a dictator game and the impartial authority becomes the receiver. This transfer is used as a measure of the subordinate's attitude toward the authority and toward the institution as well as to the outcome of the decision. In sum, it turns out that irrespective of the actual decision of the impartial authority, the attitude toward the authority is better if one is involved in the decision making process through voicing one's opinion.

Summing up, the thesis answers several questions concerning how people behave in an economic context if behavioral norms and rules are not obvious. In particular, it is investigated how subjects behave if other people in the role of third parties are involved or if the subjects themselves have the role of third parties. The main findings are that people value fair and efficient processes and, if they have the chance to be responsible for the outcomes of the process, they are happy to contribute to it. People are willing to forego own earnings for "saving", "rewarding", or "punishing" strangers even if this does not lead to potential future monetary gains for themselves. This behavior is correctly anticipated by others and prompts them to behave according to the desired norms. Whether in the role of impartial authorities or in the shoes of subordinates, people are happy to use communication to share their normative views with others even if they do not expect to meet again. And, apparently, this communication can influence the behavior of others in an efficient way.

This thesis applies the toolbox of experimental economics to examine the aforementioned issues from different angles. Each chapter evolves a separate study. In the four individual chapters the behavioral changes are explained by different norms. The norms are applied depending on the different conditions; the different beliefs about the desired and implemented relevant norms by others; and potential own influence on the implementation of norms by others. In the following four chapters the separate studies are presented. My contribution to these studies is as follows: Chapter 2 is single-authored while the remaining chapters have been written in collaboration with co-authors. Christoph Engel had the initial idea for the first chapter while I collected the data. The remaining work (experimental design, hypothesis leading, literature review, data analysis and writing) on this chapter has been conducted jointly and in equal proportion. The same applies to the other joint paper (chapter 3). The last chapter of the thesis has been written jointly with Marco Kleine and Pascal Lagenbach. While Marco Kleine and I contributed the most to the postulating of hypotheses and the data analysis, we all contributed in equal shares to the set-up of the experimental design. The same applies to the collection of the data and the writing. As Pascal Langenbach is a lawyer, he had the initial idea and he contributed the most to the literature review.

The thesis ends with a conclusion which summarizes the design of the studies and the most important findings of each study, and discusses the possible policy implications.

REFERENCES

Bicchieri, Christina. 2006. *The Grammar of Society: The Nature and Dynamics of Social Norms.* New York: Cambridge University Press.

Balliet, Daniel. 2010. "Communication and Cooperation in Social Dilemmas: A Meta-Analytic Review." *Journal of Conflict Resolution*, 54(1): 39-57.

Camerer, Colin F., and Ernst, Fehr. 2004. "Measuring Social Norms and Preferences Using Experimental Games: A Guide for Social Scientists." In J. Henrich, R. Boyd, S. Bowles, C. F. Camerer, E. Fehr, and H. Gintis (eds.), *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*, New York: Oxford University Press.

Chen, Yan, and Sherry Xin Li. 2009. "Group Identity and Social Preferences." *American Economic Review*, 99(1): 431-57.

Crawford, Vincent. 1998. "A Survey of Experiments on Communication via Cheap Talk." *Journal of Economic Theory*, 78(2): 286-298.

Edgeworth, Ysidro Francis. 1881. *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences.* London: Kegan Paul.

Engel, Christoph, and Lilia Zhurakhovska. 2012. "You are in Charge. Experimentally Testing the Motivating Power of Holding a (Judicial) Office." *MPI Collective Goods Discussion Paper*.

Fischbacher, Urs, and Simon Gächter. 2010. "Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Good Experiments." *American Economic Review*, 100(1): 541-556.

Fischbacher, Urs, Simon Gächter, and Ernst Fehr. 2001. "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment." *Economics Letters*, 71(3): 397-404.

Folger, Robert. 1977. "Distributive and procedural justice: Combined impact of voice and improvement on experienced inequity." *Journal of Personality and Social Psychology*, 35(2): 108–119.

Gibbons, Robert, and John Roberts. 2013. *Handbook of Organizational Economics* (eds.), Princeton University Press.

Katz, Daniel. 1960. "The functional approach to the study of attitudes." *Public Opinion Quarterly*, 24(2): 163-204.

Lind, E. Allan, Ruth Kanfer, and P. Christopher Earley. 1990. “Voice, control, and procedural justice: Instrumental and noninstrumental concerns in fairness judgments.” *Journal of Personality and Social Psychology*, 59(5): 952-959.

Loewenstein, George, Samuel Issacharoff, Colin Camerer, and Linda Babcock. (1993). “Self-Serving Assessments of Fairness and Pretrial Bargaining.” *The Journal of Legal Studies*, 22(1): 135-159.

Nikiforakis, Nikos, Charles N. Noussair, and Tom Wilkening. 2012. “Normative conflict and feuds: The limits of self-enforcement.” *Journal of Public Economics*, 96(9-10): 797-807.

Rapoport, Anatol, and Albert M. Chammah. 1965. *Prisoner's Dilemma. A Study in Conflict and Cooperation*. Ann Arbor: University of Michigan Press.

Sally, David. 1995. “Conversation and Cooperation in Social Dilemmas. A Meta-analysis of Experiments from 1958 to 1992.” *Rationality and Society*, 7(1): 58-92.

Seinen, Ingrid, and Arthur Schram. 2006. “Social status and group norms: Indirect reciprocity in a repeated helping experiment.” *European Economic Review*, 50: 581–602.

Smith, Adam. 1776. *The Wealth of Nations*. Edited by Edwin Cannan, 1904. Reprint. New York: Modern Library, 1937.

Stanca, Luca, Luigino Bruni, and Luca Corazzini. 2009. “Testing Theories of Reciprocity: Do Motivations Matter?” *Journal of Economic Behavior & Organization*, 71(2): 233-245.

Tyler, Tom R. 1987. “Conditions leading to value-expressive effects in judgments of procedural justice: A test of four models.” *Journal of Personality and Social Psychology*, 52(2): 333-344.

Tyler, Tom R. 2006. *Why People Obey the Law*. Princeton: Princeton University Press.

Tyler, Tom R., Kenneth Rasinski, and Nancy Spodick. 1985. “Influence of voice on satisfaction with leaders.” *Journal of Personality and Social Psychology*, 48(1): 72-81.

Chapter 1

Good News: Harm To Outsiders Reduces Cooperation With Insiders – An Experiment

CHRISTOPH ENGEL[*] AND LILIA ZHURAKHOVSKA[*+#]

[+]UNIVERSITY OF COLOGNE,
[#]UNIVERSITY OF ERLANGEN-NUREMBERG,
[*]MAX PLANCK INSTITUTE FOR RESEARCH ON COLLECTIVE GOODS, BONN

Keywords: conditional cooperation, social preferences, negative externalities, prisoner's dilemma, beliefs, efficiency.

JEL codes: C90, D01, D03, D62, D63

I. INTRODUCTION

Morally, imposing harm on an innocent outsider is bad. Reproach is even stronger if the harm does not correspond to a direct benefit for insiders. Such action is not even selfish; it is purely spiteful. Now what if imposing harm does not benefit insiders individually, but makes them better off as a group? This is the case if insiders face a dilemma, in which they impose harm on a bystander whenever at least one of them cooperates. In such a situation, the moral balance becomes more complicated.

In the field, the conflict between kindness at the interior and meanness at the exterior is not infrequent. Sometimes, being mean is the very purpose of cooperation, as in a military coalition or in a trade union. At other instances, the harm is more a side-effect which is deliberately taken into account. Those closer to the source of a river build a dam, knowing that this deprives those closer to the estuary of the benefits of the river. Or a municipality builds a landfill to keep garbage off its streets, knowing that this puts the groundwater of neighboring municipalities at risk.

The most obvious motivation of our paper, however, is oligopoly. Viewed from inside the supply side of the market, competition may be interpreted as a prisoner's dilemma. In this perspective, collusion is the equivalent of cooperation, competitive behavior is defection. Individually, each supplier is best off if the other suppliers are faithful to the cartel, and she undercuts the collusive price or, for that matter, surpasses her quota. Yet if they cooperate, suppliers impose a distributional loss on the demand side; and they generate a deadweight loss, to the detriment of society.

If both players of a symmetric one-shot prisoner's dilemma game hold standard preferences, both players defecting prescribes the unique equilibrium. Yet, as has been shown long ago, to a remarkable degree this prediction is violated in the laboratory, i.e., subjects cooperate to a remarkable extend in such games (see already Rapoport and Chammah, 1965). Literature on discrimination and group identity (e.g. Chen and Li, 2009) has taught us that subjects have a general tendency to cooperate more with insiders than with outsiders even if this can reduce their own payoffs. This literature would predict that there is no reason to reduce cooperation with an insider if cooperation harms outsiders. In fact, a recent study by Engel and Rockenbach (2011) shows that in a linear public good game with passive outsiders insiders do not cooperate less if cooperation makes outsiders worse off. Their results are rather in line with the explanation that insiders try to increase the payoff gap between themselves and outsiders.

Following the argument of Chen and Li (2009) and Engel and Rockenbach (2011) when testing the power of morally grounded reticence to impose harm on a passive outsider in a laboratory experiment, we were therefore surprised that subjects behave more pro-social towards outsiders than towards insiders. We tested participants on a one-shot symmetric prisoner's dilemma. In the main the *Harm* treatments, insiders impose harm on a third, passive participant whenever at least one of them cooperates. In a control condition, harm is imposed only in case of joint cooperation. We gradually vary the level of harm in the treatments.

Contrary to the described argument, we find that active players are significantly less likely to cooperate with insiders if cooperative moves impose harm on outsiders and that they cooperate less the higher the level of harm is. If we allow in our design for harm being avoided as soon as only one player defects, even with a very small level of harm cooperation is statistically significantly lower in the *Harm* treatments compare to the *Baseline*. Subjects' behavior is not simply driven by conditional cooperation (e.g., Fischbacher et al., 2001; Fischbacher and Gächter, 2010) or by aversion to the sucker's payoff (e.g., Rapoport and Chammah, 1965), i.e., results do not change if we control for subjects' beliefs.

In a sense, our results are good news: subjects do not want to let down innocent passive outsiders. However, it is surprising that they prefer to spare the outsiders even low levels of harm by forgoing a joint cooperative payoff for the group of insiders to which they belong.

The remainder of the paper is organized as follows: Section 2 relates the paper to the existing literature. Section 3 introduces the design. Section 4 makes theoretical predictions. Section 5 presents and discusses results. Section 6 concludes.

II. LITERATURE

The effects of externalities on passive outsiders have only rarely been studied. To the best of our knowledge, they have not been tested in a standard prisoner's dilemma. Most related is a paper by one of us with another co-author. Engel and Rockenbach (2011) study a standard repeated four-person linear public good game with three passive outsiders. They vary the direction of the externality and the endowment of the outsiders. Insiders do not cooperate more if this has the additional advantage of making outsiders better off, and they do not cooperate less if this has the additional disadvantage of making outsiders worse off. Rather results are in line with insiders trying to increase the payoff gap between themselves and outsiders. We build on this design, but focus on the most interesting effect, the apparent absence of reticence to impose harm on passive outsiders. Our design differs in the following respects: we implement a one-shot game. This excludes the shadow of the future as a potential confounding factor. We use two-person games. This excludes expectations and experiences about heterogeneity as a possible explanation. We use various levels of harm. This allows us to test whether the effect is confined to certain, in particular to fairly small levels of harm. Finally, and most importantly, we elicit beliefs. That way we can disentangle cognitive and motivational effects of imposing harm on passive outsiders.

Other relevant studies are for example Güth and van Damme (1998). They present an ultimatum game with an externality on an inactive third player who has no say. The proposer decides how to divide the pie between three players. The division is executed if and only if the responder accepts. Otherwise, all three players receive nothing. In this game, the outsider receives very little. If the responder only learns the fraction the proposer wants to give the outsider, proposers keep almost everything for themselves. In anticipation, responders are very likely to reject the (mostly unknown) offer. Bolton and Ockenfels (2010) study lottery choice tasks in which the actor's choice also influences the payoff of a non-acting second player. This induces participants to take larger risks, provided the safe option yields unequal payoffs. Abbink (2005) plays a two-person bribery game in which corruption negatively affects passive workers. He concludes that reciprocity between briber and official overrules concerns about distributive fairness towards other members of the society. Ellman and Pezanis-Christou (2010) study how a firm's organizational structure influences ethical behavior towards passive outsiders. A firm of two players decides on its production strategy, which influences a passive third player. They find that horizontally organized firms in which the firm's decision corresponds to the average of both individual decisions are less likely to harm the outsider than consensus-based firms or firms in which one of both members is the boss. There is a rich experimental literature on oligopoly (see the meta-study by Engel, 2007), yet it does not focus on the fact that oligopoly is socially embedded.

III. Design

In our experiment, we have a *Baseline* with just two active players, and three treatments with an additional passive outsider who is negatively affected by insiders choosing a cooperative move.

III.1. The Game

TABLE 1
Payoff Matrix

	C	D
C	$R€,\ R€,–h€$	$S€, T€,–h€$
D	$T€, S€,–h€$	$P€,\ P€,0€$

C cooperative move, D defective move
In each cell, left payoff is for the row player, and middle payoff is for the column player, the right payoff is for the outsider (if there is one)

Our game is a standard symmetric two-choice prisoner's dilemma with two active players and *n* passive players, as in Table 1. If both players cooperate, each of them earns *R€*, and the passive players earn –*h*€. If one cooperates and the other defects, the cooperator earns *S€*, while the defector earns *T€*, and the passive players earn –*h*€. If both defect, each of them earns *P€*, and the passive players earn *0€*. Following the labels originally introduced by Rapoport and Chammah (1965), *R* stands for reward, *S* for sucker, *T* for temptation and *P* for punishment.

We choose the following parameters: *R=5*, *S=0*, *T=10*, *P=2.45*. In the *Baseline* the number of passive players is *n=0* (and thus no harm to passive players is implemented) and in the treatments we have *n=1*. In three treatments we vary the level of harm, i.e., in *Small* the level of harm h=.3; in *Middle h=2.1*; in *High h=4.8*.

III.2. Considerations Motivating the Design

In a stylized way, our game captures a one-shot Bertrand market with constant marginal cost where two firms individually decide whether to set the collusive price (C) or to engage in a price war (D). Our introduction of harm to a passive outsider is meant to capture the loss in consumer welfare, and in total welfare, inherent in anticompetitive behavior. If both engage in (tacit or explicit) collusion, both set the monopoly price and split the monopoly profit evenly (R=T/2). If only one of them starts a price war, it undercuts the collusive price by the smallest possible decrement. As is standard in the theoretical literature, in this interpretation of our design we assume the decrement to be infinitesimally small, which implies that the aggressive firm cashes in the entire monopoly profit (T), while the firm that is faithful to the cartel receives nothing (S). Therefore, in the experiment, we do not confine harm to the situation where both active players cooperate. Yet if both firms start fighting, they end up in the Nash equilibrium. This removes harm to the opposite market side, and the deadweight loss (harm only if both cooperate).

We deliberately avoid a market frame. This not only makes sure that our results are not driven by the frame. It is also necessary to isolate the effects of externalities. In a market setting, from their world knowledge, subjects would know that collusion is illegal and might be motivated by this social and legal norm, rather than by their reticence to impose harm.

Our choice of parameters is primarily driven by experimental concerns. We create the maximum difference between the sucker payoff S=0 and the temptation payoff T=10. That way, both the premium for beating one's opponent and the penalty for losing in competition are largest. By contrast, the payoff in case both players defect almost holds the middle between the reward for cooperation and the penalty for being outperformed. For this payoff, we deliberately have not chosen either extreme.

III.3. Robustness Check

As mentioned above, in the oligopoly application that triggered this research, it suffices for a single firm to set the collusive price in order to impose externalities. Yet in other applications, like a union going on strike, unilateral action is not harmful, while coordinated action is. In the main experiment one cannot be sure that by defecting one avoids harm to the passive player. As a consequence, if we do not find a higher defection rate in the treatments as compared to the *Baseline* this could be attributed to uncertainty about the effectiveness of one's own action. To test the robustness of our findings, in the treatments in a subsequent stage we also implement a prisoners' dilemma with harm imposed on a passive player only in the case of joint cooperation. The content of this stage is not announced in the beginning of the experiment. Group composition differs and all subjects remain in their roles, i.e., an active player remains an active player and a passive player remains a passive player. Table 2 presents the payoffs in this design.

TABLE 2
Payoff Matrix Robustness Check

	C	D
C	$R€,\ R€,–h€$	$S€, T€, 0€$
D	$T€, S€, 0€$	$P€,\ P€, 0€$

C cooperative move, D defective move
In each cell, left payoff is for the row player, and middle payoff is for the column player, the right payoff is for the outsider (if there is one)

III.4. Beliefs

After each prisoner's dilemma we elicit beliefs about the cooperativeness of active players in the game, i.e., each active player must estimate the number of active players in her session, who chose the cooperative move (labeled neutrally). If a participant gets the number exactly right, she earns an additional 2€. If her estimate is within a range of +/- 2 around the true number, she earns an additional 1€.

III.5. Procedures

Subjects know that the experiment has several parts,[4] but receive specific information about the content of each part only immediately before playing the relevant part. Group composition varies between the parts. No information about other participants' decisions and therefore about any earnings is given to the subjects before the end of the entire experiment so that independence is preserved. All instructions are read out aloud by the experimenter immediately before the relevant part to achieve common knowledge about the procedure.[5]

The experiment was run in February 2014 at the University of Bonn with a computerized interaction using z-Tree (Fischbacher, 2007). ORSEE (Greiner, 2004) was used to invite subjects from a subject pool of approximately 3500 subjects. Each subject participated only in one session. We collected 48 independent observations from active players in the prisoners' dilemma in almost each

[4] We ensure that the number of stages is the same in all conditions to exclude any behavioral changes caused simply by differences in expectations about the duration of the experiment. The data from subsequent stages of the experiment are not relevant for the present paper and are therefore not presented here.
[5] See section I.1. in the Appendix for an English translation of the instructions.

condition (we have only 44 independent observations in the treatment with h=2.1[6]). In total 258 subjects participated in the experiment (70 in the role of passive players). In each session of the treatments, at the beginning of the experiment the active and the passive players were randomly picked from the pool of participants present in the laboratory. Subjects were on average 23.54 year old (range 17-55). 56.57% were female. Almost all of them were students, with various majors. Each session lasted about one and a half hours. Participants in each session received a show-up fee of 10€ that suffices to cover potential losses. Subjects earned on average 20.61€ (equivalent to $28.02 on the last day of the experiment, range 5.2€-45€).[7]

IV. HYPOTHESIS

Since our game is a one-shot prisoner's dilemma, money-maximizing agents defect in the *Baseline* as well as in each treatment.

Empirically, many experimental participants have been found to be conditional cooperators (Fischbacher et al., 2001; Fischbacher and Gächter, 2010). Pure conditional cooperators (at least weakly) prefer cooperation over defection if they expect their counterpart to cooperate with certainty. This implies that they resist the temptation to exploit their counterpart. If conditional cooperators are perfectly optimistic, they do not expect to run a risk. Consequently, perfectly optimistic conditional cooperators cooperate.

In line with previous experiments, we expect conditional cooperation to be more prevalent than outright selfishness. Yet we expect participants to be less than perfectly optimistic. If their beliefs make them less optimistic, conditional cooperators run the risk of not getting gains from cooperation. If they are neutral to risk and losses, they compare the expected payoff from cooperation with the expected payoff from defection. If they are pure conditional cooperators in the sense of not desiring gains from exploitation, they discount gains from cooperation by their subjective degree of pessimism, and compare them with the minimum payoff in case they defect.

If an actor defects while the other actor cooperates, two effects combine. Payoffs are unequal, with an advantage for the defecting actor (as modelled in Fehr and Schmidt, 1999; and in Bolton and Ockenfels, 2000). If the first actor expects the second to cooperate, she also violates the second actor's expectation of reciprocal action (as modelled in Rabin, 1993 and in Dufwenberg and Kirchsteiger, 2004). The reciprocity motive is not affected by adding a third player in the treatments. Since the third player is inactive, she has no chance to reciprocate kind or unkind behavior. By contrast, in the treatments the inequity balance is more complicated. If both active players defect, they are symmetrically favored with respect to the inactive player. If both cooperate, they are favored even

[6] In two sessions, not all invited participants showed up, so that we could not fill one group of three.
[7] The average payoff was 23.25€ in the *Baseline.* In the *Small* subjects earned on average 20.70€ (15.95€ for passive players), in *Middle*, 20.34€ (15.06€ for passive players), and in *High* 19.02€ (15.30€ for passive players).

more. If one defects while the other cooperates, the defecting one is strongly favored in comparison with both other players, while the cooperating one has a payoff of 0€, and the inactive player incurs a loss of –h€.[8]

This line of argument, however, neglects that in case both active players defect, the payoff difference in comparison with the inactive players "is not their fault". Actually if they want to be kind to the inactive player, defecting is the best thing both can do. In situations that are structurally similar to the one tested here, it has been shown that intentions matter in the assessment of fairness (Falk et al., 2008). Taking this into account, the treatments expose active players to a conflict between fairness with the inactive player (calling for both defecting) and the motives behind conditional cooperation (calling for cooperation, provided the player is sufficiently optimistic about cooperativeness in this population). However, defection has a double dividend in this game: the defecting active player at least secures the payoff she expects for herself if both players defect, and she acts as best she can to protect the inactive player from harm. The effect should be stronger the more severe the harm to the outsider. We therefore predict:

> **Hypothesis:** Cooperation will be the highest in *Baseline,* it will be lower in *Small*, even lower in *Middle* and the lowest in *High*.

V. RESULTS

Figure 1 (panel a) collects choices in the *Baseline* and panel b choices in the main treatments. In the main experiment the degree of cooperation monotonically decreases in the harm inflicted on a passive outsider. Yet, non-parametrically we do not find a significant difference between the *Baseline* and the *Small* (h=.3€) (Chi2: $p = 0.149$). By contrast, the difference between the *Baseline* and the two remaining treatment is significant (Chi2: $p \leq 0.003$)[9].

[8] That is not the case in the second prisoner's dilemma in the play treatments. Here, if one active player cooperates only she receives a lower payoff than the other active player, but the passive player does not suffer harm. Nonetheless, even here the passive player does not earn more than the cooperating active player.
[9] Comparing the treatment with each other the only statistically significant difference is between *Small* and *High* (*Small* vs. *Middle*: Chi^2: $p = 0.111$; *Middle* vs. *High*: Chi^2: $p = 0.302$; *Small* vs. *High*: Chi^2: $p = 0.009$).

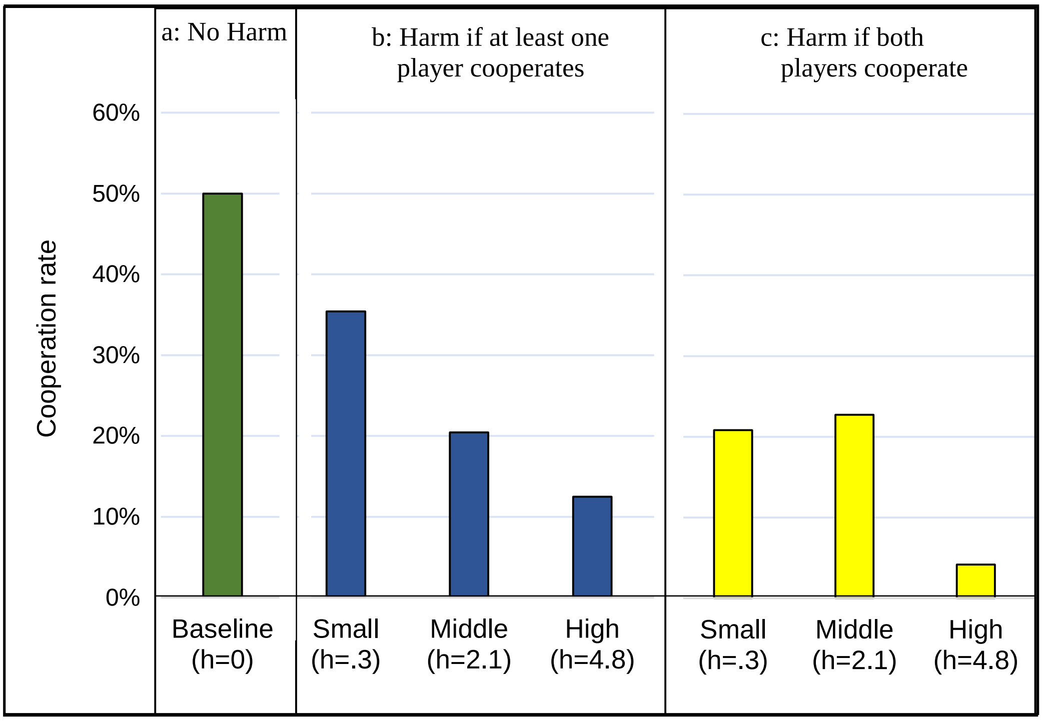


FIGURE 1

Degree of Cooperation in Prisoner's Dilemma with Single Level of Harm

On the vertical axis, one can see the cooperation rate in percent of active players. The harm imposed on outsider (in €) is presented on the horizontal axis. The scale goes from 0 (green bar panel a: corresponds to the data in the *Baseline*) to 4.8 (panel b: treatments with single cooperation leading to harm; panel c: joint cooperation leads to harm).

If we run a simple OLS model with robust standard errors (Table 3, Model 1), we find the same results as with our non-parametric test, i.e., we do not find a negative significant effect on cooperation rates in *Small* (h=.3€) but we find this effect in *Middle* (h=2.1€) and in *High* (h=4.8€). When controlling for beliefs (Table 3, Model 2), the results do not change and we find a positive significant effect of the belief.[10]

[10] We find similar results if we run separate models comparing the *Baseline* with only one of treatments or if we use ordered probit models.

TABLE 3

Explaining cooperation rate – comparison Baseline and treatments (h=.3; 2.1; 4.8) with single cooperation sufficient for harm

linear probability model (OLS), robust standard errors

Dependent variable: cooperation rate

	Model 1	Model 2
Small (.3€)	-.1458	-.0793
	(.101)	(.076)
Middle (2.1€)	-.2955***	-.2114***
	(.095)	(.074)
High (4.8€)	-.375***	-.3243***
	(.087)	(.070)
Belief		.9706***
		(.093)
Constant	.5***	.0189
	(.073)	(.067)
N	188	188
P model	<.001	<.001
R^2	0.1000	0.4303

OLS regressions. Robust standard errors are presented in parentheses. *Small*, *Middle* and *High* are treatment dummies that equal 1 for observations in these treatments. *Baseline* is the reference category. Significance at the 10%, 5% and 1% by *, ** and ***.

In the field, harm to outsiders comes in two different conditions: if insiders effectively overcome their dilemma, or if only one of them attempts to do so by setting a cooperative move. The previous data has focused on the latter situation. In conclusion, we compare subjects' behavior in the *Baseline* with the former situation. As Figure 1 panel c demonstrates, if harm requires successful coordination, the difference to the *Baseline* is even more pronounced. If harm is pronounced (h=4.8€), only 4% of the active participants cooperate. For all levels of harm, the treatment effect is significant (Chi2: $p \leq 0.003$).[11]

Compared to a situation, in which harm is imposed if at least one subject cooperates, cooperation is lower (Wilcoxon: *Small* (h=.3€): p = .0522; *Middle* (h=2.1€): p = 0.7389; *High* (h=4.8€): p = 0.0455). This supports our intuition that by allowing for harm only in case of joint cooperation we put our results to a more conservative test.

[11] Comparing the treatment with each other only *Small* and *Middle* are not statistically significantly different, yet all other comparisons are statistically significantly different (*Small* vs. *Middle*: Chi^2: p = 0.826; *Middle* vs. *High*: Chi^2: p = 0.008; *Small* vs. *High*: Chi^2: p = 0.014).

TABLE 4

Explaining cooperation rate – comparison Baseline and treatments (h=.3; 2.1; 4.8) with joint cooperation as requirement for harm

linear probability model (OLS), robust standard errors

Dependent variable: cooperation rate

	Model 1	Model 2
Small (.3€)	-.2917***	-.2024***
	(.094)	(.071)
Middle (2.1€)	-.2727***	-.1735**
	(.097)	(.096)
High (4.8€)	-.4583***	-.3205***
	(.079)	(.068)
Belief		.9075***
		(.097)
Constant	.5***	.0502
	(.073)	(.069)
N	188	188
P model	<.001	<.001
R^2	0.1492	0.4635

OLS regressions. Robust standard errors are presented in parentheses. *Small*, *Middle* and *High* are treatment dummies that equal 1 for observations in these treatments. *Baseline* is the reference category. Significance at the 10%, 5% and 1% by *, ** and ***.

Using a simple OLS regression with robust standard errors we find even more support for the results presented in Table 3, i.e., in this specification even small harm is sufficient to reduce cooperation significantly as compared to the *Baseline* (Table 4, Model 1).[12] Again, we find a positive significant effect of optimism about the cooperativeness of others on own cooperation (Table 4, Model 2). Of course, these results should be treated with caution since choices were elicited after the participants have played the prisoners' dilemma with harm becoming effective as soon as only one active player cooperates. Note that we did not give feedback on any payoffs or choices of others before all parts of the experiment were completed. Nonetheless, we cannot exclude order effects. Therefore, we only present these data as a robustness-check.

In sum, first, from our results we learn that with harm imposed on outsiders subjects become less cooperative than without harm. Second, this result cannot be explained by simple conditional cooperation, i.e., controlling for pessimism about the cooperativeness of others does not change the result. Third, if we impose the structure that harm can be easily avoided as soon as at least one subject defects our result becomes even more clear, i.e., even small harm leads to a significant reduction of cooperation. Summing up the findings from all tests we have converging evidence for our hypothesis:

[12] We find similar results if we run separate models comparing the *Baseline* with only one of treatments or if we use ordered probit models.

Result: If active participants who play a simultaneous symmetric prisoner's dilemma know that attempted or effective cooperation imposes harm on a passive outsider, they cooperate less the higher the harm is.

VI. Conclusion

We find that if cooperation imposes harm on an innocent outsider, this makes cooperation less likely in a symmetric one-shot-two-person prisoner's dilemma. Given our hypothesis, our results are plausible. Yet, given previous literature (e.g., Engel and Rockenbach, 2011; Chen and Li, 2009) one could have expected the opposite result, i.e., we could have found that insiders try to maximize the difference in payoffs between themselves and the outsides. In a different context Neugebauer et al. (2009) and Fischbacher and Gächter (2010) have shown that most participants who are in principle willing to cooperate nonetheless desire to have a higher payoff than other active players. Through adding a third passive player, our design gives even more scope for payoff comparisons. Therefore, it is remarkable that relatively small harm (*Middle* (h=2.1€)) is enough to significantly reduce cooperation and that in a situation, in which avoiding cooperation is extremely easy, even smaller harm (*Small* (h=.3€)) effects subjects' choices in the described way.

From a policy perspective, our findings are welcome news. The prediction of the Bertrand model (with homogenous goods) seems to be true (see e.g. the discussion in Tirole, 1988). We find that the mere structure of the game is suffice to deter collusion. The fact that the suppliers' dilemma is embedded in a market mitigates the otherwise pronounced ability to overcome the dilemma. Of course, one should be careful in extrapolating from our laboratory results to predicting behavior of profit maximizing firms in the market. Nonetheless, antitrust has reason to doubt the pure willingness of suppliers to incur the risk of cooperation. Making salient that others could suffer from a person's choices can make this person think twice about what is the right thing to do.

VII. References

Abbink, Klaus. 2005. "Fair Salaries and the Moral Costs of Corruption. In B. N. Kokinov (ed.), *Advances in Cognitive Economics*, Sofia, NBU Press.

Bolton, Gary E., and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity and Competition." *American Economic Review*, 90(1): 166-193.

Bolton, Gary E., and Axel Ockenfels. 2010. "Betrayal Aversion. Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States. Comment." *American Economic Review*, 100(1): 628-633.

Chen, Yan, and Sherry Xin Li. 2009. "Group Identity and Social Preferences." *American Economic Review*, 99(1): 431-57.

Dufwenberg, Martin, and Georg Kirchsteiger. 2004. "A Theory of Sequential Reciprocity." *Games and Economic Behavior*, 47(2): 268-298.

Ellman, Matthew, and Paul Pezanis-Christou. 2010. "Organisational Structure, Communication and Group Ethics." *American Economic Review*, 100(5): 2478-91.

Engel, Christoph. 2007. "How Much Collusion? A Meta-Analysis on Oligopoly Experiments." *Journal of Competition Law and Economics*, 3(4):491-549.

Engel, Christoph, and Bettina Rockenbach. 2011. "We Are Not Alone. The Impact of Externalities on Public Good Provision." MPI Collective Goods Preprint No. 2009/29.

Falk, Armin, Ernst Fehr, and Urs Fischbacher. 2008. "Testing Theories of Fairness – Intentions Matter." *Games and Economic Behavior*, 62(1): 287-303.

Fehr, Ernst, and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, 114(3): 817-868.

Fischbacher, Urs. 2007. "z-Tree. Zurich Toolbox for Ready-made Economic Experiments." *Experimental Economics*, 10(2): 171-178.

Fischbacher, Urs, and Simon Gächter. 2010. "Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Good Experiments." *American Economic Review*, 100(1): 541-556.

Fischbacher, Urs, Simon Gächter, and Ernst Fehr. 2001. "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment." *Economics Letters*, 71(3): 397-404.

Greiner, Ben. 2004. "An Online Recruitment System for Economic Experiments." In K. Kremer, and V. Macho (eds.), *Forschung und Wissenschaftliches Rechnen.* Gesellschaft für Wissenschaftliche Datenverarbeitung Bericht, Göttingen: Datenverarbeitung, 63: 79–93.

Güth, Werner, and Eric Van Damme. 1998. "Information, Strategic Behavior, and Fairness in Ultimatum Bargaining. An Experimental Study." *Journal of Mathematical Psychology*, 42: 227-247.

Neugebauer, Tibor, Javier Perote, Ulrich Schmidt, and Malte Loos. 2009. "Selfish-biased Conditional Cooperation. On the Decline of Contributions in Repeated Public Goods Experiments." *Journal of Economic Psychology*, 30(1): 52-60.

Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." *American Economic Review*, 83(5): 1281-1302.

Rapoport, Anatol, and Albert M. Chammah. 1965. *Prisoner's Dilemma. A Study in Conflict and Cooperation*. Ann Arbor: University of Michigan Press.

Tirole, Jean. 1988. *The Theory of Industrial Organization*. Cambridge, Mass.: MIT Press.

Chapter 2

Strategic Trustworthiness via Unstrategic Third-Party Reward – An Experiment

LILIA ZHURAKHOVSKA

UNIVERSITY OF COLOGNE,
UNIVERSITY OF ERLANGEN-NUREMBERG,
MAX PLANCK INSTITUTE FOR RESEARCH ON COLLECTIVE GOODS, BONN

Key words: strong indirect reciprocity, third-party reward, trust game, helping game, anticipation, norms

JEL: C90, D03, D63

I. INTRODUCTION

> *"Virtually every commercial transaction has within itself an element of trust, certainly any transaction conducted over a period of time. It can be plausibly argued that much of the economic backwardness in the world can be explained by the lack of mutual confidence."*
>
> (Kenneth J. Arrow, 1972, p. 357)

The concept of trust has always played a crucial role in economics (cf. Arrow, 1972). It can be a money-maximizing strategy to trust someone more the higher that person's incentives are to be trustworthy. In repeated settings, one is trustworthy (or cooperates) because one expects others to reciprocate (positively or negatively) one's behavior, and thus mutual trustworthiness leads to higher earnings for the actors. Strategically motivated reciprocity can be carried out by the person who is directly affected by someone's act himself (on direct reciprocity, see, e.g., Fehr and Gächter, 2000; Andreoni and Miller, 1993) or by another person (on indirect reciprocity, see, e.g., Rockenbach and Milinski, 2006; Seinen and Schram, 2006; Engelmann and Fischbacher, 2009). In modern societies, however, many contacts are anonymous, indirect, and rarely (if ever) repeated. One example of such transactions is anonymous online trades between private persons via platforms such as www.ebayclassifieds.com/ or www.craigslist.org/. Consequently, it is important to study how trustworthiness can be enhanced by the anticipation of a reward or punishment by an impartial stranger who does not have an incentive to reciprocate (strong indirect reciprocity). I.e., in this paper, "positive strong indirect reciprocity" is defined as a third-party reward in situations in which any strategic concerns for the third parties can be excluded.[13] Consider the following example, which illustrates the meaning and meaningfulness of anticipated (positive) "strong indirect reciprocity" for trustworthiness. Think of a politician who runs for a post. The voters invest their trust in the candidate by voting for him. Assume an institutional design in which reelection is not possible, and thus in which the politician has no strategic incentives to reciprocate that trust. Typically, not everything that is done behind the curtains of a public institution is transparent to the public. Therefore, it is not easy fully to judge how well the politician does his job. Now image, after the politician has been elected, his institution implements a new policy. This new policy implies that reports on the effort provided by the politicians are publicly available. Imagine you learn from the reports that the politician does a good job. Imagine further that you are the head of a bank in which the very same politician applies for a loan. First of all, beyond all money-maximizing concerns from your perspective: would you give him the loan with a higher probability, compared to a situation in which you had not learned that he

[13] Carpenter and Matthews (2004) and Carpenter et al. (2004) relate to "*strong indirect reciprocity*" as "*social reciprocity*". Camerer and Fehr (2004, p. 56) simply define "*Reciprocity means that people are willing to reward friendly actions and to punish hostile actions although the reward or punishment causes a net reduction in the material payoff of those who reward or punish.*"

has reciprocated his voters' trust (positive strong indirect reciprocity)? Second, would the probability of your giving him the loan be even higher if you knew that while the politician was doing his job he did not know that the report might become publically available (motivational crowding out/in of strong indirect reciprocity)? Third, would the politician put more effort in his work if he were able to anticipate that this could be rewarded by others who have no extrinsic incentives to do so (higher trustworthiness in anticipation of positive strong indirect reciprocity)? Finally, would more voters elect the politician if they knew in advance that he would have an incentive to invest in his good reputation in anticipation of a reward by others (higher trust)? This paper attempts to answers all these questions. It provides evidence that positive strong indirect reciprocity exists; it is anticipated by potential recipients; and it can change these recipients' previous behavior in an efficient way.

To the best of my knowledge, this is the first paper explicitly to study the effect of anticipated positive strong indirect reciprocity on trustworthy behavior.[14] Following the call by Almenberg et al. (2011),[15] it is one of the few papers to study positive strong indirect reciprocity. Additionally, in this paper, the impartial stranger has a richer strategy set to condition her[16] rewards on the history of her beneficiary than in previous studies. So far, only one study analyzes whether there is a crowding out of strong indirect positive reciprocity, if a potential recipient of this reward can act strategically (Stanca et al., 2009). Interestingly, in contrast to Stanca et al. (2009), the present study finds no crowding out.

In the present experimental paper, a trust game is implemented and it is followed by a helping game[17] with a different group composition, i.e., the trustee of the trust game in one group becomes a receiver in a helping game in a different group. The helper has the possibility to reward a co-player conditional on his performance in the trust game. In the *Baseline,* subjects first play the trust game and receive the instructions for the helping game only afterwards, while in the *Anticipation* treatment, subjects are informed about the content of both games at the beginning of the experiment. Since the experiment is one-shot, any strategic concerns for the helpers are ruled out in both treatments.

Nonetheless, many helpers make positive transfers to trustees and send significantly more money as the trustees' return transfers increase. Helpers apparently care more about socially desirable behavior than about the motives behind trustees' transfers, i.e., on average, helpers' transfers are the same, regardless of whether the helping game is announced or not. Trustees anticipate helpers' behavior if the helping game is announced, i.e., the absolute level of return transfers as well as the

[14] In Stanca (2009) and Stanca et al. (2009), the anticipation effect of a reward by an impartial stranger on the behavior of a first-mover (instead of an effect on a second-mover in anticipation of a reward by an impartial third-party) is reported. However, this aspect is not the focus of these papers.

[15] Almenberg et al. (2011) state "*While costly punishment has received the lion's share of attention, costly rewarding also plays an important role in human prosociality.*" They continue: "*[A] sizeable amount of evidence exists for the importance of rewarding in human cooperation. Yet the reward-based analog to third party punishment, where I reward you in an anonymous one-shot interaction because you have cooperated with somebody else, remains largely unexplored.*" (Almenberg et al., 2011, p. 75, p. 77).

[16] Throughout the paper, the female form "she" is used for the third parties (and for investors, i.e., for players A in the experimental design) and the male form "he" for the other players.

[17] The helping game is, in fact, a simple dictator game with an efficiency factor. The name of the game, which is often used in the literature, might be misleading, since the game does not necessarily have anything to do with help for a person in need.

relative level to the investments are higher in the *Anticipation* treatment as compared to the *Baseline*. Investments, on average, do not differ between the *Anticipation* treatment and the *Baseline*.

The remainder of the paper is organized as follows: first, an overview of the relevant literature is presented; afterwards, the design of the experiment is explained. Next, hypotheses are explored. The results and the statistical analysis of the data are presented in the penultimate section. In the last section, the conclusions are drawn.

II. LITERATURE

I am aware of only three studies on positive strong indirect reciprocity.[18] In contrast to the present study, none of these three studies analyzes strategic versus unstrategic trustworthiness (due to an anticipation of a reward by an impartial stranger). The most closely related papers to the present study are Stanca (2009) and Stanca et al. (2009).[19] In both studies, the return transfer in a one-shot variant of the trust game either comes from the recipient of the investment or from a stranger.[20] In contrast to the present study, both papers focus on the strong indirect reciprocity (return transfers) and do not analyze in detail the effect on behavior of a player in *anticipation* of strong indirect reciprocity (a change in investment). Stanca (2009), however, does report not finding significant differences in the investments. This means that in his experiment the strategic motives for the players awaiting direct or indirect reciprocity do not seem to matter. Stanca et al. (2009) hypothesize that the motives behind the reciprocated action can crowd out strong indirect reciprocity (rewards by impartial strangers). The treatment difference here is whether the first mover knows that the second stage will follow. The results confirm their hypothesis, i.e., if the strong indirect reciprocator knows that the first mover was aware of the second stage, she transfers a smaller amount compared to a situation in which the second game was not announced. Notably, the results in the present study are not in line with their hypothesis.

Almenberg et al. (2011) implement a one-shot dictator game where a player can transfer either none, half, or all of her endowment to another player, and a third party can either punish or reward the dictator. Furthermore, the number of recipients, the achievable share, and the effectiveness of the reward and punishment given by the third party are varied across treatments. In all treatments, dictators are aware of the presence of the third party. Thus, an effect of anticipated strong indirect reciprocity cannot be studied in their setting. The authors' main findings show that selfish behavior is punished while generous behavior is rewarded, and that rewards are at least as common as punishments.

[18] In fact, the basic design in Almenberg et al. (2011) is very similar to Stanca (2009) and Stanca et al. (2009). In all three studies presented in this section, player A transfers an amount of money to player B which can be observed and rewarded and/or sanctioned by player C.

[19] Here only the papers on the topic of downstream/social indirect reciprocity (A acts towards B and C acts as a reaction to this in a certain way towards B) are discussed, since these papers are most relevant for the present study. Nevertheless, it is important to mention that there are also interesting papers on generalized/upstream indirect reciprocity (A acts towards B and B acts as a reaction to that in a certain way towards A). Notable examples are Dufwenberg et al. (2001); Boyd and Richerson (1989); Greiner and Levati (2005); Güth et al. (2001).

[20] He calls it a gift-exchange game.

Quite a few studies analyze negative strong indirect reciprocity. Only one recent paper examines the anticipation effect of punishment. Balafoutas, Nikiforakis, and Grechenig (2014) demonstrate, in a one-shot, three-player taking game, that taking rates decrease in anticipation of unstrategic punishment by an impartial third party. Furthermore, their paper shows that third-party punishment significantly decreases if counter-punishment directed towards the third party is allowed. However, the focus of their paper lies on counter-punishment, which could be interpreted as an emotional reaction or revenge.

Carpenter and Matthews (2004) run a repeated public-goods experiment with punishment. In one treatment, only members of the own group can be punished, while in the other treatment, members of the own as well as of another group can be punished. The authors find evidence for the existence of negative strong indirect reciprocity, i.e., members of stranger-groups are punished. Fehr and Fischbacher (2004) suggest that Carpenter and Matthews (2004) "*could not rule out third-party punishment for reasons of self-interest*". There is also a strong disciplining component in their design, i.e., punishing someone should lead to more norm compliance in future periods. Fehr and Fischbacher (2004) find strong evidence for third-party punishment in their one-shot, three-person dictator experiment. Bernhard et al. (2006) run one-shot dictator games with third-party punishment in Papua New Guinea. They find that in-group members are avenged more than out-group members, while the affiliation of the punished person does not play a role for the punishers' decisions. On the contrary, norm violators expect to be punished less if the third party belongs to their peer group than if she belongs to a different group.

III. Design

This section is divided into three subsections. In the first subsection, the experimental design is explained; in the second, the motivation behind this design is discussed; and finally, the experimental procedures followed in the experiment are reported.

III.1. Game and Treatments

The game consists of two parts and subjects were aware of that.[21] In the *Baseline*, they receive specific information about the content of each part only immediately before playing the relevant part of the experiment. In the *Anticipation* treatment, instructions for both parts are handed out at the beginning of the experiment.[22] Subjects are explicitly told that they cannot lose the money they have earned in a previous part in any of the subsequent parts. In the experiment, the experimental currency unit (ECU)

[21] In total, there are three parts. Part 1 and Part 2 are described in this section. Part 3 is a standard risk aversion elicitation experiment (Holt and Laury, 2002). The results of the risk aversion measure are not reported, as these are not relevant for the present study.

[22] A treatment that comprises only the trust game was run as well. The results of the *trust game-only* treatment do not differ significantly from the results of the trust game in the *Baseline*; therefore, they are not reported here.

is used. All instructions are read out aloud by the experimenter to achieve common knowledge about the procedure.[23]

At the beginning of the experiment, each subject is randomly assigned one of the three roles A, B, or C. Players keep their roles for the two parts of the experiment. The roles A and B are assigned to 11 subjects each and the role C is assigned to 2 players per session. The distribution of the roles is not made explicit to the subjects. They only know which roles there are, their own role, and, at the relevant point in time, the role of their co-player. The group composition differs between the parts, i.e., players from part 1 do not meet in part 2. The game is played only once.

Part 1:

In part 1, the reduced trust game (TG) (first introduced by Berg et al., 1995) is played by two players, A and B, who move sequentially. The players are endowed with $E^{TG} = 100$ ECU each. At first, player A (from now on called investor) decides how many ECU she wants to send to B (called trustee from now on). Her transfer $t^{TG}{}_A$ = X can be 0, 10, 20, 30, 40, 50, or 60 (from now on called player A's investment).

The investment is tripled by the experimenter and then transferred to B. In case of a positive investment, B can make a return transfer to A. His return transfer is $t^{TG}{}_B$ = Y*X, where Y can be 0, 1, 2, or 3. The decision by player B is elicited via the strategy method (Selten, 1967), i.e., B decides about Y for each possible X.[24] Table 1 displays player B's strategy table. After player B has made his decision, all players are informed about their payoffs from the trust game ($\pi^{TG}{}_i$). These are:

$$\pi^{TG}{}_A = E^{TG} - t^{TG}{}_A + t^{TG}{}_B \text{ for player A and } \pi^{TG}{}_B = E^{TG} + t^{TG}{}_A*3 - t^{TG}{}_B \text{ for player B.}$$

Players C do not take any decision in the TG, nor are they informed about the decisions of the other players. Players C receive a fixed payoff of $\pi^{TG}{}_C = 100$.[25]

[23] See section II.1 in the Appendix for an English translation of the instructions.
[24] Brandts and Charness (2011) show that, if the focus lies on the comparison of decisions within strategies, using the strategy method can be problematic. For the comparison between treatments, the main limitation of that method is that the "*strategy method provides a lower bound for testing for treatment effects*" (p. 392). A similar argument is made in Fischbacher et al. (2012).
[25] Please note that, in case of zero investment, $\pi^{TG}{}_A = \pi^{TG}{}_B = \pi^{TG}{}_C = 100$.

TABLE 1

Experimental Design – Trustees' Strategy

In case player A has sent me…			
… **0 Taler**, my income has thereby increased by **0 Taler**.			
… **10 Taler** and my income has thereby increased by **30 Taler**, I will now send	ʘ nothing ʘ the transfer ʘ double the transfer ʘ triple the transfer	(0) (10) (20) (30)	to player A.
… **20 Taler** and my income has thereby increased by **60 Taler**, I will now send	ʘ nothing ʘ the transfer ʘ double the transfer ʘ triple the transfer	(0) (20) (40) (60)	to player A.
… **30 Taler** and my income has thereby increased by **90 Taler**, I will now send	ʘ nothing ʘ the transfer ʘ double the transfer ʘ triple the transfer	(0) (30) (60) (90)	to player A.
… **40 Taler** and my income has thereby increased by **120 Taler**, I will now send	ʘ nothing ʘ the transfer ʘ double the transfer ʘ triple the transfer	(0) (40) (80) (120)	to player A.
… **50 Taler** and my income has thereby increased by **150 Taler**, I will now send	ʘ nothing ʘ the transfer ʘ double the transfer ʘ triple the transfer	(0) (50) (100) (150)	to player A.
… **60 Taler** and my income has thereby increased by **180 Taler**, I will now send	ʘ nothing ʘ the transfer ʘ double the transfer ʘ triple the transfer	(0) (60) (120) (180)	to player A.

In the box, the screen for the elicitation of trustees' choices via the strategy method is depicted. In the first column, the trustee can see how high the investment could have been. In the second column, the participant sees a radio button, on which he can click – for each possible investment X>0, he can choose how much he wants to send back to his investor, i.e., he can choose for each investment X>0 his return transfer X*Y, where Y can be 0 ("nothing"), 1 ("the transfer"), 2 ("double the transfer"), or 3 ("triple the transfer"). Only the transfer decision for the relevant situation will become payoff-relevant.

Part 2:

TABLE 2

Experimental Design – Helpers' Strategy

1st Screen

If, in Stage 2, my co-player is …

- … a **Player C**, …

  … I will now send ____ Taler.

- … a **Player B** who has been sent **0 Taler** in the first experiment, …

  … I will now send ____ Taler.

2nd-6th Screens

If, in Stage 2, my co-player is …

- … a **Player B** who has been sent **X Taler** in the first part of the experiment and…

  … sent back **0*X Taler**, I will now send ____ Taler.

  … sent back **1*X Taler**, I will now send ____ Taler.

  … sent back **2*X Taler**, I will now send ____ Taler.

  … sent back **3*X Taler**, I will now send ____ Taler.

In the first box, the first screen of the elicitation of helpers' choices via the strategy method is depicted. Here, the helper can choose how much she wants to send to her co-player in case her co-player is a player C or in case her co-player is a player B who has not received an investment. The second box shows the remaining screens of the helpers' strategy method. Here, the helper can choose how much she wants to transfer to her co-player in case he is a player B and has received an investment of X (X=10 on the 2nd screen, X=20 on the 3rd screen, etc.) and in case he has then sent back Y*X (Y can be 0, 1, 2, or 3). On the actual screens of the players, instead of "X", "0*X", or "1*X" etc., the absolute numbers of the respective transfers are written. The helper can insert in each line a number between 0 and 100. Only the transfer decision for the relevant situation will become payoff-relevant.

For the helping game (HG) (similar to the dictator game, as in Forsythe et al., 1994), played in part 2, new groups with two players are formed. Each group consists of a player A and a co-player. The randomly selected co-player is either player C or player B, who has not been matched with this player A in the trust game, i.e., absolute stranger matching is implemented. Player A (from now on called the helper) is endowed with $E^{HG}_A = 100$ ECU. She can transfer any natural number of ECU (t^{HG}_A) from 0 to 100 to her co-player. The transfer is tripled by the experimenter and then transferred to the relevant co-player. Player A's decision is elicited using the strategy method (Selten, 1967).[26] She has to make a

[26] Again, one could argue that the strategy method prompts subjects to take different decisions for different situations. However, the results show that helpers indeed condition their transfers on the relative return transfers of the trustees, but less on the investments that trustees receive. Furthermore, helpers' transfers are more strongly correlated with trustees' history in the *Anticipation* treatment compared to the *Baseline*, which again cannot be explained by the use of the strategy method.

decision for every possible composition, i.e., she has to state how much she wants to transfer in case her co-player is player B or in case her co-player is player C. Additionally, supposing B is her co-player, she can make her decision conditional on the history of the player B in the TG, i.e., she can make her transfer conditional on any possible return transfer t^{TG}_{B} of her co-player to his investor given any possible prior investment t^{TG}_{A} her co-player might have received. This means that the helper makes 26 transfer decisions. Table 2 gives an overview of helpers' strategy space. Only the payoff-relevant transfer decision is realized, based on whether the co-player is in fact a player B or a player C and, if applicable, based on player B's actual return transfer t^{TG}_{B} and on the investment he has received. After the decisions have been made, the players are informed about the actual group composition and the relevant transfer.[27] The payoffs in the helping game (π^{HG}_{i}) are:

$\pi^{HG}_{A} = E^{HG}_{A} - t^{HG}_{A}$ for the helper and $\pi^{HG}_{co\text{-}player} = t^{HG}_{A}*3$ for the co-player.

After the subjects have finished the experiment, the subjects complete questionnaires regarding their attitudes towards trust, risk, and reciprocity, as well as demographics.

III.2. Motivation for the Design

A restricted version of the trust game is used for reasons of simplification, i.e., this allows helpers to have a complete overview of every possible situation they might be facing. This method allows for testing for strong indirect reciprocity of helpers, i.e., it provides detailed data for helpers' transfers conditional on previous behavior of their co-players.

The inclusion of a player C, who does not take any actions, helps to identify an individual benchmark for the helpers' general willingness to help. Helpers' transfers to C players cannot be interpreted as a reward for any previous action. These players have the same income as players B, who are passive, i.e., who do not receive an investment and who therefore cannot make a return transfer.

In the experiment, a player A becomes the helper in the helping game. Charness and Rabin (2002) have developed a theory that analyzes disinterested social preferences (non-self-interested distributional preferences). One potential critique is that in the present study the helping game involves a mixture of self-interested and disinterested preferences. Giving each player feedback about the payoffs after each game and controlling for these payoffs in a regression analysis (instead of trying to elicit beliefs about their earnings) helps to disentangle these motives. Furthermore, participation in the first part of the experiment facilitates the understanding of the strategy method in the helping game.

[27] Subjects are informed in the instructions that the players who are not randomly selected to become the helper's co-players, can earn money *m* in an additional task. It is not made explicit in the instructions how many subjects have to perform that task, nor how much money can be earned in that task. Only the players who are chosen to perform that task receive additional instructions for the task on their computer screen. The task is to count the number of zeros in tables that consisted of 150 randomly ordered zeros and ones (similar to the task used in Abeler et al., 2011). Each correct answer increases *m* by 50 ECU. The payoff for these players is therefore $\pi^{HG}_{other\ player} = m$. In each session, 10 of the helpers were matched with one player B each and one helper was matched with one player C. Therefore, one B player and one C player were not chosen to become a co-player of the helper.

III.3. Experimental procedure

Experiments were run at the University of Bonn in May 2010 and in January 2012. The experiment was programmed and conducted using z-Tree (Fischbacher, 2007). Four sessions with 96 subjects were held, leading to 22 independent observations per treatment for the roles A and B, respectively. Subjects were invited from the University of Bonn using ORSEE software (Greiner, 2004) and had no experience with related experiments, i.e., they had neither participated in trust games nor in dictator or helping games in the past. Most of the subjects were students. 19 of the subjects were economics students, 20 were law students. The remaining subjects came from various different disciplines. 44 participants (46%) were female. The average earning was 12.52 Euro.[28] The sessions lasted 70 minutes on average.

IV. HYPOTHESES

In the helping game, under the assumption of pure payoff maximization, the theoretical prediction is zero transfers. A self-interested, payoff-maximizing trustee has no monetary incentive for a positive return transfer, irrelevant of the information about the helping game. A rational, self-interested, payoff-maximizing investor anticipates this and does not invest any points in any treatment. From the theoretical point of view, under the assumption of pure payoff maximization and common knowledge, the unique Nash Equilibrium predicts zero transfers in all games and in all treatments.

However, it has been shown that, in helping games, positive transfers are observed (cf., e.g., Forsythe et al., 1994), and in trust games positive transfers are made in both directions (see, e.g., Berg et al., 1995). The first is explained by social preferences such as warm glow (Andreoni, 1990). The latter is explained by strong direct reciprocity modeled by, e.g., Rabin (1993), Dufwenberg and Kirchsteiger (2004), Falk and Fischbacher (2006), or Levine (1998). These models assume that an actor has a reciprocity parameter and that the person has a positive utility from punishing unkind action and rewarding kind action. Which actions are perceived as kind depends on the particular model. In fact, the intuition of the model by Levine (1998) is not restricted to two-player direct interactions. In the following, this model will be used to derive (most of the) behavioral hypotheses in the present study. In Levine (1998) a player i = 1, …., n receives a *direct utility* of u_i and has a coefficient of altruism $-1 < a_i < 1$. Therefore, he receives an *adjusted utility* of

$$v_i = u_i + \sum_{j \neq i} \frac{a_i + \lambda a_j}{1+\lambda} u_j,$$

where $0 \leq \lambda \leq 1$. In other words, when $\lambda > 0$, a person derives a positive utility from his own direct utility (which can be his payoff) and, in addition, a positive utility from rewarding another person for

[28] That corresponded approximately to \$15.25 during the first wave of the experiment and approximately to \$15.96 during the second wave of the experiment.

his altruism, i.e., *i*'s utility from *j*'s utility is greater the greater *j*'s coefficient of altruism is.[29] In the model, a player *i* maximizes her utility given her preferences and her beliefs about the preferences of her co-players. In the present experiment, the helper does not need to form her own beliefs about the preferences of her co-player, since she can infer them from the trustee's return transfers to his investor. Assuming that at least some helpers have a coefficient of altruism $a_i > 0$ and $\lambda > 0$, it follows:

Hypothesis 1a: The higher a helper's co-player's return transfers in the trust game are, the more help will be transferred.

The second key research question is whether helpers' evaluation of trustees' altruism differs between the *Baseline* and the *Anticipation* treatment. Levine's model does not make specific assumptions about intentions behind the displayed altruism. Therefore, from Levine (1998) it does not follow that helpers' transfers should differ between the *Baseline* and the *Anticipation* treatment. I.e., if helpers in the experiment care mainly about socially desirable behavior, they should not make different transfers in the *Baseline* compared to the *Anticipation* treatment. However, Falk et al. (2008) show that intentions matter.[30] If a trustee anticipates that he will be rewarded by a helper in the *Anticipation* treatment, and therefore increases his return transfer for strategic reasons, his act might be perceived as less kind. Falk et al. (2008) argue that players acting out of intrinsic motives will be rewarded more than other players. This finding suggests that, in the present experiment, helpers' transfers should be higher in the *Baseline* than in the *Anticipation* treatment because, in the *Baseline*, trustees' return transfers cannot be strategically motivated. Similarly, Stanca et al. (2009) find that more strong indirect reciprocity is displayed when strategic motivations can be ruled out. This leads to:

Hypothesis 1b: For every possible return transfer of trustees, helpers' transfers are lower in the *Anticipation* treatment than in the *Baseline*.

Following the model in Levine (1998), one can derive the following prediction about trustees' return transfers in the *Baseline*:

Hypothesis 2a: The trustees in the trust game make higher positive return transfers the higher the investment they receive.

[29] Depending on the parameters, the model can also make predictions on negative reciprocity and on unconditional altruism. However, since punishment is not possible in the present experiment, these predictions will not be discussed in this paper. Transfers to players C can be explained by unconditional altruism of helpers.

[30] Some theories on reciprocity incorporate intentions as well (e.g., Rabin, 1993; Charness and Rabin, 2002).

In the *Anticipatio*n treatment, "*predictive power of the theory is about what we would expect from a signaling model*" (Levine, 1998, p. 605). Since the trustee not only cares about being altruistic (reciprocal) to others (to the investor), but also cares about his personal utility (his payoffs), he has to form beliefs about whether signaling that he is altruistic (trustworthy) can be beneficial to him. In the *Baseline*, the expected probability of being rewarded for revealed altruism (trustworthiness) is zero.[31] In the *Anticipation* treatment, however, the trustee knows that a helper has the opportunity to reward him. If at least some trustees expect some helpers to have a coefficient of altruism $a_i > 0$ and $\lambda > 0$, they should expect to receive higher transfers in the helping game the more trustworthy they are, i.e., they anticipate the prediction in Hypothesis 1a. If a positive number of trustees expect their helper's a_i to be large enough to offset the monetary loss from higher return transfers, these trustees have an incentive to make higher return transfers in the *Anticipation* treatment than in the *Baseline*. Consequently, one can derive the following hypothesis:

Hypothesis 2b: The trustees in the trust game make higher positive return transfers in the *Anticipation* treatment than in the *Baseline*.

Costa-Gomes et al. (2012) find a positive correlation between an investors' investment and his optimism about a return transfer. Hence, if at least some investors anticipate the predictions about return transfers, which follows from Levine (1998), this leads to:

Hypothesis 3: Investors will make higher investments in the *Anticipation* treatment than in the *Baseline*.

V. RESULTS

This section is organized as follows: first, the hypotheses concerning the transfers are tested non-parametrically[32] and then robustness of the results with parametric tests including further control variables is checked. The results are presented in the same order as the respective hypotheses.

V.1. Helpers' Transfers

This section first investigates whether helpers' transfers depend on relative return transfers made by trustees in the trust game and on the investment received by the trustees. Then, treatment differences on helpers' transfers are tested. In addition, it is analyzed which level of the return transfers is

[31] One could argue that by not announcing the helping game in the *Baseline*, but by informing the players that a second game will follow, some subjects might expect their actions in the trust game to have an influence of their later payoffs. However, the expected probability for that is centennially greater in the *Anticipation* treatment than in the *Baseline*.
[32] Throughout this paper, reported p-values are always two-sided, unless stated otherwise.

particularly strongly rewarded, i.e., what level of return transfers is regarded as especially altruistic and therefore worth to be rewarded.

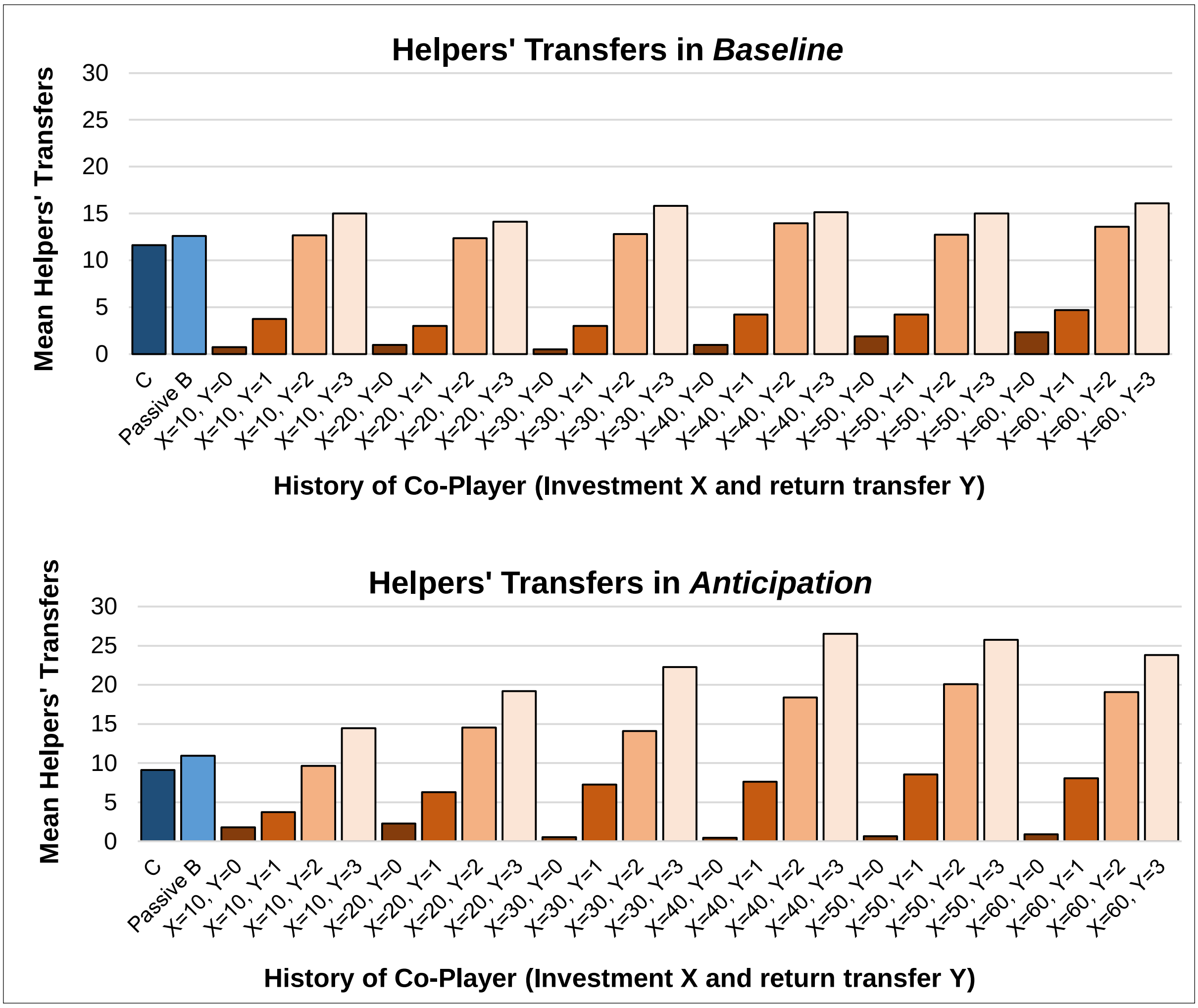


FIGURE 1

Helpers' Transfers by Treatment

The upper figure displays helpers' transfers in the *Baseline* and the lower figure in the *Anticipation* treatment. On the x-axis, the exact condition is displayed, i.e., one can see if the co-player is either Player C or a passive Player B (no investment and thus no opportunity for a return transfer) or an active Player B who has received an investment of X and has made a relative return transfer of Y (Y=0: trustees keeps full transfer; Y=1: trustee returns transfer and keeps rent; Y=2: equal split; Y=3 full return). On the y-axis, mean helpers' transfers are displayed for the particular situation. Standard errors are indicated.

One quarter of helpers (11 out of 44) always make zero transfers in the helping game (7 in *Baseline* and 4 in *Anticipation*).[33] As one can see in Figure 1, higher relative return transfers are rewarded by higher helpers' transfers (Spearman's Rho: r_S=0.4815, p=0.0000; *Baseline*: r_S= 0.4352, p=0.0000; *Anticipation*: r_S=0.5286, p=0.0000).[34] Figure 1 furthermore suggests that helpers' transfers

[33] The likelihood for a helper to make any positive transfer is not different between the treatments: 1-sided Fisher's exact = 0.244.
[34] This finding is further supported by the Wilcoxon signed-rank tests (see Table 5 in the Appendix II.2.).

for a particular return transfer do not depend on investments (Spearman's Rho: r_S=0.0152, p=0.6211; *Baseline*: r_S=0.0118, p=0.7875; *Anticipation*: r_S=0.0221, p=0.6123).[35] Overall, transfers are very similar in the *Baseline* and in the *Anticipation* treatment. The simple non-parametric analysis of all average transfers of the helpers yields no significant treatment difference (Mann-Whitney rank-sum |z|= 1.112, p-value = 0.2661).[36]

TABLE 3

Explaining helpers' transfers – comparison Baseline and Anticipation treatment

Random effects Tobit regressions ("helpers" as group)

Dependent variable: Helpers' transfers in the helping game to active trustees

	Model 1	Model 2	Model 3	Model 4
Anticipation	10.05	5.66	9.87	-1.24
	(10.59)	(10.38)	(8.52)	(9.43)
Relative Return Transfers	16.43***	16.43***	16.41***	14.82***
	(.72)	(.72)	(.72)	(1.00)
Investment	.12***	.12***	.12***	.044
	(.04)	(.04)	(.04)	(.06)
*Anticipation**Relative Return Transfer				2.94**
				(1.37)
*Anticipation**Investment				.14*
				(.08)
Own Investment		.03	-.23	-.23
		(.26)	(.23)	(.26)
Own Profit in TG		.36**	.12	.12
		(.17)	(.15)	(.15)
Transfer to passive B			.34	.71
			(.47)	(.45)
Transfer to C			.71	.32
			(.45)	(.47)
Constant	-47.22***	-82.30***	-62.71***	-56.65***
	(7.99)	(20.41)	(17.12)	(17.25)
N	1056	1056	1056	1056
N of group	44	44	44	44
P model	<.001	<.001	<.001	<.001
Wald Chi2	528.48	530.67	539.19	550.51

Random effects Tobit regressions. Standard errors are presented in parentheses. The *Anticipation* dummy equals 1 for all observations of the *Anticipation* treatment, *relative return transfers* controls for the relative return of a trustee (Y) for a given investment, *investment* controls for the investment the trustee has received, *own investment* is the investment the helper has transfers in the trust game himself to his trustee, *Anticipation*relative return transfer* and *Anticipation*investment* are interaction terms, *own profit in TG* controls for the helper's profit from part 1 of the experiment. *Transfer to passive B* and *Transfer to C* are the levels transferred to passive players. Significance at the 10%, 5%, and 1% level is denoted by *, **, and ***, respectively. Left-censored = 577; right-censored = 14.

[35] This finding is further supported by the Wilcoxon signed-rank tests (see Table 7 in the Appendix II.2.).
[36] See Table 8 in the Appendix II.2. for all pairwise comparisons.

Using the random effects Tobit regression model, one can confirm the visual impression.[37] The dependent variable is the helpers' transfers in the helping game. In no model in Table 3 is the dummy variable for the *Anticipation* treatment significant. Model 1 in Table 3 shows that the main determining factor for helpers' transfers is the relative return transfer of trustees, i.e., higher relative return transfers yield higher helpers' transfers.[38] The investment the trustees received (resulting in a higher or lower absolute return transfer) has a significant, but very small, positive influence on helpers' transfers. The coefficient is not significant once one controls for an interaction between the treatment and the effect of the investment (see Models 4).

Model 2 controls for helpers' own experience in the trust game. It shows that pure willingness to send positive transfers does not lead to higher helpers' transfers (variable *own investment*). However, there could be a small wealth effect: the more a helper has earned in the previous trust game, the more willing she is to help in the helping game.[39] The significance vanishes once one controls for helpers' transfers to passive players (see Model 3), which can be used as a measure of a helpers' individual benchmark of altruism.

In Model 4, an interaction term between the *Anticipation* treatment and the *investment* to the trustee and an interaction term between the *Anticipation* treatment and the *relative return transfer* of the trustee are added. The coefficient of the investment is not significant in this model. On the contrary, coefficients of both interaction terms are significant and positive. The most important result derived from Model 3 in the Tobit regressions is that trustees are treated somewhat differently in the *Anticipation* treatment than in the *Baseline.* In contrast to the prediction in Hypothesis 1b, they receive, on average, for a given history (a particular investment followed by a particular return transfer) a higher transfer from a helper, if they can invest in their good reputation knowing that the helping game will follow. This result contradicts the findings in Stanca et al. (2009). It also calls for a new model of (positive strong indirect) reciprocity in which other factors besides intentions are considered. In the present experiment, helpers seem to care more about socially desirable behavior of trustees than about intrinsically motivated intentions behind the return transfers.[40]

---

[37] The Tobit regression is used because, in the helping game, helpers' transfers are exogenously restricted with an upper and a lower bound; the lower bound is usually zero-giving. Bardsley (2008) shows that subjects also take money if they have the opportunity in similar situations. In the present setting, this seems plausible, since, as stated before, relative return transfers are rewarded a lot by helpers, while very low relative return transfers lead to very low helpers' transfers and often to transfers of zero. Tobit regressions account for the possibility that (some) subjects might have even taken money instead of giving nothing by controlling for censoring. Moreover, as there are 24 transfer decisions per individual (due to the strategy method – only taking into account transfers to active players), random effects models which take individual specific effects into account are in order. The coefficient of the treatment dummy is directly interpretable in the sense that it gives exactly the value of the average marginal effect of the independent variable.

[38] The Spearman correlation analysis is used to see if any personal traits of participants influenced their helpers' transfers. Negative reciprocity is negatively correlated with helpers' transfers (Spearman's rho = -0.3247; p-value = 0.0315). Besides that, neither the gender nor any other personality trait is significantly correlated with helpers' transfers.

[39] One could also interpret that coefficient as a proxy for generalized/upstream indirect reciprocity – the more a helper has received in return from his own trustee, the more does she help in the helping game. With the help of the regression, one can disentangle that motive from the social/downstream indirect reciprocity motive of the helper, which is captured in the variable *relative return transfers*.

[40] One should be cautious in interpreting this result, since in the present experiment helpers arise from the population of investors. Therefore, they might feel they belong to the group of investors. As a result, they might care more about the total earnings resulting from high return transfers of investors than about potential strategic motives of trustees.

In the regressions, there is strong left-censoring, which indicates that helpers would possibly like not just to give less money to, but even to take money from greedy trustees. The results from the random effects Tobit regression mainly support Hypothesis 1a, while there is no clear support of Hypothesis 1b. This leads to:

Result 1a: The higher the relative return transfers a trustee makes in the trust game, the more he receives from a helper. Generally, the investment that a trustee has received does not seem to determine a helper's transfer to the trustee.

and

Result 1b: Helpers' transfers are, on average, not lower if the helping game is announced. The transfers are more positively correlated with relative as well as with absolute return transfers, if the helping game is announced.

The strategy of helpers includes passive players (Player C and Player B, who do not receive an investment and can therefore not make any return transfers). Helpers' transfers to passive players can be regarded as a benchmark of how much a helper is willing to transfer to a player who does not have a history.[41] Furthermore, it can be identified which distributive norms achieved in the trust game helpers reward (higher transfer than to passive players) and which they punish (lower transfer than to passive players). A relative return transfer of Y=3 leaves the trustee with his endowment only and reciprocates the investment completely; a relative return transfer of Y=2 leads to an equal split between the trustee and the investor; by a relative return transfer of Y=1, the investor is compensated for his investment and earns as much as his endowment would have been without an investment, while the trustee keeps the complete rent from the investment; the least generous possible relative return transfer in the experiment is Y=0, i.e., this relative return transfer makes the investor worse off than if he had not made an investment and leaves the trustee with the highest possible earning. A Wilcoxon signed-rank test shows that the passive Players C are not treated differently than passive Players B ($|z|=0.653$, $p = 0.5138$; *Baseline* only: $|z|=0.899$, $p = 0.3688$; *Anticipation* only $|z|=0.196$, $p = 0.8444$). Figure 1 indicates that, for any investment, if a trustee makes a relative return transfer of Y = 2, he receives on average the same transfer by the helper as a passive player; for each investment for a relative return transfer of Y<2 he receives a lower transfer than a passive player; and, respectively, for each investment a relative return transfer of Y=3>2 earns him a higher transfer by a helper as to a passive player. Indeed, this visual impression is supported by Wilcoxon signed-rank tests (see Table 6 in the Appendix II.2.) in both treatments.[42] Given these observations on can state:

[41] Please note that each Player C and each Player B receives an endowment of 100 ECU. Only after receiving an investment from Player A can a Player B become active and thereby increase the investor's and his own payoffs. In case Player B returns the entire transfer, he ends up with the same payoff as a Player C or a passive Player B (with 100 ECU).
[42] As an exception from that pattern, in the *Anticipation* treatment, if a trustee makes a relative return transfer of Y=1 and has previously received an investment of X≥20, he receives on average the same transfer from a helper as a passive player.

Result 1a′: Helpers transfer to trustees, who implemented an equal split in the trust game, does not significantly differ from the transfer to passive players. Relative return transfers that leave the investor with more money than the trustee lead to higher helpers' transfers, while lower return transfers lead to lower helpers' transfers.

V.2. Return Transfers

This section analyzes whether trustees correctly anticipate helpers' transfer decisions and therefore reciprocate investors' transfers more in the *Anticipation* treatment than in the *Baseline*. In addition, a positive correlation between investments and return transfers is examined.

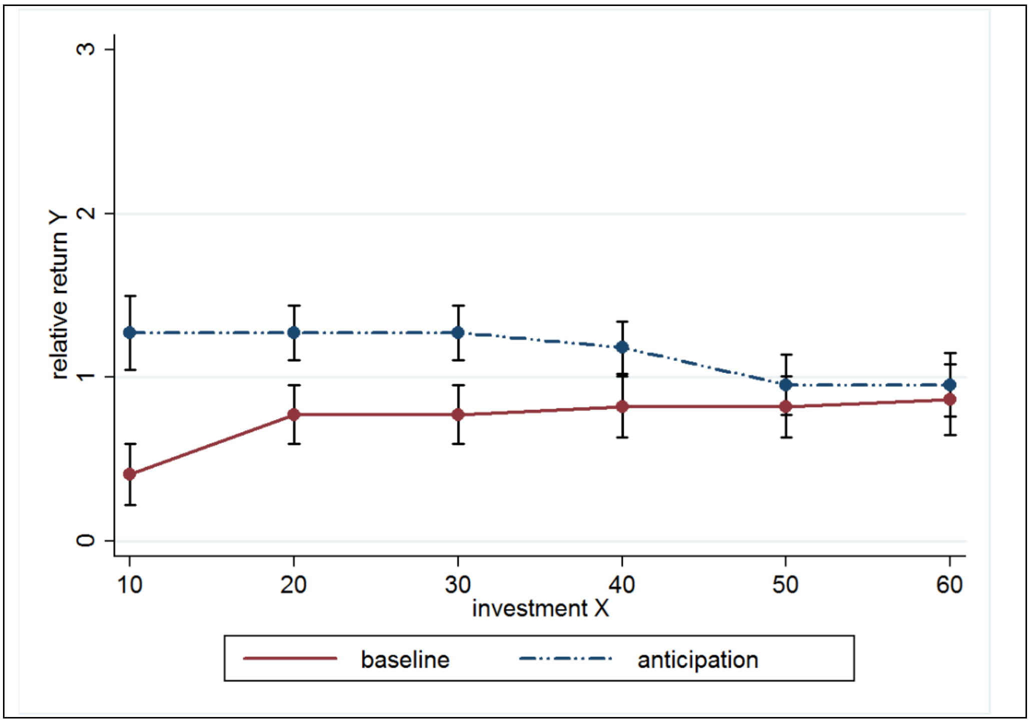


FIGURE 2

Relative Return Transfers by Treatment

On the x-axis, the investment the trustee has received is depicted; on the y-axis, mean relative return transfers are displayed. Standard errors are indicated.

There are 12 out of 44 trustees who do not make a positive return transfer (9 in *Baseline* and 3 in *Anticipation*).[43] For the average return transfers, a significant difference between the *Anticipation* treatment and the *Baseline* (Mann-Whitney rank-sum $|z|=2.135$, $p = 0.0327$) is found. In addition, the pattern of the results presented in Figure 2 shows a difference between the treatments. Specifically, in the *Baseline*, one can observe a typical outcome for the trust game (see Falk et al., 2013): the higher

Furthermore, in the *Anticipation* treatment a relative return transfer of Y=3 for a given investment of X=10 leads to the same transfer as to player C.

[43] The likelihood of any positive return transfer is significantly higher in the *Anticipation* treatment: 1-sided Fisher's exact = 0.044. This is the first indication of the predicted treatment effect by Hypothesis 2b.

the investment, the higher the relative return transfer (Spearman's Rho: r_S=0.1496, p= 0.0870). In the *Anticipation* treatment, the reverse occurs: low investments are reciprocated a lot, while the relative returns of high investments do not differ significantly between the treatments (Spearman's Rho: r_S= -0.1505, p= 0.0849).[44] The relative return transfers differ significantly between the treatments for investments lower than 50 (Mann-Whitney rank-sum |z|=1.681, p ≤ 0.0927), while they are not statistically different for the highest two possible investments (Mann-Whitney rank-sum |z|=0.605, p ≥ 0.5453).[45] Furthermore, the motives behind the return transfers seem to change in the *Anticipation* treatment compared to the *Baseline*, i.e., subjects "invest" in a reputation of being trustworthy when it is cheap (when a high relative return transfer results in a comparatively low absolute return transfer) and do not reciprocate high investments more than in the *Baseline*. This finding is especially interesting given that studies involving reward by second parties (direct reciprocity) in one-shot, public-good games have not found an increase in socially desirable behavior (e.g., Dufwenberg et al., 2001; Walker and Halloran, 2004). Thus, the findings support Hypothesis 2b, while Hypothesis 2a is only supported in the *Baseline*.

Here, a random effects ordered Probit regression is used to test which influence factors determine the return transfers.[46] The most important insight from the models in Table 4 is that controlling for different additional influence variables, the data confirm Hypothesis 2b, i.e., the coefficient of the treatment dummy (*Anticipation*) remains highly significant. At first glance, the investment alone (Model 2) does not appear to have a correlation with the return transfer. Yet, one can see a positive significant correlation of the investment and the return transfer when including the interaction of the investment and the treatment dummy (Models 3). The interaction effect is negative and statistically significant. Thus, the results generally support Hypothesis 2a, but the effect is reversed in the *Anticipation* treatment. Summing up, the results of the random effects ordered Probit regression support the findings from the non-parametric data analysis. Overall, the results support Hypothesis 2b, while Hypothesis 2a can only be supported in the *Baseline*. This leads to:

Result 2a: In the *Baseline*, relative return transfers increase as the investments increase. The opposite pattern of the return transfers is found in the *Anticipation* treatment.

and to:

Result 2b: The return transfers are, on average, higher in the *Anticipation* treatment than in the *Baseline*.

---

[44] Comparing for each investment the relative return transfers (return transfer for X=10 vs. for X=20; for X=20 vs. for X=30, etc.) overall (and in the *Baseline*), only the first comparison is significantly different (|z|=1.728, p = 0.0839; *Baseline*: (|z|=2.178, p = 0.0294), while the rest is not statistically significantly different (|z|=1.000, p ≥ 0.3173; *Baseline*: (|z|=1.414, p ≥ 0.1573). In the *Anticipation* treatment, only an investment of X=50 is reciprocated significantly more than an investment of X=40 (|z|=1.651, p = 0.0987), while no other comparison is statistically significantly different (|z|=1.414, p ≥ 0.1573). See Table 9 in the Appendix II.2. for all comparisons.

[45] See Table 10 in the Appendix II.2. for all non-parametric comparisons.

[46] The random effects control for the fact that each trustee takes six decisions (one for each investment) and an ordered Probit model suits these data the best, since trustees can only chose between four different relative return transfer per investment (Y=0, Y=1, Y=2 or Y=3).

TABLE 4

Explaining trustees' transfers – comparison Baseline and Anticipation treatment

Random effects ordered Probit regressions ("trustees" as group) Dependent variable: Relative return transfers in the trust game			
	Model 1	Model 2	Model 3
Anticipation	2.45***	2.45***	4.16***
	(.27)	(.27)	(.50)
Investment		-.00	.02***
		(.00)	(.01)
Investment*Anticipation*			-.04***
			(.01)
Cut 1 Constant	.29	.25	1.13***
	(.18)	(.25)	(.34)
Cut 2 Constant	2.29***	2.25***	3.28***
	(.21)	(.27)	(.38)
Cut 3 Constant	4.54***	4.50***	5.65***
	(.39)	(.41)	(.52)
Rho Constant	.86***	.86***	.87***
	(.02)	(.02)	(.02)
N	264	264	264
N of groups	44	44	44
P model	0.033	0.101	<.001
LR Chi2	4.52	4.58	23.73

Random effects ordered Probit regressions. Standard errors are presented in parentheses. The *Anticipation* dummy equals 1 for all observations of the *Anticipation* treatment, *investment* controls for the investment (X) the trustee has received. Significance at the 10%, 5%, and 1% level is denoted by *, **, and ***, respectively.

V.3. Investments

This section tests whether investors correctly anticipate the different return transfers between the treatments and therefore invest more in the *Anticipation* treatment compared to the *Baseline*.

Statistically, the investments do not, on average, differ between the treatments (Mann-Whitney rank-sum: $|z| = 0.465$, $p = 0.6419$), but the cumulative distribution function of the investments in Figure 3 shows that the distribution of the investments differs. Specifically, in the *Anticipation* treatment, zero-investments are never made, whereas in the *Baseline*, zero-transfers are present.[47] Nonetheless, the data do not support Hypothesis 3 and lead to:

> **Result 3:** On average, investments do not statistically differ significantly between treatments.

[47] Using Fisher's exact test for a positive investment does not lead to a statistically significant treatment difference (1-sided Fisher's exact = 0.116). Testing the data parametrically with an ordered Probit regression model does not lead to any significant results either.

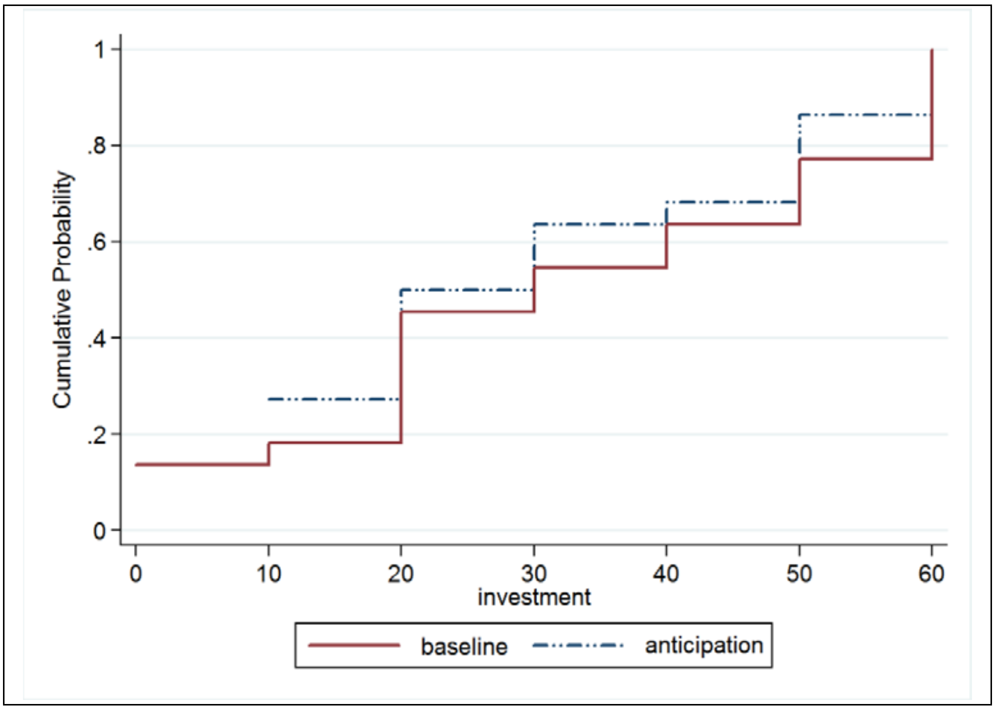


FIGURE 3

Investments by Treatment

The graph shows an empirical cumulative distribution function of the investments for both treatments. On the x-axis, the investment is depicted; on the y-axis, the estimated probability for each investment is displayed.

V.4. Earnings

As mentioned above, helpers seem to care more about socially desirable outcomes than about the motives behind trustees' transfers. Those preferences raise the question of whether the monetary efficiency is, in fact, higher in the *Anticipation* treatment. As shown above, helpers make, on average, the same transfers in the *Anticipation* treatment as in the *Baseline* given a particular return transfer and trustees anticipate that. I.e., trustees make higher relative return transfers in the *Anticipation* treatment especially if such an "investment" in a good reputation is cheap in absolute terms. It turns out that the money-maximizing strategy for trustees (given the strategy of the helpers) in the *Anticipation*[48] treatment is fully to return (Y=3) each investment (X) (see Table 11 in the Appendix II.2. for a calculation of possible payoffs for each strategy of a trustee). Thus, trustees do not anticipate helpers' positive strong indirect reciprocity enough to maximize their profits.

Given the decisions of trustees investors' money-maximizing strategy is to invest X=0 in *Baseline* and X=30 in the *Anticipation* treatment (see Table 12 in the Appendix II.2. for a calculation of possible payoffs for each investment of an investor). However, the quantitative difference in

[48] Obviously, in the *Baseline* (as discussed in chapter IV), in expectations the money-maximizing strategy for trustees is to send zero transfers (since in the *Baseline* trustees do not know that the helping game will follow). Similarly, (as well discussed in chapter IV) the money-maximizing strategy in both treatments for helpers is to send zero transfer. Quantitatively, trustees follow the optimal strategy (see Table 12 in the Appendix II.2.).

payoffs is very small (100 compared to 108.18). If investors are at least slightly risk-averse, they make the correct decision by not investing differently in the *Anticipation* treatment compared to the *Baseline*. An alternative explanation for the underinvestment could be that the level of reasoning is too high. I.e., investors have to anticipate not only what trustees will do, but also what they think what trustees think the helpers will do in the later game. [49]

Now, the actually realized transfers are presented. Since average realized return transfers in the trust game are statistically significantly higher in the *Anticipation* treatment than in the *Baseline* (*Baseline:* 26.36 (sd=34.98); *Anticipation*: 41.81 (sd=34.03); Mann-Whitney rank-sum |z|= 1.906, p = 0.0566) realized helpers' transfers are (descriptively) higher in the *Anticipation* treatment than in the *Baseline* (*Baseline*: 17.72 (sd=29.86); *Anticipation*: 28.63 (sd=50.26); Mann-Whitney rank-sum: |z| = 0.404, p = 0.6861). As a result, Players B (trustees in the trust game and helpers' co-players in the helping game) have, on average, the same total earning in the *Anticipation* treatment as in the *Baseline* (*Baseline*: 188.18 (sd=55.02); *Anticipation*: 173.63 (sd=75.15); Mann-Whitney rank-sum: |z| = 1.506, p = 0.1321).[50] In total, players A have, on average, the same profits in the *Baseline* and in the treatment (*Baseline*: 187.72 (sd=30.96); *Anticipation*: 197.27 (sd=32.61); Mann-Whitney rank-sum: |z| = 0.969, p = 0.3324). This finding, however, should be treated with caution, since the earnings of the players arise from the particular design and the matching in the experiment, i.e., trustees become helpers' co-players in the helping game not only in the *Anticipation* treatment, but also in the *Baseline*, and Players A have the role of investors as well as that (the role) of helpers.

VI. CONCLUSION

From a welfare point of view, rewards are better for the society than punishment since they do not lead to an efficiency loss. Rand et al. (2009) show in a repeated public goods game that "*reward is as effective as punishment for maintaining public cooperation and leads to higher total earnings. Moreover, when both options are available, reward leads to increased contributions and payoff, whereas punishment has no effect on contributions and leads to lower payoff*" (Rand et al., 2009, p. 1272). In a world in which anonymous interactions become more and more frequent (e.g., via the internet), it is important to pay more attention to economic consequences of reward systems by impartial strangers. Surprisingly, until now only a small number of papers analyzes rewards rather than punishments and even less literature considers rewards given by non-strategically motivated third parties. Most notably, there exists no literature on the question: How can the anticipation of a reward

[49] See the level-k literature, as in Nagel (1995), Stahl et al. (1995), Ho et al. (1998), Costa-Gomes et al. (2001), and Costa-Gomes and Crawford (2006), etc.

[50] Here, only trustees who become co-players of helpers in the helping game are included. The result does not change if trustees who do not become helpers' co-players in the second part of the experiment are included. The total earnings for the C-players do not differ significantly between the treatments either, irrespectively of whether they become co-players of helpers (Mann-Whitney rank-sum: |z| = 0.408, p = 0.6831) or not (Mann-Whitney rank-sum: |z| = 0.298, p = 0.7660).

from an impartial stranger enhance trustworthiness (and cooperation)? The current paper attempts to close this gap.

This paper shows that positive strong indirect reciprocity exists and, moreover, that it is correctly anticipated by potential recipients. Helpers' transfers are surprisingly more positively correlated with relative return transfers in the *Anticipation* treatment than in the *Baseline*. Trustees anticipate this behavior correctly and make higher return transfers in the *Anticipation* treatment. In particular, they make higher relative return transfers in response to lower investments, signaling higher trustworthiness when such signals are cheap in absolute terms. Despite the fact that the earnings of players do not differ between the treatments, trustees become more trustworthy in the *Anticipation* treatment than in the *Baseline* and they would have earned even more money than their counterparts in the *Baseline* if they had fully anticipated the complete strategy of helpers.

The experiment still leaves some questions open. There is high zero-censoring in the helping game. This could be interpreted as helpers' willingness to punish greedy trustees. An experiment where real punishment is used instead of denying help may be used to understand the helpers' actions further. Dohmen et al. (2009) show that positive and negative reciprocity are different concepts. This finding might be true for strong negative and positive indirect reciprocity as well. However, without encountering ethical concerns, one could implement only the announced version of the punishment game.

This paper helps to explore the motivations behind and the consequences of third-party reward of trustworthiness. The findings from this paper are applicable to many different settings. The results show that, in all situations in which people display trustworthy behavior for intrinsic reasons, they might in fact do a better job if they could expect strangers to reward them for their actions. That is even true if the strangers do not have any strategic incentives to do so and if their actions are costly. From a policy perspective, it turns out that it can be socially beneficial to promote greater publicity on socially desirable acts.

VII. References

Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman. 2011. "Reference points and effort provision." *American Economic Review*, 101(2): 470–492.

Almenberg, Johan, Anna Dreber, Coren L. Apicella, and David G. Rand. 2011. "Third Party Reward and Punishment: Group Size, Efficiency and Public Goods." In N. M. Palmetti, and J. P. Russo (eds.), *Psychology and Punishment,* New York: Nova Science Publishers: 73-92.

Andreoni, James. 1990. "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving." *The Economic Journal*, 100(401): 464–477.

Andreoni, James, and John H. Miller (1993) "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence." *The Economic Journal*, 103(418): 570-585.

Arrow, Kenneth. J. 1972. "Gifts and Exchanges." *Philosophy and Public Affairs*, 1(4): 343-362.

Balafoutas, Lucas, Nicos Nikiforakis, and Kristoffel Grechenig. 2014. "Third-party punishment and counter-punishment in one-shot interactions." *Economics Letters*, 122(2): 308–310.

Berg, Joyce, John Dickhaut, and Kevin McCabe. 1995. "Trust, reciprocity, and social history." *Games and Economic Behavior*, 10: 122–142.

Bernhard, Helen, Urs Fischbacher, and Ernst Fehr. 2006. "Parochial altruism in humans." *Nature,* 442: 912-915.

Brandts, Jordy, and Gary Charness. 2011. "The strategy versus the direct-response method: A first survey of experimental comparisons." *Experimental Economics*, 14(3): 375–398.

Bardsley, Nicolas. 2008. "Dictator Game Giving: Altruism or Artefact?" *Experimental Economics*, 11(2): 122-133.

Boyd, Robert, and Peter J. Richerson. 1989. "The evolution of indirect reciprocity." *Social Networks*, 11: 213–236.

Camerer, Colin F., and Ernst Fehr. 2004. "Measuring Social Norms and Preferences Using Experimental Games: A Guide for Social Scientists." In J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr, and H. Gintis (eds.), *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies,* New York: Oxford University Press: 55-95.

Carpenter, Jeffrey P., and Peter Hans Matthews. 2004. "Social reciprocity." *IZA Discussion Paper No. 1347*.

Carpenter, Jefftery P., Peter. Hans Matthews, and Okomboli Ong'ong'a. 2004. "Why Punish? Social reciprocity and the enforcement of prosocial norms." *Journal of Evolutionary Economics*, 14: 407–429.

Charness, Gary, and Matthew Rabin. 2002. "Understanding social preferences with simple tests." *The Quarterly Journal of Economics*, 117(3): 817-869.

Costa-Gomes, Miguel, and Vincent P. Crawford. 2006. "Cognition and Behavior in Two-Person Guessing Games: An Experimental Study." *American Economic Review*, 96(5): 1737-1768.

Costa-Gomes, Miguel, Vincent P. Crawford, and Bruno Broseta. 2001. "Cognition and Behavior in Normal-Form Games: An Experimental Study." *Econometrica*, 69(5): 1193-1235.

Costa-Gomes, Miguel, Steffen Huck, and Georg Weizsäcker. 2012. "Beliefs and actions in the trust game: Creating instrumental." *WZB Discussion Papers SP II 2012-302*.

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde. 2009. "Homo Reciprocans: Survey Evidence on Behavioural Outcomes." *The Economic Journal*, 119(536): 592–612.

Dufwenberg, Martin, Uri Gneezy, Werner Güth, and Eric van Damme. 2001. "Direct versus indirect reciprocity: An experiment." *Homo Oeconomicus*, 18: 19–30.

Dufwenberg, Martin, and Georg Kirchsteiger. 2004. "A theory of sequential reciprocity." *Games and Economic Behavior*, 47: 268–298.

Engelmann, Dirk, and Urs Fischbacher. 2009. "Indirect reciprocity and strategic reputation building in an experimental helping game." *Games and Economic Behavior*, 67: 399–407.

Falk, Armin and Urs Fischbacher. 2006. "A Theory of Reciprocity." *Games and Economic Behavior*, 54(2): 293-315.

Falk, Armin, Ernst Fehr, and Urs Fischbacher. 2008. "Testing theories of fairness – Intentions matter." *Games and Economic Behavior*, 62(1): 287–303.

Falk, Armin, Stephan Meier, and Christian Zehnder. 2013. "Do Lab Experiments Misrepresent Social Preferences? The Case of Self-Selected Student Samples." *Journal of the European Economic Association*, 11(4): 839-852.

Fehr, Ernst, Urs Fischbacher. 2004. "Third-party punishment and social norms." *Evolution and Human Behavior*, 25: 63-87.

Fehr, Ernst, Urs Fischbacher, and Simon Gächter. 2002. "Strong reciprocity, human cooperation and the enforcement of social norms." *Human Nature*, 13: 1–25.

Fehr, Ernst, and Simon Gächter. 2000. “Cooperation and punishment in public goods experiments.” *American Economic Review*, 90: 980-994.

Fischbacher, Urs. 2007. “z-Tree. Zurich Toolbox for Ready-made Economic Experiments.” *Experimental Economics*, 10(2): 171-178.

Fischbacher, Urs, Simon Gächter, and Simone Quercia. 2012. “The Behavioural Validity of the Strategy Method in Public Good Experiments.” *Journal of Economic Psychology*, 33(4): 897-913.

Forsythe, Robert, Joel L. Horowitz, Nathan E. Savin, and Martin Sefton. 1994. “Fairness in Simple Bargaining Games.” *Games and Economic Behavior*, 6(3): 347–369.

Greiner, Ben. 2004. “An Online Recruitment System for Economic Experiments.” In K. Kremer, and V. Macho (eds.), *Forschung und Wissenschaftliches Rechnen.* Gesellschaft für Wissenschaftliche Datenverarbeitung Bericht, Göttingen: Datenverarbeitung, 63: 79–93.

Greiner, Ben, and M. Vittoria Levati. 2005. “Indirect reciprocity in cyclical networks. An experimental study.” *Journal of Economic Psychology*, 26: 711–731.

Güth, Werner, Manfred Konigstein, Nadège Marchand, and Klaus Nehring. 2001. “Trust and reciprocity in the investment game with indirect reward.” *Homo Oeconomicus*, 18: 241–262.

Ho, Teck-Hua, Colin Camerer, and Keith Weigelt. 1998. “Iterated Dominance and Iterated Best Response in Experimental „p-Beauty Contests“.” *American Economic Review*, 88(4): 1313–1326.

Holt, Charles A., and Susan K. Laury. 2002. “Risk Aversion and Incentive Effects.” *American Economic Review*, 92(5): 1644-55.

Levine, David K. 1998. “Modeling Altruism and Spitefulness in Experiments.” *Review of Economic Dynamics*, 1: 593-622.

Nagel, Rosemarie. 1995. “Unraveling in Guessing Games: An Experimental Study.” *American Economic Review*, 85(5): 1313–1326.

Rabin, Matthew. 1993. “Incorporating fairness into game theory and economics.” *American Economic Review*, 83: 1281–1302.

Rand, David G., Anna Dreber, Tore Ellingsen, Drew Fudenberg, and Martin A. Nowak. 2009. “Positive Interactions Promote Public Cooperation.” *Science*, 325: 1272-1275.

Rockenbach, Bettina, and Manfred Milinski. 2006. “The efficient interaction of indirect reciprocity and costly punishment.” *Nature*, 444: 718–723.

Seinen, Ingrid, and Arthur Schram. 2006. “Social status and group norms: Indirect reciprocity in a repeated helping experiment.” *European Economic Review*, 50: 581–602.

Selten, Reinhard. 1967. “Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments.” In E. Sauermann (ed.), *Beiträge zur experimentellen Wirtschaftsforschung*, Tübingen, Mohr: 136-168.

Stahl, Dale O., and Paul W. Wilson. 1995. “On Players’ Models of Other Players: Theory and Experimental Evidence.” *Games and Economic Behavior*, 10(1): 218-254.

Stanca, Luca. 2009. “Measuring Indirect Reciprocity: Whose Back Do We Scratch?” *Journal of Economic Psychology*, 30(2): 190-202.

Stanca, Luca, Luigino Bruni, and Luca Corazzini. 2009. “Testing Theories of Reciprocity: Do Motivations Matter?” *Journal of Economic Behavior & Organization*, 71(2): 233-245.

Walker, James M., and Matthew Halloran. 2004. “Rewards and Sanctions and the Provision of Public Goods in One-Shot Settings.” *Experimental Economics*, 7(3): 235-247.

Chapter 3

Words Substitute Fists – Justifying Punishment in a Public Good Experiment

CHRISTOPH ENGEL[*] AND LILIA ZHURAKHOVSKA[*+#]

[+]UNIVERSITY OF COLOGNE,

[#]UNIVERSITY OF ERLANGEN-NUREMBERG,

[*]MAX PLANCK INSTITUTE FOR RESEARCH ON COLLECTIVE GOODS, BONN

JEL: C91, D03, D62, D63, H41, K14

Keywords: justification, authority, central intervention, public good, experiment

I. INTRODUCTION

Sometimes, punishment is the only act of communication between the authority and the subordinate. The mother just slaps the child that has broken his toy. The teacher just sends the pupil, who disturbed the class, out of the room. The abbot just excludes the monk, who has missed the morning prayer, from the high table. Punishment without reasons is even frequent at the heart of the judicial system. Juries often do not explain why they find the defendant guilty (for more examples from the legal system see Schauer, 1995, p. 634).

Yet often the subordinate comes back and asks: “but why?” There are multiple legitimate reasons for this request. The punishee may have been unaware of the normative expectation in the first place. The normative expectation may have been ambiguous, as often in law. Even if the violation of the norm is established, it may be debated whether the punishee deserves punishment and, if so, of which severity. Frequently authorities anticipate the question, and directly add justifying reasons to the sanction. The mother tells her boy: we have entrusted your toys to you. Be more heedful in the future. For this time, we will buy you a new one. But if you break it again, there will not be a new toy then.

The subordinate is not the only possible addressee of reasons. The authority may herself have a supervisor who asks her to justify the intervention. The headmaster finds the pupil walking idle in the corridor and calls upon the teacher to justify her decision. The prison warden wants the guard to explain why he used corporal punishment. The convict appeals his case. Another addressee of explicit justification is fellow subordinates. The mother punishes her elder boy and tells the younger one: be aware, this is what will happen if you do not look after your toys. Jeremy Bentham has built his entire utilitarian theory of criminal law on this point (Bentham, 1830). Finally, explicit reasons may help those who have installed the authority to assess whether she should remain in office, or they may help the general public to form an opinion, and maybe call for political intervention. A case in point is criminal judges standing for re-election.

In the last decade, experimental economics has made considerable progress in understanding punishment. The main field of application is linear public good games. If the experimenter does not provide any institutional framework, initially many participants make substantial contributions to the public project. Yet over time, contributions decay. The trend reverses if participants are given the opportunity to punish each other, despite the fact that, in the typical implementation, punishment is costly (see only Fehr and Gächter, 2000; Herrmann et al., 2008).

We use this framework to test the effects of a justification requirement. In the interest of coming closer to the real world applications that motivate our research, we randomly select one participant to be an authority for a group of four active players. The participant in the role of the authority receives a fixed income (think of the judge’s salary) and therefore does not benefit monetarily from the provision of the public good; in that sense we make the authority impartial. Yet to make her choices credible she has to pay for punishment points out of a small additional endowment. Each punishment point she does not use increases her income by a small amount. That way we incentivize choices; despite the fact that the authority receives a fixed wage (think of additional effort or hassle, the more so the more severe the sanction). For reasons of external validity, we implement a stranger design, i.e., group composition differs every period. This is analogous to a court in which not every trial is between the same judge and the same defendant.[51]

In all treatments, authorities are requested to justify their choices. Yet in the *Baseline*, the reasons they give go to the experimenter only. In the *Private* treatment, each active player only learns the reasons for the decision affecting herself. We finally implement a *Public* treatment. In this treatment, all active players see the reasons directed to themselves and to all other group members.

We have subtle, but interesting results. In the *Private* and *Public* treatments, there is significantly less punishment than in the *Baseline*. We interpret this finding such that if reasons are communicated, authorities partly substitute words for action. Contributions increase over time in the *Baseline* and in the *Public* treatment, while they do not in the *Private* treatment. Hence, if justification is to the entire group, less monetary punishment is equally effective. In that setting, words also substitute action in terms of disciplining active players. Our data suggest however that there is a mismatch between the expectations of authorities and active players if reasons are only communicated to the addressee. While active players become even more sensitive to the severity of punishment, authorities reduce punishment, arguably because they expect reasons to serve as a partial substitute. By contrast, if reasons are made public, active players become considerably more sensitive to the amount contributed by the remaining active players. Punishment combined with reasons stabilizes contributions on this indirect path.

The remainder of the paper is organized as follows: Section 2 relates the paper to the literature. Section 3 presents the design of the experiment. Section 4 derives predictions. Section 5 reports results. Section 6 concludes.

II. LITERATURE

To the best of our knowledge the effect of a justification requirement on punishment and contribution behavior in a public good has not previously been studied, neither theoretically nor experimentally.

In treatments *Private* and *Public*, justification is a form of one-way communication from the authority to the active members. Communication among active players has generally been shown to increase cooperation (see the meta-analysis by Sally, 1995; the survey by Crawford, 1998; the meta-analysis by Balliet, 2010) (from the rich literature see, e.g., Bochet et al., 2006). Our design differs from this literature in that the only player allowed to communicate is the authority. Communication can therefore not serve as a vehicle for creating trust among the active players. It may merely serve the

[51] Additional technical reasons for this design choice are discussed in the design section of the paper.

backward looking function of explaining why a player has been harmed, and the forward looking function of explaining an authority's punishment policy.

Duffy and Feltovich (2002) tested a prisoner's dilemma where active players either had a chance to send a pre-play cheap talk message, or where they could observe each other's choices in the previous period. Both had roughly the same, positive effect. We implement a stranger design. Therefore, through feedback from earlier periods participants only learn about the population, not about the individual interaction partners in the next period. If communication by an authority is equally effective, we should expect a positive effect.

If all players hold (sufficiently pronounced) Fehr-Schmidt preferences (Fehr and Schmidt, 1999), the behavioral game has the character of a coordination game with multiple equilibria. It has been shown that, in coordination games, pre-play communication facilitates coordination on the Pareto-dominant equilibrium (Blume and Ortmann, 2007). Communication by the exogenous authority might serve a similar function.

If reasons are communicated, the authority may use them to express disapproval. Masclet et al. (2003) have shown that disapproval increases contributions, even if it is not backed up by monetary sanctions. They did not study the interaction of monetary and non-monetary sanctions, which is what we implement.

In treatments *Private* and *Public*, the authority may use the reasons she gives to announce a punishment policy. In Berlemann et al. (2009), non-binding announcements had practically no effect. There was a slight effect if, afterwards, it could be checked whether (active) participants behaved as announced. Yet in our experiment, active players cannot check whether the authority implements a consistent policy, given active players and authorities are re-matched every period.

Croson and Marks (2001), in a step level public good, introduced a recommendation by the experimenter how much to contribute. This only had a significant effect on contributions if participants benefitted heterogeneously from the provision of the public good. In our design, active players are homogeneous. Yet if the authority uses justifications to fix an expected contribution level, this is not a recommendation by the experimenter, but by another participant. Moreover, the authority has power to enforce her chosen norm. We might therefore see a positive effect.[52]

If active players learn the reasons, the authority may use justification to threaten free-riders in future periods. Masclet et al. (2012) have found that threats preceding decentralized punishment increase cooperation. Contrary to our paper, they have not analyzed any substitution effects between justifications and punishment. Furthermore, justifications in their paper were mainly meant as announcements for future periods, while in our paper justifications are directly connected to chosen punishment levels in the current period.

[52] In our experiment we inform participants in the instructions about average contributions in a similar experiment; see instructions in the Appendix III.1. That information could also be regarded as a subtle form of recommendation by the experimenter. However we neither expected ex ante nor found ex post that this information had a remarkable effect on the behavior of our participants. The only purpose of that information was to provide participants with one potential plausible contribution norm.

We entrust punishment to a fifth player. Engel and Zhurakhovska (2012) run an experiment with the same structure as in this paper. Yet, in Engel and Zhurakhovska (2012), the authority is neither able nor requested to justify her decisions. Engel and Zhurakhovska (2012) find that the large majority of authorities is neither selfish nor spiteful, i.e., no anti-social punishment is found. Authorities do also not exploit punishment to equalize their own earnings with the earnings of active players. Instead, they aim at disciplining free-riders in the groups they happen to be assigned to. In Engel and Zhurakhovska (2012) punishment decisions of authorities who do not have to write any justifications do not significantly differ from the punishment decisions in the *Baseline* in the present study (in which justifications are given, but not communicated to the punished players). Engel and Zhurakhovska (2012) study whether and why authorities are willing to discipline free-riders, even if this is costly for them and yields no pecuniary benefit. By contrast in the current paper, we want to investigate whether explicit justification induces recipients to increase contributions to a public good, and if so, why.

In a sender receiver game, Xiao and Tan (2014) compare three settings: a punishment authority receives a flat fee; the authority has a straightforward monetary incentive to punish senders who have communicated the truth; this incentive is upheld, but authorities are obliged to justify their decision in a message that is communicated to the remaining two participants at the end of the experiment. With this obligation, authorities are less likely to abuse their power. Senders are less likely to lie. We test a different game. We make it impossible for authorities to be selfish. In our experiment, interest is not in taming corruption, but in improving the effectiveness of punishment. To that end we manipulate to whom reasons are communicated. We also derive hypotheses from a formal model.

In the legal literature, the obligation to justify decisions has been studied from a normative perspective (McCormac, 1994; Schauer, 1995). This literature expects explicit reasons to clarify the meaning of authoritative intervention, to authoritatively construct reality, to increase compliance, to enable control, to remove biases in addressees, to dissolve conflict (Engel, 2007) and to make authorities more accountable (Tetlock, 1983; Seidenfeld, 2001).

III. Design

III.1. The Game

We conduct a linear public good experiment with costly punishment by an additional participant who does not benefit from the provision of the public good. The additional participant provides reasons justifying her punishment for each of her decisions. All active players (players who can contribute to the public good) learn their own punishment and the punishment of all other group members as well as the contributions of all other group members. In addition, (depending on the treatment) active players do not or do receive the justification for their own punishment and the punishment of the other active group members. Specifically, the main experiment comes in three steps.[53]

Step 1:

N active players may contribute to the public good. In line with most of the experimental literature on public goods, (first stage) payoff of an active player i is defined by

$$\pi_i^1 = e - c_i + \mu \sum_{n=1}^{N} c_n \qquad (1)$$

where e is the endowment, c_i is the contribution of this player to the public good, $0 < \mu < 1 < N\mu$ is the marginal per capita rate, n is generic for any player, and N is group size.

Step 2a:

An additional player A (the authority) learns about the contributions of all four active players in her group. The authority can assign punishment points to the active players. Therefore, the second stage payoff of an active player is given by

$$\pi_i^2 = e - c_i + \mu \sum_{n=1}^{N} c_n - \tau p \qquad (2)$$

where τ measures the negative effect of receiving one punishment point p on an active player's first stage income.

The authority has the payoff

$$\pi_A = x + y - \sum_{n=1}^{N} m p_n \qquad (3)$$

where x is a fixed wage, y is an additional endowment she can use for punishing active players, $\sum_{n=1}^{N} p_n$is the total number of points which she assigns to any active player n multiplied by a marginal cost per point of m.

Step 2b:

Simultaneously to assigning the punishment points to the active players, the authority gives reasons justifying her punishment for each of her decisions.

[53] We have two post-experimental tests, for social value orientation (Liebrand and McClintock, 1988), and for relative risk aversion (Holt and Laury, 2002), which we, however, do not use for the analysis since they do not turn out informative.

Step 3a:

Each active player is informed about the contributions made and of the number of punishment points received by each member of the group (including own contributions and punishment).

Step 3b:

Simultaneous to learning the contributions and punishment levels of all active group members with probability *a*, each active player learns the reasons formulated by the authority for the received punishment. In addition, with probability *b*, each active player learns the reasons formulated by the authority for the punishment of all other active group members.

III.2. Treatments and procedure

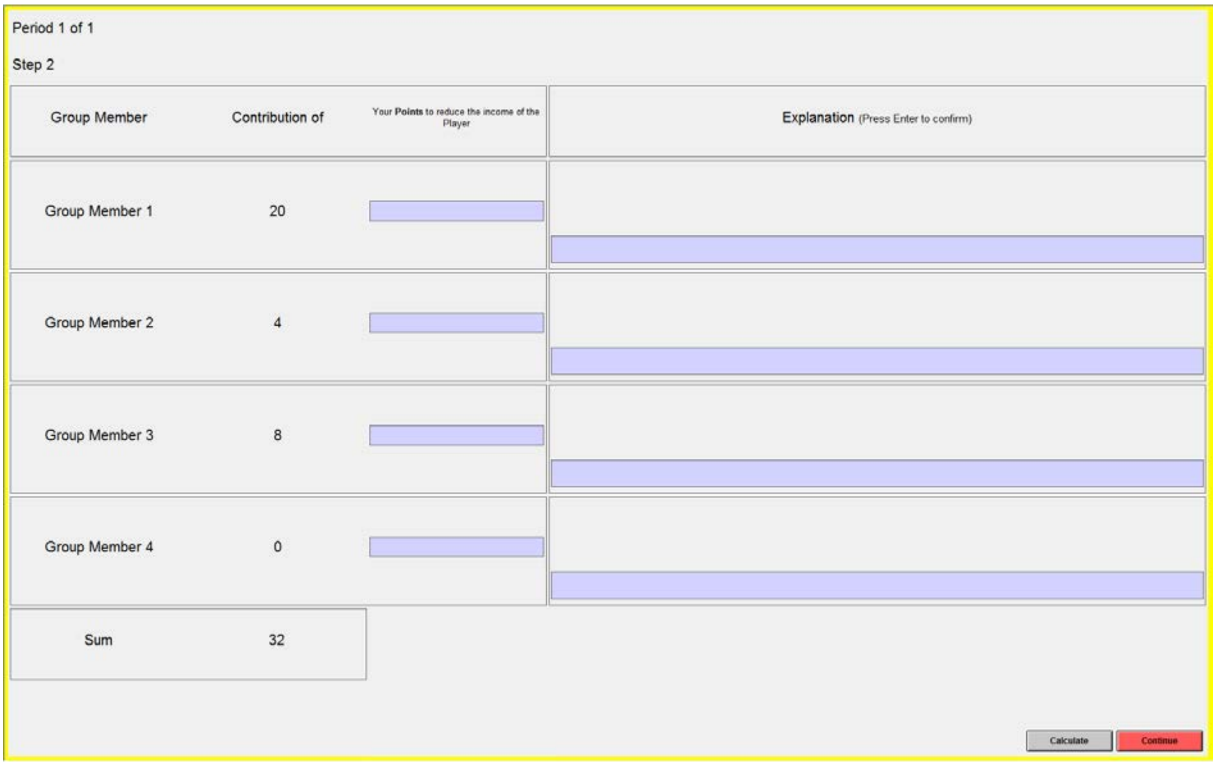


FIGURE 1

Decision screen of authority in Step 2

The authority must insert a number between 0 and 20 in each box in the column "*Your Points to reduce the income of the Player*". The total of her punishment point cannot exceed 20. She may use the "*Calculate*" button to calculate the sum of her inserted points. She is requested to type up to 500 characters in the boxes in the column "*Explanation*" to justify each of her punishment decisions. She cannot leave the stage before entering all punishment points and confirming each of her justifications by pressing the Enter key. The id number of each group member is re-shuffled each period as the group composition changes each period, as well.

At the beginning of the experiment, each subject is randomly assigned one of the two roles A (authority) or P (active player). Subjects are then matched in groups consisting of one player A in the role of an authority and N players P in the role of active players. In the experiment in all treatments, we set N=4, e=20, μ=0.4, τ=3, x=400, y=80, $m=\frac{1}{4}$.[54] In the *Public* treatment the probability to learn the justifications for the punishment of the other group members b is equal to1 and the probability to learn the justifications for own punishment a is equal to 1; in the *Private* treatment b=0 and a=1; in the Baseline b=0 and a=0. All rules and parameters are common knowledge from the beginning of the experiment.

As mentioned above, in Step 2b, the authority is requested to justify her decision regarding her punishment decision. To do so, she is asked to type her reasons into four chat boxes, each box corresponding to one active player. Each box holds a maximum of 500 characters. This is made explicit in the instructions.[55] Figure 1 displays her decision screen.

After the end of the first period, there is a surprise restart, i.e., the game is repeated for another 10 periods. Participants receive additional instructions, which inform them that from now on they will be re-matched in every of the 10 periods, but that roles are kept constant throughout the experiment. We have matching groups of size 10, composed of eight active players and two authorities. Following the procedure that is standard in the experimental literature (see, e.g., Charness, 2000; Montero et al., 2008), we only tell participants that they will be re-matched every period, not that matching groups have limited size. This procedure is meant to guarantee independent observations, without inducing participants to second guess group composition.

Our main reason for implementing a stranger design is external validity. In the legal application that has triggered our research, judges are unlikely to meet the same defendant again. Another practical concern is that previous literature has shown contributions to be rather high in repeated public goods games with partner-matching and punishment (see, e.g., Fehr and Gächter, 2000). Had we used partner-matching, a ceiling effect might have made our manipulation meaningless. Note that a stranger design puts the socially beneficial effects of justification to a stronger test. Punished players know that they will not be in the same group in the next period. Therefore, they have less reason to feel guilty if the authority explains that they have misbehaved, and they have less of a chance to predict the effect of future authorities on other active players' future choices. This same authority and this same active player also only meet again with some unknown positive probability. Consequently, justification is less able to reduce the uncertainty about the next authority's punishment policy compared to a similar setting in the partner-design. Justification only reduces the uncertainty at the population level, not at the individual level. If we nonetheless find justification effects, we know that they are very robust.

[54] In fact, in the experiment, we use two different currencies. The income of active players is expressed in Taler, the pecuniary effect of punishment for authorities is expressed in Points. The above equation translates both into Taler. A Taler is worth 4 Euro-Cents. A Point is worth 1 Euro-Cent. The punishment ratio for the authority translated into Euro-Cents is 1:12, which makes punishment substantially cheaper than in most other public good experiments with punishment. Yet in our experiment, unlike in most earlier experiments, the authority does not benefit from contributions at all. Therefore, any cost demonstrates intrinsic willingness for punishment and makes it meaningful.

[55] In the instructions it is stated that authorities are not allowed to communicate any personal information, so as to preserve anonymity.

In each session all instructions were read aloud by the experimenter before the experiment started to achieve common knowledge about the procedure.[56] The experiment only started after all subjects had completed a quiz about the rules and procedures to ensure that all subjects had understood the instructions. Interaction was completely anonymous. The experiment was conducted in the Cologne Laboratory for Economic Research in 2012. The experiment is programmed in z-Tree (Fischbacher, 2007). Participants were invited using the software ORSEE (Greiner, 2004). 340 student participants of various majors had mean age 24.31. 51.54 % were female. Participants on average earned 15.81 € (20.86 $ at the time of the experiment), 15.50 € for active players, and 17.04 € for authorities. We have 12 independent observations (matching groups of 10) in the *Baseline*, and 11 in each of the two treatments.[57]

IV. HYPOTHESES

In this experiment, we investigate in which ways differently specified justification requirements affect contribution choices. Obviously, punishment choices of the authorities and contribution choices of active players are related. Engel (2014) demonstrates theoretically as well as experimentally that in a linear public good with centralized punishment participants increase contributions even if severity of punishment had been insufficient to deter a profit-maximizing individual. Engel (2014) as well as Engel and Zhurakhovska (2012) demonstrate that authorities do not act selfishly or anti-social, i.e., they punish free-riders if those contribute less than the average active player in their current group. Both studies find that despite the stranger-matching active players, on average, react by increasing their contributions in subsequent periods. Our experiment differs from Engel (2014) and from Engel and Zhurakhovksa (2012) only in that here justifications are written and, depending on the treatment, communicated to some or to all active players. For simplicity, we assume that in the present experiment, punishment behavior as well as contributions as a reaction to punishment is similar as in Engel (2014) and in Engel and Zhurakhovska (2012). Based on this assumption, in this section we derive predictions on how active players react in their choices of contribution levels to reasons provided by the authorities and how these reactions differ in the treatments.

As in our *Baseline* no justification is communicated to active players, subjects cannot react to the reasons written by the authorities. On the contrary, in the *Private* treatment as well as in the *Public* treatment subjects learn the justifications. We define three channels on which we expect the justification requirement to affect behavior: [1] justification gives recipients disutility on top of the monetary loss; [2] justification reduces the uncertainty about authorities' punishment policies; [3] if either effect is anticipated, justification reduces the uncertainty of active players about the behavior of

[56] See section III.1. in the Appendix for an English translation of the instructions.
[57] In *Private* and *Public*, we could not fill one matching group since invited participants did not show up.

other active players, and thereby makes cooperation less risky.[58] For [1] we need active players to have disutility from monetary sanctions as well as from and guilt. For [2] and [3], we need preference heterogeneity and preference uncertainty.

For active players holding standard preferences, payoff is given by (2), while the authority's payoff is given by (3).[59]

As stated above, Engel and Zhurakhovska (2012) show experimentally that in a setting similar to the one presented here the large majority of our authorities aims at disciplining free-riders in the groups to which they are randomly assigned, despite the fact that this is costly. This resonates with earlier findings on third party punishment (Carpenter et al., 2004; Fehr and Fischbacher, 2004; Charness et al., 2008; Leibbrandt and López-Pérez, 2012; Almenberg et al., 2011). First, assume that all authorities have this desire and that this is common knowledge.

Note that as we have a one-shot version of the experiment in the first period, subjects can receive an impression of possible punishment strategies of authorities in a setting, in which interaction is not repeated. Using this information, active players can form believes on possible punishment strategies of authorities in the repeated public-goods experiment with stranger-matching. However, there could be heterogeneity of authorities. Before making the contribution decision, an active player might expect to be matched either with an authority, who wants to maximize her payoff, or who wants discipline her group. I.e., with common prior η, the authority wants to discipline her group and with counter-probability $1-\eta$, the authority maximizes payoffs. At this point, we can expect treatment effects. In the *Baseline*, active players might reason: "if authorities have to explain their choices, they are induced to develop a deliberate punishment policy, even if I do not learn this policy". In *Private* and *Public*, active players additionally receive an individual signal about the punishment policies prevalent in this population of authorities. This helps them update their beliefs about the certainty and the severity of punishment. Therefore, we should have $\eta_{base} < \eta_{priv} < \eta_{pub}$. The larger η, the less authorities (who want to discipline their groups) must react to low contributions by increasing severity. This is the first (direct) channel on which we expect the justification requirement to matter.

As is well established, experimental participants are sensitive to framing manipulations (Kahneman and Tversky, 2000). One powerful type of frame labels the opportunity structure such that it triggers normative expectations (Elliott et al., 1998), like the labeling of a public good as a "community" rather than a "Wall Street" game (Ross and Ward, 1996; Liberman et al., 2004). One explanation for the effect is guilt aversion. If the authority may make its disapproval of a participant's choice explicit, she may stress her normative expectations, and thereby increase the effect of guilt in active players. In the framework of public goods, guilt aversion has been formalized as disutility from falling below a normative expectation. In the *Baseline*, the authority does not have the possibility to

[58] Note that channels [2] and [3] are also present in an experiment with stranger matching. Participants learn at the population level.

[59] We explain below in which ways our notion of guilt aversion differs from Battigalli and Dufwenberg (2007).

stigmatize free-riding, while she does have this opportunity in the remaining two treatments. Yet in the *Private* treatment, an additional element of shaming is missing. Shaming should increase guilt. Let γ denote guilt (and shame) of active players and assume that γ decreases their utility. We therefore expect $0 \leq \gamma_{base} < \gamma_{priv} < \gamma_{pub}$. This constitutes the second (direct) channel on which we expect the justification requirement to matter. Since monetary punishment is costly, if available, the authority replaces it by guilt. Severity of punishment is reduced in increased guilt.

Finally, we expect the justification requirement to matter on an indirect channel: The more this player believes the authorities to credibly deter free-riding (the larger η) and the more she believes the authorities to trigger guilt (the larger γ), the higher her expected propability (q) other players to contribute the level, which the population of authorizes tries to enforce (the larger q). As discussed above, a large majority of active players in public-goods experiments have been shown to act as conditional cooperators. Since we have explained how our treatments affect the former two parameters, these direct effects should translate into an additional indirect effect, such that $q_{base} < q_{priv} < q_{pub}$. If correctly anticipated by authorities, the indirect effect reduces punishment p accordingly.

Taken together, we only expect treatment differences in punishment, not in contributions. There is no plausible reason to expect that, on average, authorities in our setting have different intentions than those in Engel and Zhurakhovska (2012). I.e., some authorities will act selfishly (not punish at all) while some authorities will try to discipline the active players and try to implement a contribution norm. Authorities that want to discipline their groups compensate by higher severity for uncertainty and for the impossibility to make guilt (and shame) salient. For the reasons explained we expect

> **Hypothesis 1:** Punishment is most severe in the *Baseline*, less so in *Private* and least severe in *Public*.

> **Hypothesis 2:** Active players contribute the amount that the population of authorities tries to implement.

V. RESULTS

V.1. TREATMENT EFFECTS

The first, one-shot phase of the experiment was meant to test whether active players anticipate the effects of a justification requirement. This is not the case. In non-parametric Mann-Whitney tests, we do not find any significant effects.[60] Parametrically we find a weakly significant difference between the *Baseline* and *Private* in terms of punishment (p = .069).[61] Punishment is less severe in the *Private* treatment. This translates into significantly higher profit. Since anticipation at most has a very small effect, in the following we pool the data from the first and the second phases of the experiment.

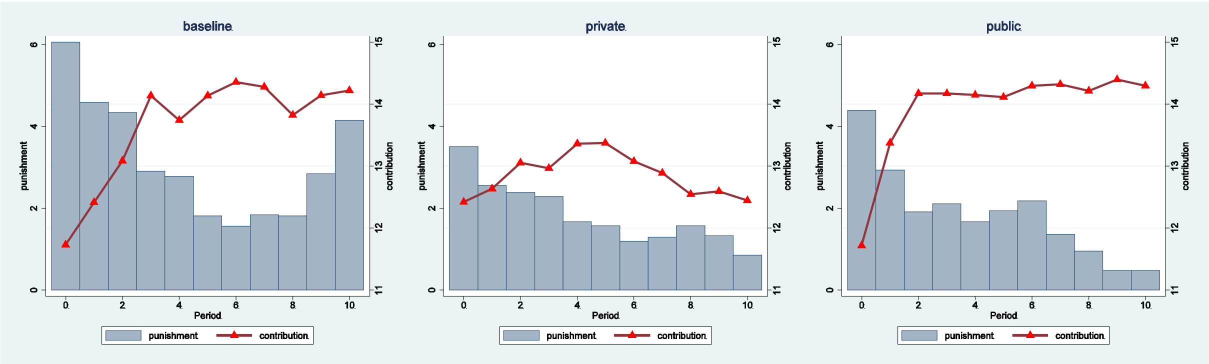


FIGURE 2

Treatment Effects

On the right vertical axis the mean contributions (in Taler) of the active players per period and treatment are displayed. On the left vertical axis one can see mean punishment (in Taler) of the authorities per period and treatment. The horizontal axis shows the periods: the one-shot game reported as period 0, the repeated game is reported as periods 1 – 10. In the graph, there is one panel per treatment.

Figure 2 reports treatment effects[62]. In all treatments, authorities use the punishment option.[63] Active players make substantial contributions to the public project.[64] The design of the experiment empowers authorities to perfectly deter free-riding. In each period each authority disposes of a

[60] In the first period, individual choices of active players are still independent. But one authority simultaneously decides about punishing four members of her first group. In this dimension, punishment decisions are thus not independent. In these tests, we therefore work with mean punishment per authority, in the first period, as the dependent variable. Results are the same if instead we work with total punishment points meted out by any one authority.

[61] In this paper we do not analyze in detail the determinants of authorities' punishment policies, and the contents of the reasons they give. Readers interested in these results are referred to our companion paper (Engel and Zhurakhovska 2012). The most prominent explanation authorities give for punishment is the fact that an active player has contributed less than the mean of her group. Some authorities also try to impose an idiosyncratic standard, typically 10 Taler. Yet others stress that the punishee has acted unfairly.

[62] In this section we present the data in Taler, which was the currency unit used in the experiment for the contributions. We also translate the punishment points into Taler, to make it easier to interpret the results.

[63] Statistical tests are complicated by the fact that the design of the experiment excludes negative punishment, i.e., rewards. Therefore, technically Hypothesis 1 calls for a test at the limit of the support. We react by reporting the highest positive amount of punishment at which a signed-rank test still rejects at conventional levels. All tests are over means at the highest level of dependence, i.e., matching groups. The test still rejects the hypothesis that mean punishment is 1.5 Taler per active player in the *Baseline* (N = 12, p = .031), 1 Taler per active player in treatment *Private* (N = 11, p = .004), and .5 Taler per active player in treatment *Public* (N = 11, p = .010).

[64] Using the same procedure as in the previous footnote we find that signed-rank tests still reject the hypothesis that mean contributions are 8 Taler in the *Baseline* (N = 12, p = .023), 7 Taler in treatment *Private* (N = 11, p = .021), and 8 Taler in treatment *Public* (N = 11, p = .016). In all treatments, mean contributions are significantly above those limits.

maximum of 20 points for punishment.[65] A complete free-rider is deterred if $p > .05$. Yet effectively in all treatments punishment is frequently non-deterrent, even if there is punishment.[66]

We now turn to hypothesis 1 that expects our treatments to matter for punishment. Descriptively there is indeed less punishment in treatments *Private* and *Public*, i.e., when active players learn reasons (see Figure 2). Moreover, in these treatments punishment decays over time, whereas it goes up again in the *Baseline*.

Non-parametrically, we find a weakly significant difference between the *Baseline* and *Public* (Mann Whitney over means per matching group, N = 23, p = .0524). This difference is significant at conventional levels for the last four periods (N = 23, p = .0354), as well as any smaller number of the final periods. For the last two periods, we also find a significant difference between the *Baseline* and treatment *Private* (N = 23, p = .0303). Parametrically, we see strong effects[67] (see Models 1 and 2 of Table 1).[68]

In these models, we not only control for the fact that, in all treatments, punishment is most pronounced in the beginning; this is captured by the time trend (*period*) and its interaction with treatment. We also take into account that in the *Baseline*, punishment is U-shaped; it goes up again in the end. This is not the case in *Private* and *Public*. The different patterns of the time trends we capture by the square of the time trend (*perdiod*^{2}), and interactions with treatments. In the mixed effects model we see a strong negative main effect of both treatments where reasons are revealed to punished players. In these treatments, punishment decays less rapidly over time (the positive interaction effects neutralize most of the negative main effect of period), and it hardly goes up in the end (the negative interaction effects neutralize most of the positive main effect of period squared). When there is room for it (since reasons are communicated), authorities partly substitute words for pecuniary punishment, as expected.

[65] A careful reader might note that in the *Baseline* punishment declines in the middle of the game but increases again in later periods, while the downward trend remains constant in the other treatments. The most plausible interpretation is that towards the end of the game the authorities apprehend a potential decline in contributions (as it is indeed usual in many public-goods experiments) and therefore try to obviate it by increasing their punishment. By contrast, in the treatments in which reasons are communicated to the active players, authorities believe that they can obviate a potential negative trend by simply warning active players to not start free-riding and informing them that they will be punished otherwise. This effect is precisely the substitution effect of verbal punishment, which we expected to find in our experiment. We discuss it in more detail below.

[66] In the *Baseline*, 27.71% of those participants who contributed less than 20 Taler to the public projects were punished such that their second stage payoff was below the payoff they would have had, had they fully contributed. In the *Private* treatment, this held for 21.39% of all participants in this situation. In the *Public* treatment, it held for 22.13%.

[67] For parametric estimation, we have challenging data. Every period each authority has power to punish four active group members. The authority stays the same over time, and she remains assigned to the same matching group (with different active players per group in each period, though). Punishment data is therefore from choices nested in periods nested in authorities nested in matching groups. This data generating process is captured by a mixed effects model. Yet most active players most of the time do not get punished at all. Therefore, the data is also left censored. This can be captured by a random effects Tobit model where the authority is the cross-section, and punishment choices directed to individual members of the current group of active participants constitute the "time" dimension. Since there is no generally acknowledged mixed effects Tobit estimator, in Table 1 we report both specifications. Note that results look similar if, instead, we estimate models with matching group fixed effects; of course then the treatment main effects are not identified, but interactions with treatment are.

[68] In Table 1, we find that even if we do not take the nested character of our data into account (see Models 2 and 4), punishment is significantly different at the 10% level in *Public* compared with the *Baseline*, while it is even more strongly significantly different if we take into account that our data are nested (see Models 1 and 3).

TABLE 1

Treatment Effects on Punishment

Dependent variable: size of deduction through punishment				
	Model 1	Model 2	Model 3	Model 4
	Mixed effects regressions	Random effects Tobit regression	Mixed effects regressions	Random effects Tobit regression
Private	-4.106***	-7.130	-6.511***	-10.338**
	(1.128)	(4.814)	(1.699)	(3.022)
Public	-3.719**	-9.052^{+}	-4.457*	-5.852^{+}
	(1.128)	(4.891)	(1.739)	(3.214)
Period	-1.733***	-5.297***		
	(.247)	(.872)		
*Private**Period	1.216**	2.935*		
	(.357)	(1.276)		
*Public**Period	1.210**	3.949**		
	(.357)	(1.347)		
Period^2	.124***	.385***		
	(.020)	(.071)		
*Private**Period^2	-.098**	-.271*		
	(.029)	(.106)		
*Public**Period^2	-.105***	-.433***		
	(.029)	(.115)		
Contribution			-.952***	-2.067***
			(.032)	(.123)
*Private**Contribution			.345***	.406*
			(.046)	(.182)
*Public**Contribution			.246***	.094
			(.052)	(.193)
Constant	7.868***	.914	16.138***	18.293***
	(.780)	(3.336)	(1.181)	(2.072)
N	2992	2992	2992	2992
Left Censored		2300		2300
P model	<.001	<.001	<.001	<.001
Wald Chi2	105.06	118.10	1525.39	470.63

Models 1 and 3 are Mixed effects regressions and Models 2 and 4 are Random effects Tobit regressions with lower censoring at 0. The dependent variable is *punishment points received*, which is nested in period nested in authority nested in matching group. The reference category is the *Baseline*. Standard errors are presented in parentheses. The *Private* dummy equals 1 for all observations in the *Private* treatment, the *Public* dummy equals 1 for all observations in the *Public* treatment. Significance at the 10%, 5%, 1% and 0.1% level is denoted by $^{+}$, *, **, and *** respectively.

TABLE 2

Treatment Effect on Contributions

	Dependent variable: contribution	
	Model 1	Model 2
	Mixed effects regressions	Random effects Tobit regression
Private	.547	.897
	(1.707)	(1.084)
Public	.551	.645
	(1.707)	(1.082)
Period	.204***	.235***
	(.035)	(.048)
*Private**Period	-.223***	-.271***
	(.051)	(.070)
*Public**Period	-.045	.036
	(.051)	(.071)
Constant	12.415***	12.970***
	(1.181)	(.747)
N	2992	2992
Left Censored		206
P model	<.001	<.001
Wald Chi2	53.31	54.24

Column 1 shows a Mixed effects Model. Here choices are nested in individuals nested in matching groups. Column 2 shows a Random effects Tobit model. Here lower censoring is at 0 and upper censoring is at 20. The reference category is the *Baseline*. Standard errors are presented in parentheses. The *Private* dummy equals 1 for all observations of the *Private* treatment, the *Public* dummy equals 1 for all observations of the *Public* treatment. Significance at the 10%, 5%, 1% and 0.1% level is denoted by $^{+}$, *, **, and *** respectively.

This interpretation is further supported by models 3 and 4 in Table 1. In these models, we control for the respective active player's contributions, i.e., we estimate authorities' empirical reaction functions. The substitution effect is directly visible in the positive interaction between contribution and treatment (*Private*contribution*; *Public*contribution*): in both treatments, the level of punishment is less sensitive to differences in contributions. This supports our hypothesis 1 and leads to[69]:

Result 1: If the reasons for punishing are communicated to punished players, punishers partly substitute them for monetary sanctions.

[69] One could argue that writing justifications is tiring and therefore in a sense costly for authorities. Consequently, authorities should save punishment costs. To avoid such an alternative interpretation, we have deliberately chosen a design, in which authorities in the *Baseline* as well as in the treatments have to write justifications. Furthermore, the scope of justifications does not differ between the *Baseline* and the treatments, which again speaks against the alternative interpretation of our results.

Descriptively, from Figure 2 we see that there is not a pronounced difference across treatments regarding the level of contributions, as predicted by our theory. Yet contributions are not stable in the *Private* treatment, while the time trend remains positive in the *Baseline* and in the *Public* treatment. The visual impression is corroborated by statistical analysis in Table 2.[70] If we compare mean contributions of the active players per matching group, in non-parametric tests we do not find any significant treatment differences. By contrast, in a parametric test of all treatments, we have a significant negative interaction between treatment *Private* and the time trend. This has the following interpretation: From the significant positive main effect of *period* it follows that contributions increase over time in the *Baseline*. Since the interaction between treatment *Public* and *period* is insignificant, the same also holds for treatment *Public.* By contrast, the significant negative interaction between treatment *Private* and *period* completely neutralizes the positive main effect of period. In treatment *Private*, contributions do not increase over time. This gives us partial support for our hypothesis 2:

> **Result 2**: If authorities are obliged to justify punishment decisions, in a linear public good contributions stabilize over time if these reasons are kept confidential or if they are made public; there is no stabilizing effect if reasons are only communicated to the punishee in private.

V.2. DRIVING FORCES

In our hypothesis section we discuss underlying forces for the behavior found. We test these potential motives in the regressions of Table 3.[71] Two forces independently and significantly explain choices: experienced severity of punishment[72], and experienced cooperativeness of the remaining active players[73].

---

[70] Again, alternative models with matching group fixed effects look very similar.

[71] Again results look similar if we add matching group fixed effects; of course the main effects of the first three regressors are not identified in fixed effects models.

[72] Our measure for severity is generated the following way: in auxiliary regressions, for each individual and period we regress received punishment on contributions, for all periods until the previous. The coefficient of this regressor is our measure for severity. For the ease of interpretation, we multiply the resulting coefficient in the auxiliary regressions by -1, so that a higher coefficient of regressor "experienced severity" in the final regression implies that participants are more sensitive to the severity of punishment.

[73] We operationalize experienced cooperativeness by the average contribution of the remaining group members, in the previous period.

TABLE 3

Driving Forces

Dependent variable: contribution		
	Model 1 Mixed effects regressions	Model 2 Random effects Tobit regression
Communication	-.029	-.958
	(1.211)	(1.135)
Transparency	-1.124	-3.029*
	(1.261)	(1.229)
Experienced Severity	.254***	.319***
	(.064)	(.083)
Communication*Exp_Sev	.354*	.584**
	(.163)	(.655)
Transparency*Exp_Sev	-.330^{+}	-.286
	(.179)	(.256)
Experienced Cooperativeness	.467***	.603***
	(.034)	(.044)
Communication*Exp_Coop	-.051	.028
	(.050)	(.064)
Transparency*Exp_Coop	.150**	.318***
	(.053)	(.071)
Period	-.012	-.005
	(.020)	(.028)
Constant	7.319***	5.954***
	(.836)	(.785)
N	2720	2720
Left Censored		187
P model	<.001	<.001
Wald Chi2	576.05	674.76

Column 1 shows a Mixed effects Model. Here choices are nested in individuals nested in matching groups. Column 2 shows a Random effects Tobit model. Here lower censoring is at 0 and the upper censoring is at 20. The reference category is the *Baseline* (no communication, no transparency). Standard errors are presented in parentheses. The *Communication* dummy equals 1 for all observations, which are not in the *Baseline*, the *Transparency* dummy equals 1 for all observations of the *Public* treatment. Experienced cooperativeness (*Exp_Coop*) is the mean contribution of other group members in previous period. Experienced severity (*Exp_Sev*) is a coefficient of local regression of received punishment on contribution, for this participant, from period 1 until previous period. The Significance at the 10%, 5%, 1% and 0.1% level is denoted by $^{+}$, *, **, and *** respectively.

To test for effects of guilt, i.e., for a possible first direct effect of justifying punishment decisions, treatments are recoded the following way: “communication” is a dummy that is 1 whenever the authority had to communicate her reasons to the punishee, i.e., in treatments *Private* and *Public*. “Transparency” is a dummy that is 1 if all reasons given by the current authority are made publicly available, i.e., in treatment *Public*. Per se, exposing punished players to higher levels of guilt is not instrumental. In the mixed effects model all effects are insignificant. In the Tobit model, the effect of transparency even turns out significantly negative. Guilt is not driving the results. The coefficient for experienced severity informs us about the second direct effect of a justification requirement.[74] As Table 3 shows, monetary punishment is most effective if reasons are communicated, but not made public (treatment *Private*). This result is supported by the regressions. The interaction between "communication" and "experienced severity" shows that severity is significantly more effective in treatment *Private*, compared with the *Baseline*.

From the strong and highly significant coefficient of experienced cooperativeness we learn that this is an important driver of cooperation even if justifications for punishment choices are not communicated. Per se, communicating reasons to the addressee does not make participants more sensitive to experienced cooperativeness (the interactions with "communication" are insignificant. This is different if reasons are made publicly known; there is a significant and strong positive interaction between "transparency" and experienced cooperativeness. This leads to:

> **Results 3**: If, in a linear public good, authorities are obliged to justify punishment decisions, this affects contributions on a direct and on an indirect channel. On the direct channel, participants become more sensitive to the severity of punishment, whenever reasons are communicated to them. On the indirect channel, participants become more sensitive to experienced cooperativeness of the remaining group members, if reasons are made publicly known.

We may now also explain why communicating justifications individually is less successful than communicating them publicly. In both treatments, authorities punish less, presumably because they see reproach as a partial substitute for monetary punishment. If reasons are made public, this strategy works, while it does not if reasons remain private. In that case active players expect others even more to be disciplined financially. This gives us:

> **Result 4:** The reasons given for punishment work as a partial substitute for monetary sanctions only if they are made public.

[74] A numeric example may help interpret the result. Assume that a participant had contributed nothing in the first period, and received 4 punishment points. She had contributed 5 Taler in the second period, and had received 3 punishment points. In the third period she had contributed 10 Taler and had received 2 punishment points. The local regression equation then becomes 4 - .2*contribution. Let's assume this participant contributes 11 Taler in period 4, and there are no other explanatory factors. The regression of contribution would then have to find very strong sensitivity to past severity of punishment. Period 4 choices of this one participant would be perfectly predicted if the coefficient for past severity was -5, and if the regression read 10 – 5 * (-.2 [severity coefficient from the local regression]) = 11.

VI. CONCLUSION

In social interaction, punishers are usually expected to justify their interventions. By contrast, the standard protocol exposes experimental punished players to sanctions without reasons. In this paper, we test in which ways punishment choices and contributions change if authorities are obliged to formulate explicit reasons for punishing active players in a linear public good. In our *Baseline*, authorities are requested to justify punishment decisions, but the reasons are kept confidential. In the first treatment, the addressee is informed about the justification of the authority's decision affecting her, but each active player only learns the reasons regarding herself. In the second treatment, all reasons are made public. Whenever reasons are communicated, there is less monetary punishment. Authorities partly substitute words for actions. However, contributions decay in later periods if the justification is only communicated to the addressee. In our *Public* treatment contributions are stabilized at a high level by a combination of low monetary punishment and justification, while in the *Baseline* without communication a high level of punishment is needed to achieve the same stable level of contributions.

In all treatments experienced cooperativeness and experienced severity significantly explain contribution choices. However, these experiences have a differently strong effect, depending on how the justification requirement is specified. Seeing the remaining participants make substantial contributions is the most important factor. This factor carries most weight if reasons are made public. If reasons are only communicated to the addressee, punishment authorities punish significantly less, but active players are even more sensitive to the severity of punishment. This suggests that there is a mismatch between the expectations of authorities (assuming reproach to be a partial substitute for monetary harm) and the expectations of active players (waiting for free-riders to be severely punished). This mismatch dissolves if reasons are made publicly available.

Some experimental designs by their very nature involve an experimenter demand effect. In a very strict sense of the term, this qualification also holds for this experiment. In all treatments, the reasons go to the experimenter. Authorities might be influenced by knowing that, at some later point, the experimenter will read these reasons. Yet for three reasons this possibility is not a relevant limitation of the design. First, in none of the motivating situations, reasons were without a recipient. Second had we designed the *Baseline* such that the reasons are not recorded, we would have had two differences between the *Baseline* and the two treatments. This would have made it impossible to interpret treatment effects. Third, the focus of the paper is on treatment effects, and this element of the design is held constant (see the discussion of this issue by Zizzo, 2010).

One should be cautious when extrapolating from the lab to the field. Lab experiments are tools for identifying effects and explaining them. In the interest of achieving this, they deliberately abstract from a host of contextual factors that are very likely to matter in the field. All of our motivating examples have features that are likely to affect the effectiveness of justification and were not present in our experiment. Specifically, in the experiment interaction was anonymous, whereas in all examples the authority and the potential recipient of punishment are identified. Moreover, in the experiment authorities and active group members were re-matched every period, whereas in many examples the relationship is stable over time. Notably this is, however, different in the legal example. In the experiment, the role of an authority was randomly assigned, whereas in all examples authorities hold a position that has been given to them by some higher authority (which is nature in the case of parents). By using a stranger matching, we not only come closer to the characteristic situation in courts. We also put our hypothesis to a harder test. As is well known, cooperation is easier to achieve in experiments with partner matching and punishment is more effective. In our design, we deliberately exclude any reputation and reciprocity-effects and thereby isolate the effect of communicating reasons. In all examples, authorities have superior competence. In the experiment, if communication is permitted it is strictly unilateral. In all examples, the potential recipient of punishment may at least explicitly ask for a justification. In many examples she even has some right to be heard.

It will be interesting, in future work, to test some of these moderating factors. Nonetheless, even based on this first experimental investigation of a justification requirement in a public good game, tentative normative conclusions can be drawn. It seems that giving reasons is not necessarily a good idea. If these reasons are not made public, the authority may overly focus on educating the addressee, whereas bystanders become skeptical that others who are tempted to misbehave are effectively disciplined. By contrast, if the authority is transparent about the reasons, words may indeed partly substitute acts, to everybody's benefit.

VII. References

Almenberg, Johan, Anna Dreber, Coren L. Apicella, and David G. Rand. 2011. “Third Party Reward and Punishment: Group Size, Efficiency and Public Goods.” In N. M. Palmetti, and J. P. Russo (eds.), *Psychology and Punishment*, New York: Nova Science Publishers: 73-92.

Balliet, Daniel. 2010. “Communication and Cooperation in Social Dilemmas: A Meta-analytic Review.” *Journal of Conflict Resolution*, 54(1): 39-57.

Battigalli, Pierpaolo, and Martin Dufwenberg. 2007. “Guilt in Games.” *American Economic Review*, 97(2): 170-176.

Bentham, Jeremy. 1830. *The Rationale of Punishment*. London: R. Heward.

Berlemann, Michael, Marcus Dittrich, and Gunther Markwardt. 2009. “The Value of Non-binding Announcements in Public Goods Experiments. Some Theory and Experimental Evidence.” *Journal of Socio-Economics*, 38(3): 421-428.

Blume, Andreas, and Andreas Ortmann. 2007. “The Effects of Costless Pre-Play Communication. Experimental Evidence from Games with Pareto-Ranked Equilibria.” *Journal of Economic Theory*, 132: 274-290.

Bochet, Oliver, Talbot Page, and Louis Putterman. 2006. “Communication and Punishment in Voluntary Contribution Experiments.” *Journal of Economic Behavior & Organization*, 60: 11-26.

Carpenter, Jeffrey P., Peter Hanns Matthews, and Okomboli Ong’Ong’a. 2004. “Why Punish? Social Reciprocity and the Enforcement of Prosocial Norms.” *Journal of Evolutionary Economics*, 14(4): 407-429.

Charness, Gary. 2000. “Self-Serving Cheap Talk. A Test of Aumann's Conjecture.” *Games and Economic Behavior*, 33: 177-194.

Charness, Gary, Ramón Cobo-Reyes, and Natalia Jiménez. 2008. “An Investment Game with Third-party Intervention.” *Journal of Economic Behavior & Organization*, 68(1): 18-28.

Crawford, Vincent. 1998. “A Survey of Experiments on Communication via Cheap Talk.” *Journal of Economic Theory*, 78(2): 286-298.

Croson, Rachel T.A., and Melanie Marks. 2001. “The Effect of Recommended Contributions in the Voluntary Provision of Public Goods.” *Economic Inquiry*, 39: 238-249.

Duffy, John, and Nick Feltovich. 2002. “Do Actions Speak Louder Than Words? An Experimental Comparison of Observation and Cheap Talk.” *Games and Economic Behavior*, 39(1): 1-27.

Elliott, Catherine S, Donald M Hayward, and Sebastian Canon. 1998. “Institutional Framing. Some Experimental Evidence.” *Journal of Economic Behavior & Organization*, 35(4): 455-464.

Engel, Christoph. 2007. “The Psychological Case for Obliging Judges to Write Reasons.” In C. Engel, and F. Strack (eds.), *The Impact of Court Procedure on the Psychology of Judicial Decision Making*, Baden-Baden: Nomos: 71-109.

Engel, Christoph. 2014. “Social preferences can make Imperfect sanctions work: Evidence from a public good experiment.” Journal *of Economic Behavior & Organization*, In Press.

Engel, Christoph, and Lilia Zhurakhovska. 2012. “You are in Charge. Experimentally Testing the Motivating Power of Holding a (Judicial) Office.” *MPI Collective Goods Discussion Paper*.

Fehr, Ernst, and Urs Fischbacher. 2004. “Third-Party Punishment and Social Norms.” *Evolution and Human Behavior*, 25: 63-87.

Fehr, Ernst, and Simon Gächter. 2000. “Cooperation and Punishment in Public Goods Experiments.” *American Economic Review*, 90: 980-994.

Fehr, Ernst, and Klaus M. Schmidt. 1999. “A Theory of Fairness, Competition, and Cooperation.” *Quarterly Journal of Economics*, 114: 817-868.

Fischbacher, Urs. 2007. “z-Tree. Zurich Toolbox for Ready-made Economic Experiments.” *Experimental Economics*, 10: 171-178.

Greiner, Ben. 2004. “An Online Recruitment System for Economic Experiments.” In K. Kremer, and V. Macho (eds.), *Forschung und Wissenschaftliches Rechnen.* Gesellschaft für Wissenschaftliche Datenverarbeitung Bericht, Göttingen: Datenverarbeitung, 63: 79–93.

Herrmann, Benedikt, Christian Thöni, and Simon Gächter. 2008. “Antisocial Punishment Across Societies.” *Science*, 319: 1362-1367.

Holt, Charles A., and Susan K. Laury. 2002. “Risk Aversion and Incentive Effects.” *American Economic Review*, 92: 1644-1655.

Kahneman, Daniel, and Amos Tversky. 2000. “Choices, Values, and Frames.” In D. Kahnemann, and A. Tversky (eds.), *Choices, Values, and Frames*, Cambridge, Cambridge University Press: 1-16.

Leibbrandt, Andreas, and Raúl López-Pérez (2012). “An Exploration of Third and Second Party Punishment in Ten Simple Games.” *Journal of Bahavioral Economics*, 84(3): 753–766.

Liberman, Varda, Steven M Samuels, and Lee Ross. 2004. “The Name of the Game. Predictive Power of Reputations versus Situational Labels in Determining Prisoner’s Dilemma Game Moves.” *Personality and Social Psychology Bulletin*, 30(9): 1175-1185.

Liebrand, Wim B., and Charles G. McClintock. 1988. "The Ring Measure of Social Values. A Computerized Procedure for Assessing Individual Differences in Information Processing and Social Value Orientation." *European Journal of Personality*, 2: 217-230.

Masclet, David, Charles Noussair, Steven Tucker, and Marie-Claire Villeval. 2003. "Monetary and Non-Monetary Punishment in the Voluntary Contributions Mechanism." *American Economic Review*, 93: 366-380.

Masclet, David, Charles Noussair, and Marie-Claire Villeval. 2012. "Threat and Punishment in Public Good Experiments." *Economic Inquiry*, 51(2): 1421–1441.

McCormac, John W. 1994. "Reason Comes Before Decision." *Ohio State Law Journal*, 55: 161-166.

Montero, Maria, Martin Sefton, and Ping Zhang. 2008. "Enlargement and the Balance of Power. An Experimental Study." *Social Choice and Welfare*, 30: 69-87.

Ross, Lee, and Andrew Ward. 1996. "Naive Realism in Everyday Life. Implications for Social Conflict and Misunderstanding." In E. S. Reed, and E. Turiel (eds.), *Values and Knowledge*, Maywah: Erlbaum: 103-135.

Sally, David. 1995. "Conversation and Cooperation in Social Dilemmas. A Meta-analysis of Experiments from 1958 to 1992." *Rationality and Society*, 7(1): 58-92.

Schauer, Frederick. 1995. "Giving Reasons." *Stanford Law Review*, 47: 633-659.

Seidenfeld, Mark. 2001. "The Psychology of Accountability and Political Review of Agency Rules." *Duke Law Journal*, 51: 1059-1095.

Tetlock, Philip E. 1983. "Accountability and Complexity of Thought." *Journal of Personality and Social Psychology*, 45: 74-83.

Xiao, Erte, and Fangfang Tan. 2014. "Justification and Legitimate Punishment." *Journal of Institutional and Theoretical Economics*, 170(1): 162-188(21).

Zizzo, Daniel John. 2010. "Experimenter Demand Effects in Economic Experiments." *Experimental Economics*, 13(1): 75-98.

Chapter 4

Voice Effects on Attitudes towards an Impartial Decision Maker – Experimental Evidence

MARCO KLEINE[*], PASCAL LANGENBACH[*] AND LILIA ZHURAKHOVSKA[*+#]

[+]UNIVERSITY OF COLOGNE,
[#]UNIVERSITY OF ERLANGEN-NUREMBERG,
[*]MAX PLANCK INSTITUTE FOR RESEARCH ON COLLECTIVE GOODS, BONN

Key words: voice, procedure, fairness, impartial decision maker, participative decision making, communication, laboratory experiment

JEL: C91, D03, D23, D63, K23, K40

I. INTRODUCTION

People care about more than their own monetary payoffs. This insight has led to the integration of distributive fairness into modern economic models (e.g., Fehr and Schmidt, 1999, Bolton and Ockenfels, 2000). Procedural fairness on the other hand, i.e., the fairness of the way an outcome is achieved has not received the same attention. However, in the last decade the interest in the question how aspects of procedural fairness shape people's behavior in economic interactions has grown (e.g., Frey and Stutzer, 2005; Konow, 2003; Frey et al., 2001, 2004; Bolton et al., 2005; Dal Bó et al., 2010; Dur and Roelfsema, 2010). An important aspect affecting people's evaluation of a procedure is the existence of "voice" (Hirschman, 1970). While voice can generally be "*some form of participation in decision making by expressing one's own opinion*" (Folger, 1977, p.109), we refer to a unilateral statement of a subordinate towards a decision maker before the decision is taken.[75] One strand of literature studies the effect of voice on decision makers (i.e., the recipients of voice) using dictator games (e.g., Charness and Rabin, 2005; Rankin, 2006; Mohlin and Johannesson, 2008; Yamamori et al., 2008; Xiao and Houser, 2009; Andreoni and Rao, 2011). Our research interest differs in two respects: First, we focus on the effects of voice on the behavior of those who state their opinion. Second, the decision maker in our setting is impartial, which means her decision is not biased by any personal monetary stakes.

Recent studies on voice towards a self-interested decision maker find that voice leads to positive or negative reciprocal reactions – depending on whether decision makers follow the voiced suggestions. Corgnet and Hernán-González (2014) study a principal-agent gift exchange setting in which the principal decides about the implementation of a transfer rule and the agent subsequently decides how much money to transfer to the principal. The authors find that principals only receive higher transfers after consulting the agents if they implement the suggested transfer rule. By contrast, for those principals who do not follow the agent's suggestion, granting voice has detrimental effects on transfers (see also Mertins, 2010). Similar effects occur in the ultimatum game: Ong et al. (2012) report that when responders have the opportunity to suggest an appropriate share towards the proposer, minimal accepted offers are higher than without voice opportunities (see also Rankin, 2003 for similar results). This "expectation effect" of voice (Ong et al., 2012) is explained as follows: people want their voice to matter and react negatively when expectations are not met. Voice is hence used to influence decisions and therefore has merely an "instrumental" function (e.g., Thibaut and Walker, 1975; Shapiro and Brett, 1993).

Contrary to these findings, we show that participation in a decision making process via voice can have a positive value in itself. In fact, we provide causal evidence that such participation can unfold unanimously positive effects on subordinates' behavior, irrespective of the actual decision by

[75] Using the ultimatum game, Xiao and Houser (2005) study the effects of the responder's opportunity to state her opinion after the offer decision has been taken. They conclude that the opportunity to let off steam can also be an important aspect of voice. A similar process seems to be at hand in one of the treatments in Ong et al. (2012), where the responder in an ultimatum game can send a message about the desired allocation to the experimenter.

the decision maker. People may derive utility from the mere fact that they can state their opinion in the decision making process (which in turn affects their behavior). Such a “value-expressive” or “non-instrumental” function of voice has been brought forward in social psychology (e.g., Katz, 1960; Tyler et al., 1985; Tyler, 1987; Lind et al., 1990). Voice procedures and outcomes resulting from these procedures are judged as fairer than their no-voice counterparts (e.g., Lind et al., 1990; van Prooijen et al., 2004 with further references). This results in positive evaluations of decision makers and an enhanced willingness to cooperate and accept decisions (e.g., Tyler, 1987; Tyler, 1988; Lind and Tyler, 1988; Tyler and Blader, 2000).[76] Apparently, in our setting, the “value-expressive” function of voice shapes behavior rather than the “instrumental” function.

In our experiment, an impartial decision maker allocates money between two subordinates. Previously, this money has been earned by the subordinates in a real-effort task in which the workload and the piece rate differed. In two *Voice* treatments, the subordinate with the higher workload and the higher piece rate expresses her opinion about a fair allocation towards the decision maker. The *Voice* treatments differ in the extent of voice: in one treatment, only the desired allocation in the form of numbers can be stated (*Narrow Voice* treatment); in another treatment, a written message can be sent in addition (*Broad Voice* treatment). No communication is possible in the *Baseline*. In a subsequent part, the subordinate player with the higher workload (and therefore the one who has been given voice) is the sender in a previously unannounced dictator game with the former decision maker being the receiver. Using a revealed preference approach, we interpret the subordinate’s transfers in the dictator game as a measure for her attitude towards the decider and hence the treatment differences in transfers as a measure for the effect of voice. At this point, the subordinate player is not yet informed about the actual decision in the first part. Applying the strategy method (Selten, 1967), she can condition her transfers on any possible allocation from the first part.

We find strong treatment differences in subordinates’ transfers. Subjects in both *Voice* treatments transfer significantly more money to the decider than in the *Baseline*. On average, transfers increase by 90%. Most interestingly, this positive effect on transfers is largely independent of allocation decisions in the first part of the experiment. This speaks in favor of a prevailing “value-expressive” effect of voice in our setting. The fact that our results stand in contrast to the results of Corgnet and Hernán-González (2014) and Ong et al. (2012) suggests that the role of the decision maker as self-interested or as impartial is crucial for the direction of the voice effect.

Furthermore, we find no differences across the two *Voice* treatments, indicating that the “value-expressive” effect is independent of the extent of voice. Interestingly, subordinates who were heard do not expect more favorable allocations in the *Voice* treatments than in the *Baseline*, although they perceive to influence the decision through voice.

[76] These studies mainly rely on surveys and experiments on self-reported fairness perception (but see, e.g., Lind et al., 1990). Although these methods have proven to be informative, they are also prone to potential biases, as there are no incentives to report truthfully. We close the gap by testing the pure effect of unilaterally voicing one’s opinion in a decision making process in a laboratory experiment and by inferring the subordinates’ attitude towards an impartial decision maker from her money allocation decisions in a subsequent dictator game.

For the remainder of the paper, we proceed as follows: We explain our experimental design in detail in section II, derive our behavioral hypotheses in section III, and report the results in section IV. Two additional treatments, which are meant to refine our findings, are introduced in section V. Section VI briefly summarizes and discusses the results.

II. Design

TABLE 2

Experimental Design

Part 1:	Real-effort task by players X and Y with asymmetric workload and piece rate ***Treatment variation:*** *player X sends/ does not send a message to player A* - *Baseline:* no message - *Narrow Voice*: statement about a fair allocation - *Broad Voice:* statement about a fair allocation plus written message (limited to 800 characters) Allocation decision by impartial player A
Part 2:	(Unannounced) dictator game with player X as dictator and player A as receiver (strategy method for all 21 possible allocations from part 1)
Part 3:	Belief elicitation of players X and Y about chosen allocation by player A in part 1

Table 1 gives an overview of the experimental design. The game consists of three parts. Subjects know that there will be several parts, but receive specific information about the content of each part only immediately before playing the relevant part of the experiment. No information about the other participants' decisions and therefore about any earnings is given to the subjects before the end of the experiment. Subjects are explicitly told that they cannot lose money they have earned in a previous part in any of the subsequent parts. In the experiment, we use an experimental currency unit (ECU). All instructions[77] are read aloud by the experimenter immediately before the relevant part to achieve common knowledge about the procedure.

[77] See section IV.1 in the Appendix for an English translation of the instructions.

At the beginning of the experiment, each subject is randomly assigned one of the three roles A, X, or Y. Players keep their roles across the three parts of the experiment. Subjects are then matched in groups of three, with one player from each role.

Part 1:

In part 1, players X and Y complete a real-effort task of counting zeros on a screen of zeros and ones.[78] Size and difficulty of the screens are identical, but the number of screens to be solved and the piece rate differs between players X and Y. Player X has to solve 12 tables, while player Y has to solve only 4 tables. Player X realizes 150 ECU per screen, while player Y realizes only 50 ECU per screen. Players cannot move to the next part unless they have completed their task.[79] Thus, player X contributes 1800 ECU and player Y contributes 200 ECU to an amount of 2000 ECU generated in total. We chose an asymmetric workload and productivity to induce a normative conflict (Konow, 2000; Reuben and Riedl, 2012; Nikiforakis et al., 2012) among players. Thus, we provide arguments for differing opinions about an appropriate allocation of the 2000 ECU between players X and Y. Focal normative rules which could be considered by the players as fair are equity output (players X and Y deserve an allocation according to the ECU they produce, i.e., 1800 ECU for player X and 200 for player Y), equity input (players X and Y deserve an allocation according to the number of tasks they solve, i.e., 1500 ECU for player X and 500 ECU for player Y), and equality (equal split of the amount of money, i.e., 1000 ECU for player X and 1000 ECU for player Y). Player A, who will later decide about the actual allocation, is not involved in the real-effort task. After completing the real-effort task, all three players indicate in private which allocation of the total amount between player X and player Y they would consider as fair.[80]

The next step is subject to our treatment variation. In the treatments *Broad Voice* and *Narrow Voice*, player X sends a message to player A. In the *Narrow Voice* treatment, player X states to player A which allocation she would consider as fair. In addition to the stated number, in the *Broad Voice* treatment, player X can send a written message (limited to 800 characters) to player A. In the *Baseline*, no messages are sent. In no treatment can players Y or A send a message.[81]

Finally, player A is asked to allocate the amount of 2000 ECU between players X and Y "in a fair way". Her decision is confined to 21 possible allocations in steps of 100 ECU (from player X receiving 2000 ECU and player Y receiving 0 ECU to player X receiving 0 ECU and player Y receiving 2000 ECU). Player A is impartial in her decision making: she receives a lump-sum payment

[78] The real-effort task is a modified version of the one used in Abeler et al. (2011).
[79] In case a table has been solved incorrectly by a subject he gets up to two additional trials. If after the third trial the table has been solved incorrectly again the subject receives a new table. Only after each subject in the session has solved the required number of tasks correctly the experiment continues. These rules (and all other rules of this part of the first part of the experiment) are public to everyone before subjects start the experiment.
[80] It is made explicit that this information will not be revealed to the other players and has no influence on the earnings of the players.
[81] We restrict the voice opportunity to player X to elicit the mere effect of voice, irrespective of other strategic considerations. If player Y also had this possibility, this would have induced beliefs about the other player's message for player X. It would have introduced a strategic component of voice and uncertainty about the effectiveness of player X's message in comparison to player Y's message.

of 5 Euro (equivalent to 1000 ECU). Therefore, her decision does not influence her payoffs in the first part.

Part 2:

In part 2, player X receives an additional endowment of 1000 ECU and plays a dictator game (Forsythe et al., 1994) with player A as the receiver.[82] Player X can transfer any integer amount up to 1000 ECU to player A. At this point, player X does not know the actual allocation decision by player A in the first part. Using the strategy method (Selten, 1967),[83] player X indicates her transfer for each of the 21 possible allocations in the first part. At the end of the experiment, only the transfer corresponding to the actual decision of player A is realized. [84]

Part 3:

In the third part, we elicit incentivized beliefs (expectations) of players X and Y about the allocation chosen by player A in the first part.[85] For a correct guess, a player receives 250 ECU; for a guess that deviates from the actual allocation by one step only, she receives 50 ECU. If the guess deviates further from the actual allocation, the player does not receive any additional earnings in this part.

The game is played only once. After part 3, subjects learn the payoff-relevant decisions of the other players and their earnings. The players' earnings are calculated as follows:

- Earnings of player A = lump-sum payment from part 1 + transfer from player X in part 2
- Earnings of player X = amount allocated by player A in part 1 + 1000 ECU - transfer to player A in part 2 + earnings from the belief elicitation
- Earnings of player Y = amount allocated by player A in part 1 + earnings from the belief elicitation

At the end of the experiment, we elicit participants' social value orientations by applying a standard test by McClintock and Liebrand (1988), in order to be able to control for differences in dictator game transfers due to differences in social preferences. Furthermore, subjects answer questionnaires including questions about the perceived fairness of the allocation and the procedure, as well as demographics.

The experiment was conducted at the Cologne Laboratory for Economic Research in May and June 2012 using z-tree (Fischbacher, 2007). 264 participants were recruited via ORSEE (Greiner,

[82] We conducted two additional treatments (*Baseline-Uninvolved* and *Narrow-Voice-Uninvolved*) in which the receiver of the second-part dictator game is not the impartial decision maker, but an uninvolved third party – in our setting, a charity. We refer to section V for more details.
[83] Brandts and Charness (2011) show that, if the main focus lies on the comparison of decisions within strategies, using the strategy method can be problematic. On the contrary, for the comparison between treatments, the main limitation of that method is that the "strategy method provides a lower bound for testing for treatment effects" (p. 392).
[84] At the end of this part, players X and Y also indicate for every possible allocation how satisfied they would be with the allocation.
[85] People may believe to influence the allocation decision in the first part if they are granted voice. Therefore, controlling for their beliefs about the actual allocation is necessary to filter out the pure effect of voice on transfers in the second part.

2004) from the subject pool of the laboratory. 87 subjects participated in the *Narrow Voice* treatment as well as in the *Baseline* (29 independent observations each) and 90 subjects participated in the *Broad Voice* treatment (30 independent observations). Participants were mainly students from various disciplines (39% majoring in economics) with a mean age of 24.80 years (sd=5.18). 56% were female. Sessions lasted approximately 90 minutes on average. The experimental currency was converted into Euro (2 ECU = 0.01 EUR) at the end of the experiment and paid out in cash. Participants earned 14.74 EUR[86] on average (sd=3.00), including a show-up fee of 4 EUR.

III. HYPOTHESES

Our main variable of interest is the transfer by player X in the dictator game as a measure for the player's attitude towards the impartial decision maker. In particular we are interested in testing if subjects only value voice as an instrument to influence outcomes or if voice has a value in itself irrespective of outcomes. Applying the standard assumptions of rational and self-interested players, predictions for transfers in the dictator game are straightforward: independent of the treatment and the possible allocation in the first part, player X will not transfer any money to the impartial decision maker. However, a vast body of experimental literature on dictator games has shown that transfers are common in such a non-strategic setting (for a meta-study, see Engel, 2011). We are particularly interested in how our treatment variations of voice towards the impartial decision maker affect the giving behavior and if this effect depends on the possible allocation chosen in part 1.

On the one hand, it is claimed that voice has a "value-expressive" function, i.e., that people derive utility from the mere fact that they can state their opinion. Positive voice effects may therefore occur irrespectively of the outcomes of the decision (e.g., Tyler et al., 1985; Tyler, 1987; Ong et al., 2012). We assume that these effects concerning the attitude towards the decision maker translate into actual behavior in the dictator game. If this "value-expressive" function of voice prevails in our setting, we should observe the following hypothesis to hold:

> **Hypothesis 1a**: Transfers in part two are generally higher in both *Voice* treatments than in the *Baseline*. Transfers in the *Voice* treatments are higher than in the *Baseline* for every possible allocation chosen in part one.

On the other hand, some authors stress an "instrumental" function of voice in decision making processes. People want to have voice in order to control the outcomes of the decision (e.g., Thibaut and Walker, 1975). If voice does not lead to more favorable outcomes, people might react negatively, as they do not perceive the voice procedure as an honest opportunity to influence the decision making process ("expectation effect", Ong et al., 2012). Experimental evidence on subordinates' behavior in

[86] 14.74 EUR corresponded to 18.39 USD at the time of the experiment.

interaction with self-interested decision makers is in line with a prevailing “instrumental” function of voice (Corgnet and Hernán-Gonzáles, 2014; Ong et al., 2012, Mertins, 2010). If voice predominantly has an “instrumental” function, its effect will depend on the favorability of the allocation in the first part of the experiment. Hence, the competing hypothesis is as follows:

> **Hypothesis 1b**: Transfers in part two are higher in the *Voice* treatments than in the *Baseline* only if the allocations for players X are favorable. If rather unfavorable outcomes are reached in part one, transfers in part two in the *Voice* treatment are lower than in the *Baseline*.

With the two different *Voice* treatments, we test the effect of the extent of voice. In the *Narrow Voice* treatment, communication between player X and player A is restricted to the indication of a fair allocation. The *Broad Voice* treatment allows for a greater opportunity to express one’s opinion, in that player X additionally sends a written message to player A. Based on a monotonicity argument, we expect that the more voice is granted the more pronounced are the voice effects that are hypothesized in 1a and 1b. Hence, we expect the following:

> **Hypothesis 2**: Voice effects are stronger in the *Broad Voice* treatment than in the *Narrow Voice* treatment.

Apart from the behavior of players X towards the impartial decision makers, we are interested in the question to what extent expectations about the actual outcome in the allocation part are influenced by the voice opportunity. In the strategic environment of the ultimatum game, the opportunity for voice towards the proposer seems to raise expectations for outcomes (Ong et al., 2012). If we assume that statements towards the impartial decision maker are biased by self-interest (e.g., Babcock and Loewenstein, 1997) and that people expect their voice to influence outcomes, we should observe similar patterns in our setting:

> **Hypothesis 3**: Players X expect higher outcomes from the allocation part in the *Voice* treatments than in the *Baseline*.

IV. RESULTS

IV.1. Main Effect of Voice

We directly turn to our main Hypotheses 1a and 1b. We thereby analyze the general effect of voice on transfers from players X to players A in the (unannounced) dictator game, as well as the dependence of the voice effect on the allocation chosen by player A in part 1. Figure 1, which shows mean transfers conditional on the possible allocations in part 1 for each treatment separately, illustrates a substantial positive effect of voice on transfers.

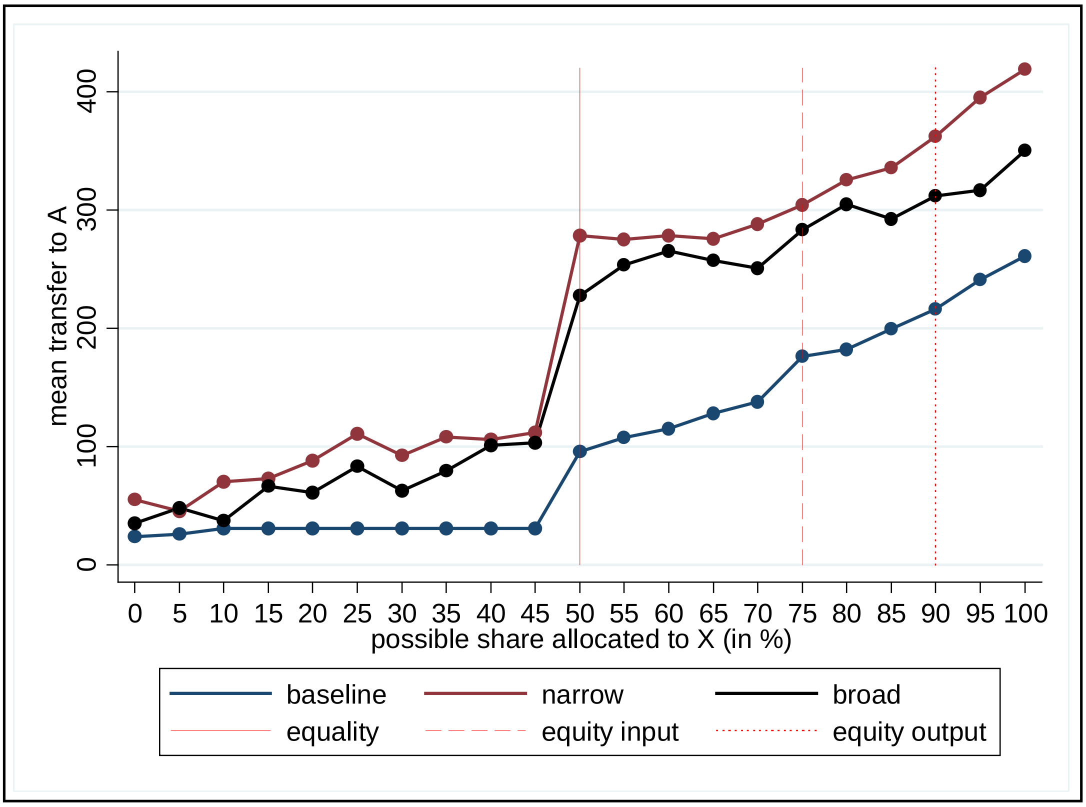


FIGURE 1

Mean transfers of players X to player A for every possible allocation allocated to her by the impartial decision makers

On the horizontal axis, the possible share of money allocated by As to Xs in part 1 of the experiment is indicated: 0% "possible share allocated to X" correspond to 0 ECU for X and 2000 ECU for Y; 5 % correspond to 100 ECU for X and 1900 ECU for Y, etc. 100% correspond to 2000 ECU allocated to player X and 0 ECU to Y. On the vertical axis, the mean transfers by players X to A in part 2 of the experiment are indicated.

Average transfers are higher in both *Voice* treatments than in the *Baseline*. Descriptively, the effect is especially pronounced for all allocations which guarantee players X more than 50% of the total amount produced in the real-effort task. A sizeable jump in transfers at the equal split is present in all three treatments. It may be explained by the fact that allocations to players X below 50% of the total amount are not supported by any of the focal normative fairness rules. However, even for these "unfair" allocations to players X, differences between the *Voice* treatments and the *Baseline* remain present and have the same (positive) algebraic sign as the differences in response to rather favorable allocations. Accordingly, mean transfers over all possible allocations differ largely between the *Voice* treatments and the *Baseline*. On average, players X transfer 210 ECU (sd=186) in the *Narrow Voice* treatment and 181 ECU (sd=137) in the *Broad Voice* treatment to the impartial decision makers (averages over all possible allocations). Transfers are the lowest in the *Baseline* with an average of 103

ECU (sd=120). Mann-Whitney tests show that mean transfers per player X are significantly higher in the *Voice* treatments than in the *Baseline* (*Narrow* vs. *Baseline*: |z|=2.444, p=0.0145; *Broad* vs. *Baseline*: |z|=2.390, p=0.0169).[87] In order to provide further evidence for the main effect of higher transfers in both *Voice* treatments, we conduct random effects Tobit regressions, considering transfers in the dictator game as the dependent variable.[88] Results are presented in Table 2. The treatment dummies *Narrow* and *Broad* explain treatment differences in comparison to the *Baseline* (level effects). In Model 1, we control by the variable *Possible allocation part 1* for effects that are due to a particular allocation that might have been implemented by decision makers. In fact, we find that the less money is allocated to players X in part 1 of the experiment, the less players X transfer in the dictator game (significant coefficient of -33.68). In Model 2, the variable *fair allocation* additionally controls for the private statement of players X about a fair allocation. In the Model 3, we add the variable *expectation* controlling for the allocation expectations by players X. In the Model 4, we further add the participants' *social value orientation score* to control for differences in dictator game transfers due to general individual differences in pro-social behavior. The variable *expectation* and the *social value orientation score* have explanatory power for transfers: the more money players expect for themselves in the initial allocation, the lower is their transfer; the more players are socially-minded, the more they give in the second-part dictator game. But most importantly, the treatment effects (i.e., the dummy coefficients *Narrow* and *Broad* with transfers in the *Baseline* being the dependent variable) are positive and significant in all models. Moreover, the effect sizes of the treatment dummies are substantial, once again indicating large differences between the *Baseline* and the *Voice* treatments. Hence, the regression analyses provide further support for the main effect of generally higher transfers in the *Voice* treatments.

As a final step to test Hypotheses 1a and 1b, we check non-parametrically if the treatment effect is robust for all possible outcomes in the first part. As indicated above, average transfers for every possible allocation are higher in the *Voice* treatments than in the *Baseline*. These differences are also statistically significant for most of the possible allocations.[89] Non-significant differences in transfers only emerge for extreme allocations: e.g., if players X receive less than 25% of the total amount. Additionally, transfers in the *Broad Voice* treatment are not significantly higher for very favorable allocations for players X (when they receive 80% or more). Extreme allocations can be considered as highly unfair (for either players X or players Y). Thus, the positive effect of voice on transfers in the dictator game is significant for all allocations, which can be supported as reasonably

[87] Throughout this paper, reported p-values are always two-sided.

[88] We use Tobit regressions because, in dictator games, giving possibilities are exogenously restricted with an upper and a lower bound; the lower bound is usually zero-giving. Bardsley (2008) shows that subjects also take money if they have the opportunity. In our setting, this seems plausible, since transfers show a general downward trend from favorable to unfavorable allocations to players X and often stop at the zero transfer level for the most unfavorable allocations. The downward trend is therefore stopped artificially. Tobit regressions account for the possibility that (some) subjects might have even taken money instead of giving nothing by controlling for censoring. Moreover, as we have 21 transfer decisions per individual (due to the strategy method), random effects models which take individual specific effects into account are in order.

[89] See Table 3 in the Appendix IV.2 for exact values.

fair. It even persists for rather unfair allocations, i.e., when a player receives less than 50% (but more than 25%) of the total amount. In particular, in no situation are transfers lower in the *Voice* treatments than in the *Baseline*.

TABLE 3

Treatment effects on transfers – comparison of Baseline and Voice treatments

Random effects Tobit regressions
Dependent variable: transfers in the dictator game

	Model 1	Model 2	Model 3	Model 4	Model 5
Narrow	236.22***	217.71***	198.28***	191.92***	275.86***
	(80.37)	(77.51)	(75.32)	(72.56)	(90.93)
Broad	211.97***	161.87**	155.40**	149.59**	246.63***
	(79.75)	(78.79)	(76.21)	(73.42)	(89.54)
Fair allocation		-27.49***	-15.86	-1.18	-1.35
		(10.36)	(11.11)	(12.12)	(11.16)
Expectation			-27.02**	-27.93**	-27.84***
			(11.44)	(11.02)	(10.00)
Social value orientation score				4.65**	4.20**
				(1.82)	(1.67)
Narrow*Gender					-222.89*
					(133.22)
Broad*Gender					-267.92**
					(131.00)
Gender					-32.25
					(94.38)
Possible allocation part 1	-33.68***	-33.68***	-33.68***	-33.67***	-33.67***
	(1.08)	(1.08)	(1.08)	(1.08)	(1.08)
Constant	217.26***	664.22***	862.25***	608.26***	636.15***
	(58.18)	(176.43)	(190.23)	(208.09)	(205.66)
N	1848	1848	1848	1848	1848
P model	<.001	<.001	<.001	<.001	<.001
Wald Chi2	981.76	985.97	988.88	993.65	1004.88

Random effects Tobit regressions. Standard errors are presented in parentheses. The *Narrow* dummy equals 1 for all observations of the *Narrow Voice* treatment, the *Broad* dummy equals 1 for all observations of the *Broad Voice* treatment, *fair allocation* controls for the allocation players X consider as fair, *expectation* controls for players X's expectations about the actual allocation by the impartial decision maker, *social value orientation score* controls for differences in players' social value orientation, the *gender* dummy equals 1 for male players, X *possible allocation part 1* controls for possible allocations that can be implemented by the impartial decision maker. Significance at the 10%, 5%, and 1% level is denoted by *, **, and ***, respectively. Left-censored = 865; right-censored =21.

To sum up, we find strong evidence for a general positive voice effect. Subjects' behavior is in line with a prevailing "value-expressive" function of voice (Hypothesis 1a). We do not find support for Hypothesis 1b: voice does not have a predominant "instrumental" function in our setting. We state the main result of our paper as follows:

Result 1: Transfers are substantially higher in both *Voice* treatments than in the *Baseline*. This positive voice effect is largely independent of the actual allocation decision.

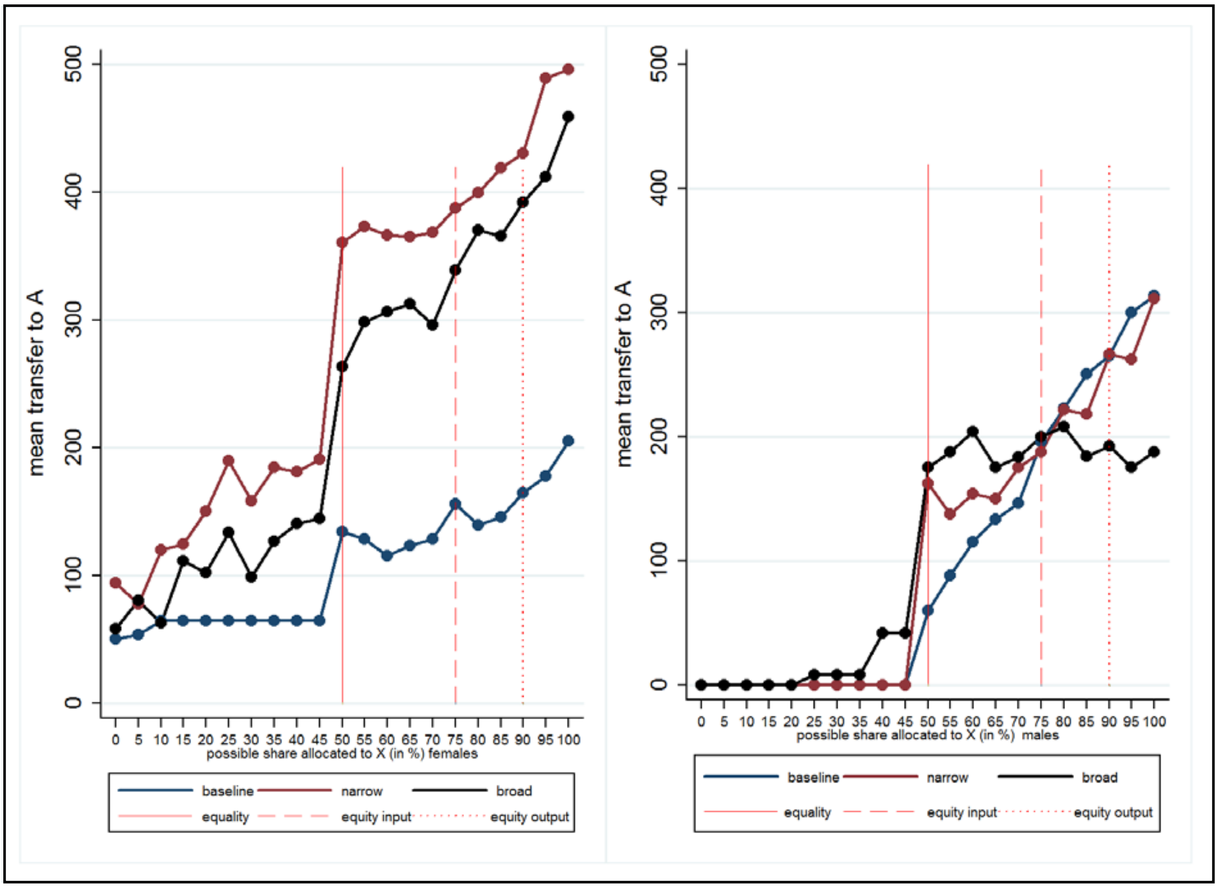


FIGURE 2

Mean transfers of players X to player A for every possible allocation allocated to her by the impartial decision makers – by gender

The left panel presents the data for the females, while the right panel presents the data for the males.
On the horizontal axis, the possible share of money allocated by As to Xs in part 1 of the experiment is indicated: 0% "possible share allocated to X" correspond to 0 ECU for X and 2000 ECU for Y; 5 % correspond to 100 ECU for X and 1900 ECU for Y, etc. 100% correspond to 2000 ECU allocated to player X and 0 ECU to Y. On the vertical axis, the mean transfers by players X to A in part 2 of the experiment are indicated.

Hence, the opportunity for voice in a decision making process positively affects the behavior of subordinates towards an impartial decision maker. In a further explorative data analysis, we unexpectedly find that the treatment effects on transfers reported above are driven by the behavior of female participants in the experiment (see Table 2 Model 5). Indeed, as depicted in Figure 2 transfers of male participants are not statistically different between treatments (average transfers: *Baseline* =100 (sd = 105), *Narrow* = 107 (sd = 92), *Broad* = 104 (sd = 86); Mann-Whitney tests of average transfers per player X: *Narrow* vs. *Baseline*: |z|= 0.196, p= 0.8449; *Broad* vs. *Baseline*: |z|= 0.294, p= 0.7689), whereas transfers of female participants are (average transfers: *Baseline* = 107 (sd = 139), *Narrow* = 282 (sd = 203), *Broad* = 232 (sd = 143); Mann-Whitney tests: *Narrow* vs. *Baseline*: |z|= 2.503, p= 0.0123; *Broad* vs. *Baseline*: |z|= 2.736, p= 0.0062). Due to the fact that we had no ex-ante hypothesis on gender differences, and due to problems of reporting and interpreting purely explorative results, we refrain from further elaboration. Nevertheless, we think that, since we are the first – to the best of our knowledge – to report such gender differences in behavior due to voice procedures, it might be interesting for future research to explore the robustness of this finding and possible explanations for it.[90]

IV.2. Extent of Voice

In the following, we analyze whether the extent of voice leads to differences in transfers by comparing the two *Voice* treatments. As Figure 1 already suggests, descriptively, differences between the *Voice* treatments seem to be small. Indeed, comparing the mean transfers per player X in the dictator game, we do not find significant differences (Mann-Whitney test, |z|=0.250, p=0.8024). This is further supported by a Wald test, which tests whether the coefficients *Narrow* and *Broad* in the regression analyses presented in Table 2 are significantly different. The test exhibits no significant differences between the *Voice* treatments ($p \geq 0.48$ for all models). Finally, we compare transfers in the *Voice* treatments for every possible allocation separately. Applying the Mann-Whitney test, none of the transfer comparisons exhibits significant differences on conventional levels.[91] Hence, we find no evidence for Hypothesis 2:

> **Result 2:** Transfers are not significantly different across the *Narrow Voice* and the *Broad Voice* treatment.

The result that the possibility to express oneself in a written message in addition to a numerical statement does not have a significant impact on the subsequent behavior is in line with the findings of Corgnet and Hernán-González (2014).[92] The numerical statement alone seems to be sufficient for the effect of voice.

[90] A possible explanation could be that women lay more emphasis on procedural fairness than men (e.g., Sweeney and McFarlin, 1997), but the robustness of this finding remains questionable, as other studies do not succeed in showing this difference (e.g., Kulik et al., 1996, Cohen-Charash and Spector, 2001, for a meta-study).
[91] See Table 3 in the Appendix IV.2 for exact values.
[92] Corgnet and Hernán-González (2014) find similar behavior in agents who could only send a simple statement to the principal, compared to the behavior of those who could chat with the principal for three minutes. Similarly, when studying the effect of voice on the recipient of the message, Andreoni and Rao (2011) report that the numerical statement influences the recipients' behavior, but an additional written message does not.

IV.3. Expectations

Presumably, people expect their voice to matter for the outcome of the decision (e.g., Ong et al., 2012). First we report results of the post-experimental questionnaire to elicit the perceived influence of players X on the decision of players A. Players X indicate on a 11-item Likert scale to what extent they perceive to have influenced the decision (0 = “no influence at all”; 10 = “very strong influence”). Players X in the *Voice* treatments perceive to have a higher influence on the decision of the impartial decision maker than those in the *Baseline* differences being highly significant (Mann-Whitney test *Broad* vs. *Baseline* |z|=3.642, p=0.0003; *Narrow* vs. *Baseline* |z|=3.989, p=0.0001).[93] Since the *Baseline* provides no means to players X to influence the decision, perceived higher influence in the *Voice* treatments seems obvious. But it raises the question if the perceived higher influence on the decision of players A also means that players X expect a more favorable allocation decision in the *Voice* treatments than in the *Baseline*. We now examine to what extent this is true for subjects in our experiment and turn to the analysis concerning our third hypothesis. On average, players X expect to receive 1228 ECU (sd=271) (of the 2000 ECU) in the *Narrow Voice* treatment, 1230 ECU (sd=261) in the *Broad Voice* treatment and 1334 ECU (sd=359) in the *Baseline* (Mann-Whitney test – *Narrow* vs. *Baseline* |z|=1.652, p=0.0985; *Broad* vs. *Baseline* |z|=1.597, p=0.1103). So, despite the more highly perceived influence on the allocation decision by players A in the *Voice* treatments, players X do not expect higher allocations to themselves:

> **Result 3:** Players X in the *Voice* treatments perceive that they have more influence on the decision in the allocation stage than players in the *Baseline*. However, at the same time, they do not expect more favorable decisions.

This result is striking and counterintuitive. It is not obvious, how one can interpret it. One potential explanation could be that, indeed, the perception of the process changes through voice while rational expectations about the outcomes of the process remain realistic (see section IV.4. for an analysis of the allocations chosen by players A). Players X apparently feel involved in the decision making process (see section IV.5. for an analysis of player X’s perception of the decision making process). However, due to the between-subject design of the experiment we cannot infer, whether players X who have voice actually think that they would have received a lower share of the total amount if they had no voice (the expected reference point changes through voice) or if they are aware that they cannot change the outcome and nonetheless appreciate having the opportunity to state their opinion.

[93] Averages – *Narrow*: 5.55, sd=3.56; *Broad*: 4.97, sd=3.19; *Baseline*: 1.90; sd=2.55. Moreover, players X messages towards the impartial decision makers are highly correlated with the expected allocation (Spearman's Rho – *Narrow*: r_S=0.5575, p=0.0017; *Broad*: r_S=0.6781, p=0.0000), which is another indication that the players believe their voice to matter. Averages of statements of fair allocations towards the impartial decision makers – *Narrow*: 1359, sd=298; *Broad*: 1260, sd=353.

IV.4. Earnings of Subordinates

Using the strategy method, we show in section IV.1 that players X value the voice opportunity irrespectively of outcomes. But it could well be that, in the actually realized decisions, players X do profit from the voice procedure in monetary terms. We therefore analyze the actual earnings of players X in the *Voice* and in the *Baseline*s.

In total, players X earn significantly less in the *Voice* treatments than in the *Baseline* (average earnings part 1 and 2 *Baseline*: 2144, sd=290; *Narrow*: 1868, sd=378, *Broad*: 1893, sd=311; Mann-Whitney test *Broad* vs. *Baseline*: |z|=2.652, p=0.0080; *Narrow* vs. *Baseline*: |z|=2.904, p=0.0037). This difference is due to two facts: First, in the *Voice* treatments, players A allocate lower amounts of money to players X in the first part of the experiment than in the *Baseline*. On average, players X in the *Baseline* receive 1286 ECU (sd=243) (of 2000 ECU to be distributed) in the first part. This is more than the earnings of those in the *Broad Voice* treatment (average of 1160 ECU, sd=192; Mann-Whitney test *Broad* vs. *Baseline*: |z|=2.127, p=0.0334) and of those in the *Narrow Voice* treatment (average of 1183 ECU, sd=267; Mann-Whitney test *Narrow* vs. *Baseline*: |z|=1.553, p=0.1204).[94] Second, despite the fact that players X in the *Voice* treatments receive less favorable allocations in the first part of the experiment, actually realized transfers from the dictator game are significantly higher in the *Voice* treatments than in the *Baseline* (averages: *Baseline*: 142 ECU, sd=180; *Narrow*: 315 ECU, sd=247; *Broad*: 267 ECU, sd=233; Mann-Whitney test *Narrow* vs. *Baseline*: |z|=2.993, p=0.0028; *Broad* vs. *Baseline* |z|=2.378, p=0.0174). Both aspects lead to the fact that players X do not profit from the voice procedures in monetary terms.

IV.5. Voice Effects on Perceived Fairness

Although we focus on subjects' actual behavior, we also aim at understanding if voice shapes subjects' perception of fairness with regard to procedures and outcomes. We again turn our analysis to players X. At the end of the experiment, subjects answer questions about the perceived fairness of the procedure,[95] namely (1) how fair they perceive the procedure to be in which the decision about the allocation came about in general, (2) how fair they perceive the procedure to be in which the decision about the allocation has been made from the viewpoint of player X in particular and (3) to what extent they personally feel treated in a fair way in the decision making process. Furthermore, they are asked to state (4) the extent to which they perceive the outcome as fair and (5) the extent to which they accept the decision.[96] In line with procedural justice literature, we predict a positive effect of voice on perceived fairness of the procedure and of the outcomes, as well as on the acceptance of the decision.

[94] In this paper, we almost exclusively focus on the effect of voice on the subordinate. We refer to a companion paper by Kleine, Langenbach, and Zhurakhovska (2013) for a detailed analysis of the influence of voice on the fairness decision by the impartial decision maker.

[95] As seen in section IV. 4, realized outcomes for players X are lower in the *Voice* treatments than in the *Baseline* treatments. Higher perceptions of fairness in voice treatments can therefore not be explained by higher outcomes.

[96] All questions had to be answered on an 11-point Likert scale ranging from 0="not fair at all" to 10="completely fair" (0="not at all" to 10="completely" for question (5)).

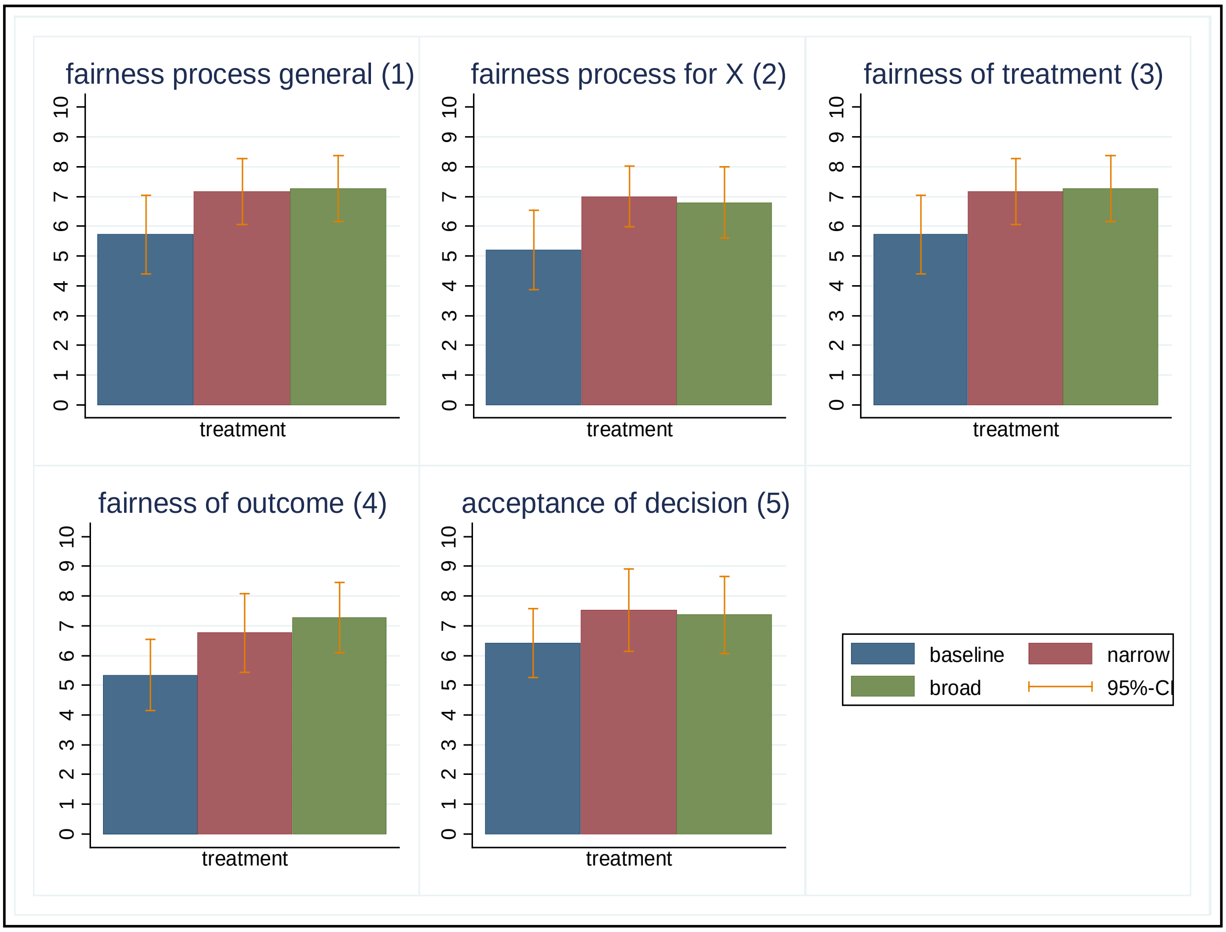


FIGURE 3

Players X' average ratings of: fairness of the process (1), fairness of the process from the point of view of a player X (2), fairness of the personal treatment (3), fairness of the outcome (4) and extent to which they can accept the decision (5)

The mean ratings are presented on the vertical axis. The scale goes from 0 (not at all) to 10 (completely). On the horizontal axis, one can see the different treatments. "95%-CI" is the 95% confidence interval.

Figure 3 illustrates the results. Indeed, the answers to the questions regarding the procedural fairness, i.e., questions (1)-(3), all point to the direction that players X consider the voice procedure as fairer than the no-voice procedure. The results are (weakly) significant for all comparisons between the *Broad Voice* and the *Baseline* and significant for comparisons of ratings concerning question (3) between the *Narrow Voice* and the *Baseline*.[97] The higher fairness ratings in question (1) (general procedural fairness) are remarkable: In fact, the *Voice* treatments could be considered as rather unfair in that they provide only one player with the opportunity for voice. However, players X seem to appreciate the voice opportunity when rating procedural fairness. Furthermore, the outcomes are also

[97] Question (1): averages – *Baseline*: 5.24, sd=3.31; *Narrow*: 5.83, sd=3.14; *Broad*: 7.10, sd=3.06; Mann-Whitney tests – *Narrow* vs. *Baseline*: |z|=0.682, p=0.4952; *Broad* vs. *Baseline*: |z|=2.241, p=0.0250; Question (2): averages – *Baseline*: 5.72, sd=3.45; *Narrow*: 7.17, sd=2.89; *Broad*: 7.27, sd=2.98; Mann-Whitney tests – *Narrow* vs. *Baseline*: |z|=1.527, p=0.1267; *Broad* vs. *Baseline*: |z|=1.685, p= 0.092; Question (3): averages – *Baseline*: 5.21, sd=3.50; *Narrow*: 7.00, sd=2.69; *Broad*: 6.80, sd=3.18; Mann-Whitney test – *Narrow* vs. *Baseline*: |z|=1.957, p=0.0504; *Broad* vs. *Baseline*: |z|=1.781, p=0.0749.

judged as significantly fairer in the *Voice* treatments than in the *Baseline*.[98] Finally, players X in the *Voice* treatments are somewhat more willing to accept the decision made by players A than those in the *Baseline*.[99] We therefore conclude that from the point of view of players X in this setting voice increases perceived fairness in regard to procedure and outcomes.

> **Result 4:** Players X in the *Voice* treatments generally perceive the procedure and outcomes as fairer than those in the *Baseline* and are more willing to accept the decision by the impartial decision makers.

V. Extension – Voice Effects on Behavior towards Uninvolved Parties?

So far, we have shown that voice has a positive effect on generous behavior towards an impartial decision maker. Moreover, the pattern we have observed has shed some light on the underlying mechanisms of how voice affects the players who are granted voice. Yet, we make one further attempt at improving our understanding of the mechanisms behind the voice effect. Therefore, we examine the question whether voice effects can only be shown in a direct interaction with the impartial decision maker, or whether it also affects behavior towards third, uninvolved parties. If voice effects are merely due to a positive change in the emotional state or due to an activation of a general sensibility for acting in a fair way (e.g., because participants feel treated more fairly), we conjecture that voice should also lead to more generosity towards uninvolved parties. If no such effect towards uninvolved parties occurs, this would not prove wrong the conjecture that emotional states are influenced by voice. Rather, it would favor the notion that we are correct with our interpretation that voice changes the attitude towards the impartial decision maker (i.e., a directed expression of gratitude for being treated fairly due to voice).

Therefore, we conducted two additional treatments. In terms of design and instructions, these treatments are identical with the *Baseline* and the *Narrow Voice* treatment described in section II, with the exception that in part two the recipient of the dictator game is an uninvolved third party rather than the impartial decision maker.[100] By assigning the role of the uninvolved party to a charity,[101] we keep the roles in the laboratory constant. We refer to these additional treatments as *Baseline-Uninvolved* and *Narrow Voice-Uninvolved*.

90 subjects participated in each of the treatments, which were conducted in March 2013. Subjects who participated in the treatments described in section II of this paper were not invited to

---

[98] Question (4): averages – *Baseline*: 5.34, sd=3.12; *Narrow*: 6.76; sd=3.49; *Broad*: 7.27, sd=3.15; Mann-Whitney tests *Narrow* vs. *Baseline*: |z|=1.853, p=0.0638; *Broad* vs. *Baseline*: |z|=2.467, p=0.0136.

[99] Question (5): averages – *Baseline*: 6.41, sd=3.04; *Narrow*: 7.52, sd=3.62; *Broad*: 7.37, sd=3.48; Mann-Whitney tests – *Narrow* vs. *Baseline*: |z|=2.013, p=0.0441; *Broad* vs. *Baseline*: |z|=1.610, p=0.1074. In related questions, ratings about the satisfaction with the outcome conditional on every possible allocation do not differ between treatments.

[100] See section IV.1.3 in the Appendix for an English translation of the instructions.

[101] The chosen charity "Deutsche Welthungerhilfe e.V." supports development projects against hunger and poverty worldwide. Subjects were informed about the goal of the charity and that it is certified by the "Stiftung Deutsches Zentralinstitut für soziale Fragen" (German Central Institute for Social Issues).

these sessions.[102] Our analysis is based on 30 independent observations for *Baseline-Uninvolved* and 29 independent observations for *Narrow-Uninvolved*. Participants were mainly students from various disciplines (34% majoring in economics) with a mean age of 25.21 (sd=6.15). 48% were female. As in the other treatments described in section II, sessions lasted approximately 90 minutes on average. Participants earned 16.19 EUR on average (sd=3.19).[103]

Again, we focus our attention on the main variable of interest – the transfers in the second part of the experiment.[104] In the *Baseline-Uninvolved* treatment, players X transfer on average 184 ECU (sd=221) to the uninvolved party (average over all possible allocations). In the *Narrow Voice-Uninvolved* treatment, the average transfer is 197 ECU (sd=234). The Mann-Whitney test that compares average transfers is insignificant ($|z|=0.076$, $p=0.9394$), which confirms that average transfers are very similar across the treatments. Also, when we compare transfers for every possible allocation the impartial decision maker might have implemented in the first part, we find no evidence for significant differences in transfers (for all respective Mann-Whitney tests: $p \geq 0.44$).[105] Moreover, we conduct a number of Tobit regressions (see Table 5 as an example in the Appendix IV.2.); in no model specification can we reject the null hypothesis that transfers in both treatments are the same. We therefore state the following result:

> **Result 5:** Transfers in the *Baseline-Uninvolved* and the *Narrow-Uninvolved* treatment are not statistically significantly different. Hence, in our setting, there seems to be no spillover effects of voice on behavior towards an uninvolved party.

VI. CONCLUSION

We show that the opportunity for voice ameliorates the attitude of a subordinate towards an impartial decision maker. In our experiment, the implementation of voice leads to substantively higher transfers to the impartial decider in a subsequent dictator game. This effect is largely independent of the outcomes of a previous decision by the impartial decision maker. Hence, our results stress the importance of a "value-expressive" effect of voice. Participants seem to appreciate the voice procedure not primarily for instrumental reasons. They rather value the mere fact of stating their opinion in the decision making process. In that sense, our results are distinct from the behavioral effects of voice

[102] One independent observation from the analysis of the *Narrow Voice-Uninvolved* treatment had to be excluded from the analysis, as one subject erroneously participated in both *uninvolved* treatments.
[103] 16.19 EUR corresponded to 21.06 USD at the time of the experiment.
[104] Since we implemented the uninvolved treatments to test whether voice also affects the subordinate's behavior towards an uninvolved party, we limit our analysis on the transfers of players X. For results on the other variables (expectations, perceived fairness ratings), we refer to the Table 6 in the Appendix IV.2. But note that comparisons of non-incentivized perceived fairness ratings across these treatments do not yield the same results as those in the main experiment. Applying Mann-Whitney tests, all comparisons between *Baseline-Uninvolved* and *Narrow Voice-Uninvolved* are insignificant. This indicates that the fairness ratings to some extent are influenced by the setting in which players take their decisions.
[105] See Table 4 in the Appendix IV.2. for exact values.

towards a self-interested decision maker (Corgnet and Hernán-Gonzalez, 2014; Ong et al., 2012; Mertins, 2010).

We find no differences in transfers between the two *Voice* treatments, which indicates that the opportunity to state one's opinion does not need to be extensive. Even a restricted voice opportunity is sufficient to improve the relationship between the impartial decision maker and the subordinate. Surprisingly, voice towards an impartial decision maker does not seem to increase subordinates' expectations of favorable outcomes, although it does increase perceived influence on the decision. Further, voice leads to positive effects on subjects' fairness perceptions: subjects in our main treatments who can voice their opinion generally perceive procedures and outcomes as fairer and are more willing to accept these decisions. Finally, we do not find that positive effects of voice on subordinates' generous behavior spill over to interactions with an uninvolved charity. They seem to unfold in direct interactions with the impartial decision maker only.

Our results may have important implications for all those situations in which impartial decision makers take actions that affect subordinates. Examples of such settings can be found in courts between the parties and the judge, in administrative procedures between citizens and authorities, in regulated markets between regulators and market participants, or in the institution of the ombudsman. In a broader sense, it applies generally to decision making which is not perceived as being biased by self-interest. This may also include decision making situations in firms and organizations between principals and agents. According to our results, such decision makers may benefit from granting subordinates the possibility to present their views in the decision making process, even if rather unfavorable decisions have to be made.

Of course, important questions remain unanswered. In our setting, only one out of two subordinates has a voice opportunity. This way we can identify pure voice effects. It would be interesting to test voice effects when more than one subordinate may express herself. Further, it remains open how voice is affected by a more powerful impartial decision maker who may grant or deny voice opportunities at her discretion. As already mentioned above, it might also be interesting for future research to explore further the robustness of gender differences in reactions to voice and potential reasons for it. Along these lines, more experimental evidence is needed to understand the importance of voice procedures in economic, social, and legal interactions. Our experimental design may serve as a general framework to address these questions.

VII. REFERENCES

Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman. 2011. "Reference points and effort provision." *American Economic Review*, 101(2): 470–492.

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

Babcock, Linda, and George Loewenstein. 1997. "Explaining bargaining impasse: The role of self-serving biases." *The Journal of Economic Perspectives*, 11(1): 109-126.

Bardsley, Nicholas. 2008. "Dictator game giving. Altruism or artefact?" *Experimental Economics*, 11(2): 122-133.

Brandts, Jordy, and Gary Charness. 2011. "The strategy versus the direct-response method: A first survey of experimental comparisons." *Experimental Economics*, 14(3): 375–398.

Bolton, Garry E., Jordi Brandts, and Axel Ockenfels. 2005. "Fair procedures: Evidence from games involving lotteries." *The Economic Journal*, 115: 1054-1076.

Bolton, Garry E., and Axel Ockenfels. 2000. "A theory of equity, reciprocity, and competition." *American Economic Review*, 90(1): 166–93.

Charness, Gary, and Matthew Rabin. 2005. "Expressed preferences and behavior in experimental games." *Games and Economic Behavior*, 53(2): 151–169.

Cohen-Charash, Yochi, and Paul E. Spector. 2001. "The role of justice in organizations: A meta-analysis." *Organizational Behavior and Human Decision Processes*, 86(2): 278-321.

Corgnet, Brice, and Roberto Hernán-González. 2014. "Don't ask me if you will not listen: The dilemma of participative decision making." *Management Science*, 60(3): 560-585.

Dal Bó, Pedro, Andrew Foster, and Louis Putterman. 2010. "Institutions and behavior: Experimental evidence on the effects of democracy." *American Economic Review*, 100(5): 2205-2229.

Dur, Robert, and Hein Roelfsema. 2010. "Social exchange and common agency in organizations." *Journal of Socio-Economics*, 39(1): 55-63.

Engel, Christoph. 2011. "Dictator games: a meta study." *Experimental Economics*, 14(4): 583-610.

Fehr, Ernst, and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, 114(3): 817-868.

Fischbacher, Urs. 2007. "z-Tree. Zurich Toolbox for Ready-made Economic Experiments." *Experimental Economics*, 10(2): 171-178.

Folger, Robert. 1977. “Distributive and procedural justice: Combined impact of voice and improvement on experienced inequity.” *Journal of Personality and Social Psychology*, 35(2): 108–119.

Forsythe, Robert, Joel L. Horowitz, Nathan E. Savin, and Martin Sefton. 1994. “Fairness in Simple Bargaining Games.” *Games and Economic Behavior*, 6(3): 347–369.

Frey, Bruno S., Marcel Kucher, and Alois Stutzer. 2001. “Outcome, process and power in direct democracy.” *Public Choice*, 107: 271–293.

Frey, Bruno S., and Alois Stutzer. 2005. “Beyond outcomes – Measuring procedural utility.” *Oxford Economic Papers*, 57(1): 90-111.

Frey, Bruno S., Matthias Benz, and A. Stutzer. 2004. “Introducing procedural utility: Not only what, but also how matters.” *Journal of Institutional and Theoretical Economics*, 160(3): 377-401.

Greiner, Ben. 2004. “An Online Recruitment System for Economic Experiments.” In K. Kremer, and V. Macho (eds.), *Forschung und Wissenschaftliches Rechnen.* Gesellschaft für Wissenschaftliche Datenverarbeitung Bericht, Göttingen: Datenverarbeitung, 63: 79–93.

Hirschman, Alberto O. 1970. *Exit, voice and loyalty: Responses to decline in firms, organizations, and states*. Cambridge, MA: Harvard University Press.

Katz, Daniel. 1960. “The functional approach to the study of attitudes.” Public *Opinion Quarterly*, 24(2): 163-204.

Kleine, Marco, Pascal Langenbach, and Lilia Zhurakhovska. 2013. “Fairness and persuasion. Experimental evidence on the influence of suggestions by stakeholders on fairness decisions by impartial decision makers.” *MPI Collective Goods Preprint No. 2014/03.*

Konow, James. 2000. “Fair shares: Accountability and cognitive dissonance in allocation decisions.” *American Economic Review*, 90(4): 1072-1091.

Konow, James. 2003. “Which is the fairest one of all? A positive analysis of justice theories.” *Journal of Economic Literature*, 41(4): 1188-1239.

Kulik, Carol T., E. Allan Lind, Maurice Ambrose, and Robert J. MacCoun. 1996. “Understanding gender differences in procedural and distributive justice.” *Social Justice Research*, 9(4): 351-369.

Lind, E. Allan, Ruth Kanfer, and P. Christopher Earley. 1990. “Voice, control, and procedural justice: Instrumental and noninstrumental concerns in fairness judgments.” *Journal of Personality and Social Psychology*, 59(5): 952-959.

Lind, E. Allan, and Tom R. Tyler. 1988. The *social psychology of procedural justice*. New York: Plenum Press.

McClintock, Charles G., and Wim B. Liebrand. 1988. "”The role of interdependence structure, individual value orientation and other’s strategy in social decision making: A transformational analysis.” *Journal of Personality and Social Psychology*, 55(3): 396-409.

Mertins, Vanessa. 2010. “The value of voice in an experimental labor market.” *Beiträge zur Jahrestagung des Vereins für Socialpolitik: Ökonomie der Familie – Session: Incentives and Contracts*, G10-V2.

Mohlin, Erik, and Magnus Johannesson. 2008. “Communication: Content or relationship? *Journal of Economic Behavior & Organization*, 65(3-4): 409-419.

Nikiforakis, Nikos, Charles N. Noussair, and Tom Wilkening. 2012. “Normative conflict and feuds: The limits of self-enforcement.” *Journal of Public Economics*, 96(9-10): 797-807.

Ong, Qiyan, Yohanes E. Riyanto, and Steven M. Sheffrin. 2012. “How does voice matter? Evidence from the ultimatum game.” *Experimental Economics*, 15(4): 604-621.

Rankin, Frederick W. 2003. “Communication in ultimatum games.” *Economics Letters*, 81(2): 267-271.

Rankin, Frederick W. 2006. “Request and social distance in dictator games.” *Journal of Economic Behavior & Organization*, 60(1): 27-36.

Reuben, Ernesto, and Arno Riedl. 2013. “Enforcement of contribution norms in public good games with heterogeneous populations.” *Games and Economic Behavior*, 77(1): 122-137.

Selten, Reinhard. 1967. „Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments.“ In E. Sauermann (ed.), *Beiträge zur experimentellen Wirtschaftsforschung*, Tübingen: Mohr: 136-168.

Shapiro, Debra L., and Jeanne M. Brett. 1993. “Comparing three processes underlying judgments of procedural justice: A field study of mediation and arbitration.” *Journal of Personality and Social Psychology*, 65: 1167–1177.

Sweeney, Paul D., and Dean B. McFarlin. 1997. “Process and outcome: Gender differences in the assessment of justice.” *Journal of Organizational Behavior*, 18: 83-98.

Thibaut, John W., and Laurens Walker. 1975. *Procedural justice: A psychological analysis*. Hillsdale, NJ and New York: L. Erlbaum.

Tyler, Tom R. 1987. “Conditions leading to value-expressive effects in judgments of procedural justice: A test of four models.” *Journal of Personality and Social Psychology*, 52(2): 333-344.

Tyler, Tom R. 1988. “What Is Procedural Justice?” *Law and Society Review*, 22(1): 103-135.

Tyler, Tom R., and Steven L. Blader. 2000. *Cooperation in groups: Procedural justice, social identity, and behavioral engagement. Essays in Social Psychology*. Philadelphia: Psychology Press.

Tyler, Tom R., Kenneth Rasinski, and Nancy Spodick. 1985. “Influence of voice on satisfaction with leaders.” *Journal of Personality and Social Psychology*, 48(1): 72-81.

van Prooijen, Jan-Willem, Kees van den Bos, and Henk A. M. Wilke. 2004. “Group Belongingness and Procedural Justice: Social Inclusion and Exclusion by Peers Affects the Psychology of Voice.” *Journal of Personality and Social Psychology*, 87(1): 66-79.

Xiao, Erte, and Daniel Houser. 2005. “Emotion expression and punishment.” *Proceedings of the National Academy of Science*, 102(20): 7398-7401.

Xiao, Erte, and Daniel Houser. 2009. “Avoiding the sharp tongue: Anticipated written messages promote fair economic exchange.” *Journal of Economic Psychology*, 30(3): 393-404.

Yamamori, Tetsuo, Kazuhiko Kato, Toshiji Kawagoe, and Akihiko Matsui. 2008. “Voice matters in a dictator game.” *Experimental Economics*, 11(4): 336-343.

Conclusion

This thesis focuses on the impact of institutions, incentives, and communication on decision making in various contexts using laboratory experiments. In particular, how people behave in the presence of third parties is investigated. The introduction of third parties into a situation has three potential effects: it can change the behavioral norms of active players; it can introduce a dilemma between competing behavioral norms for the players; or it can help subjects to behave according to a particular behavioral norm.

One focus of this work involves passive third parties who are affected by the behavioral choices of active players. In particular, it is analyzed how and why people cooperate with each other in a dilemma situation when their cooperation does or does not impose negative externalities on a passive third party. Beliefs about the cooperativeness of active co-players as well as the own desire not to harm passive players are investigated as explanatory factors for subjects' choices both in the presence and in the absence of negative externalities.

Another strand of the thesis concerns third parties in the role of impartial authorities who act non-strategically. It is analyzed how the introduction of an authority into a situation can help subjects behave according to particular norms. In the present thesis, impartial authorities can use sanction mechanisms (in particular cases delayed rewarding or immediate punishing) to enforce norm compliance. Using these mechanisms is costly for the authority and does not lead to any potential (monetary) benefit in the future. One important question is whether the authorities indeed use these mechanisms and whether their behavior is effective in the sense that it can help to improve norm compliance by subordinates.

Furthermore, it is tested whether subjects anticipate the behavior of impartial authorities and therefore change their own behavior to align with what the authority wants them to do. In addition to this, the transparency of sanctioning norms by an impartial authority via communication as a means to facilitate the understanding of norms is investigated. In particular, it is examined whether this transparency can lead to a high level of cooperation from subordinates with less punishment compared to a situation without communication. The thesis also considers communication between subordinates and an impartial authority as a method to involve the subordinate in the decision making process. It is tested whether giving subjects the possibility to inform the authority of their desired distributive norm increases their satisfaction with the outcome of the decision and improves their attitude toward the authority. Especially, it is controlled whether these effects depend on the actual decision of the authority, i.e., if the effects hinge on whether the decision of the authority corresponds to the subordinates' requests.

Answers to the questions presented in the previous paragraphs are relevant to many different fields within economics, law, and psychology. This research can, for example, help to better understand cartel formation or the role of judges. Moreover, this thesis informs the reader of the psychological effects on subjects facing various procedures and administrative orders.

In each chapter of this thesis, an individual study that aims at answering the aforementioned questions is presented. In the following paragraphs the content of the studies is summarized, conclusions are drawn, and potential outlooks for future research are provided. The last two paragraphs provide an overall conclusion of the thesis.

Chapter 1 (with Christoph Engel) analyzes how subjects behave when facing a multi-sided dilemma situation, i.e., a situation in which subjects face several competing behavioral norms. A one-shot prisoner's dilemma in the presence (or absence) of a passive third party, who suffers a loss of cooperation with the active players is used. The data suggest that moral intuition leads to a correct conclusion: Subjects cooperate significantly less with insiders if their cooperation harms outsiders. This result holds if one controls for subjects' beliefs about the cooperativeness of co-players. In a sense, this finding is surprising as by sparing the passive outsider harm one hurts the insider. This finding is in stark contrast to a recent study by Engel and Rockenbach (2011). They show that trying to distance oneself from the outsider, instead of not trying not to harm him, is the main motive behind high levels of cooperation in a public goods experiment.

In the experiment, subjects are put in the roles of firms who can collude and thereby impose harm on the demand side of the market, which is presented by the passive player. One might have doubts as to whether subjects that act anonymously in a laboratory experiment behave as cruelly toward innocent subjects as huge firms managed by experienced CEOs do when an incredible amount of money is on the table and the demand side is not a specific person but an anonymous mass. Nonetheless, the findings of this study are useful from a policy perspective: Antitrust authorities could make the negative effects more salient and hope that this reduces collusion instead of simply increasing audit probabilities and punishment.

One potential critique of the present study is that the chance of subjects managing to avoid harming the passive player is relatively low, i.e., as soon as at least one party cooperates harm is implemented. In a robustness check this argument is addressed by imposing harm only if both active participants cooperate. Surprisingly, the aforementioned effect is even more pronounced in this setting.

Another possible extension of the study would be to change the game to a repeated interaction allowing for a swap in roles. This might be more realistic since in the real world CEOs are also consumers themselves and because cartels can last for a very long time. An even more complicated setting could be implemented in which groups consist of more than two active players and in which subjects can form or reject cartels and, at the same time, be passive players in other groups. All these extensions would help to better understand cartel formation. However, the more complicated a setting is the more difficult it is to interpret behavior and to link it to possible motives. Consequently, the simple presented work is relevant and is a good start to studying subjects' behavior in dilemma-situations in which negative externalities on third parties are involved. To the best of our knowledge, the empirical literature on this topic is scarce.

In chapter 2 (single-authored paper) a different dilemma situation is studied. Again, subjects can try to act according to behavioral norms. In this study, the particular norm for the affected players is trustworthiness. The difficulty for the subjects in this study is not to find out which behavioral norm to follow but rather to form beliefs about whether a third party in the role of an impartial authority is willing to reward compliance with this norm. In other words, subjects must anticipate whether a third party, who has no monetary interest in rewarding them, will do so or if she will rather try to maximize own income (by doing nothing). For this purpose, a trust game followed by a variant of a dictator game with a different group composition is implemented. The trustee in the trust game becomes the receiver in the dictator game. The information provided on the second game is varied between the treatments.

The study shows that the anticipation of monetary reward from an impartial stranger can increase one’s trustworthiness, even though the stranger’s rewards are costly and do not lead to any potential future benefits for her. The findings suggest that allowing for reputation building via public information (e.g., evaluations on Google maps, on Yelp or on other review sides by anonymous customers) can improve people’s behavior and lead to more socially desirable actions, even if the potential future counterparts cannot profit from punishing or rewarding one’s actions.

In the study positive (direct and indirect) reciprocity is analyzed. In most real-life situations one has the possibility to reciprocate the action of others both positively and negatively. In a potential extension, one might allow for a combination of both reciprocity forms. The trust game could be replaced by a moonlighting game (Abbink et al., 2000) in which the trustee can give money to, as well as take money from the investor. In addition to this, the impartial authority in the second game could have the option to use her endowment either to reward or to punish the substitute. In that context, one could investigate in more depth the behavior of subjects when anticipating a potential future reward and punishment from stranger co-players.

In a setting with repeated interaction between strangers and in the absence of clearly defined norms, it is arguably difficult to form correct beliefs about behavioral norms. Thus, if a person wants to comply with the norm, he faces an almost unsolvable task. The next study (chapter 3 with Christoph Engel) investigates how an impartial authority that has the possibility to pair monetary punishment with justifications succeeds in enforcing norm compliance under such conditions. To answer this question under controlled conditions, a multi-period public goods experiment with central punishment by an authority is conducted. The authority is impartial, i.e., she does not benefit from contributions to the public good and punishment is costly for her. Along with the punishment decisions she has to write justifications for her decisions. Treatments vary in how public the justifications are, i.e., whether justifications are communicated to active players and, if so, whether one receives the justifications for own punishment or for all group members in a particular period. Despite the incentive structure, authorities punish the active players for low contributions. No antisocial punishment takes place.

Whenever reasons are communicated, there is less monetary punishment while contributions are on average on a high level in all conditions.

The results lead to the following conclusions: First, reasons either serve as a verbal punishment and therefore as a substitute for monetary punishment or to increase the understanding of norms. Second, even authorities who are unaffected in monetary terms by public good contributions feel the need to intervene when observing injustice. These findings are important for designers of institutions. Accompanying direct punishment with justifications can be time-consuming and thus costly. Nonetheless, justification costs (e.g., opportunity costs of time) are often lower than the costs of direct punishment. Using a combination of both punishment forms can turn out to be more efficient. Making justifications publically available can further help to decrease total punishment costs. A justification can serve not only as an explanation of a punishment rule but also as a threat to those who were not punished at a particular point of time. As a consequence, one learns that making, for example, verdicts in court mandatory can reduce crime rates and can therefore be beneficial for the society. However, the results of the present study are also applicable to all institutions and even to the interaction between and within companies on any level.

In the future, it could be interesting to investigate similar settings in a partner matching and to analyze how effective and credible the reasons provided for punishment are if the authority can benefit from the public good. In addition, one could investigate how effective pure communication (without monetary punishment) is. In addition, the effectiveness of a potential shame-effect on an increase of cooperation could be investigated, i.e., whether only being observed by an authority, even if the authority cannot react to the contributions, can influence the behavior of affected subjects.

The study presented in the fourth chapter (with Marco Kleine and Pascal Langenbach) again investigates subjects' behavior in an interaction with an impartial authority in the presence and in the absence of communication. However, in this study it is the subordinate who can communicate with the authority before taking a decision. Here the question answered is whether subordinates behave compliantly with an institution (in the form of an impartial decision maker) when they can contribute to the decision making process by voicing their views. A two-step experiment is used. In the first step two subordinates perform a real-effort task. One subordinate has to solve more tasks and receives a higher piece rate than the second subordinate. After the task has been executed an impartial authority has to allocate the total money produced between the two subordinate players. The authority receives a fixed wage. The treatments differ in that the subordinate with the higher workload and piece rate has the possibility to proffer his opinion of a fair allocation to the authority before the decision is made. In the second step the subordinate with (or without) the opportunity of voice is the sender and the impartial authority is the receiver in a dictator game. The transfers in this game are used as a measure of the subordinate's positive attitude toward and compliance with the authority. The results show that if voice is possible people care about giving their opinion, even if this does not influence their payoff in any way, i.e., transfers in the second step are significantly higher for most possible allocations (and never lower). This is true irrespective of what the subordinate perceives to be a fair allocation. Furthermore, players who are given the opportunity of voice perceive outcomes and procedures as fairer. A robustness check demonstrates that the transfers in the dictator game are indeed a measure of the attitude toward the impartial authority and are not just a measure of general satisfaction.

Previous studies (e.g., Corgnet and Hernán-González, 2014; Ong et al., 2012) show that people use cheap talk to influence the decisions of others and thus their payoffs and react negatively reciprocal if their requests are not answered in the desired way. One possible interpretation of why in the present setting voice has a non-instrumental value is that here subjects are interacting with an impartial authority whose payoffs are not affected by her choice. The results highlight the importance of voice for the design of decision procedures when impartial authorities and subordinates interact. In practice, impartial decision makers such as referees in sports, judges and juries in courts, ombudsmen in public administration and business organizations or editors of journals may rely on voice procedures to get support for their decisions even when unfavorable outcomes have to be pushed through.

In future work, one could investigate the underlying processes that lead to the presented results. In particular, one could investigate how an authority's feelings and decisions are affected when she receives a message after or before taking her decision. Another potentially interesting extension of that project could be to analyze what happens if the subordinate has to pay for voice. Additionally, in the present study only one participant is given voice and the effectiveness to influence the decision has been found to be irrelevant. If voice was given to both subordinates they might regard their requests as being competitive and their beliefs about the effectiveness of own voice compared to the requests by the competing party might matter. Thus, a study on competing voice would be interesting and it would represent many real-world situations.

Overall, the thesis demonstrates how decisions and the behavior of single individuals can be shaped when they have the possibility to interact with each other and if their choices affect not only themselves. It is elaborated on how behavioral norms evolve and which decisions subjects take when facing conflicts between competing norms. In each of the studies described in this thesis, third parties are involved. In the first chapter, the third party is a passive player who suffers harm if two active players cooperate. Even when controlling for subjects' beliefs about the cooperativeness of others, they cooperate less with insiders if this imposes harm on outsiders. In the studies in the remaining part of the thesis the third party has the role of an impartial authority that does not directly benefit in monetary terms from her decisions. Chapter 2 demonstrates that an authority makes use of costly reward to positively reciprocate the trustworthiness of a stranger toward another stranger. The potential recipient of the reward increases his trustworthiness if he can thereby send a signal to the authority that he is norm compliant. Chapter 3 displays that costly punishment by an impartial authority is most effective in enforcing cooperativeness between other subjects when paired with justifications for the punishment. The fourth chapter illustrates that subordinates appreciate being involved in a distributive decision by an impartial authority by receiving voice and that they value this involvement irrespective of the outcome of the decision.

As stated above, the findings described in the present thesis are applicable to a fair number of situations in economic and private life as well as to interactions with courts and administrations in which third parties are involved. In none of the described studies can the third parties make strategic choices in order to gain monetary benefit from their choices. Most economic work is focused on subjects' behavior in situations in which they can at least in the long run monetarily benefit from own choices. If this condition is not met, the direct reactions of subjects to own experiences with others are analyzed. However, in the real world many situations involve third parties that are real outsiders and who nonetheless have an influence on our social and economic life. On the one hand, third parties might not be directly affected by the behavior of others and might nonetheless intervene. On the other hand, third parties might be very much directly influenced by our decisions but might not have the opportunity to take actions. It is important to analyze both these extreme cases in the context of economic decision making. The thesis highlights important motives involved in behavioral decision making and elaborates on several channels through which the behavior of subjects can be influenced. Much more research is needed to investigate these channels in detail. And even more studies are needed to elaborate further influence factors on economically relevant choices.

REFERENCES

Abbink, Klaus, Bernd Irlenbusch, and Elke Renner. 2000. “The moonlighting game. An experimental study on reciprocity and retribution.” *Journal of Economic Behavior & Organization*, 42 (2): 265–277.

Corgnet, Brice, and Roberto Hernán-González. 2014. “Don’t ask me if you will not listen: The dilemma of participative decision making.” *Management Science,* 60(3): 560-585.

Engel, Christoph, and Bettina Rockenbach. 2011. “We Are Not Alone. The Impact of Externalities on Public Good Provision.” *MPI Collective Goods Preprint No. 2009/29.*

Ong, Qiyan, Yohanes E. Riyanto, and Steven M. Sheffrin. 2012. “How does voice matter? Evidence from the ultimatum game.” *Experimental Economics*, 15(4): 604-621.

Appendix

I. Chapter 1

I.1. Instructions

The Instructions for the *Baseline* and the treatments differ only in Part 1 and in Part 2. The rest is identical. Therefore, we report first the full instructions of the *Baseline* and afterwards only Part 1 and Part 2 of the treatment *Small*. The difference between *Small*, *Middle* and *High* lies only in the level of harm for the passive player. Note that in the *Baseline* Part 2, 3 and 4 (and in the treatments Parts 3 and 4) were a post-experimental test of risk and loss aversion and social value orientation. We use neither test for this paper. Recall that subjects were not aware of the content of the subsequent parts when making their choices in the first part (the prisoners' dilemma).

I.1.1. *Baseline*

Welcome to our experiment. Please remain quiet and do not talk to the other participants during the experiment. If you have any questions, please give us a signal. We will answer your queries individually.

Course of Events

The experiment is divided into four parts. We will distribute separate instructions for each of the four parts of the experiment. Please read these instructions carefully and make your decisions only after taking an appropriate amount of time to reflect on the situations, and after we have fully answered any questions you may have. Only when all participants have decided will we move on to the next part of the experiment. All of your decisions will be treated anonymously.

Your Payoff

At the end of the experiment, we will give you your payoff in cash. Each of you will receive the earnings resulting from the decisions you will have made in the course of the experiment. It is possible to make a loss in one part of the experiment. These losses will be subtracted from the earnings in the other parts and from your show-up fee.

Thus:

Total payment =

+ Earnings from Part 1

+ Earnings from Part 1a

+ Earnings from Part 2

+ Earnings from Part 3

+ Earnings from Part 4

+ 10 €

In Part 2, however, losses are possible, too. Should you incur losses, these will be deducted from your earnings from Part 1, Part 3, or Part 4 and from your show-up fee of 10€.

We will explain the details of how your payoff is made up for each of the four parts separately. In each of the four parts, possible payoffs are given in Euro, which is the currency you will be paid in.

<u>Part 1</u>

The basic idea of this part of the experiment is as follows: you are anonymously paired by us with another participant. You and the other participant will make one decision.

We will show you <u>one tables that look as follows</u>:

		Type B	
		Above	*Below*
Type A	*Above*	5€, 5€	0€, 10€
	Below	10€, 0€	2.45€, 2.45€

We will let you know at the start whether you are a Type A or a Type B participant. (You will probably notice that the payments given to both types are symmetrical; the distinction between Type A and Type B is solely for the purpose of explaining the experiment.)

The decisions *Above* or *Below* determine the payoffs to you and the other participant. In each of the four cells of the table, <u>the figure on the left denotes A's profit, while the figure on the right denotes B's profit</u>.

For instance, if Type A chooses the option *Above* and Type B chooses the option *Above*, then both receive a payment of 5€. If Type A chooses *Above* and Type B chooses *Below*, then Type A receives zero profit and Type B gets 10€. The same is valid for a *Below*/*Above* constellation. Finally, if Type A chooses *Below* and Type B chooses *Below*, then both receive a payment of 2.45€.

Let us first begin with some test questions. (The aim of these questions is merely to verify whether all participants have fully understood the instructions. Neither the questions nor the answers have anything to do with your final payment.) Then the screen on which your actual decisions are marked will appear.

Do you have any further questions?

Part 1a

This part of the experiment refers to the previous part where you made eleven decisions, “Above” or “Below”. The number of participants in the roles A and B who participated in this task will be presented to you on the screen. We ask you to estimate how many participants of the experiment selected “Above”. In case you make a precise estimation, you can gain 2€ in addition. If your estimation deviates by +/-2, you still gain 1€ in addition. Otherwise, you gain nothing in addition.

Part 2

The basic idea of this part of the experiment is as follows. In the following, you will be requested to make six decisions. In this part of the experiment, no other participant is paired with you. The payoffs therefore relate only to you. In each of your six decisions, you may therefore choose to play a "lottery" or decline.

What are these "lotteries" then? In these lotteries, a computer-simulated random toss of a coin determines whether you win or lose money. If the coin shows "tails" (i.e., a number), you win 6€; if it is "heads", you lose. How much you lose depends on the particular lottery. Losses vary between 2€ and 7€. If losses occur, they are subtracted from the earnings from the other parts of the experiment at the end of the experiment.

You can accept or refuse these lotteries on an individual basis, just as you can accept or refuse all. If you refuse, you will make no profit and lose nothing, i.e., your payoff will be zero. If you accept, the toss of the coin determines your payoff, as described above.

In the end, one of the six lotteries is randomly chosen, and then the payment is determined according to your decision and the coin throw for this particular lottery. Thus, once again the lot decides twice in a row: first, one of the lotteries is drawn by lot, and then the toss of a coin decides whether or not you win in this lottery – on condition that you have decided to go for the lottery.

Let us first begin with some test questions. (The aim of these questions is merely to verify whether all participants have fully understood the instructions. Neither the questions nor the answers have anything to do with your final payment.) Then the screen on which your actual decisions are marked will appear.

Part 3

This part of the experiment is as follows: one Type X participant has to decide between two situations (1 or 2). His decision influences his own payoff, and the payoff of one other randomly paired Type Y participant, as follows:

Situation 1: Type X receives a payoff, determined by lot, of 5€ or 10€, Type Y receives a payoff of zero Euro. The likelihood with which Type X either receives 5€ or 10€ is systematically varied in the following table. Type X must make a decision for each of the eleven constellations (a total of 11 decisions).

Situation 2 remains the same for all 11 constellations: Type X and Type Y both receive 5€.

In this part, all participants must initially make their decisions in the role of Type X.

We will proceed with the payoff as follows:

- The lot is drawn to determine whether your payments, following your own decisions, classify you as a Type X or a (passive) Type Y. We will draw one half of the group as Type X and the other as Type Y.
- The next draw pairs each Type Y participant with a Type X participant.
- Finally, the third draw determines one single payoff-relevant situation out of the total of eleven situations.

Therefore, one out of the eleven decisions emerges as the basis for payoff. With a probability of ½, it will be your own decision, and with the same likelihood it will be another participant's decision.

Example for Part 3

	Profit	With likelihood of
You	10€	30%
	5€	70%
Other participant	0€	100%
	1	
Your decision		
	2	
Both	5€	100%

As stated above, all participants will make eleven decisions of this kind. Please mark your decision by clicking on the appropriate box.

Part 4

In this part of the experiment, no other participant is paired with you. The payoffs therefore relate only to you. The decisions of the other participants only have an influence on their own respective payoffs.

In this part of the experiment, you are asked to decide in 10 different situations (lotteries) between option A and B. These situations will be presented to you on consecutive screens. The two lotteries each comprise 2 possible monetary payoffs, one high and one low, which will be paid to you with different probabilities.

The options A and B will be presented to you on the screen as in the following example:

The computer uses a random draw program, which assigns you payments exactly according to the denoted probabilities.

For the above example, this means:
Option A obtains a payoff of 2 Euro with a probability of 10% and a payoff of 1.60 Euro with a probability of 90%.
Option B obtains a payoff of 3.85 Euro with a probability of 10% and a payoff of 0.10 Euro with a probability of 90%.
Now you have to click on the particular option you decide for.

Please note that at the end of the experiment only one of the 10 situations will eventually be paid. Yet, each of the situations can be randomly chosen with equal probability to be the payoff-relevant one. After this, a draw will determine whether for the payoff-relevant situation the high payoff (2.00 Euro or 3.85 Euro) or the low payoff (1.60 Euro or 0.10 Euro) will be paid.

I.1.2. TREATMENT *SMALL*

<u>Part 1</u>

The basic idea of this part of the experiment is as follows: you are anonymously paired by us with two other participants. There exist Type A, Type B and Type C players. Type C is passive in that experiment. If you are not Type C, you and one other participant will make one decision.

We will show you eleven tables that look as follows:

		Type B	
		Above	Below
Type A	Above	5€, 5€, -.3€	0€, 10€, -.3€
	Below	10€, 0€, -.3€	2.45€, 2.45€, 0€

We will let you know at the start whether you are a Type A or a Type B participant. (You will probably notice that the payments given to both types are symmetrical; the distinction between Type A and Type B is solely for the purpose of explaining the experiment.)

The decisions Above or Below determine the payoffs to you and the other participants. In each of the four cells of the table, the figure on the left denotes A's profit, while the figure on the right denotes B's profit. Type C receives either –.3€ or 0€, depending on the decisions of Type A and B.

For instance, if Type A chooses the option Above and Type B chooses the option Above, then both receive a payment of 5€ and Type C receives –.3€. If Type A chooses Above and Type B chooses Below, then Type A receives zero profit, Type B gets 10€, and Type C receives –.3€. The same is valid for a Below/Above constellation. Finally, if Type A chooses Below and Type B chooses Below, then both receive a payment of 2.45€ and Type C receives 0€.

Let us first begin with some test questions. (The aim of these questions is merely to verify whether all participants have fully understood the instructions. Neither the questions nor the answers have anything to do with your final payment.) Then the screen on which your actual decisions are marked will appear.

Part 1a

This part of the experiment refers to the previous part where you made eleven decisions, “Above” or “Below”. The number of participants in the roles A and B who participated in this task will be presented to you on the screen. We ask you to estimate how many participants in the roles A and B of the experiment selected “Above”. In case you make a precise estimation, you can gain 2€ in addition. If your estimation deviates by +/-2, you still gain 1€ in addition. Otherwise, you gain nothing in addition.

Part 2

The basic idea of this part of the experiment is similar to the on in the previous part: you are anonymously paired by us with two other participants. Again, there exist Type A, Type B and Type C players. Type C is passive in that experiment. If you are not Type C, you and one other participant will make one decision. You have the same role in this part as in the previous part. Yet, it is impossible that you are paired with the same participants in a group as in the previous part. Note that the difference to the previous part is under which conditions, Type C receives which payoffs.

We will show you eleven tables that look as follows:

		Type B	
		Above	Below
Type A	Above	5€, 5€, -.3€	0€, 10€, 0€
	Below	10€, 0€, 0€	2.45€, 2.45€, 0€

We will let you know at the start whether you are a Type A or a Type B participant. (You will probably notice that the payments given to both types are symmetrical; the distinction between Type A and Type B is solely for the purpose of explaining the experiment.)

The decisions Above or Below determine the payoffs to you and the other participants. In each of the four cells of the table, the figure on the left denotes A's profit, while the figure on the right denotes B's profit. Type C receives either –.3€ or 0€, depending on the decisions of Type A and B.

For instance, if Type A chooses the option Above and Type B chooses the option Above, then both receive a payment of 5€ and Type C receives –.3€. If Type A chooses Above and Type B chooses Below, then Type A receives zero profit, Type B gets 10€, and Type C receives 0€. The same is valid for a Below/Above constellation. Finally, if Type A chooses Below and Type B chooses Below, then both receive a payment of 2.45€ and Type C receives 0€.

Let us first begin with some test questions. (The aim of these questions is merely to verify whether all participants have fully understood the instructions. Neither the questions nor the answers have anything to do with your final payment.) Then the screen on which your actual decisions are marked will appear.

Part 2a

This part of the experiment refers to the previous part where you made eleven decisions, “Above” or “Below”. The number of participants in the roles A and B who participated in this task will be presented to you on the screen. We ask you to estimate how many participants in the roles A and B of the experiment selected “Above”. In case you make a precise estimation, you can gain 2€ in addition. If your estimation deviates by +/-2, you still gain 1€ in addition. Otherwise, you gain nothing in addition.

II. CHAPTER 2

II.1. Instructions

Here, the English translation of the instructions is presented. The original German version can be passed on upon request. The only difference between the *Anticipation* treatment and the *Baseline* is that, in the *Anticipation* treatment, first only the general instructions and the instructions for the trust game are presented and only after the trust game is player, the instructions for the helping game are handed out. On the contrary, in the *Baseline* the full instructions are handed out immediately at the beginning of the experiment. The instructions and the control questions are presented here in the order in which they are read out aloud by the experimenter (and filled out by the participants).

General Instructions for Participants

You are about to take part in economic experiments. Depending on the decisions that you and others make, you can earn a substantial amount of money. It is therefore very important that you read these instructions carefully.

The written statements you have received from us serve your own private information only. **During the experiments, any communication whatsoever is forbidden.** If you have any questions, please ask only us. Please raise your hand and we will come to you. Disobeying these rules will lead to exclusion from the experiments and from all payments.

During the experiments, we speak not of Euro, but instead of Taler. Your entire income is hence initially calculated in Taler. The total number of Taler you earn during the experiments is converted into Euro at the end and paid to you in cash, at the rate of

1 Taler = 2 Eurocent.

In addition, each participant is paid a lump sum of 4 Euro for showing up today.

You will take part in several experiments today. The instructions to each experiment will be handed out to you one by one, just before the respective experiment is about to begin. On the following pages, we will describe the exact procedure of the first experiment.

In this experiment, there are three different roles: Player A, Player B, and Player C. At the beginning of the experiment, you are randomly assigned one of these three roles.

In the experiment, you are required to make your decision once only; i.e., the experiment is conducted only once. You will thus make no repeated decisions today.

Information on the **First Experiment**

In this experiment, one Player A and one Player B are randomly paired. Each player initially receives 100 Taler, which we shall refer to in the following as the initial endowment.

This experiment consists of 2 stages.

Stage 1: In the first stage, **Player A** decides which transfer of X Taler from the initial endowment he or she wishes to send to B. X may be one of the following values: 0, 10, 20, 30, 40, 50, or 60. The Taler sent by Player A are **tripled** and transferred to the Player B who has been assigned to him/her.

Stage 2: If Player A has sent more than 0 Taler in Stage A, **Player B** now has to decide, in Stage 2, how many Taler Y*X he or she wishes to transfer back to A. Player B may choose from the following options: no Taler (Y = 0), the same amount of Taler as in the transfer (Y = 1), double the transferred amount of Taler (Y = 2), or three times the transferred amount of Taler (Y = 3).

The income at the end of the first part of the experiment is therefore:

Player A:

If he/she has sent 0:	100
If he/she has sent X:	100 **-** X + (X * Y)

Player B:

If he/she receives 0:	100
If he/she receives X:	100 + 3 * X - (X * Y)

Here, X may be equal to 0, to 10, to 20, to 30, to 40, to 50, or to 60.

Y may be equal to 0, to 1, to 2, or to 3.

Examples:

1.) If, for example, Player A sends a sum of X = 10 Taler and B decides to send back three times the amount of the transfer (Y = 3), then both players have the following income:

A: 100 - 10 + (10 * 3) = 120, and B: 100 + 3 * 10 - 3 * 10 = 100.

2.) If, for example, Player A sends a sum of X = 60 Taler and B decides to send back the amount of the transfer (Y = 1), then both players have the following income:

A: 100 - 60 + (60 * 1) = 100, and B: 100 + 3 * 60 - 1 * 60 = 220.

3.) If, for example, Player A sends a sum of X = 40 Taler and B decides to send back twice the amount of the transfer (Y = 2), then both players have the following income:

A: 100 - 40 + (40 * 2) = 140, and B: 100 + 3 * 40 - 2 * 40 = 140.

However, only at the end of Stage 2 are you told how high your income is from the experiment.

Player C does not have to make any decision in this experiment. Player C’s payoff in this experiment is 100 Taler.

The experiment ends here. You are then told how high your income is from the experiment. Further experiments follow – **however, it is impossible for you to be assigned once again to a group with the same players**. **Further, you cannot lose your payoffs from the experiments**. Following the final experiment, you will be given a questionnaire. Once you have filled in the questionnaire, you will receive your payoff from us in cash. In order to receive your payoff, please bring all documents you have received from us with you.

Control Questions for Experiment One

1.) Assuming that Player A sends 50 Taler to Player B and Player B sends 50 Taler back to Player A,

   a. How high is Player A's income? ___________
   b. How high is Player B's income? ___________

2.) Assuming that Player A sends 20 Taler to Player B and Player B sends 60 Taler back to Player A,

   a. How high is Player A's income? ___________
   b. How high is Player B's income? ___________

3.) Assuming that Player A sends 0 Taler to Player B,

   a. Can Player B send something back to Player A? ______________
   b. How high is Player A's income? ___________
   c. How high is Player B's income? ___________

4.) How high is Player C's income? _________________

Information on the **Second Experiment**

Participants in this experiment have the same roles as in the first experiment. Hence, this means that a participant who had role A in the first experiment will still have role A in this experiment; a participant who had role B in the first experiment will still have role B in this experiment; and a participant who had role C in the first experiment will still have role C in this experiment. Random assignation is conducted anonymously, and we ensure that **you are not assigned once again to a group with the same players as in the first experiment**.

This second experiment also consists of two stages.

Stage 1: In the first stage, **Player A** receives an endowment of 100 Taler. He or she now has to decide how many of these 100 Taler to send his or her co-player (full numbers between 0 and 100). Every Taler sent is **tripled** and credited to the other player's account.

Whether the co-player is a Player B or a Player C is decided randomly only in Stage 2. However, Player A must make a binding decision in Stage 1 on how many Taler to send if the co-player is a Player B, and how many Taler to send if the co-player is a Player C. In addition, if it is a Player B, Player A may also decide how many Taler to send depending on Player B's behavior in the first experiment. In Stage 1, Player A therefore has to fill in the following seven decision tables for all possible transfers *X* and *Y* * *X*:

This is what the decision tables look like:

1st Screen

If, in Stage 2, my co-player is …

- … a **Player C**, …

  … I will now send _____Taler.

- … a **Player B** who has been sent **0 Taler** in the first experiment, …

  … I will now send _____Taler.

2nd Screen

If, in Stage 2, my co-player is …

- … a **Player B** who has been sent **10 Taler** in the first experiment and…

  … sent back **0** **Taler**, I will now send _____Taler.

  … sent back **10** **Taler**, I will now send _____Taler.

  … sent back **20** **Taler**, I will now send _____Taler.

  … sent back **30** **Taler**, I will now send _____Taler.

3rd Screen

If, in Stage 2, my co-player is …

- …a **Player B** who has been sent **20 Taler** in the first experiment and …

  … sent back **0** **Taler**, I will now send _____Taler.

  … sent back **20** **Taler**, I will now send _____Taler.

  … sent back **40** **Taler**, I will now send _____Taler.

  … sent back **60** **Taler**, I will now send _____Taler.

4th Screen

If, in Stage 2, my co-player is …

- …a **Player B** who has been sent **30 Taler** in the first experiment and …

  … sent back **0** **Taler**, I will now send _____Taler.

  … sent back **30** **Taler**, I will now send _____Taler.

  … sent back **60** **Taler**, I will now send _____Taler.

  … sent back **90** **Taler**, I will now send _____Taler.

5th Screen

If, in Stage 2, my co-player is …

- … a **Player B** who has been sent **40 Taler** in the first experiment and …

… sent back **0 Taler**, I will now send _____Taler.

… sent back **40 Taler**, I will now send _____Taler.

… sent back **80 Taler**, I will now send _____Taler.

… sent back **120 Taler**, I will now send _____Taler.

6th Screen

If, in Stage 2, my co-player is …

- … a **Player B** who has been sent **50 Taler** in the first experiment and …

… sent back **0 Taler**, I will now send _____Taler.

… sent back **50 Taler**, I will now send _____Taler.

… sent back **100 Taler**, I will now send _____Taler.

… sent back **150 Taler**, I will now send _____Taler.

7th Screen

If, in Stage 2, my co-player is …

- … a **Player B** who has been sent **60 Taler** in the first experiment and …

… sent back **0 Taler**, I will now send _____Taler.

… sent back **60 Taler**, I will now send _____Taler.

… sent back **120 Taler**, I will now send _____Taler.

… sent back **180 Taler**, I will now send _____Taler.

You may enter any number between 0 and 100 in each line. It goes without saying that the lines do **not** have to add up to 100 either, for **only that line is decision-relevant which actually corresponds to the situation drawn by lot in Stage 2!** The decisions in the other lines (not drawn) do not influence your payoff. Do please note, however, that while filling in the table you do not yet know if your co-player is a Player C or B (and you do not know, in case a Player B is assigned to you, how the he has behaved in the first part of the experiment). **In each line, you therefore have to consider your decision carefully, for every one can become relevant for you.**

Stage 2: In the second stage, it is decided by draw whether Player A's co-player is a Player B or C. The player who is drawn then receives the amount of Taler according to the corresponding decision table from Stage 1. He or she therefore does not have to make any decision in this second experiment.

The other player – who is not drawn – is given the possibility to increase his or her own income individually, by way of a small task. (Precise instructions for this task will appear later on this player's screen.)

The income from the second experiment is therefore:

Player A:

100 - (what A sent the co-player)

Co-player (B or C, depending on the draw in Stage 2):

3 * (what A sent the co-player)

Player not drawn (B or C, depending on the draw in Stage 2):

Income from the individual small task

Examples:

1.) If Player A should hence decide to enter the following numbers in the second table: top line 1, second line 13, third line 17, bottom line 0; and if the co-player assigned to Player A is a Player B who was sent 10 Taler by his co-player in the first experiment, and who in turn decided to send nothing back, then the payoffs from this second experiment are as follows for Player A, who made the decision here:

100 - 1 = 9. The payoff for the co-player assigned to Player A is: 1 * 3 = 3.

2.) If Player A should decide to enter the following numbers in the first table: top line 99, second line 14; and if the co-player assigned to Player A is a Player C, then the payoffs from this second experiment are as follows for Player A, who has made the decision here:

100 - 99 = 1. The payoff for the co-player assigned to Player A is: 99 * 3 = 297.

The experiment ends here. You will then hear about your payoff from the first and second experiment. Further experiments will follow – **however, it is impossible for you to be assigned once again to a group with the same players**. **Further, you cannot lose your payoffs from the experiments**. Following the final experiment, you will be given a questionnaire. Once you have filled in the questionnaire, you will receive your payoff from us in cash. In order to receive your payoff, please bring all documents you have received from us with you.

Control Questions for Experiment Two

1.) We assume that Player A in Experiment 2 is assigned to a Player B who was sent 10 Taler by Player A in Experiment 1 and sent back 30 Taler to Player A. If Player A fills in the table in Experiment Two in the following way:

> If, in Stage 2, my co-player is…
>
> - … a **Player B** who was sent **10 Taler** in the first experiment and …
>
>   … sent back **0** **Taler**, I will now send __5__Taler.
>
>   … sent back **10** **Taler**, I will now send _17__Taler.
>
>   … sent back **20** **Taler**, I will now send __0__Taler.
>
>   … sent back **30** **Taler**, I will now send __21_Taler.

a. How high is Player A's income from Experiment 2? ____________

b. How high is Player B's income from Experiment 2? ____________

2.) We assume that Player A in Experiment 2 is assigned to a Player C. If Player A fills in the table in Experiment Two in the following way:

> If, in Stage 2, my co-player is…
>
> - … a **Player C**,
>
>   … I will now send __50__Taler.
>
> - … a **Player B** who was sent **0 Taler** in the first experiment,…
>
>   … I will now send __50__Taler.

a. How high is Player A's income from Experiment 2? ____________

b. How high is Player C's income from Experiment 2? ____________

Information on the **Third Experiment**

In this part of the experiment, no other participant is paired with you. The payoffs therefore relate only to you. The decisions of the other participants only have an influence on their own respective payoffs.

In this part of the experiment, you are asked to decide in 10 different situations (lotteries) between option A and B. These situations will be presented to you on consecutive screens. The two lotteries each comprise 2 possible monetary payoffs, one high and one low, which will be paid to you with different probabilities.

The options A and B will be presented to you on the screen, as in the following example:

The computer uses a random draw program, which assigns you payments exactly according to the denoted probabilities.

For the above example, this means:
Option A obtains a payoff of 2 Euro with a probability of 10% and a payoff of 1.60 Euro with a probability of 90%.
Option B obtains a payoff of 3.85 Euro with a probability of 10% and a payoff of 0.10 Euro with a probability of 90%.
Now you have to click on the particular option you decide for.

Please note that, at the end of the experiment, only one of the 10 situations will eventually be paid. Yet, each of the situations can be randomly chosen with equal probability to be the payoff-relevant one.

After this, a draw will determine whether for the payoff-relevant situation the high payoff (2.00 Euro or 3.85 Euro) or the low payoff (1.60 Euro or 0.10 Euro) will be paid.

II.2. Tables

TABLE 5

Pairwise comparison of transfers to active Player B in the helping game per increased return transfer by one for any given possible investment

Wilcoxon signed-rank tests: p-values, |z| in parentheses				
Investment (X)	*Return Transfer (Y)*	*Baseline* & *Anticipation*	*Baseline*	*Anticipation*
X = 10	Y = 0 vs. Y = 1	0.0008, (3.346)	0.0050, (2.810)	0.0546, (1.922)
X = 10	Y = 1 vs. Y = 2	0.0000, (4.886)	0.0007, (3.398)	0.0004, (3.515)
X = 10	Y = 2 vs. Y = 3	0.0002, (3.753)	0.0125, (2.498)	0.0049, (2.811)
X = 20	Y = 0 vs. Y = 1	0.0009, (3.315)	0.0627, (1.861)	0.0050, (2.808)
X = 20	Y = 1 vs. Y = 2	0.0000, (4.434)	0.0025, (3.019)	0.0012, (3.251)
X = 20	Y = 2 vs. Y = 3	0.0007, (3.409)	0.0642, (1.851)	0.0034, (2.928)
X = 30	Y = 0 vs. Y = 1	0.0000, (4.087)	0.0084, (2.636)	0.0018, (3.122)
X = 30	Y = 1 vs. Y = 2	0.0000, (4.581)	0.0007, (3.396)	0.0021, (3.080)
X = 30	Y = 2 vs. Y = 3	0.0000, (4.473)	0.0125, (2.496)	0.0002, (3.733)
X = 40	Y = 0 vs. Y = 1	0.0000, (4.087)	0.0085, (2.633)	0.0018, (3.123)
X = 40	Y = 1 vs. Y = 2	0.0000, (4.966)	0.0007, (3.394)	0.0003, (3.625)
X = 40	Y = 2 vs. Y = 3	0.0000, (4.131)	0.0750, (1.781)	0.0001, (3.908)
X = 50	Y = 0 vs. Y = 1	0.0001, (3.847)	0.0146, (2.441)	0.0030, (2.971)
X = 50	Y = 1 vs. Y = 2	0.0000, (4.748)	0.0037, (2.906)	0.0002, (3.726)
X = 50	Y = 2 vs. Y = 3	0.0000, (4.376)	0.0276, (2.204)	0.0001, (3.824)
X = 60	Y = 0 vs. Y = 1	0.0025, (3.018)	0.0600, (1.881)	0.0171, (2.384)
X = 60	Y = 1 vs. Y = 2	0.0000, (4.591)	0.0052, (2.794)	0.0003, (3.627)
X = 60	Y = 2 vs. Y = 3	0.0003, (3.622)	0.0864, (1.715)	0.0014, (3.202)

The shaded area highlights all comparisons that do **not** lead to a significant effect at the 10-percent level.

TABLE 6

Pairwise comparison of transfers to active Player B vs. to passive players in the helping game for any possible investment and return transfer

Wilcoxon signed-rank tests: p-values, |z| in parentheses							
Investment (X)	*Return Transfer (Y)*	*To active Player B* vs. *to passive Player B (X=0)*			*To active Player B* vs. *to Player C*		
		Baseline & Anticipation	*Baseline*	*Anticipation*	*Baseline & Anticipation*	*Baseline*	*Anticipation*
X = 10	Y = 0	0.0000, (4.282)	0.0080, (2.653)	0.0050, (2.809)	0.0001, (4.041)	0.0018, (3.121)	0.0072, (2.686)
X = 20	Y = 0	0.0000, (4.270)	0.0061, (2.742)	0.0052, (2.792)	0.0000, (4.085)	0.0030, (2.970)	0.0053, (2.787)
X = 30	Y = 0	0.0000, (4.425)	0.0011, (3.266)	0.0026, (3.015)	0.0000, (4.611)	0.0018, (3.122)	0.0007, (3.392)
X = 40	Y = 0	0.0000, (4.394)	0.0011, (3.264)	0.0032, (2.946)	0.0000, (4.611)	0.0030, (2.970)	0.0004, (3.514)
X = 50	Y = 0	0.0000, (4.066)	0.0011, (3.266)	0.0026, (3.015)	0.0000, (4.148)	0.0132, (2.480)	0.0007, (3.392)
X = 60	Y = 0	0.0001, (3.885)	0.0011, (3.263)	0.0045, (2.843)	0.0001, (3.911)	0.0161, (2.408)	0.0015, (3.176)
X = 10	Y = 1	0.0053, (2.791)	0.0168, (2.391)	0.0906, (1.692)	0.0029, (2.974)	0.0748, (1.781)	0.0170, (2.387)
X = 20	Y = 1	0.0058, (2.760)	0.0044, (2.847)	0.2203, (1.226)	0.0032, (2.946)	0.0029, (2.975)	0.1655, (1.387)
X = 30	Y = 1	0.0081, (2.649)	0.0044, (2.850)	0.3075, (1.020)	0.0130, (2.484)	0.0050, (2.810)	0.3412, (0.952)
X = 40	Y = 1	0.0358, (2.099)	0.0230, (2.273)	0.3892, (0.861)	0.0607, (1.875)	0.0252, (2.238)	0.4972, (0.679)
X = 50	Y = 1	0.0529, (1.935)	0.0230, (2.273)	0.5367, (0.618)	0.0679, (1.825)	0.0252, (2.238)	0.5403, (0.612)
X = 60	Y = 1	0.0698, (1.813)	0.0239, (2.259)	0.5703, (0.568)	0.0185, (2.356)	0.0264, (2.220)	0.2357, (1.186)
X = 10	Y = 2	0.3227, (0.989)	0.3785, (0.881)	0.5543, (0.591)	0.5473, (0.602)	0.5087, (0.661)	0.8261, (0.220)
X = 20	Y = 2	0.7258, (0.351)	0.2242, (1.215)	0.6396, (0.468)	0.5026, (0.670)	0.9546, (0.057)	0.3614, (0.913)
X = 30	Y = 2	0.8972, (0.129)	0.6173, (0.500)	0.8150, (0.234)	0.4039, (0.835)	0.7412, (0.330)	0.3911, (0.858)
X = 40	Y = 2	0.9013, (0.124)	0.3887, (0.862)	0.4521, (0.752)	0.4279, (0.793)	0.9094, (0.114)	0.3287, (0.977)
X = 50	Y = 2	0.7783, (0.281)	0.4479, (0.759)	0.3673, (0.902)	0.1415, (1.470)	0.4864, (0.696)	0.1601, (1.405)
X = 60	Y = 2	0.4420, (0.769)	0.7548, (0.312)	0.2798, (1.081)	0.1494, (1.442)	0.3825, (0.873)	0.2177, (1.233)
X = 10	Y = 3	0.0167, (2.393)	0.0993, (1.648)	0.0836, (1.730)	0.0208, (2.311)	0.0333, (2.128)	0.2406, (1.173)
X = 20	Y = 3	0.0018, (3.125)	0.0835, (1.731)	0.0104, (2.563)	0.0089, (2.616)	0.0807, (1.747)	0.0464, (1.992)
X = 30	Y = 3	0.0006, (3.411)	0.0257, (2.230)	0.0094, (2.598)	0.0005, (3.498)	0.0333, (2.128)	0.0055, (2.775)
X = 40	Y = 3	0.0004, (3.510)	0.0257, (2.231)	0.0060, (2.747)	0.0012, (3.244)	0.0779, (1.763)	0.0057, (2.765)
X = 50	Y = 3	0.0007, (3.394)	0.0257, (2.230)	0.0083, (2.638)	0.0001, (3.811)	0.0216, (2.298)	0.0024, (3.039)
X = 60	Y = 3	0.0002, (3.690)	0.0573, (1.901)	0.0022, (3.062)	0.0001, (4.013)	0.0085, (2.632)	0.0025, (3.021)

The shaded area highlights all comparisons that do **not** lead to a significant effect at the 10-percent level.

TABLE 7

Pairwise comparison of transfers to active Players B in the helping game per return transfer by a given investment

Wilcoxon signed-rank tests: p-values, |z| in parentheses				
Investment (X)	*Return Transfer (Y)*	*Baseline & Anticipation*	*Baseline*	*Anticipation*
X = 10 vs. X = 20	Y = 0	0.5638, (0.577)	0.3173, (1.000)	0.9738, (0.033)
X = 10 vs. X = 30	Y = 0	1.0000, (0.000)	1.0000, (0.000)	1.0000, (0.000)
X = 30 vs. X = 40	Y = 0	0.5729, (0.564)	0.3173, (1.000)	0.9738, (0.033)
X = 40 vs. X = 50	Y = 0	0.5458, (0.604)	0.3173, (1.000)	0.9738, (0.033)
X = 50 vs. X = 60	Y = 0	0.5547, (0.591)	0.3173, (1.000)	0.9738, (0.033)
X = 10 vs. X = 20	Y = 1	0.9300, (0.088)	0.3300, (0.974)	0.3697, (0.897)
X = 10 vs. X = 30	Y = 1	1.0000, (0.000)	1.0000, (0.000)	1.0000, (0.000)
X = 30 vs. X = 40	Y = 1	0.3353, (0.963)	0.5275, (0.632)	0.4944, (0.683)
X = 40 vs. X = 50	Y = 1	0.5547, (0.591)	1.0000, (0.000)	0.5455, (0.605)
X = 50 vs. X = 60	Y = 1	0.7264, (0.350)	0.3173, (1.000)	0.9834, (0.021)
X = 10 vs. X = 20	Y = 2	0.2889, (1.061)	0.6717, (0.424)	0.1027, (1.632)
X = 10 vs. X = 30	Y = 2	1.0000, (0.000)	1.0000, (0.000)	1.0000, (0.000)
X = 30 vs. X = 40	Y = 2	0.0608, (1.875)	0.5924, (0.535)	0.0478, (1.9799
X = 40 vs. X = 50	Y = 2	0.1452, (1.457)	0.3547, (0.925)	0.2273, (1.207)
X = 50 vs. X = 60	Y = 2	0.2854, (1.068)	0.2838, (1.072)	0.6218, (0.493)
X = 10 vs. X = 20	Y = 3	0.2689, (1.106)	0.5455, (0.605)	0.1048, (1.622)
X = 10 vs. X = 30	Y = 3	1.0000, (0.000)	1.0000, (0.000)	1.0000, (0.000)
X = 30 vs. X = 40	Y = 3	0.1545, (1.424)	0.3064, (1.023)	0.0179, (2.367)
X = 40 vs. X = 50	Y = 3	0.4364, (0.778)	0.3547, (0.925)	0.7970, (0.257)
X = 50 vs. X = 60	Y = 3	0.8501, (0.189)	0.3805, (0.877)	0.6069, (0.514)

The shaded area highlights all comparisons that do **not** lead to a significant effect at the 10-percent level.

TABLE 8

Pairwise comparison of transfers to active Players B in the helping game per return transfer and per investment by treatment

Mann-Whitney rank-sum tests: p-values, |z| in parentheses		
Investment (X)	*Return Transfer (Y)*	*Baseline vs. Anticipation*
X = 10	Y = 0	0.7333, (0.341)
X = 20	Y = 0	0.3346, (0.965)
X = 30	Y = 0	0.9812, (0.024)
X = 40	Y = 0	0.3113, (1.012)
X = 50	Y = 0	0.6242, (0.490)
X = 60	Y = 0	0.3116, (1.012)
X = 10	Y = 1	0.5311, (0.626)
X = 20	Y = 1	0.3373, (0.960)
X = 30	Y = 1	0.3213, (0.992)
X = 40	Y = 1	0.2961, (1.045)
X = 50	Y = 1	0.4058, (0.831)
X = 60	Y = 1	0.5307, (0.627)
X = 10	Y = 2	0.6731, (0.422)
X = 20	Y = 2	0.3165, (1.002)
X = 30	Y = 2	0.5060, (0.665)
X = 40	Y = 2	0.4055, (0.832)
X = 50	Y = 2	0.3594, (0.916)
X = 60	Y = 2	0.4473, (0.760)
X = 10	Y = 3	0.7100, (0.372)
X = 20	Y = 3	0.3164, (1.002)
X = 30	Y = 3	0.3699, (0.897)
X = 40	Y = 3	0.2288, (1.203)
X = 50	Y = 3	0.2430, (1.167)
X = 60	Y = 3	0.2332, (1.192)

The shaded area highlights all comparisons that do **not** lead to a significant effect at the 10-percent level.

TABLE 9

Pairwise comparison of relative return transfers by active Players B in the trust game by a given increase in the investment by 10

Wilcoxon signed-rank tests: p-values, |z| in parentheses			
Investment (X)	*Baseline* & *Anticipation*	*Baseline*	*Anticipation*
X = 10 vs. X = 20	0.0839, (1.728)	0.0294, (2.178)	1.0000, (0.000)
X = 10 vs. X = 30	0.3173, (1.000)	0.1573, (1.414)	1.0000, (0.000)
X = 30 vs. X = 40	0.5637, (0.577)	0.3173, (1.000)	0.1573, (1.414)
X = 40 vs. X = 50	0.1005, (1.642)	1.0000, (0.000)	0.0987, (1.651)
X = 50 vs. X = 60	0.5073, (0.663)	0.4644, (0.732)	1.0000, (0.000)

The shaded area highlights all comparisons that do **not** lead to a significant effect at the 10-percent level.

TABLE 10

Pairwise comparison of relative return transfers for a given investment by active Players B in the trust game by treatment

Mann-Whitney rank-sum tests: p-values, |z| in parentheses	
Investment (X)	*Baseline* vs. *Anticipation*
X = 10	0.0022, (3.056)
X = 20	0.0107, (2.551)
X = 30	0.0279, (2.199)
X = 40	0.0927, (1.681)
X = 50	0.4370, (0.605)
X = 60	0.6622, (0.437)

The shaded area highlights all comparisons that do **not** lead to a significant effect at the 10-percent level.

TABLE 11

Possible payoffs for Players B including trust game given the played strategy of helpers in the Anticipation Treatment

Investment (X)	*Return Transfer (Y)*	*Possible Payoff in Trust Game for Player B*	*Mean Help Decision (standard errors in parentheses)*	*Mean Help received (helpers' decision tripled)*	*Total possible Payoff for Player B*
10	0	130	1.27, (4.4584)	3.82	133.82
10	1	129	3.73, (5.4404)	11.18	140.18
10	2	128	11.16, (16.259)	33.48	161.48
10	3	127	14.73, (18.775)	44.18	171.18
20	0	160	1.61, (7.7556)	4.84	164.84
20	1	159	4.64, (9.8408)	13.91	172.91
20	2	158	13.45, (19.069)	40.36	198.36
20	3	157	16.68, (21.851)	50.05	207.05
30	0	190	0.52, (2.1184)	1.57	191.57
30	1	189	5.14, (9.6727)	15.41	204.41
30	2	188	13.45, (19.906)	40.36	228.36
30	3	187	19.05, (25.356)	57.14	244.14
40	0	220	0.70, (2.5479)	2.11	222.11
40	1	219	5.93, (10.500)	17.80	236.80
40	2	218	16.18, (24.779)	48.55	266.55
40	3	217	20.84, (28.356)	62.52	279.52
50	0	250	1.27, (4.9526)	3.82	253.82
50	1	249	6.39, (12.289)	19.16	268.16
50	2	248	16.41, (24.831)	49.23	297.23
50	3	247	20.39, (27.252)	61.16	308.16
60	0	280	1.61, (6.7967)	4.84	284.84
60	1	279	6.39, (13.113)	19.16	298.16
60	2	278	16.34, (24.240)	49.02	327.02
60	3	277	19.95, (25.231)	59.86	336.86

The orange areas highlight the maximal profit a player B could have earned given an investment.

TABLE 12

Possible payoffs for investors in the trust game given the played strategy of trustees

	Baseline			*Anticipation*		
Investment (X)	*Mean Trustees Relative Transfer (Y) (standard errors in parentheses)*	*Mean Trustees Absolut Return (Y*X)*	*Total possible Payoff for investor*	*Mean Trustees Relative Transfer (Y) (standard errors in parentheses)*	*Mean Trustees Absolut Return (Y*X)*	*Total possible Payoff for investor*
0	----	----	100.00	----	----	100.00
10	0.4091, (0.8541)	4.09	94.09	1.27, (1.0320)	12.72	102.72
20	0.6818, (0.6463)	13.63	93.63	1.27, (0.7673)	25.45	105.45
30	0.7727, (0.8125)	23.18	93.18	1.27, (0.7673)	38.18	108.18
40	0.8182, (0.8528)	32.72	92.72	1.18, (0.7327)	47.27	107.27
50	0.8182, (0.8528)	40.90	90.90	0.95, (0.8439)	47.72	97.72
60	0.8636, (0.9902)	51.81	91.81	0.95, (0.8985)	57.27	97.27

The orange areas highlight the maximal profit an investor could have earned in the trust game in the two conditions.

III. Chapter 3

III.1. Instructions

The instructions for the *Baseline* and the other treatments differ only in Step 2 of Part One and in Part Two of the Experiment. The rest is identical. Therefore, we report first the full instructions of the *Baseline* and afterwards only Step 2 of Part One and in Part Two of the Experiment of the other treatments. To make it easier to see the changes across treatments, the parts of the instructions that differ across treatments are shaded her; they were not shaded in the original instructions.

III.1.1. *BASELINE*

General Instructions

In the following experiment, you can earn a substantial amount of money, depending on your decisions. It is therefore very important that you read these instructions carefully.

During the experiment, any communication whatsoever is forbidden. If you have any questions, please ask us. Disobeying this rule will lead to exclusion from the experiment and from all payments.

You will in any case receive 4 € for taking part in this experiment. In the first two parts of the experiment, we do not speak of €, but instead of Taler. Your entire income from these two parts of the experiment is hence initially calculated in Taler. The total number of Taler you earn during the experiment is converted into € at the end and paid to you in cash, at the rate of

1 Taler = 4 Eurocent.

The experiment consists of four parts. We will start by explaining the first part. You will receive separate instructions for the other parts.

Part One of the Experiment

In the first part of the experiment, there are two roles: A and B. Four participants who have the role A form a group. One participant who has the role B is allocated to each group. The computer will randomly assign your role to you at the beginning of the experiment.

On the following pages, we will describe to you the exact procedure of this part of the experiment.

Information on the Exact Procedure of the Experiment

This part of the experiment has two steps. In the first step, role A participants make a decision on contributions to a project. In the second step, the role B participant can reduce the role A participants' income. At the start, each **role A** participant receives **20 Taler**, which we refer to in the following as the **endowment**. **Role B** participants receive 20 points at the start of step 2. We explain below how role B participants may use these points.

Step 1:

In Step 1, **only the four role A participants** in a group make a decision. Each role A member's decision influences the income of all other role A players in the group. The income of player B is not affected by this decision. As a role A participant, you have to decide how many of the 20 Taler you wish to invest in a **project** and how many you wish to keep for yourself.

If you are a **role A** player, **your income** consists of two parts:

(1) the Taler you have kept for yourself (**"income retained from endowment")**

(2) the **"income from the project"**. The income from the project is calculated as follows:

Your income from the project =
0.4 *times* the total sum of contributions to the project

Your **income** is therefore calculated as follows:

(20 Taler – your contribution to the project) + 0.4* (total sum of contributions to the project).

The income **from the project** of all role A group members is calculated according to the same formula, i.e., each role A group member receives the same income from the project. If, for example, the sum of the contributions from all role A group members is 60 Taler, then you and all other role A group members receive an income from the project of 0.4*60 = 24 Taler. If the role A group members have contributed a total of 9 Taler to the project, then you and all other role A group members receive an income from the project of 0.4*9 = 3.6 Taler.

For every Taler that you keep for yourself, you earn an income of 1 Taler. If instead you contribute a Taler from your endowment to your group's project, the sum of the contributions to the project increases by 1 Taler and your income from the project increases by 0.4*1 = 0.4 Taler. However, this also means that the income of all other role A group members increases by 0.4 Taler, so that the total group income increases by 0.4*4 = 1.6 Taler. In other words, the other role A group members also profit from your own contributions to the project. In turn, you also benefit from the other group members' contributions to the project. For every Taler that another group member contributes to the project, you earn 0.4*1 = 0.4 Taler.

Please note that the role B participant cannot contribute to the project and does not earn any income from the project.

Step 2:

In Step 2, **only the role B participant** makes decisions. As role B participant, you may **reduce or maintain** the income of **every** participant in Step 2 by distributing **points**.

At the beginning of Step 2, the four role A participants and the role B participant are told how much each of the role A participants has contributed to the project.

As a role B player, you now have to decide, for **each** of the four role A participants, whether you wish to distribute points to them and, if so, how many points you wish to distribute to them. You are obliged to enter a figure. If you do not wish to change the income of a particular role A participant, please enter 0. Should you choose a number greater than zero, you reduce the income of that particular participant. **For each point that you allocate to a participant, the income of this participant is reduced by 3 Taler.**

The total Taler income of a role A participant from both steps is hence calculated using the following formula:

Income from Step 1 – 3* (sum of *points* received)

Please note that Taler income at the end of Step 2 can also be negative for role A participants. This can be the case if the income-subtraction from points received is larger than the income from Step 1. However, the role B participant can distribute a maximum of 20 points to all four role A members of the group. 20 points are the maximum limit. As a role B participant, you can also distribute fewer points. It is also possible not to distribute any points at all.

If you have role B, please state your reasons for your decision to distribute (or not to distribute) points, and why you distributed a particular number of points, if applicable. In doing this, please try to be factual. Please enter your statement in the corresponding space on your screen. You have 500 characters max. to do this. Please note that, in order to send your statement, you will have to press

“Enter” once each time. As soon as you have done this, you will no longer be able to change what you have written.

The reasons you give will remain confidential. This means that only the experimenter knows them. Of course, the reasons will remain anonymous – the experimenter will therefore not know which of the participants gave what reason.

The income of the role B participant does not depend on the income of the other role A participants, nor on the income from the project. For taking part in the first part of the experiment, he or she receives a fixed payment of

1 €.

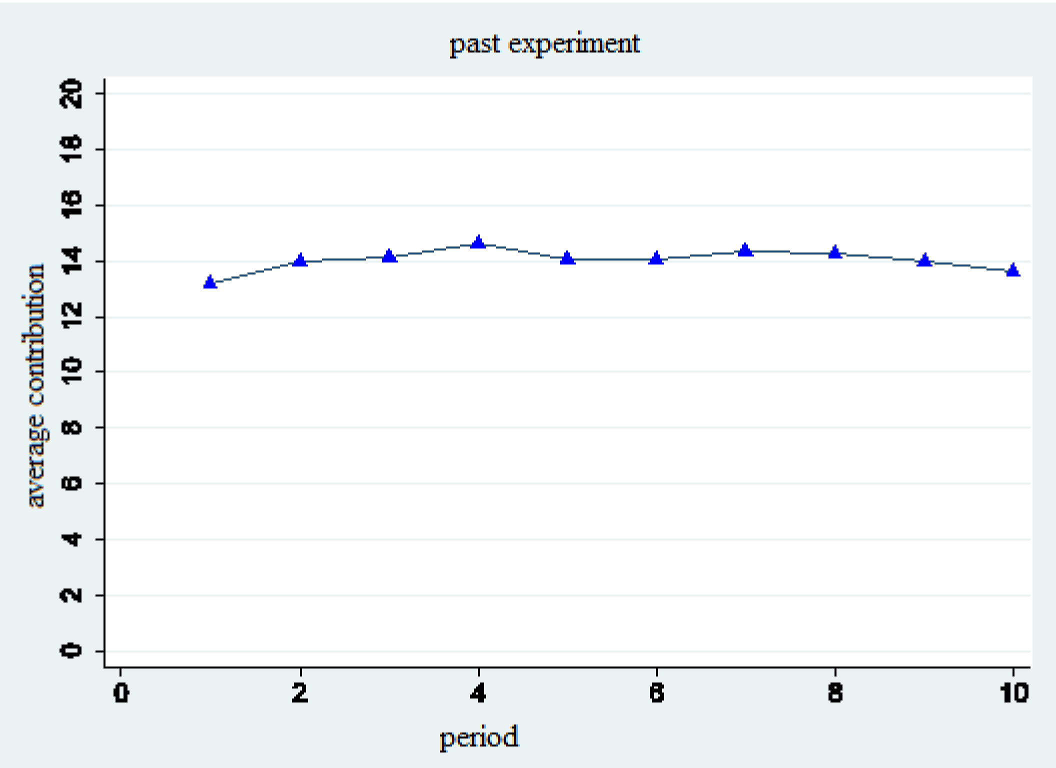
In addition, the role B participant receives the sum of 0.01 € for each point that he or she did not distribute. Once all participants have made their decisions, your screen will show your income for the period and your total income so far.

After this, the first part of the experiment ends. You will then be told what your payment is for this part of the experiment. Hence, you will also know how many points you and all other participants have been given by player B.

Experiences from an Earlier Experiment

For your information, we give you the following graph, which tells you the average contributions made in a very similar experiment that was conducted in this laboratory.

In this experiment, too, there were groups of 4 role A participants and one role B participant each. The role A participants' income was calculated in exactly the same way. The experiment had 10 equal periods. The role B participant also had 20 points at his disposal in each period. At the end of each period, the role A participants were told how much each of the other participants had contributed and how the role B participant had reacted to this.



Part Two of the Experiment

The second part of the experiment consists of 10 repetitions of the first part. **Throughout the entire second part, all participants keep the role they had in the first part of the experiment. The computer randomly re-matches the groups of four in every period. In each period, the computer randomly assigns a role B participant to each group.**

As a reminder:

In each period, each role A participant receives 20 Taler, which may be contributed to the project entirely, in part, or not at all. For each period, calculating the income from the project for the role A participants in a group happens in exactly the same way as it did in the first part of the experiment. In each period, each role B participant receives 20 points, which may be used to reduce the income of the players A in the group. For each point that a role A participant receives in a period, 3 Taler are subtracted. For each point that a role B participant does not use, he or she is given the sum of 0.01 €. In addition to the income from the points retained, each role B participant receives a flat fee of 10 € for participating in this second part of the experiment.

At the beginning of Step 2 of each period, the four role A participants and the role B participant are told how much each of the role A participants contributed to the project.

Please note that the groups are re-matched anew in each period.

After each period, you are told about your individual payoff. You are therefore also informed how many points you and the other participants have been assigned by the role B participant.

Part Three of the Experiment

We will now ask you to make some decisions. In order to do this, **you will be randomly paired with another participant.** In several distribution decisions, you will be able to allocate points to this other participant and to yourself by repeatedly **choosing between two distributions, 'A' and 'B'**. The points you allocate to yourself will be paid out to you at the end of the experiment at a rate of **500 points = 1 €**. At the same time, you are also randomly assigned to **another** participant in the experiment, who is, in turn, also able to allocate points to you by choosing between distributions. This participant is **not the same participant** as the one to whom you have been allocating points. The points allocated to you are also credited to your account. The **sum** of all points you have allocated to yourself and those allocated to you by the other participant are paid out to you at the end of the experiment at a rate of 500 points = 1 €.

Please note that the participants assigned to you in this part of the experiment are **not the members of your group** from the preceding part of the experiment. You will therefore be dealing with other participants.

The individual decision tasks will look like this:

Possibility A:		**Possibility B:**	
Your points	The points of the experiment participant allocated to you	Your points	The points of the experiment participant allocated to you
0	500	304	397

A	B

In this example: If you click 'A', you give yourself 0 points and 500 points to the participant allocated to you. If you click 'B', you give yourself 304 points and 397 points to the participant allocated to you.

Part Four of the Experiment

In this part of the experiment, you **do not form a pair** with another participant. Your decisions are therefore only significant to you and **only influence your own payoff**. The other participants' decisions only influence their own payoffs.

In this part of the experiment, you are requested to decide, **in 10 different cases (lotteries)** between **Option a and Option b**. Both options consist of **two possible payments** (one high and one low), which are paid with varying possibilities.

Options a and b are presented to you on your screen, as in the following example:

Lottery	**Option a**	**Option b**	**Your decision**
1	2.00 Euro with a chance of 10%, or	3.85 Euro with a chance of 10%, or	Option a □
	1.60 Euro with a chance of 90%	0.10 Euro with a chance of 90%	Option b □

The computer will ensure that these payments occur with exactly the possibilities that have been indicated.

For the above example, this means:

If option a is chosen, the winnings of 2 € have a 10 % chance of occurring, and the winnings of 1.60 € have a 90 % chance of occurring.

If option b is chosen, the winnings of 3.85 € have a 10 % chance of occurring, and the winnings of 0.10 € have a 90 % chance of occurring.

In the right-hand column, please indicate which option you would like to choose.

Please note that at the end of the experiment **only one** of the 10 cases becomes relevant for your payment. All cases are **equally possible**. The computer will randomly choose **one payment-relevant case**.

After this, the computer determines, for the payment-relevant case and with the possibilities indicated above, whether the higher (2 € or 3.85 €) or the lower winnings (1.60 € or 0.1 €) will be paid to you.

III.1.2. Treatment *Private*

<u>Step 2:</u>

In Step 2, **only the role B participant** makes decisions. As role B participant, you may **reduce or maintain** the income of **every** participant in Step 2 by distributing **points**.

At the beginning of Step 2, the four role A participants and the role B participant are told how much each of the role A participants has contributed to the project.

As a role B player, you now have to decide, for **each** of the four role A participants, whether you wish to distribute points to them and, if so, how many points you wish to distribute to them. You are obliged to enter a figure. If you do not wish to change the income of a particular role A participant, please enter 0. Should you choose a number greater than zero, you reduce the income of that particular participant. **For each point that you allocate to a participant, the income of this participant is reduced by 3 Taler.**

The total Taler income of a role A participant from both steps is hence calculated using the following formula:

Income from Step 1 – 3* (sum of *points* received)

Please note that Taler income at the end of Step 2 can also be negative for role A participants. This can be the case if the income-subtraction from points received is larger than the income from Step 1. However, the role B participant can distribute a maximum of 20 points to all four role A members of the group. 20 points are the maximum limit. As a role B participant, you can also distribute fewer points. It is also possible not to distribute any points at all.

If you have role B, please state your reasons for your decision to distribute (or not to distribute) points, and why you distributed a particular number of points, if applicable. In doing this, please try to be factual. Please enter your statement in the corresponding space on your screen. You have 500 characters max. to do this. Please note that, in order to send your statement, you will have to press “Enter” once each time. As soon as you have done this, you will no longer be able to change what you have written.

Each role A participant is informed of the reasons that you have given him/her for your decision. Of course, the reasons will remain anonymous – neither the experimenter nor the participants will therefore know which of the participants gave what reason.

The income of the role B participant does not depend on the income of the other role A participants, nor on the income from the project. For taking part in the first part of the experiment, he or she receives a fixed payment of

1 €.

In addition, the role B participant receives the sum of 0.01 € for each point that he or she did not distribute. Once all participants have made their decisions, your screen will show your income for the period and your total income so far.

After this, the first part of the experiment ends. You will then be told what your payment is for this part of the experiment. Hence, you will also know how many points you and all other participants have been given by player B.

In addition, you will be told player B's reason for distributing whatever amount of points you got. This information goes only to you. The other players do not know this reason. They are only aware of the reasons they have been given for their own allocation of points.

Part Two of the Experiment

The second part of the experiment consists of 10 repetitions of the first part. **Throughout the entire second part, all participants keep the role they had in the first part of the experiment. The computer randomly re-matches the groups of four in every period. In each period, the computer randomly assigns a role B participant to each group.**

As a reminder:

In each period, each role A participant receives 20 Taler, which may be contributed to the project entirely, in part, or not at all. For each period, calculating the income from the project for the role A participants in a group happens in exactly the same way as it did in the first part of the experiment. In each period, each role B participant receives 20 points, which may be used to reduce the income of the players A in the group. For each point that a role A participant receives in a period, 3 Taler are subtracted. For each point that a role B participant does not use, he or she is given the sum of 0.01 €. In addition to the income from the points retained, each role B participant receives a flat fee of 10 € for participating in this second part of the experiment.

At the beginning of Step 2 of each period, the four role A participants and the role B participant are told how much each of the role A participants contributed to the project.

Please note that the groups are re-matched anew in each period.

After each period, you are told about your individual payoff. You are therefore also informed how many points you and the other participants have been assigned by the role B participant.

In addition, you will be told player B's reason for distributing whatever amount of points you got. This information goes only to you. The other players do not know this reason. They are only aware of the reasons they have been given for their own allocation of points.

III.1.3. TREATMENT *PUBLIC*

<u>Step 2:</u>

In Step 2, **only the role B participant** makes decisions. As role B participant, you may **reduce or maintain** the income of **every** participant in Step 2 by distributing **points**.

At the beginning of Step 2, the four role A participants and the role B participant are told how much each of the role A participants has contributed to the project.

As a role B player, you now have to decide, for **each** of the four role A participants, whether you wish to distribute points to them and, if so, how many points you wish to distribute to them. You are obliged to enter a figure. If you do not wish to change the income of a particular role A participant, please enter 0. Should you choose a number greater than zero, you reduce the income of that particular participant. **For each point that you allocate to a participant, the income of this participant is reduced by 3 Taler.**

The total Taler income of a role A participant from both steps is hence calculated using the following formula:

Income from Step 1 – 3* (sum of *points* received)

Please note that Taler income at the end of Step 2 can also be negative for role A participants. This can be the case if the income-subtraction from points received is larger than the income from Step 1. However, the role B participant can distribute a maximum of 20 points to all four role A members of the group. 20 points are the maximum limit. As a role B participant, you can also distribute fewer points. It is also possible not to distribute any points at all.

If you have role B, please state your reasons for your decision to distribute (or not to distribute) points, and why you distributed a particular number of points, if applicable. In doing this, please try to be factual. Please enter your statement in the corresponding space on your screen. You have 500 characters max. to do this. Please note that, in order to send your statement, you will have to press “Enter” once each time. As soon as you have done this, you will no longer be able to change what you have written.

All reasons are told to all role A participants in the group. Of course, the reasons shall remain anonymous – neither the experimenter nor the participants will therefore know which of the participants gave what reason.

The income of the role B participant does not depend on the income of the other role A participants, nor on the income from the project. For taking part in the first part of the experiment, he or she receives a fixed payment of

1 €.

In addition, the role B participant receives the sum of 0.01 € for each point that he or she did not distribute. Once all participants have made their decisions, your screen will show your income for the period and your total income so far.

After this, the first part of the experiment ends. You will then be told what your payment is for this part of the experiment. Hence, you will also know how many points you and all other participants have been given by player B.

In addition, you will be told player B's reasons for distributing whatever amount of points you and the other participants got. The other players also know these reasons.

Part Two of the Experiment

The second part of the experiment consists of 10 repetitions of the first part. **Throughout the entire second part, all participants keep the role they had in the first part of the experiment. The computer randomly re-matches the groups of four in every period. In each period, the computer randomly assigns a role B participant to each group.**

As a reminder:

In each period, each role A participant receives 20 Taler, which may be contributed to the project entirely, in part, or not at all. For each period, calculating the income from the project for the role A participants in a group happens in exactly the same way as it did in the first part of the experiment. In each period, each role B participant receives 20 points, which may be used to reduce the income of the players A in the group. For each point that a role A participant receives in a period, 3 Taler are subtracted. For each point that a role B participant does not use, he or she is given the sum of 0.01 €. In addition to the income from the points retained, each role B participant receives a flat fee of 10 € for participating in this second part of the experiment.

At the beginning of Step 2 of each period, the four role A participants and the role B participant are told how much each of the role A participants contributed to the project.

Please note that the groups are re-matched anew in each period.

After each period, you are told about your individual payoff. You are therefore also informed how many points you and the other participants have been assigned by the role B participant.

In addition, you will be told player B's reasons for distributing whatever amount of points you and the other participants got. The other players also know these reasons.

IV. Chapter 4

IV.1. Instructions

The instructions for the *Baseline* and the *Voice* treatments only differ in one regard. In the first experiment, the *Baseline* consists only of Steps 1 and 2. In the *Voice* treatments, an additional intermediate Step between these two is introduced. Therefore, we report first the full instructions of the *Baseline* and afterwards only the new Step 2 of the *Voice* treatments.

IV.1.1. *BASELINE*

General Instructions for Participants

Please begin by reading these instructions carefully. **Communication during the experiments is prohibited.** If you have any questions, please raise your hand. We will then come to you. **Disobeying this rule will lead to exclusion from the experiment and all payments.**

The experiments are conducted anonymously, i.e., nobody is told with which other participant he or she has interacted. The analysis of the experiment results will also be conducted anonymously.

You will take part in several experiments today. You can earn money during the experiments, depending on the decisions you and the other participants make. In the first experiment, we speak not of €, but of Taler. Your income from this experiment is therefore initially calculated in Taler. At the end of the experiment, the Taler earned are **converted into Euro at a rate of 2 Taler = 1 Cent and paid out to you**. In addition, each participant receives a lump sum payment of 4 Euro for showing up today.

The instructions for the individual experiments will be handed out to you just before each respective experiment. On the following pages, we will first describe the exact procedure of the first experiment.

Then, there will be more experiments. It will be **impossible** for you to lose your earnings from one of the experiments in a later experiment.

After the final experiment, you will be given a questionnaire. Once you have filled in this questionnaire, the total sum you have earned will be paid to you in cash.

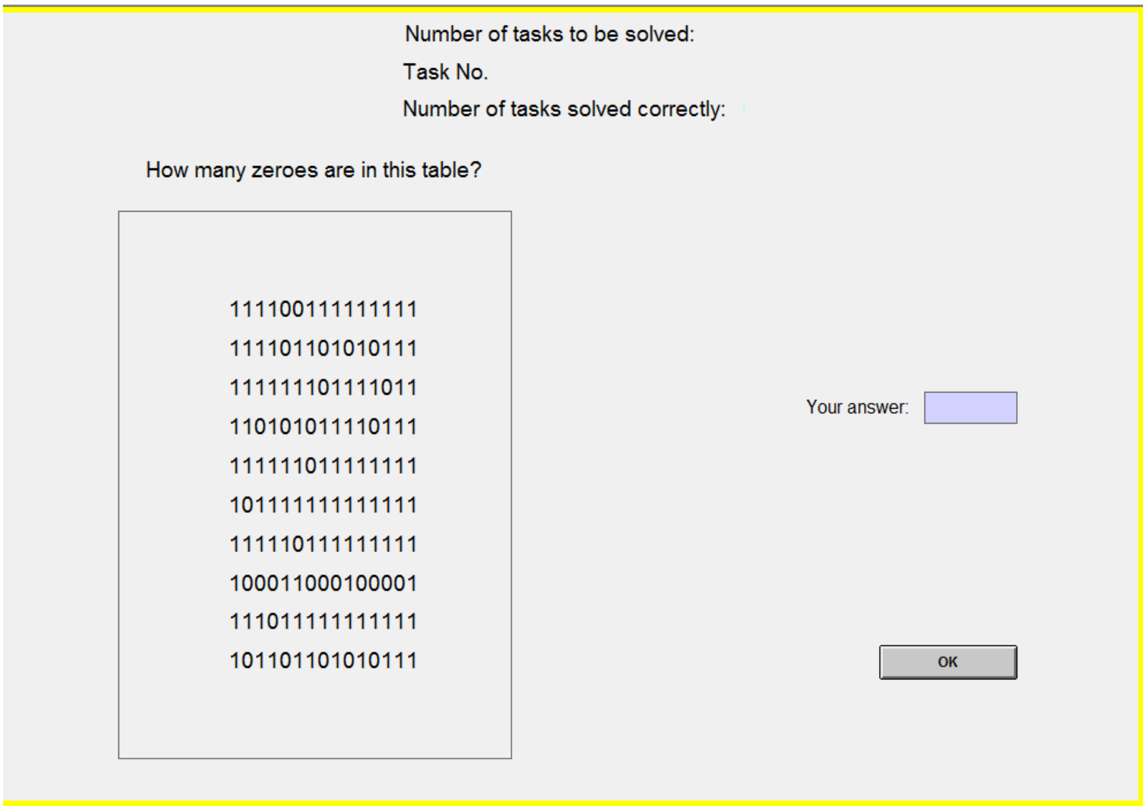
Information on the First Experiment: Part 1

In this experiment, there are three roles: **A**, **X**, and **Y**. At the beginning of the experiment, you are assigned a role **at random**. One participant A, one participant X, and one participant Y form a **group** in this experiment. In this part of the experiment, participant **A** receives a **fixed lump sum of 5 Euro**, which remains unaltered regardless of the decisions taken by A or the other participants. The earnings of participants **X and Y are determined** by the **decisions** made in the course of the experiment. We shall now explain how exactly this works.

This experiment consists of **several parts**. First, we explain and conduct the **first part of the experiment**. You will receive further information separately for the other parts. **Here it is also impossible for you to lose what you have earned in a previous part of the experiment**.
The first part of the experiment consists of **two** steps.

Step 1:

Participants **X** and **Y** **each solve** a predetermined number of **tasks**. Each task consists of determining the correct amount of zeroes in a table consisting of the numbers 0 and 1. If an incorrect number is given, the participant has up to two more attempts to find the correct number. If the number given is still incorrect after three attempts, the task is considered not completed, and the participant is given a new task. The format of the table (i.e., the number of lines and columns) is the same for all tasks and participants. The tasks are presented to participants X and Y on the screens, as in the following example:



The respective participant is shown new tasks until the predetermined number of tasks that are to be solved has been reached. The number of tasks to be solved and the Taler earned per task correctly solved **are different** for participant **X** and participant **Y**, as the following table shows:

	Number of tasks to be solved correctly	Taler earned per task correctly solved
Participant **X**	**12**	**150**
Participant **Y**	**4**	**50**

In total, thus, participants X and Y together accumulate 2000 Taler in this step. The Taler earned are added up.

In the second step, participant **A** will determine the **definitive distribution** of the 2000 Taler amongst the participants X and Y. Participant A does not solve any tasks.

At the end of this step, all participants state which Taler distribution amongst participants X and Y they would consider fair. 100-Taler increments are possible here. Each participant hence indicates one of the following distributions:

- X receives 2.000 Taler, Y receives 0 Taler
- X receives 1.900 Taler, Y receives 100 Taler
- X receives 1.800 Taler, Y receives 200 Taler
- X receives 1.700 Taler, Y receives 300 Taler
- X receives 1.600 Taler, Y receives 400 Taler
- X receives 1.500 Taler, Y receives 500 Taler
- X receives 1.400 Taler, Y receives 600 Taler
- X receives 1.300 Taler, Y receives 700 Taler
- X receives 1.200 Taler, Y receives 800 Taler
- X receives 1.100 Taler, Y receives 900 Taler
- X receives 1.000 Taler, Y receives 1.000 Taler
- X receives 900 Taler, Y receives 1.100 Taler
- X receives 800 Taler, Y receives 1.200 Taler
- X receives 700 Taler, Y receives 1.300 Taler
- X receives 600 Taler, Y receives 1.400 Taler
- X receives 500 Taler, Y receives 1.500 Taler
- X receives 400 Taler, Y receives 1.600 Taler
- X receives 300 Taler, Y receives 1.700 Taler
- X receives 200 Taler, Y receives 1.800 Taler
- X receives 100 Taler, Y receives 1.900 Taler
- X receives 0 Taler, Y receives 2.000 Taler

Please note: **This information is not shown to any other participant and has no consequences on the payoffs** – neither on the own payoffs nor on those of the other participants.

Step 2:

Participant **A now decides how to distribute fairly** among participants X and Y the Taler earned by these two participants.

100-Taler increments are possible here. Participant **A** hence opts for one of the following distributions:

- X receives 2.000 Taler, Y receives 0 Taler
- X receives 1.900 Taler, Y receives 100 Taler
- X receives 1.800 Taler, Y receives 200 Taler
- X receives 1.700 Taler, Y receives 300 Taler
- X receives 1.600 Taler, Y receives 400 Taler
- X receives 1.500 Taler, Y receives 500 Taler
- X receives 1.400 Taler, Y receives 600 Taler
- X receives 1.300 Taler, Y receives 700 Taler
- X receives 1.200 Taler, Y receives 800 Taler
- X receives 1.100 Taler, Y receives 900 Taler
- X receives 1.000 Taler, Y receives 1.000 Taler
- X receives 900 Taler, Y receives 1.100 Taler
- X receives 800 Taler, Y receives 1.200 Taler
- X receives 700 Taler, Y receives 1.300 Taler
- X receives 600 Taler, Y receives 1.400 Taler
- X receives 500 Taler, Y receives 1.500 Taler
- X receives 400 Taler, Y receives 1.600 Taler
- X receives 300 Taler, Y receives 1.700 Taler
- X receives 200 Taler, Y receives 1.800 Taler
- X receives 100 Taler, Y receives 1.900 Taler
- X receives 0 Taler, Y receives 2.000 Taler

This distribution by participant **A** determines the earnings of participants **X** and **Y** in this part of the experiment.

The first part of this experiment ends with participant A making the decision described above.

Participants **X** and **Y** are told about the distribution decided upon by participant **A** and about their earnings from the first part of the experiment **after the end of this experiment**.

It is impossible for you to lose, in a later part of the experiment, the earnings you have accumulated in the first part of the experiment.

You will now be shown some control questions on your screen. After you have answered these questions correctly, the first experiment will begin.

Information on the First Experiment: Part 2

The participants in this part of the experiment have the same roles as in the first part of the experiment. This means that a participant who had role A in the first part of the experiment will also have role A in this part of the experiment; a participant who had role X in the first part of the experiment will also have role X in this part of the experiment; and a participant who had role Y in the first part of the experiment will also have role Y in this part of the experiment. The constellation of the groups also remains the same as in the first part of the experiment. This means participants are always allocated to the same two participants as in the first part of the experiment.

In this part of the experiment, participant **X** receives an endowment of **1000 Taler**. The participant decides how many of these 1000 Taler to send to participant **A** (any full number between 0 and 1000). Each Taler sent is credited to participant A.

Participant X can make the decision on how many Taler to send **dependent on every possible Taler distribution chosen by participant A in the first part of the experiment**. The actual distribution from the first part of the experiment is told to participant **X** only after the experiment has ended. However, X has to decide in **this** part of the experiment how many Taler he or she wishes to send **for every possible distribution**.

In this part of the experiment, participant **X** must therefore fill in the following decision table:

For every possible case, you can send participant A between 0 and 1000 Taler from your endowment. **In case participant A has chosen the following distribution,...**

You receive 2.000 Taler,	Y receives 0 Taler:	you send now	Taler.
You receive 1.900 Taler,	Y receives 100 Taler:	you send now	Taler.
You receive 1.800 Taler,	Y receives 200 Taler:	you send now	Taler.
You receive 1.700 Taler,	Y receives 300 Taler:	you send now	Taler.
You receive 1.600 Taler,	Y receives 400 Taler:	you send now	Taler.
You receive 1.500 Taler,	Y receives 500 Taler:	you send now	Taler.
You receive 1.400 Taler,	Y receives 600 Taler:	you send now	Taler.
You receive 1.300 Taler,	Y receives 700 Taler:	you send now	Taler.
You receive 1.200 Taler,	Y receives 800 Taler:	you send now	Taler.
You receive 1.100 Taler,	Y receives 900 Taler:	you send now	Taler.
You receive 1.000 Taler,	Y receives 1.000 Taler:	you send now	Taler.
You receive 900 Taler,	Y receives 1.100 Taler:	you send now	Taler.
You receive 800 Taler,	Y receives 1.200 Taler:	you send now	Taler.
You receive 700 Taler,	Y receives 1.300 Taler:	you send now	Taler.
You receive 600 Taler,	Y receives 1.400 Taler:	you send now	Taler.
You receive 500 Taler,	Y receives 1.500 Taler:	you send now	Taler.
You receive 400 Taler,	Y receives 1.600 Taler:	you send now	Taler.
You receive 300 Taler,	Y receives 1.700 Taler:	you send now	Taler.
You receive 200 Taler,	Y receives 1.800 Taler:	you send now	Taler.
You receive 100 Taler,	Y receives 1.900 Taler:	you send now	Taler.
You receive 0 Taler,	Y receives 2.000 Taler:	you send now	Taler.

Continue

Participant **X** may enter any full number between 0 and 1000 in every line. **The only line that becomes payoff-relevant is the one that corresponds to the distribution actually chosen by A in the first part of the experiment.** The decisions in the other lines do not influence the participants' payments. The lines do not have to add up exactly to 1000 either, as the only decision-relevant line is the one that corresponds to the actual situation.

As participant **X**, please note that at the time of filling in the table you do not yet know which decision participant A has made in the first part of the experiment. **You therefore have to consider your decision very carefully in every line, because any of the lines could become payoff-relevant for you.**

After filling in the table, participants X and Y will be given a brief questionnaire.

Participant **Y** makes no decisions in this part of the experiment and receives **no earnings**. Participant **A** makes no decisions either in this part of the experiment.

The **earnings** from this experiment add up as follows:

Participant **A**	Taler sent by X to A (in the payoff-relevant situation)
Participant **X**	1.000 Taler **-** Taler sent by X to A (in the payoff-relevant situation)

After filling in the questionnaire, the second part of the experiment ends.

Only after the experiment has ended are participants given information on the **actual distribution decision** of participant **A** in the first part of the experiment and on the sum corresponding to this distribution, which participant **X** has sent to participant A in this part of the experiment.

It is impossible for you to lose, in a later part of the experiment, the earnings you have accumulated in the first and second part of the experiment.

You will now be shown some control questions on your screen. After you have answered these questions correctly, the second part of the experiment will begin.

Information on the First Experiment: Part 3

The participants in this part of the experiment have the same roles and the same group as in the first two parts of the experiment.

In this part, participants **X** and **Y** estimate which distribution they think participant **A** has opted for in the first part of the experiment.

Each participant who correctly estimates the exact distribution chosen by participant A receives 200 Taler. If the estimate is incorrect, albeit straying merely by one decision possibility from participant A's actual decision, then the participant earns 50 Taler. A deviation by one decision possibility means that participant A actually gave participant X 100 Taler more (or less) and participant Y 100 Taler less (or more) than estimated. If the estimate deviates even more from participant A's actual decision, then the participant receives no further earnings from the estimate. The earnings from this part of the experiment are hence calculated in the following manner:

Possible scenario	**Earnings from Estimate**
Correct estimate	200 Taler
Overrated X's earnings by 100 Taler and underrated Y's earnings by 100 Taler	50 Taler
Underrated X's earnings by 100 Taler and overrated Y's earnings by 100 Taler	50 Taler
Stronger deviation (more than 100 Taler) from the estimate	0 Taler

In this part of the experiment, the earnings of a participant who makes an estimate depend only on the correct estimate. No participant can influence in any way, in this part of the experiment, the earnings of another participant. The earnings of a participant in this part of the experiment do not depend on the earnings of another participant in this part of the experiment.

In this part of the experiment, participant **A** makes no decision and receives **no earnings**.
The first experiment ends after these decisions.

All participants are then told the decisions made by the members of their group **that are payoff-relevant to them, as well as their earnings from the individual parts of the experiment.** After this, we would then ask you please to fill in a **brief questionnaire**.

A further experiment will follow. **Here, it will not be possible for participants to interact once again with the same participants from the first experiment**. Further, as before, participants will not be able to lose their earnings from previous experiments in the following experiments.

Information on the Second Experiment

In the following, we would ask you please to make your own decisions. In order to do this, you will be **randomly matched with another participant.** In several distribution decisions, you can give this other participant and yourself points. For this to happen, you will have to choose repeatedly **between two distributions, "A" and "B"**. The points you give yourself are paid out to you at the end of the experiment, at a rate of **250 points = 1 €**. At the same time, you are also randomly matched with **another** experiment participant, who also distributes points to you by choosing distributions. This participant is **not the same participant** as the one to whom you may distribute points. The points given to you are also credited to your account. The **sum** of all points you allocate to yourself and those allocated to you is paid out to you at the end of the experiment, at a rate of 250 points = 1 €.

Please note that the participants matched with you in this part of the experiment are **no members of your group** from the preceding part of the experiment. You are hence matched with other participants in this case.

The individual decision tasks will look like this:

Possibility A:		**Possibility B:**	
Your points	The points of the experiment participant allocated to you	Your points	The points of the experiment participant allocated to you
0	500	304	397

A B

In this example: If you clicked "A", you would give yourself 0 points and 500 points to the experiment participant allocated to you. If you clicked "B", you would give yourself 304 points and 397 points to the experiment participant allocated to you.

Subsequently we will ask you please to fill in some questionnaires. While you do this, we will prepare your payments.

IV.1.2. *VOICE* MANIPULATION – ADDITIONAL STEP

In Step 2, participant **X** has the chance to send participant **A** a **message**. If you are a participant **X**, please follow the instructions about this on your screen. The participants **Y** and **A** have no possibility to send a message.

IV.1.3. Treatments *Baseline-Uninvolved* and *Narrow-Uninvolved*

The instructions of the both uninvolved treatments only differ in the second part of the experiment:

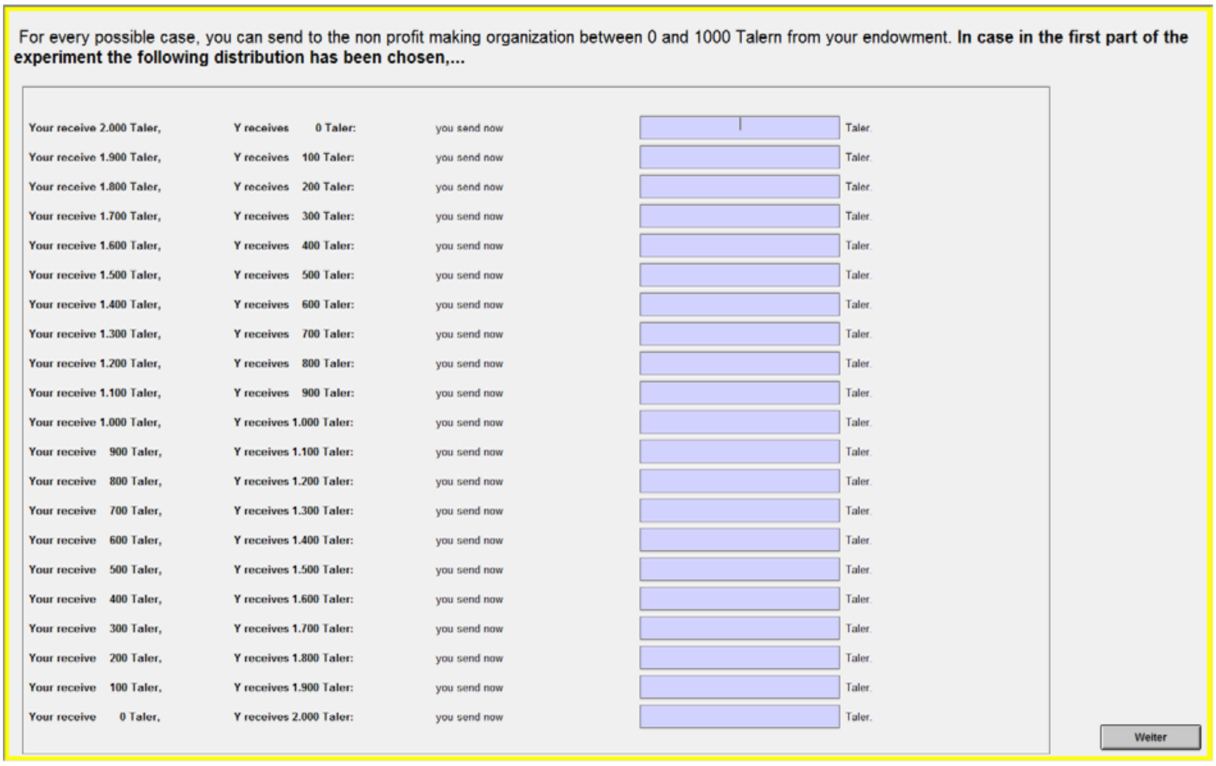
Information on the First Experiment: Part 2

The participants in this part of the experiment have the same roles as in the first part of the experiment. This means that a participant who had role A in the first part of the experiment will also have role A in this part of the experiment; a participant who had role X in the first part of the experiment will also have role X in this part of the experiment, and a participant who had role Y in the first part of the experiment will also have role Y in this part of the experiment. Additionally, the non-profit organization "Deutsche Welthungerhilfe e.V." will be relevant in this part of the experiment. This organization is active in the field of development organization and is officially certified by the German Central Institute for Social Issues.

In this part of the experiment, player **X** receives an endowment of **1000 Taler**. The player decides how many of these 1000 Taler to send to the non-profit institution (any full number between 0 and 1000). Each Taler sent is credited to the non-profit institution.

Player X can make the decision on how many Taler to send **dependent on every possible Taler distribution chosen by player A in the first part of the experiment**. The actual distribution from the first part of the experiment is told to player **X** only after the experiment has ended. However, X has to decide in **this** part of the experiment how many Taler he or she wishes to send **for every possible distribution**.

In this part of the experiment, player **X** must therefore fill in the following decision table:


For every possible case, you can send to the non profit making organization between 0 and 1000 Talern from your endowment. **In case in the first part of the experiment the following distribution has been chosen,...**

Your receive 2.000 Taler,	Y receives 0 Taler:	you send now	Taler.
Your receive 1.900 Taler,	Y receives 100 Taler:	you send now	Taler.
Your receive 1.800 Taler,	Y receives 200 Taler:	you send now	Taler.
Your receive 1.700 Taler,	Y receives 300 Taler:	you send now	Taler.
Your receive 1.600 Taler,	Y receives 400 Taler:	you send now	Taler.
Your receive 1.500 Taler,	Y receives 500 Taler:	you send now	Taler.
Your receive 1.400 Taler,	Y receives 600 Taler:	you send now	Taler.
Your receive 1.300 Taler,	Y receives 700 Taler:	you send now	Taler.
Your receive 1.200 Taler,	Y receives 800 Taler:	you send now	Taler.
Your receive 1.100 Taler,	Y receives 900 Taler:	you send now	Taler.
Your receive 1.000 Taler,	Y receives 1.000 Taler:	you send now	Taler.
Your receive 900 Taler,	Y receives 1.100 Taler:	you send now	Taler.
Your receive 800 Taler,	Y receives 1.200 Taler:	you send now	Taler.
Your receive 700 Taler,	Y receives 1.300 Taler:	you send now	Taler.
Your receive 600 Taler,	Y receives 1.400 Taler:	you send now	Taler.
Your receive 500 Taler,	Y receives 1.500 Taler:	you send now	Taler.
Your receive 400 Taler,	Y receives 1.600 Taler:	you send now	Taler.
Your receive 300 Taler,	Y receives 1.700 Taler:	you send now	Taler.
Your receive 200 Taler,	Y receives 1.800 Taler:	you send now	Taler.
Your receive 100 Taler,	Y receives 1.900 Taler:	you send now	Taler.
Your receive 0 Taler,	Y receives 2.000 Taler:	you send now	Taler.

Weiter

Participant **X** may enter any full number between 0 and 1000 in every line. **The only line that becomes payoff-relevant is the one that corresponds to the distribution actually chosen in the first part of the experiment.** The decisions in the other lines do not influence the participants' payments. The lines do not have to add up exactly to 1000 either, as the only decision-relevant line is the one that corresponds to the actual situation.

As participant **X**, please note that at the time of filling in the table you do not yet know which allocation decision has been made in the first part of the experiment. **You therefore have to consider your decision very carefully in every line, because any of the lines could become payoff-relevant for you.**

After filling in the table, participants X and Y will be given a brief questionnaire.

Player **Y** and player **A** make no decisions in this part of the experiment and receive **no earnings**.

The **earnings** from this experiment add up as follows:

Participant **X**	1.000 Taler **–** Taler sent by X to the non-profit institution (in the payoff-relevant situation)

After filling in the questionnaire, the second part of the experiment ends.

Only after the experiment has ended are participants given information on the **actual distribution decision** of player **A** in the first part of the experiment and on the sum corresponding to this distribution, which player **X** has sent to the non-profit institution in this part of the experiment.

It is impossible for you to lose, in a later part of the experiment, the earnings you have accumulated in the first and second part of the experiment.

The Taler sent to the non-profit institution will be converted into Euro at the rate stated above and will be transferred by the experimenters after the end of all experimental sessions. After the end of today's experiments, a web address will be shown to all participants, where proof of the total amount donated in all experimental sessions will be displayed from 17 March 2013 onwards.

You will now be shown some control questions on your screen. After you have answered these questions correctly, the second part of the experiment will begin.

IV.2. Tables

TABLE 3

Pairwise treatment comparison of transfers in the dictator game conditional on every possible allocation

Mann-Whitney tests: p-values, |z| in parentheses				
Possible allocation		*Baseline* vs. *Narrow*	*Baseline* vs. *Broad*	*Narrow* vs. *Broad*
X: 2000	Y: 0	0.0757, (1.776)	0.5432, (0.608)	0.3530, (0.929)
X: 1900	Y: 100	0.0374, (2.081)	0.4766, (0.712)	0.2340, (1.190)
X: 1800	Y: 200	0.0311, (2.156)	0.2190, (1.229)	0.4235, (0.800)
X: 1700	Y: 300	0.0340, (2.120)	0.1382, (1.483)	0.5224, (0.640)
X: 1600	Y: 400	0.0337, (2.123)	0.0501, (1.959)	0.7716, (0.290)
X: 1500	Y: 500	0.0484, (1.974)	0.0753, (1.779)	0.6913, (0.397)
X: 1400	Y: 600	0.0289, (2.185)	0.0301, (2.168)	0.7254, (0.351)
X: 1300	Y: 700	0.0284, (2.191)	0.0130, (2.484)	0.9573, (0.054)
X: 1200	Y: 800	0.0108, (2.549)	0.0059, (2.753)	0.9390, (0.077)
X: 1100	Y: 900	0.0108, (2.549)	0.0133, (2.475)	0.7529, (0.315)
X: 1000	Y: 1000	0.0021, (3.080)	0.0096, (2.590)	0.4416, (0.769)
X: 900	Y: 1100	0.0566, (1.906)	0.0133, (2.476)	0.6550, (0.447)
X: 800	Y: 1200	0.0592, (1.887)	0.0140, (2.458)	0.6733, (0.422)
X: 700	Y: 1300	0.0592, (1.887)	0.0175, (2.376)	0.8068, (0.245)
X: 600	Y: 1400	0.0661, (1.838)	0.0554, (1.916)	0.9313, (0.086)
X: 500	Y: 1500	0.0592, (1.887)	0.0883, (1.705)	0.7865, (0.271)
X: 400	Y: 1600	0.1126, (1.587)	0.1534, (1.428)	0.8221, (0.225)
X: 300	Y: 1700	0.1270, (1.526)	0.2304, (1.199)	0.7215, (0.356)
X: 200	Y: 1800	0.1374, (1.486)	0.2548, (1.139)	0.6809, (0.411)
X: 100	Y: 1900	0.2214, (1.223)	0.3631, (0.909)	0.7760, (0.284)
X: 0	Y: 2000	0.6928, (0.395)	0.7803, (0.279)	0.8984, (0.128)

TABLE 4

Pairwise treatment comparison of transfers in the dictator game conditional on every possible allocation – uninvolved treatments

Mann-Whitney tests: p-values, |z| in parentheses		
Possible allocation		*Baseline-Uninvolved* vs. *Narrow-Uninvolved*
X: 2000	Y: 0	0.7252, (0.351)
X: 1900	Y: 100	0.5268, (0.633)
X: 1800	Y: 200	0.5566, (0.588)
X: 1700	Y: 300	0.5365, (0.618)
X: 1600	Y: 400	0.5616, (0.580)
X: 1500	Y: 500	0.7367, (0.336)
X: 1400	Y: 600	0.9086, (0.115)
X: 1300	Y: 700	0.8363, (0.207)
X: 1200	Y: 800	0.8303, (0.214)
X: 1100	Y: 900	0.8124, (0.237)
X: 1000	Y: 1000	0.8532, (0.185)
X: 900	Y: 1100	0.7334, (0.341)
X: 800	Y: 1200	0.6400, (0.468)
X: 700	Y: 1300	0.5643, (0.577)
X: 600	Y: 1400	0.9937, (0.008)
X: 500	Y: 1500	0.7945, (0.260)
X: 400	Y: 1600	0.9608, (0.049)
X: 300	Y: 1700	0.9151, (0.107)
X: 200	Y: 1800	0.9346, (0.082)
X: 100	Y: 1900	0.5816, (0.551)
X: 0	Y: 2000	0.4427, (0.768)

TABLE 5

Treatment effects on transfers – comparison of Baseline-Uninvolved and Narrow-Uninvolved

Random effects Tobit regression
Dependent variable: transfers in the dictator game

	Model 1	Model 2	Model 3	Model 4	Model 5
Narrow-Uninvolved	22.56	41.65	74.68	76.93	88.20
	(111.82)	(106.62)	(103.73)	(97.41)	(131.15)
Fair allocation		-36.40**	2.12	18.35	16.70
		(17.83)	(25.16)	(24.53)	(24.30)
Expectation			-52.86**	-53.28**	-54.58**
			(25.31)	(23.77)	(23.51)
Social Value Orientation				8.23**	7.35**
				(3.19)	(3.17)
Possible allocation part 1	-22.37***	-22.35***	-22.36***	-22.35***	-22.35***
	(0.99)	(0.99)	(0.99)	(0.99)	(.99)
Narrow- uninvolved*Gender					-12.86
					(191.85)
Gender					-133.72
					(135.43)
Constant	294.79***	841.05***	968.92***	618.16**	735.55**
	(79.05)	(276.46)	(272.18)	(288.47)	(295.17)
N	1239	1239	1239	1239	1239
P model	<.001	<.001	<.001	<.001	<.001
Wald Chi2	513.11	516.63	520.02	525.61	527.42

Random effects Tobit regression. Standard errors are presented in parentheses. The *Narrow-Uninvolved* dummy equals 1 for all observations of the *Narrow Voice-Uninvolved* treatment, *fair allocation* controls for the differences in the allocation players X consider as fair, *expectation* controls for the differences in players X's expectation about the actual allocation by the impartial decision makers, *social value orientation score* controls for differences in players' social value orientation, the *gender* dummy equals 1 for male players, *possible allocation part 1* controls for possible allocations that can be implemented by the impartial decision makers. Significance at the 10%, 5% and 1% level is denoted by *, ** and ***, respectively. Left-censored = 432; right-censored = 68.

TABLE 6

Expectations and perceived fairness in Baseline-Uninvolved and Narrow-Uninvolved

		Baseline-Uninvolved	*Narrow-Uninvolved*	Mann-Whitney tests *Baseline-Uninvolved* vs. Narrow-*Uninvolved*
		Mean (sd)	Mean (sd)	p-values (|z|)
Expecations	Perceived influence by players X	1.27 (2.53)	4.28 (3.52)	0.0003 (3.640)
	Expectation of players X concerning allocation in part 1	1223 (310)	1341 (298)	0.117 (1.568)
Perceived fairness by players X	Fairness process general (1)	6.47 (2.98)	5.62 (3.19)	0.3158 (1.00)
	Fairness process for X (2)	5.90 (3.26)	6.45 (3.77)	0.3771 (0.883)
	Fairness of treatment (3)	6.63 (2.89)	6.07 (3.67)	0.7303 (0.345)
	Fairness of outcome (4)	6.60 (3.40)	6.00 (3.99)	0.4675 (0.727)
	Acceptance of decision (5)	7.30 (3.37)	6.72 (4.01)	0.5849 (0.546)

Perceived influence as well as all *perceived fairness* ratings on a Lickert scale from 0 ("not at all") to 10 ("completely"); expectations in ECU.

Declaration of Authorship

I hereby declare that I have completed the following work without help from third parties and without means of assistance, apart from those indicated. I have cited the sources of all direct and indirect quotations, dates, and ideas that are not my own. The following persons assisted me with the selection and evaluation of research materials as described below and for payment or without payment as indicated: exclusively my co-authors, i.e., Christoph Engel in Chapters 1, and 3, as well as Marco Kleine and Pascal Langenbach in Chapter 4.

No other persons except those listed in the work's introduction were involved in preparing the contents of this work. I certify that I have not used the paid services of consultation firms, and that I have paid no one, directly or indirectly, for tasks connected to the contents of this dissertation. The work has not yet been submitted in the same or similar form to another institution in Germany or abroad. I certify that this statement is true and complete to the best of my knowledge.

Nuremberg, July 17, 2014

Curriculum Vitae

LILIA ZHURAKHOVSKA

July 18, 2014

University of Erlangen-Nuremberg
Chair of Economic Theory
Lange Gasse 20
90403 Nuremberg
Germany

Email: lilia.zhurakhovska@fau.de
Phone: +49 911 5302-9415
Fax: +49 911 5302-168

PERSONAL DETAILS

Date of birth:	7 September 1983
Place of birth:	Kyiv, Ukraine
Citizenship:	German and Ukrainian

POSITIONS IN RESEARCH

Since 2013	**Research Fellow,** University of Erlangen-Nuremberg Chair of Economic Theory, Nuremberg, Germany
Since 2013	**Guest Researcher**, Max Planck Institute for Research on Collective Goods, Bonn, Germany
2010–2013	**Research Fellow,** Max Planck Institute for Research on Collective Goods, Bonn, Germany
2008–2009	**Student Research Assistant,** Max Planck Institute for Research on Collective Goods, Bonn, Germany

ACADEMIC CAREER

2012–2013	**Visiting Researcher** University of California, San Diego
2010–2014 (expected)	**(expected) Ph.D. in Economics** University of Cologne & IMPRS (The International Max Planck Research School on Adapting Behavior in a Fundamentally Uncertain World), Germany Supervisor: Bettina Rockenbach & Christoph Engel Evaluator: Bettina Rockenbach & Dirk Sliwka
2005–2009	**Diplom-Volkswirts (equivalent to Master in Economics)** University of Bonn, Germany Supervisor: Armin Falk
2006–2007	**Exchange semester** University of Växjö, Sweden
2004–2005	**Diploma Student economics** University of Munich (LMU), Germany

RESEARCH INTERESTS

Behavioral and Experimental Economics
Law and Economics
Psychology and Economics
Public Economics

TEACHING

2014	Exercise in "Micro Economics" for Bachelor students, University of Erlangen-Nuremberg, Germany
2013	"Behavioral Economics Seminar" for Master and Bachelor students, University of Erlangen-Nuremberg, Germany
	"Experimental Design Course" for Ph.D. students, Evidence-Based Economics Graduate School (EBE), Germany
2011	Supervision of one Bachelor's and one Master's thesis in economics, University of Bonn, Germany
	Tutorial in "Economics and Financial Politics" for Bachelor and Diploma students, University of Bonn, Germany
2008	Tutorial in "Economics and Financial Politics" for Bachelor and Diploma students, University of Bonn, Germany

PROFESSIONAL CAREER

Since 1999	Assistant translator (Russian, Ukrainian, German) with the translation agency Valeryi Zhurakhovskyi, Cologne, Germany
2007	Internship at the branch office of the Ukrainian embassy in Remagen, Germany
2005	Internship at the health insurance company AOK (Controlling), Hürth and Düsseldorf, Germany

REFERENCES

Prof. Dr. Christoph Engel, MPI for Research on Collective Goods, Germany
engel@coll.mpg.de
Prof. Dr. Bettina Rockenbach, University of Cologne, Germany
bettina.rockenbach@uni-koeln.de
Prof. Dr. Uri Gneezy, University of California San Diego, USA
ugneezy@ucsd.edu
Prof. Dr. Joel Sobel, University of California San Diego, USA
jsobel@ucsd.edu

SKILLS

Languages: German (native), Russian (native), Ukrainian (fluent) English (fluent), French (school level)
Programs: z-Tree, Stata, ORSEE, MS-Office