Cooperation and Norm Enforcement

Lab and Field Evidence in Experimental Economics

Inaugural-Dissertation
zur Erlangung des Grades
Doctor oeconomiae publicae (Dr. oec. publ.)
an der Ludwig-Maximilians-Universität München

2014

vorgelegt von
Bianca Kern

Referent: Prof. Dr. Martin G. Kocher
Korreferent: Prof. Dr. Florian Englmaier
Promotionsabschlussberatung: 05.11.2014

Datum der mündlichen Prüfung:	22.10.2014
Namen der Berichterstatter:	Prof. Dr. Martin G. Kocher, Prof. Dr. Florian Englmaier, Prof. Dr. Joachim Winter

to my family

Acknowledgements

There are many people I would like to thank for sharing their knowledge with me, for their valuable and inspiring comments and their support during my doctoral studies.

First of all I am very grateful to my supervisor and co-author Martin G. Kocher for the chance to write this dissertation. I highly appreciated his introduction to the exciting field of Behavioral and Experimental Economics, his patient guidance, inspiring ideas and always very helpful advice – not only on joint projects. His uncomplicated manner led to a very fruitful and trustful cooperation in several research projects. I also want to thank him for making it easier for me to combine the parenting of Maximilian and the writing of the dissertation.

I owe thanks to Florian Englmaier and Joachim Winter for agreeing to supervise my dissertation and their very helpful comments on my research projects or Joachim Winter's great introduction in the field of econometrics.

During the years of my doctoral studies I received numerous valuable and inspiring comments and had profound discussions with colleagues, friends and participants of workshops or conferences: I want to thank especially participants of the MELESSA Brown Bag Session, the Micro Workshop, the Brown Bag Seminar at the Chair for Behavioural and Experimental Economics, all internal seminars of the University of Munich, and participants of the ESA 2011 in Luxembourg. I am grateful to colleagues, the entire staff of MELESSA, and the research assistants of the Chair for Behavioural and Experimental Economics especially for their very motivated support in the conduct of my lab-field comparison. Furthermore I thank the Cusanuswerk for their financial and ideational support.

Lastly, I am very grateful to my parents for their support in everything I do and in particular to my husband Roland for his patience and his encouragement. I want to thank them especially for their very valuable support in parenting Maximilian and thereby making it possible for me to complete this dissertation.

Bianca Kern

Contents

List of Tables

List of Figures

Preface

Cooperation is undeniably an important element of human life. From an evolutionary perspective cooperation is crucial, as groups with cooperating members have selective advantages compared to groups with less cooperative members, and therefore have a better chance to survive. In the economic context cooperation is essential for societies, organizations and teams. For instance, think about effort provision in teams, the private provision of public goods, the use of natural resources, climate protection, contributions to charities, or tax honesty. All mentioned examples represent a social dilemma, in which group members have to overcome their selfish interests to free-ride and, instead, cooperate to obtain an efficient outcome. In order to foster cooperation in social dilemmas, social norms have evolved. Frequently such norms are enforced, e.g. by peer punishment. Interestingly, peer punishment occurs, although it usually bears some cost for the punishing party. We encounter examples of norm enforcement, i.e. punishing free-riding, everyday. For instance, team members are excluded by their colleagues for little work morale, pedestrians complain about people littering in public places or people waiting in lines insult queue-jumpers.

Economic research on cooperation and norm enforcement is crucial to derive appropriate interventions for the real world, e.g. for politics and organizations. For instance, the knowledge about whether teams or individuals are more cooperative in social dilemmas can be used when deciding about parties to conduct political negotiations. Understanding the factors that increase cooperation and reduce punishment might be important for organizations to increase the efficiency of their working teams. Empirical economic research on cooperation and norm enforcement so far has shown that norm enforcement in the form of peer punishment is one of the most successful mechanisms to enhance cooperation in social dilemmas (e.g. Fehr and Gächter, 2000). Among others it has inspired the development of seminal theoretical models, namely models of other-regarding or social preferences. These approaches go beyond the traditional concept of the homo-oeconomicus and include the long-time neglected fact that people also care about others. Models of inequity aversion (Fehr and Schmidt, 1999; for a similar approach see Bolton and Ockenfels, 2000) or social welfare (Charness and Rabin, 2002) are important examples of these approaches. They strongly improved the predictions of real world behaviour in social contexts.

Regarding the methodology for the research on cooperation and norm enforcement, laboratory experiments play an important role. In contrast to standard empirical methods laboratory experiments have several advantages: They allow a high degree of control, including a controlled variation of certain independent variables, a randomized assignment of subjects to treatments and a context-free elicitation of preferences. Therefore laboratory experiments allow causally attributing differences in behavior to differences in treatments. All studies in this dissertation include laboratory experiments. However, the main criticism on laboratory experiments is their external validity. Hence, it is put into question whether the results of laboratory experiments are generalizable to the real world. To encounter this main criticism the last study of this dissertation additionally includes a natural field experiment.

In all studies of this dissertation cooperation and/or norm enforcement are examined in the lab in the context of a public goods setting or its binary version, a prisoners' dilemma game, which constitute a social dilemma. The norm enforcement possibility is implemented by an additional decision stage in which players receive information on the individual cooperation decisions of their group members and can punish them by reducing their income (whereby punishing is costly also for the punishing party). The first chapter of this dissertation explores the question how cooperation changes when decision-makers are not individuals, as commonly assumed in theoretical models and most laboratory experiments, but unitary teams, i.e. a team of several individuals has to come up with a unitary decision. Chapter 2 investigates how cooperation and norm enforcement (and therefore efficiency) are affected when groups have to reach a certain threshold of group income in order to "survive", i.e. in order to participate in the game in the future and therefore earn more money. Chapter 3 makes a contribution to the exploration of the external validity of lab experiments by combining norm enforcement behavior in a natural field experiment with behavior of the same persons elicited in the lab.

Chapter 1, which is joint work with Martin Kocher, examines how cooperation in a public goods setting changes, if the decision-makers are interacting unitary teams instead of individuals. There is a sizeable literature, especially in social psychology, stating that teams are less cooperative and, thus, more competitive than individuals, which psychologists term the "*(interindividual-intergroup) discontinuity effect*". In economics cooperation of individuals has been extensively studied. In contrast, an analysis of the aggregation of individual cooperative preferences to a single team decision when teams interact as players in a social dilemma is missing. However such an analysis seems warranted as the social psychological studies on the discontinuity effect apply a very special experimental design. Moreover, many relevant economic decisions in the real world that involve social dilemmas are very well characterized by teams acting as decision makers, from the small (e.g. families, management teams) to the large (e.g. company boards, governments, international organizations). Thus, assessing team reasoning and team decision-making seems fruitful for economists in order to derive correct predictions for real life and to draw appropriate conclusions in which settings decisions are better taken by individuals or by teams.

In a between-subjects design we implement two laboratory treatments with interacting individuals and interacting teams of three team members in a public goods setting. In contrast to most of the literature so far we implement repeated interaction that additionally enables us to study learning and the dynamics of interaction. Eliciting individual beliefs about others' contributions, asking for contribution proposals by the team members before the team decision, and adding an independent measure of cooperative preferences on the individual level allow us to analyze the mechanisms behind team decision-making. Team members anonymously communicate via real-time chat within their team, which guarantees maximal control over the interaction. At the same time the analysis of the chat communication allows valuable additional insights into the motives of team decisions.

We rather find a *reverse discontinuity effect* with teams to be more willing to cooperate than individuals, at least in the first half of the public goods game. Our results support several reasons for this finding: Firstly, contributions in the public goods game strictly increase social welfare. Furthermore the chat analysis strongly suggests that efficiency-concerns of teams might be an important reason for their enhanced willingness to cooperate. A second explanation for the reverse discontinuity effect is related to the more optimistic expectations towards other teams than towards other individuals. Especially conditional cooperators (in contrast to free-riders) have more optimistic expectations towards other teams than towards other individuals. Our chat analysis suggests that these more optimistic expectations combined with advantageous inequity aversion might lead to an increased willingness to cooperate in the team setting. A third explanation that we cannot completely rule out might be an enhanced motivation of teams to (indirectly) build up a cooperative reputation.

Concentrating on the team decision-making process, we find that the applied decision rule is affected by the cooperative preferences of team members: Teams consisting only of conditional cooperators decide more often by implementing an informal compromise than other teams. The final team decision is not only affected by the team composition in terms of cooperative preferences and the implicit decision rules applied. It also depends on the first proposal verbalized in the team discussion – we thus observe an anchoring effect – and, interestingly, the level of cooperation increases with the length of the discussion within a team.

Given the numerous examples in which cooperation decisions are taken by teams, our results have broad practical implications, e.g. for organizations and politics, and suggest that team decisions can be more cooperative than individual decisions and need not be seen too negatively.

In *chapter 2*, which is also joint work with Martin Kocher, we investigate how cooperation and norm enforcement, i.e. efficiency, in a public goods setting with punishment opportunities are affected by a group extinction mechanism. Drawing on evolutionary ideas the group extinction mechanism reflects the fact that group members can "survive", i.e. in the lab proceed in future periods and earn more income, only if the group income exceeds a certain survival threshold. In real life the existence of groups in the long run often depends on the achievement of a minimum group income, e.g. working groups in companies or organizations often fight for survival as a group, i.e. they can easily be dissolved if unsuccessful. Although group extinction often takes place in real-world settings, this study is the first one to investigate the effect of this mechanism. However, if the possibility of group extinction is not taken into account in economic research, laboratory experiments may yield wrong predictions concerning the extent of cooperation and norm enforcement in real-world public goods settings where punishment is an option. This argument is supported by the following facts: More often than not laboratory experiments with public goods settings and a punishment option show severe efficiency problems, as cost for punishment mostly exceed the benefits from increased cooperation. The efficiency problem is even worse if a retaliation option exists. Data from field experiments on norm enforcement show lower rates of punishment than the laboratory data. Besides the fear of retaliation that plays a role, efficiency concerns could be much more important in the field than conventionally believed. Given that the possibility of group extinction is an important factor in real world public goods settings with a norm enforcement option, evidence on the effect of this mechanism yields important practical implications, e.g. for working teams in organizations.

Pertaining to the design our setup implements a repeated linear public goods game with costly punishment opportunities in the lab. The design features four experimental treatments in a two-by-two factorial between-subjects design: Within the first dimension, we vary whether peer punishment is available with or without a retaliation option, i.e. whether a one-sided or two-sided punishment option is given. The second dimension that is varied is the existence of a survival threshold. If the survival threshold is present, groups have to reach a certain level of group income in order not be extinct as a group in the next period. The treatments all include four identical parts with five periods each. If a group is extinct in one part, group members are excluded from the public goods game – and therefore get an income of zero – in the following periods of that part. In order to be able to examine learning effects over parts, we allow the extinct subjects to restart the public goods game in the following part within a new group.

Referring mainly to the theoretical model of social welfare by Charness and Rabin (2002) we hypothesize that the salience of the efficiency concern created by the group extinction mechanism increases cooperation and reduces (counter-) punishment, and thus, increases efficiency. Our results mainly confirm these hypotheses: The group extinction mechanism leads to higher efficiency, partly by increasing cooperation and partly by decreasing norm enforcement. We furthermore find that groups learn to adapt to the survival threshold over time, indicated by the decreasing number of extinct groups over parts. Hence, the effect of the group extinction mechanism may be even enhanced in settings with longer interaction terms or when further forms of punishment, e.g. feuds, are possible. Our findings contribute – among others – to a better understanding of why in real-world scenarios often low levels of peer punishment have been observed. Our results further imply that the threat of group extinction can be applied, adapted or made transparent in the field in order to increase efficiency, e.g. in working groups.

In *chapter 3* I examine whether the norm enforcement behvior in the lab translates to norm enforcement behavior of the same subjects in the field. Laboratory experiments have for a long time been the fundament of behavioral economics and they undoubtedly have several advantages in comparison to standard empirical methods. However, the external validity of laboratory experiments is often put into question. This criticism is especially pronounced in the domain of social preferences, as their elicitation is particularly prone to an experimenter demand effect. In the light of this criticism I conduct a within-subjects comparison of costly norm enforcement between the lab and the field.

In a natural field experiment queuers are observed when a norm violator cuts in line in front of them. In order to minimize the possibility of an experimenter demand effect a research assistant invites the subjects from the field to the lab without revealing that they participated in a field experiment. In the lab three treatments dealing with norm enforcement are implemented: The standard treatment to examine norm enforcement in the lab – a three-person prisoners' dilemma game with a one-sided punishment option (see Fehr and Gächter, 2000; 2002; Falk et al., 2005) – and a treatment additionally allowing for counter-punishment (Nikiforakis, 2008), i.e. a three-person prisoners' dilemma game with a two-sided punishment option. Moreover I apply a newly created game with which I try to represent the situation in the queue in an abstract way. This so-called "queuing game" includes a one-sided punishment option.

Additionally to the behavioral data from the field and the lab experiment, I elicit of the same subjects the intended norm enforcement behavior stated in a survey. The latter presents three scenarios of real life situations, which describe different violations of concrete social norms. Apart from being able to investigate the external validity of stated norm enforcement intentions elicited in the questionnaire, the survey allows me to examine whether a so-called norm enforcing type exists, i.e. somebody who generally behaves in a norm enforcing way across several concrete social norms.

I find a norm enforcement rate of 32.1% in the field. Norm enforcement in the field correlates with decisions in the lab treatment including a two-sided punishment option, but not with decisions in the treatments in which no counter-punishment option exists. Specifically I find that the likelihood of norm enforcement in the field increases with the willingness to cooperate and with the sanctioning of defectors in a prisoners' dilemma game with a two-sided punishment option. This lab-field comparison is a first examination of the external validity of norm enforcement behavior in the lab and therefore generalizations have to be made with care. However the result suggests the following: Firstly, with respect to norm enforcement, it confirms prior survey evidence that weighing up the danger of being counter-punished is crucial for norm enforcement decisions in the field. Secondly, concerning the methodological aspect, the result indicates that more evidence from the field is necessary to establish the external validity of the standard treatment with a one-sided punishment option.

Furthermore I find a high external validity of the norm enforcement intentions stated in the survey, as long as the same social norm is affected in the survey scenario and in the field. I do not find evidence for a norm enforcing type. Instead the likelihood of norm enforcement seems to vary with the concrete social norm violated, and specifically with the negative externality of the social norm violation.

Overall the lab-field comparison is encouraging with respect to the generalizability of lab behavior to the field. However, it also illustrates that the relevant institutional factors of the field need to be incorporated in the lab to allow for external validity of lab experiments.

Chapter 1

Cooperation and Decision-Making of Teams in Public Goods Experiments

1.1 Introduction

There is a sizable literature in social psychology on the inclination of small groups or teams[1] to cooperate in prisoners' dilemmas compared to individuals. The general finding is that teams are less cooperative and, thus, more competitive (Charness and Sutter, 2012, as well as Kugler et al., 2012, provide recent surveys). Insko, Schopler and co-authors have coined the term "*(interindividual-intergroup) discontinuity effect*" for the difference in cooperative/competitive behavior between small teams and individuals, and they have established the effect in a large number of experimental studies.[2]

The preference for cooperation of individuals and the aggregation of cooperative preferences within teams matter commonly in social interaction: they play a role in team work, collective action, the organization of society, or the private provision of public goods. Nonetheless, economic assessments of the aggregation are still missing. Cooperation *of individuals* has been studied extensively in economics, using the prisoners' dilemma game in earlier papers and the voluntary contribution mechanism, also known as the public goods game, later (see Chaudhuri, 2011, for a recent survey). In contrast, an analysis of the aggregation of individual cooperative preferences to a single team decision (in so-called unitary teams) when teams interact as players in a social dilemma is still missing. Such a study seems warranted for at least two reasons. Firstly, the setup that Insko, Schopler, and co-authors apply is a very special one. Teams interact face-to-face both within teams and between teams before taking a decision. Hence, communication within teams and across teams is important, and effects from the communication are hard to control for. Secondly, many relevant economic decisions in the real world that involve social dilemmas are very

[1] The meaning of the two terms *group* and *team* is sometimes the same in the literature and sometimes not. In this paper we stick to the term *team*.
[2] For instance, Insko et al. (1987a; 1987b; 1988; 1990), Insko and Schopler (1987), Schopler and Insko (1992), Schopler et al. (1991; 2001), Wildschut et al. (2003; 2007), Wildschut and Insko (2007).

well characterized by teams acting as decision makers, from the small (e.g. families, management teams) to the large (e.g. company boards, governments, international organizations). Thus, more generally, assessing team reasoning and team decision making seems fruitful for economists.[3]

The objective of this paper is to fill a gap in the literature. To the best of our knowledge, we are the first in economics to systematically study the aggregation of cooperative preferences into a team decision when teams interact as decisions makers in a social dilemma and compare the outcomes with a standard social dilemma in which individuals interact. Our laboratory experiment based on a comparatively large sample provides several advantages over the existing literature: Firstly, it uses a public goods game instead of the prisoners' dilemma game, where the former allows for a larger strategy space, a more continuous measure of cooperation, and a finer-grained analysis of responses. Secondly, in contrast to most of the literature, we implement repeated interaction that enables us to study learning and the dynamics of interaction, alongside the pure preference for cooperation. Given recent results by Müller and Tan (2013), the time horizon of an interaction (even in a design that rules out reputational concerns) could affect individuals and teams differently (see also Kocher and Sutter, 2005). Despite the learning possibility we are, thirdly, able to control for initial individual cooperative preferences by using a preference elicitation method developed by Fischbacher et al. (2001) and validated for repeated interactions by Fischbacher and Gächter (2010). This allows for keeping control of the team composition with respect to individual cooperative preferences (cooperation types). Fourthly, we control for expectations with regard to cooperation of other decision makers, which turns out to be important in explaining differences between individual behavior and team decisions. And finally, fifthly, we implement a team interaction that allows for within-team communication, but the real-time chat sustains anonymity of individuals even within teams. Such a setup guarantees maximal control over the interaction. However, the analysis of the chat protocols provides valuable additional insights into the decision making dynamics within teams.

In contrast to expectations based on the literature in social psychology, our empirical results strongly indicate that teams are more willing to cooperate and less willing to compete, at least in the first half of the game. We thus report a *"reverse discontinuity effect"*. As a consequence, teams also achieve higher profits than individuals. Expectations regarding teams are in line with the higher levels of cooperation. The effect is especially driven by conditional cooperators, who – in comparison to free-riders – are more optimistic with

[3] As a consequence, Camerer (2003) mentions in his textbook *Behavioral Game Theory* the study of group decision-making among the top ten research projects in behavioral and experimental economics.

respect to the contributions of other decision makers in the team setting and are, therefore, more likely to cooperate when they act as a team member than when they act individually. Concentrating on the team decision-making process, we find that the applied decision rule is affected by the individual cooperative preferences of team members: Teams consisting only of conditional cooperators decide more often by implementing an informal compromise than other teams. The final team decision is not only affected by the team composition in terms of cooperative preferences and the implicit decision rules applied. It also depends on the first proposal verbalized in the team discussion – we thus observe an anchoring effect – and, interestingly, the level of cooperation increases with the length of the discussion within a team.

While, as mentioned above, there is no existing study that compares individual decisions in public goods games with team decision, the literature on team decision making in economics allows for some inferences with regard to the expected effects when one looks at related games that have been investigated. The general finding of these studies (table A.1.1 in appendix A.1.1 provides an overview) is that unitary teams are more selfish and more rational than individuals (Charness and Sutter, 2012; Kugler et al., 2012). In the dictator game teams as dictators send smaller amounts (Luhan et al., 2009), in the ultimatum game they send and accept smaller amounts (Bornstein and Yaniv, 1998), in the trust game teams are less trusting (Kugler et al., 2007) and less trustworthy (Cox, 2002), in the centipede game teams exit earlier (Bornstein et al., 2004), and in the gift exchange game teams as employers make lower wage offers than individuals when the communication within teams occurs via chat (Kocher and Sutter, 2007).[4]

In contrast, for instance, Bosman et al. (2006) find for the power-to-take game no differences between teams and individuals, neither in the role of the take authority, nor in the role of the responder. Cason and Mui (1997) provide evidence that for certain teams the team choices are more generous than the individual choices of the same team members in the dictator game. Kocher and Sutter (2007) find in case of face-to-face communication within teams higher effort levels for teams than for individuals. Müller and Tan (2013) implement a Stackelberg market game played either once or over 15 periods in a stranger design. In the one-shot game, they find no differences between teams and individuals, whereas in the multi-period case their results indicate that team decisions are farther away from standard game

[4] Abbink et al. (2010) find that the conflict expenditures of teams in a contest game are substantially larger than those of individuals. However, Abbink et al. (2010) do not study unitary teams, i.e. there is no phase of deliberation among team members. Similarly, Kugler and Bornstein (2013) find individuals to cooperate better with other individuals than teams with other teams. However, their teams are not unitary but characterized by intra-team conflict.

theoretic predictions than individuals: As first movers, they make lower output choices (and therefore act in a more collusive way), as second movers, they reward or punish first movers more strongly than individuals. Balafoutas et al. (2014) carry out an experiment that involves a double price-list technique for allocation decisions with individuals and teams. Teams in comparison to individuals are significantly more frequently classified as efficiency-loving and less frequently classified as spiteful or competitive.[5]

Social dilemmas in their linear version – regardless of whether talking of prisoners' dilemmas or public goods games – share the common feature that it is individually rational to free ride (contribute nothing), but collectively optimal to cooperate (contribute the entire endowment). Thus, efficiency concerns of decision makers might matter when it comes to the cooperation decision. Obviously, the existing evidence on an individual-team difference in games that involve efficiency concerns is mixed: In the trust game and the centipede game team decisions lead to less efficient outcomes than individual decisions (Cox, 2002; Kugler et al., 2007; Bornstein et al., 2004). In the power-to-take game, Bosman et al. (2006) find no effect, whereas allocation task and the Stackelberg game results seem to indicate that teams are more efficiency-oriented than individuals (Balafoutas et al, 2014; Müller and Tan, 2013). The inconclusive results on game and allocation decisions that involve potential efficiency concerns provide another reason for conducting the present study.

Taking the results from the related literature and our findings together, there are a couple of straightforward implications for managerial decisions and institutional design. Depending on whether one wants to encourage or discourage cooperation, having individual decision makers or setting up teams as decision makers can be optimal. Delegating teams in positive-sum negotiations seem to be the better choice than delegating individuals. The assessment of interactions that are bound to be organized in teams, such as global climate negotiations or coordination in international organizations, seems less pessimistic than it was, based on the previous literature. Nevertheless, more evidence on the aggregation of individual preferences within teams and on team decisions compared to individual decisions is clearly required in order to substantiate these implications for real-world situations.

The remainder of this paper is organized as follows: In section 2 we present our experimental design and the hypothesis for our main research question. Section 3 provides

[5] Charness et al. (2007) present an economic study in which they implement a prisoner's dilemma game. They vary the saliency of group membership and find that it leads to more defection. However, in their study decisions on behalf of groups are taken by individual group representatives and not by teams that have to come up with a joint decision. Furthermore, unitary team decision making has been studied in interactive games that require some rational reasoning (e.g., Cox and Hayne, 2006; Cooper and Kagel, 2005; Kocher and Sutter, 2005; Kocher et al., 2006; Gillet et al., 2009; Sutter et al., 2009; Zhang and Casari., 2012; Feri et al., 2010).

the empirical results, and section 4 concludes the paper.

1.2 Experimental Design, Procedures and Theoretical Predictions

In the following we explain the setup of the experiment and derive theoretical predictions.

1.2.1 Experimental Design

Our basic decision situation is a standard linear public goods game (PGG, also known as the voluntary contribution mechanism; see Marwell and Ames, 1981; Isaac et al., 1985; Isaac and Walker, 1988). The participants are randomly assigned to units of three members.[6] Each unit member receives an endowment E of 20 points. The members of each unit then decide simultaneously how much of their endowment to contribute to the public good, c_i, with $0 \le c_i \le 20$, or to a private account. The payoff function is given by

$$\pi_i = 20 - c_i + 0.5 * \sum_{j=1}^{3} c_j \tag{1.1}$$

where the value of the public good is equal to the sum of contributions by all unit members. If a participant contributes one point to the public good, the private marginal return a is 0.5 points, and the social marginal benefit is 1.5 points. Hence, $a < 1 < 3a$ applies, and it implies a social dilemma: If we assume pure money-maximizing preferences, the dominant strategy for each unit member is to free ride, i.e. $c_j = 0$ for all j. However, the socially efficient outcome is reached with full contribution of all unit members.

We implement two treatments in a between-subjects design: a treatment in which individuals interact with individuals (IND) and a treatment in which teams interact with teams (TEAM) in the public goods game. IND is mainly based on the experimental design by Fischbacher and Gächter (2010). TEAM lets three-person teams decide on contribution levels as one unit member.

Subjects in both treatments know that the experiment consists of two parts and some questionnaires and that the parts are independent of each other. The first part is identical for both treatments and elicits the individual cooperative preferences in a one-shot linear public goods game according to Fischbacher et al. (2001). Using the strategy vector method we ask participants to indicate their contribution to the public good for each period average

[6] Note that we denote three interacting parties in the PGG as a *unit*, in order to differentiate from the term *team*, which represents one interacting party within the unit.

contribution level of the other two unit members. In order to be able to incentivize these conditional decisions, we ask participants to make an unconditional contribution to the public good before. Subjects know that at the end of the first part one unit member is randomly drawn for whom the conditional contribution, based on the unconditional contributions by the other two unit members, is payoff-relevant. For the other two unit members the unconditional contribution is payoff-relevant. Payoffs are calculated according to Eq. (1.1). Subjects are told that the feedback on the behavior of the other two unit members and on the own income from the first part is only given after the second part.

In the second part unit members in both treatments are asked to make unconditional contribution decisions in the linear PGG described above, which is played over 10 periods in a stranger matching. In the second part, either three individuals (IND) or three teams (TEAM) interact in the PGG. We first outline IND and then explain the specifics of our setup in TEAM. In IND, subjects know that they are reassigned at the beginning of the second part to two other individuals than in the first part, and that re-shuffling of the unit composition takes place after every period (stranger design). Subjects are told that in each period they decide on their individual contribution to the public good. Then they have to indicate their beliefs about the average contribution made by the other two unit members. Incentives for correct beliefs are kept small to minimize hedging incentives. For reasons of simplicity the linear scoring rule is implemented, meaning that subjects receive 3 points for correct guesses and 2 (1) point(s) if their estimation deviates by 1 (2) point(s) from the actual average contribution of the other two unit members. Subjects know that at the end of each period they are informed about the sum of contributions made by the other two unit members, their income from the private account, their income from the public good and their total period income. They also get feedback at the end of each period about their income from the belief elicitation.

Subjects are told that at the end of the second part one period is randomly chosen that is payoff-relevant for the income from the PGG. Irrespective thereof, a further one is chosen that is payoff-relevant for the income from the belief elicitation. Both periods are chosen for each subject individually. At the end of the second part subjects receive feedback about their final income from the second part.

In TEAM units consist of three teams, whereas each team consists of three team members. Subjects know that the team composition is kept constant over all periods, whereas teams interact in each period with other teams. We implement a team size of three members as this allows us to observe the effects of various implicit decision rules including decisions under a majority rule. In each period of the second part, participants are first asked to make an individual, non-binding proposal for the team contribution. The proposal helps us to study the aggregation of individual proposals into a team decision. Subjects know that this proposal is shown to the other two team members before the team interaction starts. They are told that they have to reach a unanimous decision within their team on the contribution to the public good. Subjects communicate within their teams (but not between teams) via a real-time chat.[7] The team communication is unrestricted, except that participants must not reveal their identity or insult each other. Messages are always sent to the two other team members automatically; directed messages to only one other team member are not feasible.

If a team does not reach a unanimous decision within five minutes in each period, the team gets a second chance within an additional minute. If no consensus is reached after this grace period, the period income from the PGG is zero points for all members of this team. For the calculation of the sum of contributions in the unit in this case, one team member is chosen randomly by the computer, and her individual proposal is used as the relevant team contribution. This procedure is common knowledge.[8] Each team member enters the team contribution on her own screen in order to ensure unanimity. Note that we do not give teams any instructions on how to reach a unanimous decision. After the common contribution decision team members indicate their individual beliefs about the average contributions by the other two teams in their unit in the incentive compatible way explained above. Subjects in TEAM know that at the end of each period they are individually informed about the sum of contributions by the other two teams, their income from the private account, their income from the public good and their total income from the PGG. They also get feedback on their income from the belief elicitation at the end of each period.

The per-capita incentives in the PGG are identical between conditions TEAM and IND. All members of a team make the same contribution c_i^t to the public good, where the superscript *t* indicates the team and $0 \le c_i^t \le 20$ holds. The income from the public good is divided by the number of members within a team. Hence, the period income of player *i* in TEAM is calculated as:

$$\pi_i = 20 - c_i^t + 0.5 * \left(\sum_{j=1}^{3} \sum_{t=1}^{3} c_j^t \right) / 3 \qquad (1.2)$$

[7] We implement a chat communication in order to eliminate possible uncontrollable influences from face-to-face communication, as e.g. effects of attractiveness.

[8] In our sample all teams reach a unanimous decision in all periods, except for two teams in one period each. Over all 10 periods and all 60 participating teams, the second chance to communicate within teams was only necessary in 13out of 600 cases.

After completion of the second part participants are asked to answer questionnaires. Participants in TEAM first fill in a questionnaire on their team decision-making experience. In both treatments the two scales extraversion and agreeableness from the NEO-Five-Factors-Inventory (Borkenau and Ostendorf, 2008) are applied to elicit traits with a social component (results not reported here). Both treatments end with a short socio-economic questionnaire.

1.2.2 Procedures

We conducted 13 sessions – 3 sessions in IND and 10 sessions in TEAM – which took place in the experimental laboratory MELESSA of the University of Munich between May 2010 and February 2011. The sessions in TEAM last for approximately two hours, the sessions in IND last approximately 30 minutes less. Altogether 246 subjects participated in our experiment. In order to parallelize between IND and TEAM the probability of meeting the same unit member(s) in the next period(s) again, subjects are told in the second part that they are in a matching group with five other potential unit members and that they are randomly re-assigned in each period to two of them. Hence, in both treatments 6 potential unit members (consisting either of individuals or teams) were reshuffled after each period. In treatment IND, 24 subjects participated in two and 18 subjects participated in one session, which yields altogether 11 statistically strictly independent observations. In treatment TEAM 18 subjects participated in each session, which yields altogether 60 teams and 10 independent observations.

At the beginning of each session the experimenter read the neutrally framed instructions for part I aloud. The instructions for part II and for the questionnaires only followed, when the preceding part was finished.[9] In order to test the participants' understanding of the incentive structure and the dilemma situation, subjects were asked to answer four control questions, which are outlined in appendix A.1.3. We can assume full understanding, as we only started the treatments as soon as all participants answered all questions correctly.

Participants were recruited via ORSEE (Greiner, 2004) from a list of voluntary participants and earned on average about 20 € each. They were German-speaking, mostly students of different academic backgrounds except economics, and had not participated in a public goods experiment at the MELESSA laboratory before. The experiment was programmed and conducted with the experiment software z-Tree (Fischbacher, 2007).

[9] Appendix A.1.2 reproduces the instructions for the team treatment. Note that we use the term *group* instead of *unit* in the instructions and contrast it with the term *team*.

1.2.3 Theoretical Predictions

Our main research question asks whether teams are more or less cooperative than individuals in the public goods game. The empirical findings, so far, outlined in the introduction are not fully conclusive. In this section we derive theoretical predictions based on several preference models. According to standard game theoretic predictions, which assume profit-maximization, subjects select $c_i = 0$. This unique Nash equilibrium applies independent of whether the decision-maker is an individual or a member of a team. Based on backwards induction, predictions for the finitely repeated public goods game are the same as for the stage game. Hence, standard theory predicts no differences in cooperation between individuals and teams.

However, it has been shown that empirical data on cooperation can be better predicted by theoretical models that incorporate social preferences, e.g. the model of inequity aversion by Fehr and Schmidt (1999) (for a similar approach see Bolton and Ockenfels, 2000). Fehr and Schmidt (1999) state that a non-selfish player *i* might have a utility loss from inequality of monetary outcomes between herself and the members of a reference group. The utility loss from disadvantageous inequality, meaning that player *i* earns less than other players in the reference group, is higher than the disutility from advantageous inequity, meaning that player *i* earns more. For a public goods game with three unit members the inequity aversion model predicts multiple Pareto-ranked Nash equilibria, given sufficiently strong inequity aversion. The reference group consists of all unit members.

In case of teams interacting with teams, we assume that the median voter model holds, i.e. the final team contribution is equal to the median of the individual contributions (the median level of social preferences). Furthermore we assume that the reference group consists of the two team members of player *i* and the six members of the other two teams in her unit. Hence, altogether player *i* compares her payoff with the payoffs of eight other persons, where the payoff of the other two team members by our unitary team design has to be the same as the payoff of player *i*. The latter implies that the utility loss from inequality of earnings related to the other unit members is mitigated in TEAM compared to IND. This entails that the median team member in TEAM must have a higher disutility from advantageous inequality than an individual in IND to make positive contributions in equilibrium.[10] To summarize, the inequity aversion model by Fehr and Schmidt (1999) predicts potentially lower average contributions in TEAM than in IND.

[10] The formal proof is available from the authors on request. See also Kugler et al. (2007).

Charness and Rabin (2002) present a model of social preferences in which they additionally include potential social welfare concerns. Social welfare concerns lead people to sacrifice own material payoff in order to increase the payoffs of all members of the reference group, i.e. to increase efficiency. As there are more people in treatment TEAM compared to treatment IND who profit from one point contributed to the public good, the efficiency gains per contributed point are higher in treatment TEAM than in treatment IND. In particular, imagine a team in TEAM (compared to an individual in IND) increases their contribution to the public good by one point. This entails, on the one hand, that the profit of two additional persons in TEAM (the other two team members) compared to IND is reduced by 0.5 points each. However, at the same time the profit of four additional persons (from the other two teams in the same unit) is increased by 0.5 points each. This yields in total an additional efficiency gain of 1 point for the whole unit in TEAM compared to IND. Hence, a model that takes into account social welfare predicts higher contributions in TEAM than in IND.

To summarize, the theoretical predictions on cooperation of individuals versus teams based on standard assumptions and on widely used models of social preferences are almost as diverse as the empirical results so far: The standard game-theoretical model of money-maximizing preferences predicts no differences between our two treatments, the model of inequity aversion predicts lower contributions in TEAM than in IND, and the model of social welfare predicts higher contributions in TEAM than in IND.

1.3. Results

In order to examine our research questions we use non-parametric and parametric methods. Among the parametric tests we calculate linear as well as non-linear regression models (tobit and probit). To control for panel effects we also calculate random effects panel regressions. None of the main effects that we observe depends on the specific parametric method used. In the following, we mostly report the results of OLS regressions, because they are based on minimal assumptions. As our dependent variables mostly have a minimum value of zero, which is chosen by a considerable number of participants, we also report hurdle models – sometimes called two-part models – if applicable. For instance, in case of unconditional contributions as the dependent variable the hurdle model captures the decision to contribute (contribution decision) and the decision of how much to contribute (contribution level) in two separate parameters (Cragg, 1971; McDowell, 2003; see also Cameron and Trivedi, 2005). Hence, the hurdle model is a parametric generalization of the tobit model. It usually captures the decision to contribute to the public good by a probit model and the decision on the level of the contribution using an OLS or a tobit regression.[11]

To start with, we report the distributions of cooperative preferences elicited in the first part, which is identical in both of our treatments (based on the classification criteria in Fischbacher et al., 2001). The results are provided in table 1.1.

Table 1.1 - Distributions of cooperative preferences.

	IND	TEAM	Total
Conditional cooperators	42 (63.6%)	120 (66.7%)	162 (65.9%)
Free riders	12 (18.2%)	24 (13.3%)	36 (14.6%)
Others	12 (18.2%)	36 (20.0%)	48 (19.5%)
Total	66	180	246

In line with Fischbacher et al. (2001) we report a considerable degree of heterogeneity in cooperative preferences: Our largest group of subjects are conditional cooperators (65.9%), whose conditional contributions show a significant (Spearman rank correlation, $p < 0.01$) positive correlation with the average contributions by the other two unit members. 14.8% of these conditional cooperators are perfect conditional cooperators; they match the contributions by others perfectly. However, most of the conditional cooperators show a self-serving bias, meaning that their conditional contributions are slightly lower than the average contributions by their unit members (see also Fischbacher et al., 2001, for a similar result). 14.6% of all our subjects fall into the category of free riders, who contribute nothing regardless of the contributions by the other unit members. 19.5% of our subjects cannot be classified in one of the two categories. We do not observe a significant difference in the distribution of cooperative preferences between treatments IND and TEAM (χ^2 test, $p = 0.351$) and can therefore assume that randomization into the treatments was successful.

We organize the presentation of the remaining results as follows: In section 1.3.1 we document differences in the unconditional contributions between individuals and teams, in section 1.3.2 and 1.3.3 we investigate the reasons and drivers behind our finding in 1.3.1, in particular with respect to beliefs and the proposals of team members. Section 1.3.4 concentrates on the team decision-making process. Section 1.3.5 focuses on the team contributions and their determinants. In section 1.3.6 we present the main arguments

[11] For other papers in experimental economics that use a hurdle model to analyze contributions in a public goods game see Botelho et al. (2009) and Nikiforakis (2008).

exchanged within teams for high and low contributions. We conclude with the results of the questionnaire on the team interaction in section 1.3.7.

1.3.1 Unconditional Contributions

In the following we present the results of two-sided Mann-Whitney U-tests (unless otherwise noted), using matching groups, which consist of 6 (in case of condition IND) or 18 subjects (in case of condition TEAM) each, as statistically strictly independent observation.

RESULT 1.1: *Teams compared to individuals are more willing to contribute a positive amount to the public good in the first half of the game.*

Support for result 1.1 is presented in figure 1.1 and table 1.2. Figure 1.1 shows the average contributions by individuals and teams over all periods (for a discussion of the belief data, see section 1.3.2). The contributions by individuals and teams start at a similar level ($p = 0.771$ for the first period), over the following seven periods team contributions seem to be higher than individual contributions, until they decrease under the contribution level of

Figure 1.1 - Contributions and beliefs by individuals and teams.

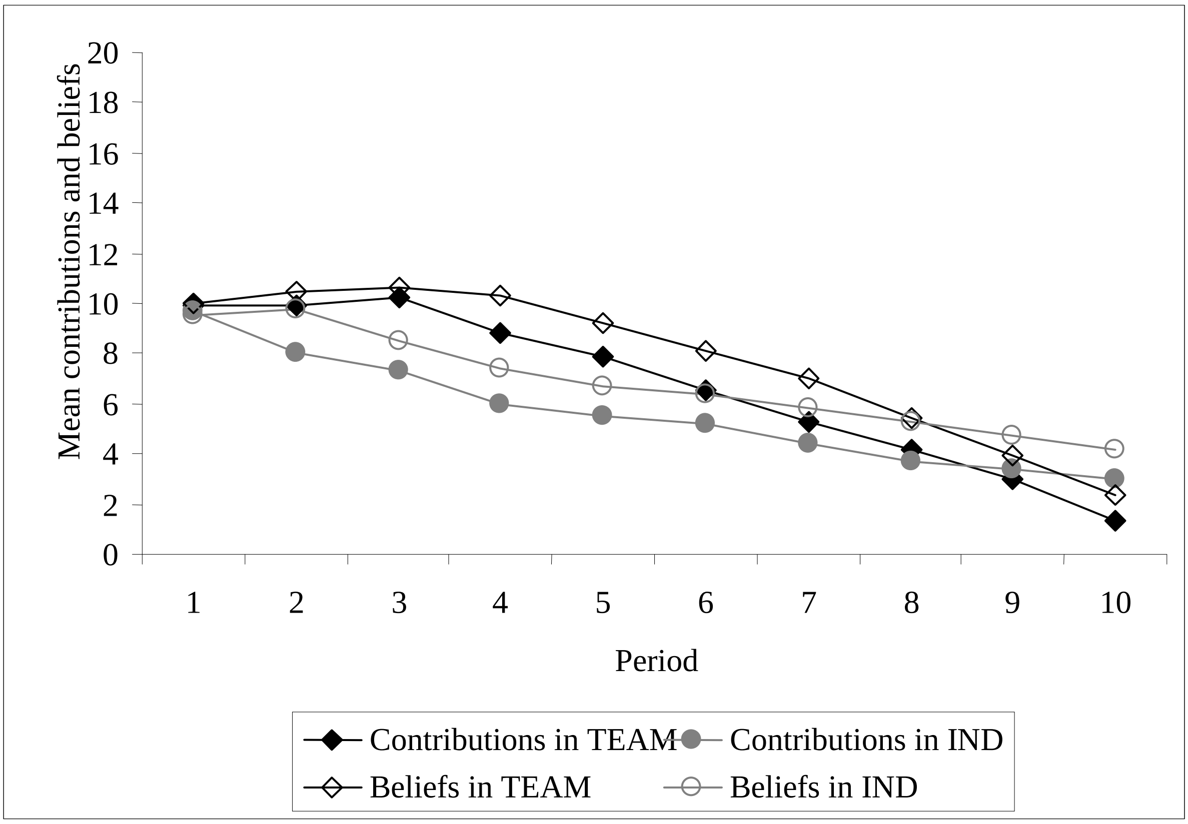


Table 1.2 – Hurdle models of unconditional contributions.

	(I) Periods 1 to 10		(II) Periods 1 to 10		(III) Periods 1 to 5		(IV) Periods 6 to 10		(V) Periods 1 to 10	
Dep. Var.:	Contrib. decision	Contrib. level	Contrib. decision	Contrib. level	Contrib. decision	Contrib. level	Contrib. decision	Contrib. level	Contrib. decision	Contrib. level
TEAM	0.064 (0.088)	0.7439 (1.730)	0.165* (0.082)	1.406 (1.152)	0.114* (0.072)	1.146 (1.389)	-0.001 (0.110)	0.130 (2.464)	0.075 (0.068)	0.126 (0.492)
Period	-0.059*** (0.009)	-0.546*** (0.155)	-0.047*** (0.010)	-0.464* (0.226)	-0.032*** (0.008)	-0.585** (0.240)	-0.083*** (0.015)	-0.538** (0.257)	-0.024** (0.010)	-0.110** (0.043)
Period × TEAM			-0.020 (0.014)	-0.175 (0.303)					-0.009 (0.013)	0.028 (0.093)
Belief									0.036*** (0.004)	0.931*** (0.047)
Constant		11.601*** (1.054)	0.888*** (0.069)	10.740*** (0.851)		11.502*** (0.895)		11.850*** (2.420)	0.517*** (0.081)	1.200** (0.484)
N	1258	825	1258	719	629	501	629	324	1258	825
Wald χ^2	55.82***				19.30***		34.13***			
Log pseudo-likelihood	-731.430				-307.405		-418.874			
R^2		0.066	0.123	0.067		0.030		0.015	0.252	0.622

Notes: This table outlines the results of hurdle models, which consist of a regression of the contribution decision in the first part and an OLS regression of the contribution level (in case of positive contributions) in the second part each. The first part is calculated by a probit regression in specifications (I), (III) and (IV). We report the marginal effects at the mean in these specifications. As marginal effects and standard errors of interaction terms in non-linear models are not correctly specified (Ai and Norton, 2003), we report in specifications (II) and (V) the results of the OLS regressions of the contribution decision. Standard errors (clustered on matching groups) are reported in parentheses. Specifications (III) and (IV) are run with data from periods 1 to 5 and periods 6 to 10 respectively. *, **, *** denote significance at the 10%, 5% and 1% level respectively.

individuals in the last two periods. The difference in the last period however is not significant (p = 0.670 for the last period). Over all periods teams interacting with teams contribute on average 6.7 points, whereas individuals interacting with individuals contribute on average 5.6 points to the public good (p = 0.291 over all periods).

Table 1.2 shows the estimates of hurdle models of unconditional contributions. We use the following independent variables: The dummy variable *TEAM* taking the value 1 if the observation stems from treatment TEAM and 0 otherwise; The variable *Period* to control for time effects; The interaction term *Period* × *TEAM* to control for potential interaction effects; And the variable *Belief* to control for the expectations of the participants.[12,13] Specification (I) shows that treatment TEAM does not have an overall effect on contributions. It further outlines that the willingness and the contribution level both decrease over time, which is in line with numerous existing studies on cooperation in public goods games. Specification (II) reveals, however, a significant effect of the treatment variable *TEAM* on the contribution decision at the beginning: Teams are by 16.5% more likely to contribute a positive amount to the public good than individuals in the beginning. We run with specifications (III) and (IV) two separate hurdle models over the first and the second half of the game. We find that our treatment variable TEAM has an effect on the willingness to contribute in the first half of the game, but not in the second half anymore. In the first half, teams are by 11.4% more willing to make positive contributions than individuals. We observe again a clear trend for the willingness to contribute and the contribution level to decrease over time. If we include beliefs as an additional regressor in specification (V) the effect of the treatment variable TEAM disappears. It seems that our treatment variable has an effect on the beliefs, which serve as a mediating variable between the treatment variable and the contributions. We examine this question in the next section in greater detail. To summarize, we cannot confirm the existence of a discontinuity effect; we rather find a reverse discontinuity effect: Teams interacting with other teams in a public goods game act more cooperatively than individuals interacting with individuals, at least in the first half of the interaction.

1.3.2 Beliefs

Section 1.3.1 indicates that beliefs play an important role when it comes to differences between individual and team cooperation levels. Insko, Schopler, and their colleagues (e.g.,

[12] In case of treatment TEAM we include the average beliefs of the three team members.
[13] The number of observations N in the first part of the hurdle models is N = 1258. This number consists of N = 66 individuals in IND * 10 periods + 60 teams in TEAM * 10 periods – 2 teams in TEAM * 1 period = 1258. Two observations are subtracted, as two teams (once each) did not come to a unanimous decision on the unconditional contribution.

Schopler and Insko, 1992; Wildschut et al. 2003) explain the discontinuity effect with the outgroup schema-based distrust hypothesis, which stipulates that interacting with a team evokes more competitive expectations about the other team's behavior than interacting with an individual. As we rather observe a reverse discontinuity effect, we also expect the opposite link to beliefs.

RESULT 1.2: *Teams are expected to make higher contributions than individuals in the beginning.*

Figure 1.1 displays the expectations of team members with regard to other teams and the expectations of individuals with regard to other individuals over time. Remember that also in the treatment TEAM the beliefs are collected individually.[14, 15] Figure 1.1 shows that – analogous to the dynamics of unconditional contributions – teams are expected to make higher contributions than individuals in the first eight periods, before in the last two periods teams are expected to make lower contributions than individuals ($p = 0.573$ for the last period). Over all periods teams are expected to contribute on average 7.7 points and individuals to contribute on average 6.8 points to the public good ($p = 0.229$ for the first period; $p = 0.307$ over all periods). Note that in line with other studies (e.g. Fischbacher and Gächter, 2010) that compare beliefs and contributions decision makers' beliefs are over-optimistic regarding the cooperation level of others.

Table 1.3 shows the results of OLS regressions of beliefs. The variable *Previous others' contribution* refers to the contribution by the other two unit members (teams in TEAM and individuals in IND) in the previous period. In line with the results on contributions, we find that teams are not expected to make higher contributions than individuals over all periods. However, they are correctly expected in the beginning of the game to make by 2.5 points higher contributions than individuals. Furthermore expectations decrease over time, which is not surprising, as contributions also decrease. Specification (III) further outlines that beliefs increase with the contributions of the other unit members in the previous period – despite the stranger matching protocol. Hence, we cannot confirm the outgroup schema-based distrust hypothesis. In contrast, our results indicate higher expectations with regard to teams than to

[14] Hoyle et al. (1989) find that the crucial element of the outgroup schema-based distrust hypothesis is not whether the self is in a group or alone, but whether the opponent is a group or an individual.

[15] We cannot rule out that the individual belief elicitation in treatment TEAM might be influenced by the chat discussion within teams before the elicitation.

individuals at least at the beginning of the public goods game.[16] Specification (IV) is discussed in the next section.

Table 1.3 - OLS regressions of beliefs.

Dep. Var.: Beliefs	(I)	(II)	(III)	(IV)
TEAM	0.924 (1.515)	2.519** (1.134)	1.553*** (0.393)	-1.371 (2.080)
Period	-0.844*** (0.108)	-0.634*** (0.137)	-0.102 (0.072)	-0.844*** (0.108)
Period $\times$ TEAM		-0.288 (0.191)	-0.269*** (0.077)	
Previous others' contribution			0.372*** (0.016)	
Conditional contributor				-1.194 (1.274)
Others				2.655 (1.329)
Conditional contributor $\times$ TEAM				2.654* (1.458)
Others $\times$ TEAM				2.923* (1.626)
Constant	11.461*** (1.233)	10.306*** (0.878)	2.742*** (0.443)	12.449*** (1.821)
N	2460	2460	2214	2460
R^2	0.208	0.212	0.699	0.216

Notes: This table shows the results of OLS regressions of beliefs. Specification (III) is calculated for data from periods 2 to 10. Standard errors (clustered on matching groups) are reported in parentheses. *, **, *** denote significance at the 10%, 5% and 1% level respectively.

1.3.3 Individual Proposals of Team Members

In each period we ask team members in condition TEAM to make an initial individual proposal for the team contribution. In the following we first examine whether there is an effect of the team situation on these individual proposals of team members. Therefore we compare the individual proposals of team members in treatment TEAM with the actual individual contributions in IND. We come to the following result:

RESULT 1.3.1: *The team setting has an effect on the individual proposals of team members: The individual proposals of team members are more often greater than zero than the actual contributions in IND.*

[16] We also examine the question whether the team situation has an effect on the variance of beliefs within teams over time. An OLS regression of the standard deviations of beliefs within teams on the time periods shows that the beliefs within teams converge over time ($p < 0.05$). We will see in the next section that this result is in line with the progress of the variance of individual proposals within teams over time.

In order to illustrate result 1.3.1 we outline in figure 1.2 the average contributions in IND and in TEAM and the average individual proposals of team members in TEAM. The graph shows that the proposals and contributions in TEAM are close to each other, the contributions in IND seem to be lower than the proposals of team members in TEAM in each period ($p = 0.121$ over all periods). In line with the team contributions the proposals of team members also decline over time.

Figure 1.2 - Contributions and team members' proposals.

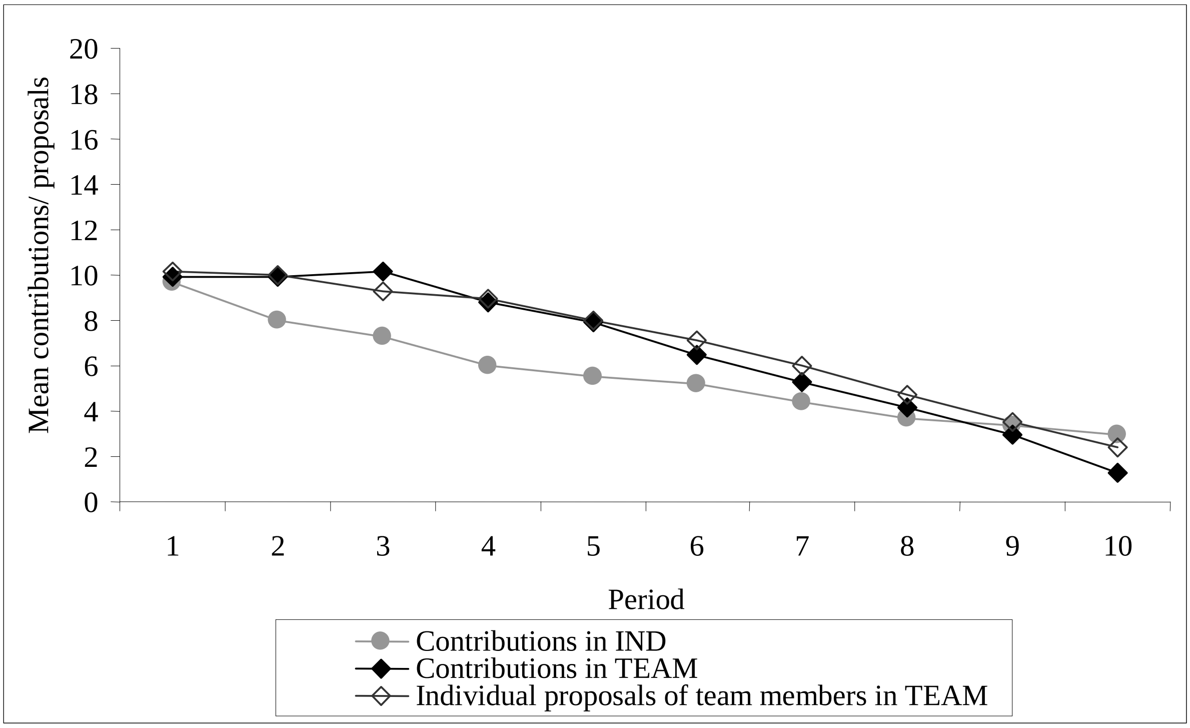


Table 1.4 shows the results of hurdle models on the effect of the team situation on the individual proposals of team members. These results correspond heavily to the results on the effect of our treatment variable TEAM on contributions: According to specification (I) we do not find an overall effect of the team setting on the individual proposals of team members. However, specification (II) shows that there is an effect in the beginning, meaning that team members in TEAM are by 16.1% more likely to *propose* a contribution greater than zero compared to individuals. Note that the proposals of team members at the beginning of the game can not be influenced by the communication within teams. We run separate hurdle models and find an effect of the team setting on the proposals of team members in the first 8 periods (see specification III). Only in the last two periods there are no differences anymore between team members' proposals and individual contributions.

Table 1.4 - Hurdle models of contributions in in IND and team members' proposals in TEAM.

	(I) Periods 1 to 10		(II) Periods 1 to 10		(III) Periods 1 to 8		(IV) Periods 9 to 10	
Dep. Var.:	Likelihood of positive contri-bution/	Contri-bution/ proposal level	Likelihood of positive contri-bution/	Contri-bution/ proposal level	Likelihood of positive contri-bution/	Contri-bution/ proposal level	Likelihood of positive contri-bution/	Contri-bution/ proposal level
TEAM	0.099 (0.086)	0.807 (1.602)	0.161** (0.074)	0.916 (1.002)	0.124* (0.079)	0.926 (1.528)	-0.041 (0.116)	0.139 (2.458)
Period	-0.060*** (0.008)	-0.481*** (0.123)	-0.047*** (0.010)	-0.464* (0.225)	-0.041*** (0.008)	-0.512*** (0.165)	-0.147*** (0.040)	0.586 (0.787)
Period × TEAM			-0.013 (0.013)	-0.022 (0.269)				
Constant		11.286*** (1.250)	0.888*** (0.069)	11.205*** (0.832)		11.303*** (1.245)		1.896 (7.484)
N	2460	1705	2460	1705	1968	1507	492	198
Wald χ^2	95.28***				72.23***		16.88***	
Log pseudo-likelihood	-1346.002				-1000.833		-325.762	
R^2		0.053	0.134	0.054		0.045		0.002

Notes: This table outlines the results of hurdle models, which consist of a regression of the contribution decision (likelihood of a positive proposal) in the first part and an OLS regression of the contribution (proposal) level (in case of positive contributions/ proposals) in the second part. Except for specification (II) the first part is calculated by a probit regression respectively. For the first part we report the marginal effects at the mean in these specifications. As marginal effects and standard errors of interaction terms in non-linear models are not correctly specified (Ai and Norton, 2003), we report in specification (II) the results of the OLS regression of the contribution decision (likelihood of a positive proposal). Standard errors (clustered on matching groups) are reported in parentheses. Specifications (III) and (IV) are run with data from periods 1 to 8 and periods 9 to 10 respectively. *, **, *** denote significance at the 10%, 5% and 1% level respectively.

In the following we examine whether certain types of players drive result 1.3.1 and therefore also the reverse discontinuity effect. We come to the following result:

RESULT 1.3.2: *The reverse discontinuity effect is driven by conditional cooperators, who – in comparison to free-riders – are more optimistic with respect to the contributions of other parties in the team setting and therefore intend to make even higher contributions in the team setting than in the individual setting.*

Figure 1.3 shows the individual contributions in IND and the proposals of team members in TEAM for the different types of cooperative preferences. First of all, we see that also among free riders there are subjects who make positive contributions or contribution proposals. This is in line with the findings by Dariel and Riedl (2010), who find that free riders do not necessarily contribute less than other cooperation types in the repeated PGG (with a stranger matching applied). Figure 1.3 indicates that in condition IND contributions are very similar between cooperation types (Kruskal Wallis rank test for the first period, p = 0.264), whereas in condition TEAM proposals depend heavily on the cooperation type (Kruskal Wallis rank test for the first period, p = 0.035). The proposals of free riders in TEAM are slightly lower than the contributions in IND, whereas the proposals of conditional cooperators are higher than the corresponding contributions in IND (Mann-Whitney U-tests for the first period: among free riders p = 0.237; among others p = 0.192; among conditional cooperators p = 0.042).

Figure 1.3 - Contributions in IND and team members' proposals in TEAM for each cooperation type.

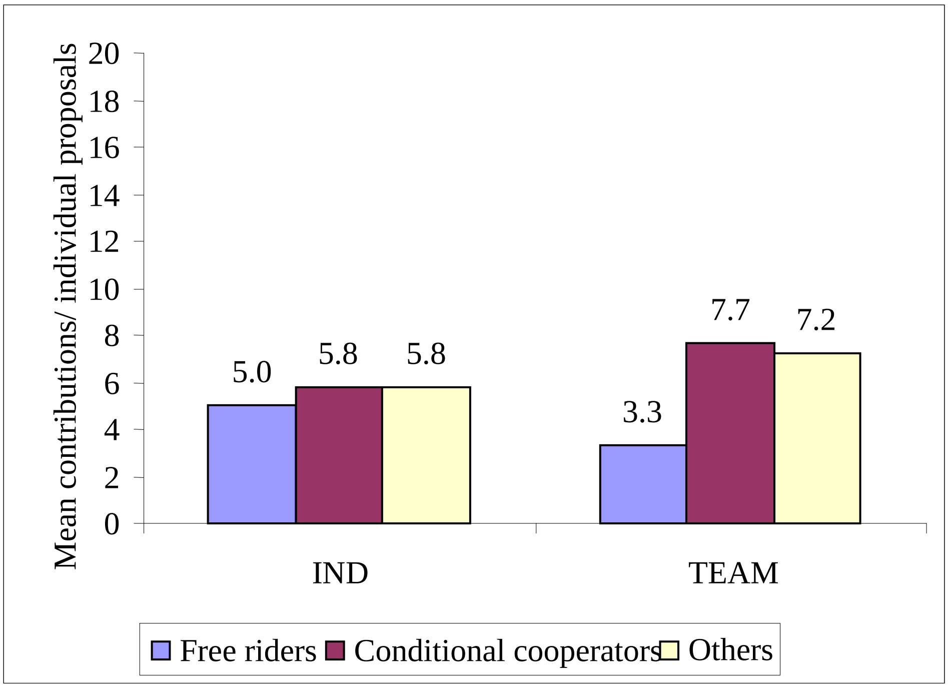


Table 1.5 displays the results of OLS regressions of contributions in condition IND and proposals of team members in condition TEAM. As regressors we insert: The dummy variable *Conditional cooperator*, which takes value 1 if the observation comes from a decision maker classified as conditional cooperator and zero otherwise; A dummy variable for the residual category *Others*; And the interaction terms between each of these dummy variables and our treatment variable *TEAM*. Specification (I) confirms that conditional cooperators as well as subjects classified as others make by about 3 points higher contributions to the public good than free riders.

Table 1.5 - OLS regressions of contributions in IND and individual team members' proposals in TEAM

Dep. Var.: Contributions in IND and team members' proposals in TEAM	(I)	(II)
TEAM	1.239 (1.511)	-1.700 (2.666)
Period	-0.840*** (0.077)	-0.840*** (0.077)
Conditional cooperator	3.262*** (0.953)	0.780 (2.045)
Others	2.903*** (1.007)	0.817 (1.574)
Conditional cooperator $\times$ TEAM		3.645[#] (2.236)
Others $\times$ TEAM		3.108 (1.954)
Constant	7.633*** (1.849)	9.593*** (2.476)
Conditional cooperator - Others	0.359	-0.037
Conditional ooperator $\times$ TEAM - Others $\times$ TEAM		0.537
N	2460	2460
R^2	0.174	0.182

Notes: This table outlines the results of OLS regressions of contributions in IND and individual proposals of team members in TEAM. Standard errors (clustered on matching groups) are reported in parentheses. *, **, *** denote significance at the 10%, 5% and 1% level respectively, [#] denotes p = 0.119.

Specification (II) shows that there are no significant differences between the different types in condition IND respectively (captured by the variables *Conditional cooperator*, *Others* and *Conditional cooperator - Others*). Variable TEAM indicates that there are no differences between the individual contributions and proposals of team members among free-riders. However, we find a weak interaction effect between the treatment variable and the conditional contributors: Compared to free-riders the team situation increases the proposals of conditional cooperators. Hence, the effect of higher contribution proposals in TEAM in contrast to the individual contributions in IND – which leads to the reverse discontinuity effect – is driven by conditional cooperators.

Why do conditional cooperators make higher contribution proposals in the team setting than they actually contribute if they play as an individual? Specification (IV) in table 1.3 suggests the following explanation: In contrast to the beliefs of free-riders the team setting has a positive effect on the beliefs of conditional cooperators: Other than free-riders, conditional cooperators are more optimistic in the team setting than in the individual setting with respect to the contributions of the other parties.[17]

1.3.4 Team Decision-Making Process

In the following we concentrate on the decision-making process within teams. We start by examining whether team members adapt to the other team members over time.

RESULT 1.4.1: *Over time the individual proposals of team members converge within teams.*

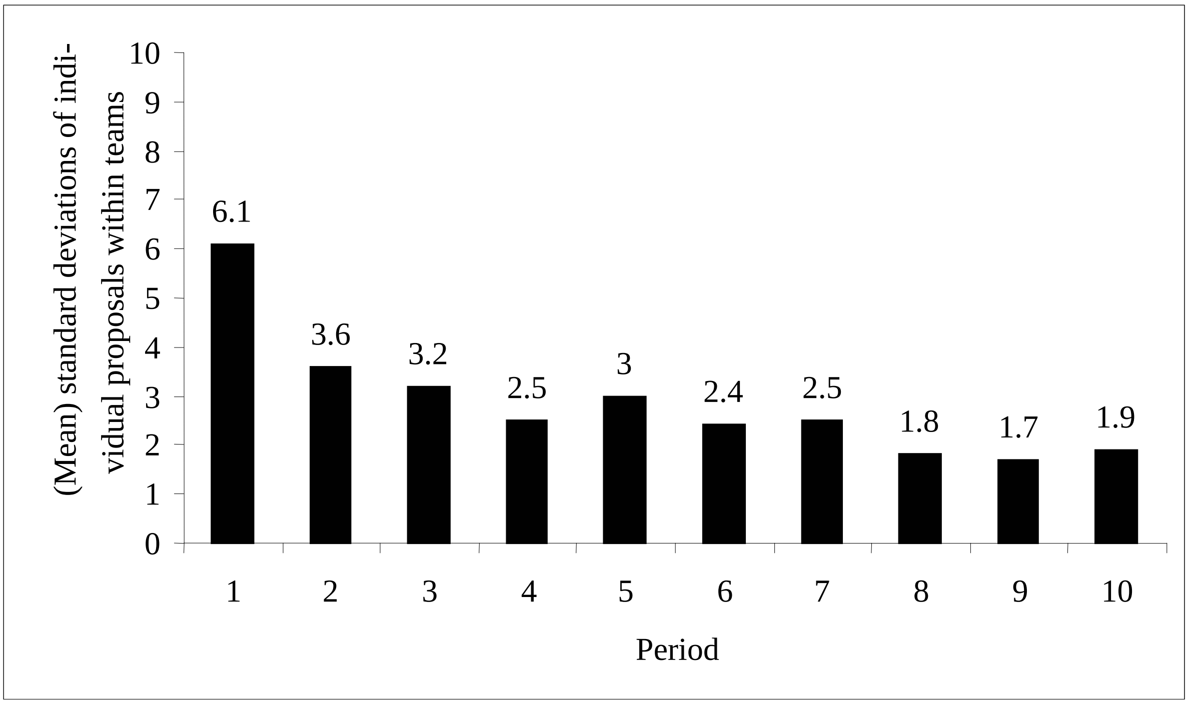
Support for result 1.4.1 is provided by figure 1.4 illustrating the mean standard deviations of individual proposals within teams for each period. It shows that the standard deviations of team members' proposals decrease over time.

In table 1.6 we present different team compositions with respect to cooperative preferences of team members. The abbreviation *cc* stands for conditional cooperator, *fr* for free rider and *other* for the residual category. We find that most teams either consist of three conditional cooperators (*3cc*), or of two conditional cooperators and one free rider (*2cc/1fr*) or one person classified as other (*2cc/1other*). There are only few mixed teams (*1cc/1fr/1other*) or teams in which other types have the majority (*at least 2others*). Hence, conditional cooperators constitute a majority in most teams. Table 1.6 further outlines the decision rules that teams implicitly or explicitly apply to aggregate individual proposals and to come to a unanimous decision.

RESULT 1.4.2: *Teams mainly come to a decision concerning the team contribution by making a compromise or by deciding according to the majority rule if the individual proposals differ a priori.*

[17] For an illustration of the beliefs in IND and TEAM for the different cooperation types see Figure A.1.1 in appendix A.1.4

Figure 1.4 - Standard deviations of individual proposals within teams.



Only six teams determine a constant contribution in the first periods to which they stick until the last period. Most teams, however, bargain on the team contribution in each period. Hence, we report in the following the decision rules for each case, i.e. for each team in each period. Table 1.6 outlines the six categories of decision rules that we can distinguish in our data and the corresponding average contributions. In 30.6% of the cases teams find a compromise, meaning that they either take the average of individual proposals, the median or they decide to contribute another amount that lies between the highest and the lowest proposal within the team. In 27.3% of the cases two team members make the same proposal and the third team member agrees to this proposal. We define this as a decision under majority rule. Our third category of decision-making is a decision with identical proposals, meaning that all three team members make identical individual proposals and make this proposal the team decision. A decision with identical proposals occurs in 22.2% of all cases. In some cases the least cooperative proposal or the most cooperative proposal within a team is taken as the team decision. The residual category summarizes the cases that cannot be assigned to one of the decision rules so far. Table 1.6 shows that the average team contributions differ quite profoundly depending on which decision rule is applied. If one of the extreme proposals is chosen as the team contribution, the contribution levels are correspondingly rather high or low. Identical proposals also lead to rather low contributions, which at least partly might be related to the fact that identical proposals occur more often in later periods in which cooperation has already decayed. We come to a more detailed analysis of the effect of different decision rules on team contributions in section 1.3.5.

Table 1.6 - Team compositions, contributions and decision rules applied.

	Number of teams	Average contribut.	Frequency of decision rules applied					
			Compro-mise	Majority	Identical proposals	Least coop. proposal chosen	Most coop. proposal chosen	Others
3cc	16 (26.7%)	8.7	65	41	21	14	14	4
2cc/1fr	16 (26.7%)	4.9	47	46	44	15	7	0
2cc/1other	15 (25.0%)	8.1	33	48	35	22	6	6
1cc/1fr/1other	6 (10.0%)	2.1	13	11	22	9	1	4
At least 2others	7 (11.7%)	7.3	25	17	11	8	6	3
Total	60		183 (30.6%)	163 (27.3%)	133 (22.2%)	68 (11.4%)	34 (5.7%)	17 (2.8%)
Average contribution			8.5	7.1	3.9	3.4	12.8	5.3

Note: *cc* stands for conditional cooperator, *fr* for free rider and *other* for a subject from the residual category.

We can now examine whether the team composition has an influence on the applied decision rule.

RESULT 1.4.3: *The decision-making process in teams is affected by the cooperative preferences of team members. In contrast to other teams, teams consisting of three* *conditional cooperators are most willing to accept a compromise and are least likely to have identical proposals.*

Table 1.7 outlines the results of probit regressions which examine the effect of different team compositions on the likelihood of the three most frequently applied decision rules. Teams consisting of three conditional cooperators are most willing to make a compromise in their decision-making process compared to other teams. At the same time the proposals of their team members are least frequently identical. Pertaining to the majority rule we do not find any significant effects of different team compositions. In line with the results mentioned above table 1.7 outlines that the likelihood of identical proposals increases over time.

Table 1.7 - Probit regressions of three decision rules.

Dep. Var.:	Likelihood of a compromise	Likelihood of a majority	Likelihood of identical proposals
2cc/1fr	-0.111*** (0.041)	0.031 (0.050)	0.169*** (0.063)
2cc/1other	-0.178*** (0.031)	0.061 (0.043)	0.127*** (0.059)
1cc/1other/1fr	-0.175* (0.070)	-0.078 (0.075)	0.288*** (0.134)
At least 2others	-0.046 (0.090)	-0.016 (0.061)	0.025 (0.055)
Period	-0.042*** (0.009)	0.005 (0.009)	0.051*** (0.010)
N	598	598	598
Wald χ^2	79.30***	5.96	60.88***
Log pseudo-likelihood	-340.410	-347.555	-268.050

Notes: This table outlines the results of probit regressions of the most often applied decision rules. We report the marginal effects at the mean. The team composition *3cc* is used as base category. Specifications (II) are calculated over the periods 2 to 10 each. Standard errors (clustered on matching groups) are presented in parentheses. *, **, *** denote significance at the 10%, 5% and 1% level respectively.

1.3.5 Team Contributions

What are the determinants that affect the final aggregated team decision? We are especially interested in the effects of the team composition, of the different types of team members' proposals, of the length of the chat communication and of the decision rules. We start with the effect of the team composition.

RESULT 1.5.1: *The composition of cooperative preferences within teams has an effect on the team contributions: Teams with three conditional cooperators are more willing to make positive contributions and, if so, they make higher contributions compared to other teams.*

Result 1.5.1 seems plausible. Econometric support is provided by specification (I) in table 1.8. Teams with three conditional cooperators are the ones with the highest contributions. They are by 26.8% or 55.4% more likely to make positive contributions than teams with two conditional cooperators and one free rider (*2cc/1fr)* or mixed teams *(1cc/1resid/1fr)* respectively. If they make positive contributions, their contributions are by 3.1 or 5.4 points higher than the ones made by teams with two conditional cooperators and one free rider (*2cc/1fr)* or mixed teams *(1cc/1resid/1fr)* respectively.

Which types of proposals influence the team decision?

RESULT 1.5.2: *The lowest and the median proposals within teams have a significantly positive effect on the final team decision, while the highest proposal does not. The final team decision is also affected by the first proposal that is made during the chat communication.*

In specification (II) of table 1.8 we insert the variables *Lowest proposal*, *Median proposal* and *Highest proposal* – that represent the lowest, median and highest proposal within teams in each period – and the variable *First proposal in chat*, which refers to the first specific proposal that is mentioned in the chat communication in each period. Each additional point of the lowest proposal in teams increases the willingness to contribute by 2.6%, each additional point of the median proposal increases the contribution level in those teams who make a positive contribution by 0.6 points. Furthermore, in teams who make a positive contribution the contribution level increases by 0.2 points with each additional point of the first proposal mentioned during the chat communication. The latter influence shows that teams often stick to the first proposal that was made by a team member in the chat. This is in line with the well-known *anchoring effect* in social psychology (Galinsky and Mussweiler, 2001; Tversky and Kahneman, 1974).

Note that the highest proposal has no significant effect on team contributions. One explanation might be that the highest proposals are rather made for social image concerns: team members pretend to be cooperative but quickly decline their proposal if they realize that other team members are less cooperative.

Specification (II) of table 1.8 also includes the variable *Number of words in chat*, which denotes the number of words team members exchange within one period in order to come to a team decision.

RESULT 1.5.3: *When teams make a positive team contribution, the contribution level increases with the length of the team discussion.*

On average teams use 86 words in the chat to come to a unanimous decision, whereas the range lies between zero words (which is possible if the individual proposals are identical) and 372 words.[18] Specification (II) of table 1.8 outlines that if teams make positive contributions the contribution level increases by 0.008 points with each word exchanged. Hence, teams make higher contributions if they discuss or reflect the situation more thoroughly.[19]

RESULT 1.5.4: *The applied decision rule has an effect on team contributions: In particular, decisions based on a compromise or on the majority rule lead to higher contributions than decisions based on identical proposals.*

Specification (III) of table 1.8 shows the results of an OLS regression of team contributions on decision rules.[20] Teams that decide by compromise, majority rule or choose the most cooperative proposal in the team make by 1.5 to 6.2 points higher contributions than teams whose team members' proposals concerning the team contribution are identical. The effect of choosing by compromise is not significantly different from the effect of applying the majority rule. As we control for time periods, the effect is not due to the fact that identical proposals within teams increase over time. The result indicates that if team members make identical proposals concerning the team contribution, these proposals are comparably low.

[18] Not surprisingly, OLS regressions of the length of the team discussion show that the latter increases with the variance of the individual proposals within teams ($p < 0.01$).
[19] A chat analysis shows that teams in the chat mainly discuss the team contribution and hardly ever drift to topics that have nothing to do with the team contribution. This behavior is incentivized by the time restriction.
[20] We cannot run a hurdle model to examine the effect of decision rules on team contributions, as some rules predict the probability of making a positive contribution perfectly.

Table 1.8 - Hurdle models and an OLS regression of team contributions.

	(I)		(II)		(III)
Dep. Var.:	Contribution decision	Contribution level	Contribution decision	Contribution level	Contribution level
2cc/1fr	-0.268*** (0.092)	-3.116** (1.250)			
2cc/1resid	-0.137 (0.108)	0.578 (1.744)			
1cc/1resid/1fr	-0.554*** (0.063)	-5.388*** (1.453)			
At least 2resid	-0.106 (0.164)	-0.908 (1.247)			
Lowest proposal			0.026*** (0.016)	0.033 (0.043)	
Median proposal			0.006 (0.005)	0.636*** (0.066)	
Highest proposal			0.003 (0.004)	0.116 (0.078)	
First proposal in chat			0.002 (0.002)	0.204** (0.070)	
Number of words in chat			0.000 (0.000)	0.008** (0.002)	
Compromise					2.041* (1.013)
Majority					1.483* (0.770)
Least cooperative proposal					-2.045 (1.236)
Most cooperative proposal					6.214*** (1.210)
Others					-0.032 (1.972)
Period	-0.076*** (0.013)	-0.706*** (0.170)	-0.001 (0.003)	-0.084	-0.873*** (0.092)
Constant		14.094*** (0.808)		-0.074 (0.594)	10.356*** (1.102)
Compromise - Majority					0.558
N	598	410	598	410	598
Wald χ^2	229.15***		274.39***		
Log pseudo likelihood	-290.310		-169.472		
R^2		0.183		0.825	0.272

Notes: This table outlines the results of hurdle models and an OLS regression of team contributions. The hurdle models consist of a probit regression of the decision in the first part and an OLS regression of the contribution level in the second part each. For the first part we report the marginal effects at the mean. We apply *3cc* and *Identical proposals* as base category. Standard errors (clustered on matching groups) are reported in parentheses. *, **, *** denote significance at the 10%, 5% and 1% level respectively.

1.3.6 Arguments in Teams for High and Low Contributions

In order to get further insights into the driving forces behind team decisions we conduct a qualitative content analysis (see also Cooper and Kagel, 2005). The procedure is the following: One researcher reads the whole chat communication and develops a category system in a deductive and inductive way on the reasons for high and low team contributions. We then train two undergraduate research assistants for the coding. Independently from each other the two research assistants read the whole team communication and code for each team in each period – hence, for all cases – whether an argument from the category system was mentioned or not. A significant kappa coefficient of 27.8% ($p < 0.001$) is reached, which indicates a fair inter-rater agreement according to Landis and Koch (1977). Results reported in the following are based on the average of the two codings.

The arguments for low and high team contributions are enlisted together with their frequency in tables 1.9 and 1.10 respectively. Both tables also display the average contribution over all cases, in which the corresponding argument is mentioned. Not surprisingly, contributions are clearly higher (with a range from 9.8 to 12.3 points) if an argument for high contributions is mentioned than if an argument for low contributions is advanced (range from 3.7 to 7.1 points).

Among the two most frequently used arguments for low contributions is the claim that other teams also make low contributions to the public good, which is mentioned in 18.3% of all cases. This argument is based on the fairness norm of inequity aversion (Fehr and Schmidt, 1999; see also Bolton and Ockenfels, 2000) which states that teams have to bear a utility loss if they contribute more and therefore earn less than other teams (disadvantageous inequity aversion). The second most frequently used argument for low contributions – mentioned in 13.6% of all cases – is the standard game theoretic argument: Low contributions are the best response to maximize profits independent of what others do.

Arguments most frequently discussed within teams in favour of high contributions are efficiency concerns, fairness concerns in the sense of advantageous inequity aversion and, interestingly, the reputation building argument (which can only indirectly apply). Most frequently, namely in 8.9% of all cases, teams discuss efficiency concerns in the sense of social welfare, which heavily supports the hypothesis that teams are more willing to cooperate than individuals because of efficiency reasons. Of course, we do not know, whether individuals thought about efficiency aspects to the same extent, but at least we can say that efficiency concerns are an important argument of teams in favour of high contributions. The argument that other teams make also high contributions – which reflects the utility losses of teams if they make lower contributions than other teams (advantageous inequity aversion) – is mentioned in 6.5% of all cases.

Table 1.9 - Arguments for low contributions in the within-team communication.

Category	Argument	Explanation of Argument	Frequency (relative frequency)	Average contribution
1a	Disadvantageous inequity aversion	Other teams also make low contributions	109.5 (18.3%)	4.9
1b	Standard game theoretic approach	Pure profit-maximization independent of what others do	81.5 (13.6%)	4.0
1c	Loss aversion	High contributions are risky	24 (4.0%)	7.1
1d	Trial and error	Very low contribution proposed to examine consequences	24 (4.0%)	4.9
1e	Stranger matching	In each period other teams are matched	11 (1.8%)	5.8
1f	Free-riding is not revealed	Other teams only get feedback on the average of the other parties' contributions and not on single team contributions	13.5 (2.25%)	3.7

Table 1.10 - Arguments for high contributions in the within-team communication.

Category	Argument	Explanation of argument	Frequency (relative frequency)	Average contribution
2a	Efficiency	Social welfare	53.5 (8.9%)	10.0
2b	Advantageous inequity aversion	Other teams also make high contributions	39 (6.5%)	9.8
2c	Reputation building	Building up a cooperative reputation for profit-maximization of the own team in the long run	30 (5.0%)	11.6
2d	Trial and error	Very high contribution proposed to examine consequences	25.5 (4.3%)	11.6
2e	Stranger matching	Teams are rematched in each period. Hence, if the other assigned teams make low contributions in the previous period, the own team gets a new chance in each period and can motivate the other teams to make high contributions	31.5 (5.3%)	12.3

1.3.7 Post-Experimental Questionnaire on Team Decisions

After having finished part II subjects in condition TEAM individually answered a post-experimental questionnaire on the team decision. Appendix A.1.5 lists the questions and the answers to four sets of questions. In the following we concentrate on the most interesting questions.

The first set of questions focuses on the contributions by individuals and teams. 43.9% of the subjects accurately expect teams to contribute rather more to the public good than individuals. The most frequently stated reasons are the perception of peer pressure within teams to make higher contributions and the outvoting of low contributors within teams by means of the majority rule.

The second set of questions concerns the atmosphere within teams. Question Q3 reveals that 54.5% of the subjects are rather satisfied with the decisions of their team. Among the reasons the soft factors – for instance, a similar strategy within the team, a quick agreement and a democratic decision-making process – are equally important compared to the hard factor of high profits. Q5 reveals that 89.0% of the team members perceive the atmosphere in their team as positive. Only 20.0% of team members would change into another team, if they had the chance (Q6). A probit regression of Q6 on the team profits, the closeness between the individual proposal and the team contribution, the team atmosphere (Q5) and the subjective influence of the team member on the team decision (Q7) reveals that only the latter three soft factors have a positive effect on the wish of subjects to stay within their team ($p < 0.05$ respectively). The team profit has no effect ($p > 0.10$).

In the third set of questions subjects were asked about the team decision-making process. The mentioned decision rules in Q8 confirm the results in section 1.3.4. Q12 shows that 62.2% of the subjects would prefer to take decisions rather in teams than individually, if they could participate in a similar experiment again. A probit regression of Q12 on several regressors including the team profits and the subjective team atmosphere (Q5) shows that only the team atmosphere has a significant effect ($p < 0.01$), whereas the team profit has no effect on whether a subject prefers to decide within a team or on her own.

1.4 Conclusion

We conduct a repeated public goods game with teams who interact with teams and individuals who interact with individuals. We further independently elicit individual cooperative preferences to examine their aggregation into team decisions. To date, the empirical results on team and individual decisions in games with a social component are inconclusive: Several experiments suggest that teams are more competitive, more selfish and less cooperative than individuals. However, other studies stipulate the contrast. Furthermore the theoretical predictions on team versus individual decisions are very different depending on the specific (social preference) model applied. The results in our study do not confirm the discontinuity effect that has been a standing finding in social psychology. Instead we rather find a reverse discontinuity effect with teams to be more willing to cooperate than individuals, at least in the first half of the public goods game. Teams interacting with other teams are more cooperative and less competitive than individuals interacting with other individuals.

We can speculate about the reasons for the finding of a reverse discontinuity effect based on some additional controls in our experiment. Firstly, contributions in the public goods game strictly increase social welfare. Furthermore our chat analysis strongly suggests that efficiency-concerns of teams might be an important reason for their enhanced willingness to cooperate. This interpretation is in line with Balafoutas et al. (2014), who find that teams are more efficiency-loving than individuals. A second explanation for the reverse discontinuity effect is related to the more optimistic expectations towards other teams than towards other individuals. Hence, we cannot corroborate the so-called schema-based distrust hypothesis that is a prominent explanation for the discontinuity effect in social psychology. Especially conditional cooperators (in contrast to free-riders) have more optimistic expectations towards other teams than towards other individuals. Our chat analysis suggests that these more optimistic expectations combined with advantageous inequity aversion might lead to an increased willingness to cooperate in the team setting. A third explanation that we cannot completely rule out might be an enhanced motivation of teams to (indirectly) build up a cooperative reputation, despite the stranger matching used in our experiment.

In our study we further examine the team decision-making process. We find that the individual proposals of team members converge within teams over time. The aggregation of individual proposals within teams either happens by compromise, majority rule or the proposals are identical anyway, where the applied decision rule is affected by the distribution of cooperative preferences of team members: Teams consisting only of conditional cooperators decide more often by using compromise, and their team members' proposals are less frequently identical than in other teams. The final team decision is not only affected by the team composition in terms of cooperative preferences and the implicit decision rules applied. It also depends on the first proposal verbalized in the team discussion (anchoring effect) and increases with the length of the discussion. The latter result reflects that teams who deliberate on the social dilemma more thoroughly make higher contributions.

Our results have relevant implications when thinking about the numerous public goods situations in reality in which decisions are not taken by individuals but by teams. However, it is important to highlight the stand-alone nature of our research in economics. There is much room for more studies. For instance, reputational concerns and efficiency concerns in teams should be analyzed comprehensively in the future. It would also be interesting to study team decision making “in the wild” by implementing field experiments that involve team decision making. Further, a careful study that varies the communication protocols and the communication media is desperately needed in economics. Finally, the effects of group size and asymmetric interaction (direct interaction of individuals and teams) have not been investigated in a social dilemma.

Appendix A.1.1: Literature overview

Table A.1.1 – Studies on individual and team decisions.

Authors	Game	Time horizon	Communication within teams	Results: Decisions of teams compared to individuals
Cason and Mui (1997)	Dictator game	One-shot	Face-to-face	In teams with different individual choices, the team choices are less self-regarding than the individual choices
Luhan et al., 2009	Dictator game	One-shot	Chat	Teams send smaller amounts
Bornstein and Yaniv (1998)	Ultimatum game	One-shot	Face-to-face	Teams send and accept smaller amounts
Kugler et al. (2007)	Trust game	One-shot	Face-to-face	Teams trust less
Cox (2002)	Trust game	One-shot	Face-to-face	Teams are less trustworthy
Bornstein et al. (2004)	Centipede game	One-shot	Face-to-face	Teams exit earlier
Kocher and Sutter (2007)	Gift exchange game	One-shot	Face-to-face/ chat	Teams with chat communication offer lower wages, teams with face-to-face communication make higher efforts
Bosman et al. (2006)	Power-to-take game	One-shot	Face-to-face	No significant differences
Psychological studies on the discontinuity effect by Insko and Schopler	Prisoner's dilemma game	Typically 10 periods in partner matching	Face-to-face	Teams cooperate less
Müller and Tan (2013)	Stackelberg market game	One-shot / 15 periods in stranger matching	Chat	In the treatment with 15 periods: Teams act farther away from standard game theoretic predictions
Balafoutas et al. (2014)	Kerschbamer's double price-list technique	One-shot	Chat	Teams are classified more frequently as efficiency-loving and less frequently as spiteful

Appendix A.1.2: Instructions for treatment TEAM (translated from German)

Welcome to the experiment! Thank you very much for participating!

From now on, please refrain from talking with the other participants!

General information

The purpose of this experiment is to study decision making. You can earn money during this experiment which you will be paid in cash at the end of the experiment.
During the experiment you and the other participants will take decisions. Your own decisions as well as those of the other participants will affect your payment according to the rules explained in the following.
The duration of the experiment is 120 minutes. If you should have any questions, please raise your hand. One of the experimenters will come to you and answer your question in private.
During the experiment your earnings will be calculated in **points.** At the end of the experiment your earnings in points will be converted into Euros according to the following exchange rate:

1 point = 0.33 Euros (3 points = 1 Euro)

For your punctual arrival you receive 100 points (4 Euros). At the end of the experiment you will be paid your total earnings from the experiment as well as the 4 Euros in private and in cash. For linguistic simplicity we only use male notations in these instructions.

Anonymity
Your input over the course of the experiment is anonymous. You will not receive personal information about the other participants. We will not link your name to data from the experiment. At the end of the experiment you will sign a confirmation which says that you have received the payment. This confirmation only serves accounting purposes for our sponsor. Our sponsor will not receive any data from the experiment either.

Utilities
For this experiment you are given a pen. We kindly ask you to leave the pen at your place after the experiment. While you take your decisions a clock will count down at the upper right corner of the computer screen. This countdown serves as an orientation for how much time you should need to take the decision. However there is no strict time limit. You can take more time in order to make your decision, especially in the beginning this will be the case. Only the purely informational screens without any decisions to be taken will close after the countdown.

The experiment
The experiment consists of three parts. You will receive the instructions for the second and third part after completion of the first and second part respectively. The parts are independent of each other; decisions in one part have no influence on your income in another part. The sum of your incomes from the first two parts yields your total income from the experiment. You receive information about your total income after completion of the second part.

Part I

The general decision situation

First we explain you the general decision situation. Afterwards you will find some sample exercises on the screen which serve to increase your familiarity with the decision situation. Then we will outline the procedure of part I in detail.

You are a member of a 3-person-group. Each member of the group has to decide on the division of 20 points that you receive at the beginning of part I. You can either allocate the 20 points on a private account or you can assign them completely or parts of them on a public account. Each point that you do not allocate to the public account is automatically assigned to the private account.

Income from the private account
For each point that you allocate to the private account you earn exactly one point. For example, if you allocate 20 points to the private account, you earn exactly 20 points from the private account. If you assign, for example, 6 points on the private account, you receive an income of 6 points from the private account. *Nobody except you receives an income from your private account.*

Income from the public account
From the amount that you allocate to the public account all group members profit in the same way. Hence, you also earn money from the contributions of the other two group members to the public account. The income from the public account, that each group member receives, is determined in the following way:

Individual income from the public account =
*Sum of contributions of all three group members to the public account * 0.5*

For example, if the sum of contributions in your group is 60 points, then you and the other two group members receive 60 * 0.5 = 30 points each from the public account. If the three group members allocate a total of 10 points to the public account, you and the other two group members receive 10 * 0.5 = 5 points each from the public account.

Total income
Your total income results from the sum of your income from the private account and your income from the public account. Thus:

Income from the private account (= 20 – contribution to the public account)
*+ Income from the public account (= 0.5 * sum of contributions to the public account)*
= Total income

Control questions
Before starting the experiment we ask you to complete the sample exercises on the screen. If you need to do any calculations, you can click on the calculator symbol in the lower part of the screen, which will open the Windows calculator. If you have any questions please raise your hand, one of the experimenters will then come to your cubicle and answer your questions in private.

Procedure of part I

Part I includes the decision situation as described above. You make your decision in part I only once.

<u>Your input</u>
As described above, you possess an endowment of 20 points, which you can allocate to your private account or to the public account. Each group member has to make two types of contribution decisions. In the following, these two decisions are called the **unconditional contribution** and the **contribution table**.

- If you decide about your unconditional contribution to the public account, you simply determine how many points out of the 20 points you would like to assign to the public account. You can only enter **integers** in the predetermined input field on the screen. The allocation to your private account is automatically given by the difference between 20 and your contribution to the public account. When you have chosen your unconditional contribution, please click on "OK".
- In the next step you are asked to fill in a **contribution table**. **For each possible (rounded) average contribution by the two <u>other</u> group members to the public account you are asked to indicate how much you would like to allocate to the public account**, i.e. you are asked to determine your own contribution as a function of the other members' average contribution. The contribution table looks as follows:

Your conditional contribution to the public good (contribution table)

0		7		14	
1		8		15	
2		9		16	
3		10		17	
4		11		18	
5		12		19	
6		13		20	

OK

The numbers in the respective left column are the possible (rounded) average contributions of the two **other group members** to the public account, i.e. the contribution that is allocated to the public account on average by the other group members. In each input field you have to enter the number of points that you would like to assign to the public account – given the other members' average investment as denoted by the column**. You have to make an entry in each field.** For example, you are asked how much you would like to contribute to the public account if the other group members invest on average 0 points to the public account; how much you would contribute, if the others contribute on average 1 or 2 or 3 etc. points. **You have to enter an integer contribution between 0 and 20 in each field** and, of course, it is also possible to enter the same contribution several times. Once you have filled in all fields on the screen, please click "OK".

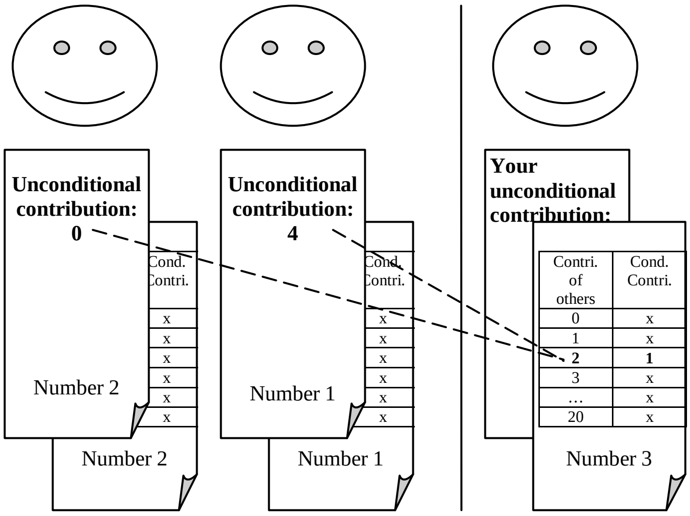
<u>Payoff-relevance of the contribution table</u>
As soon as all participants have made their decisions, one member in each group is randomly chosen**. The contribution table is payoff-relevant only for the randomly chosen group member**. For the other two group members who are not chosen, it is only the unconditional contribution that is payoff-relevant. Of course, you do not know whether you get chosen when making your contribution decision and filling in the contribution table. Therefore you have to think carefully about both types of contribution decisions as both can become relevant for you. Two examples help for illustration.

<u>Example 1:</u> Image you are chosen by chance and your contribution table is therefore payoff-relevant. For the other two group members their unconditional contributions are payoff-relevant. Imagine that their unconditional contributions are given by 2 and 6. The rounded average contribution by these two members is therefore 4 ((2 + 6) / 2 = 4).
If you have now indicated in your contribution table that you enter 1 point to the public account if the others contribute on average 4 points, then the total contribution of the group to the public account is given by 2 + 6 + 1 = 9. All group members therefore earn 0.5 * 9 = 4.5 points from the public account plus the respective income from their private accounts. If instead you have indicated in your contribution table that you invest 19 points to the public account if the others contribute on average 4 points, then the total contribution of the group to the public account is 2 + 6 + 19 = 27. All group members therefore earn 0.5 * 27 = 13.5 points from the public account plus the respective incomes from their private accounts.

<u>Example 2:</u> Imagine you **are not chosen by chance** and for you and another member of your group **the unconditional contribution is payoff-relevant**. Imagine your unconditional contribution to the public account is 16 and the unconditional contribution of the other group member is 18. The average unconditional contribution from both of you is therefore 17 = (16 + 18) / 2.
If the randomly chosen group member has indicated in the contribution table that she will allocate one point to the public account if the other two members contribute on average 17, then the total contribution of the group to the public account is 16 + 18 + 1 = 35. All group members therefore earn 0.5 * 35 = 17.5 points from the public account plus their respective incomes from their private account.
If instead the randomly chosen member has indicated in the contribution table that she will assign 19 points to the public account if the other two members contribute on average 17 points, then the total contribution of the group to the public account sums up to 16 + 18 + 19 = 53. All group members therefore earn 0.5 * 53 = 26.5 points from the public account plus their respective incomes from their private accounts.

The random choice of the group member for whom the contribution table is payoff-relevant happens in the following way: Every group member is given a number between 1 and 3. Once all participants have chosen their unconditional contributions and filled in their contribution table, one participant, who is randomly chosen by the PC, will dice a number between 1 and 3 (the six-sided dice possesses the numbers 1 to 3 twice). Afterwards he enters the number in the PC under the surveillance of the experimenter. The diced number is finally compared with your group number. If your group number is identical with the diced number, your contribution table is payoff-relevant for you and the unconditional contribution is payoff-relevant for the other two group members. If the numbers are not identical, the unconditional contribution is relevant for you. As you do not know in advance which of your decisions will be payoff-relevant, it is optimal for you to think about all decisions carefully.

The following graphic visualizes the decision situation. You are the person on the right side with the number 3. Number 3 is diced; therefore, your conditional contribution is payoff-relevant. For the two other group members the unconditional contribution is payoff-relevant.



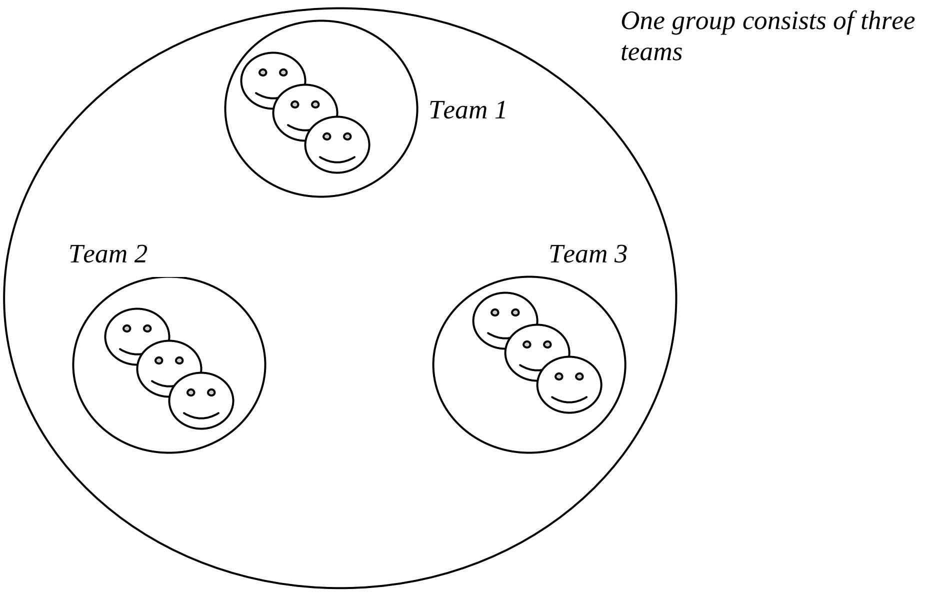
You make your decisions only **once**. At the end of part II you are informed about how much you earned in part I.

Part II

(Parts and instruction were presented sequentially)

Team membership

Part II of the experiment consists of **10 periods**. The structure of the decisions is similar to that in part I. In each period, you have to determine an unconditional contribution to the public account. This time, however, you cannot determine the contribution on your own, but you have to come to an agreement concerning the unconditional contribution **within your team**, which consists of three team members (see graphic below this paragraph). A group now consists of three teams, whereby each team consists of three team members (hence, in total there are 9 persons in one group). Each team determines a common contribution to the public account, i.e. team 1 determines a contribution, team 2 determines a contribution, and team 3 determines a contribution to the public account. The team membership is randomly assigned and remains constant throughout part II. Therefore, you have to come to an agreement concerning the contribution in all 10 periods with the same team members. In contrast, the group membership is randomly assigned and changes in each period. Hence, your team interacts in each period with other teams. There are five other teams in total. In each period your team is randomly matched to two of the five other teams.



Your decision

Each team has an endowment of 20 points at the beginning of each period. Please make the following decisions in part II:

- The private account and the public account are assigned to your whole team in part II. The income of a team member equals the income of the whole team (for details on this see below). Within your team you have to reach a unanimous decision regarding the distribution of the 20 points to the private account and to the public account. A unanimous decision in your team means that all three team members enter an identical number for the unconditional contribution to the public account on their screens. To do so you can communicate with the other team members before each decision via chat (for the chat rules see below).
- Before you make the decision on the unconditional contribution to the public account within your team, each team member is individually asked at the beginning of each period to indicate her **individual, non-binding proposal** concerning the contribution. This proposal is communicated to all team members before the chat.

Income per period

As in part I the team can decide to allocate the 20 points on the private account or to assign some or all of them to the public account. Each point that the team does not allocate to the public account, is automatically assigned to the private account of the team. Your income per period is determined analogously to part I:

Income from the private account (= 20 – contribution to the public account)
*+ Income from the public account (= 0.5 * sum of contributions of all three teams to the public account)*
= Total income

The total income of a team member in one period is identical with the total income of her team in this period. Hence, the total income of your team is not divided by 3, but each team member receives the full amount of the income of the team.

If, for example, all three teams decide to contribute 20 points to the public account, your team and, therefore, each member of your team receives 60 * 0.5 = 30 points from the public account and 0 points from the private account. Each member of the other two teams also receives 30 points. If the teams decide to allocate altogether 10 points to the public account, you and each member of your team receives 10 * 0.5 = 5 points from the public account and the respective income from the private account. Each member of the other two teams receives also 5 points each from the public account and the respective income from the private account.

If a team is not able to reach an agreement on the unconditional contribution **within 5 minutes**, the team (and, therefore, each member of the team) earns 0 points in this period. To determine the total contribution to the public account for the two other teams, one member of the team will be randomly chosen by the PC and her individual proposal will be used as contribution by her team.

Payoff-relevance of one period

At the end of each period you receive feedback on how much the other two teams contributed together to the public account and on your income in this period. At the end of part II one period is randomly chosen and the income in this period is paid out at the end of the experiment. As you do not know in advance which period is payoff-relevant, it is optimal for you and your team to carefully think about your decisions in all periods.

Rules for the chat communication

You have the possibility to discuss with your team members via electronic chat in order to come to a unanimous common decision. During the chat you have an ID (identification letter). The ID remains constant during the whole part II of the experiment. At the beginning of each communication line, the ID of the respective communicating team member is shown. The content of the chat communication is generally not restricted. However, it is not allowed to communicate personal information such as your name, age, address, gender (please use gender-neutral formulations), course of studies (including lecturers, courses or content that allows to draw conclusions on your course) or similar statements which could reveal your identity (e.g. your seat number in the laboratory). Furthermore, it is not allowed to insult other participants or to arrange side payments of any kind (in terms of gratifications or disciplines outside the experiment) of any kind. *If you violate these communication rules you do not receive any payoff.*

Each team member can send as many messages as she wishes to the other team members. Each message simultaneously appears on the screens of the other two team members. Selective messages to exclusively one team member are not possible. As soon as the three team members agree on the contribution to the public account, they enter the respective number in the input field on their individual screens. Please remember that **the inputs of all team members must be identical**. If they are not – or if a team or a team member did not make a contribution after 5 minutes – the payoff for this team will be 0 points in this period. Please be aware that the screen for the chat communication will be turned off after 5 minutes.

The screen for the decision on the unconditional contribution within your team looks as follows:

page 7 of 8

Remaining time:

Your ID is
Proposal by team member A:
Proposal by team member B:
Proposal by team member C:

Please decide within your team about the common contribution to the public account. You have 5 minutes to come to a unanimous decision within your team. Please use the chat box below.

Your team's contribution to the public account:

OK

As soon as the 10 periods of part II are completed, you are informed on how much you have earned in part I and part II. The payment is carried out after part III.

Part III

In the last part of the experiment we ask you to fill in **three short questionnaires**. They are an important element of our research. Please answer the questions honestly and spontaneously. Your answers are certainly anonymous.

Payment

Once all participants have filled in the questionnaires, the experiment is finished and you receive your income individually and in cash. Your income consists of:

Your income from the first part
\+ Your income from the second part

= Total income from both parts
\+ 4€ for your punctual appearance

page 8 of 8

Appendix A.1.3: Control questions

Please answer the following sample exercises.
The purpose of these exercises is to familiarize you with the income calculations in the experiment.
The sample exercises are not part of the actual experiment!

If you have questions concerning the sample exercises or do cannot find the right solution, please raise your hand. The experimenter will come to your cubicle and help you.

Exercise 1
You are a member of a 3-person group in which each group member possesses an endowment of 20 points.
Imagine no group member (including you) makes a contribution to the public account.

a) What will be your total income?
b) What will be the total income of each of the other group members?

Exercise 2
You are a member of a 3-person-group in which each group member possesses an endowment of 20 points.
Imagine you allocate 20 points to the public account. Each of the other group members also assigns 20 points to the public account.

a) What will be your total income?
b) What will be the total income of each of the other group members?

Exercise 3
You are a member of a 3-person-group in which each group member possesses an endowment of 20 points.
The other group members contribute in total 20 points to the public account. What will be your total income if

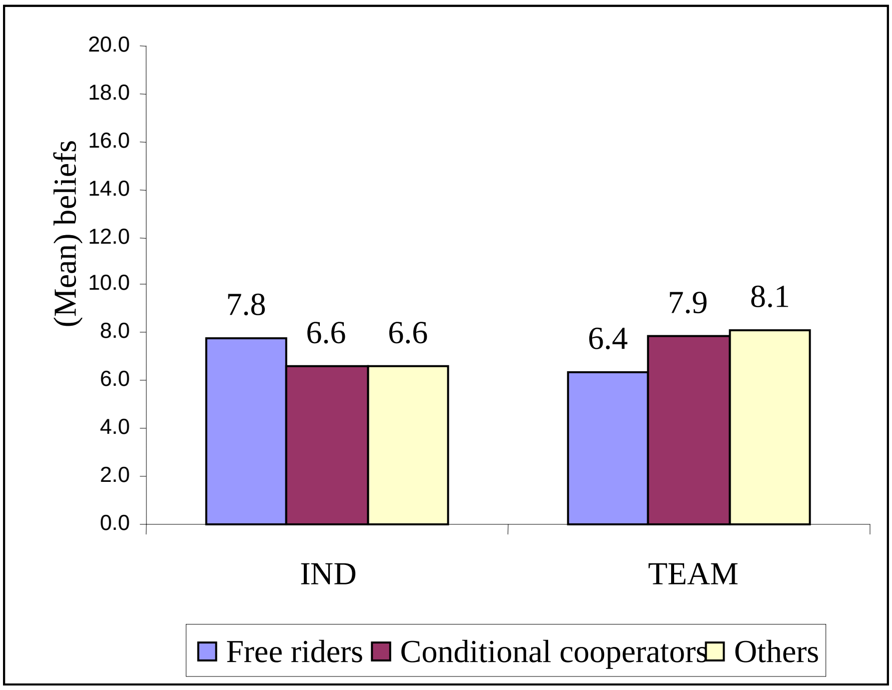
a) you contribute – in addition to the 20 points – 0 points to the public account?
b) you contribute – in addition to the 20 points – 8 points to the public account?
c) you contribute – in addition to the 20 points – 16 points to the public account?

Exercise 4
Consider a 3-person-group in which each group member possesses an endowment of 20 points.
Imagine you allocate 8 points to the public account. What will be your total income if

a) the other two group members contribute together – in addition to your 8 points – further 6 points to the public account?
b) the other two group members contribute together – in addition to your 8 points – further 14 points to the public account?
c) the other two group members contribute together – in addition to your 8 points – further 22 points to the public account?

Appendix A.1.4: Figures

Figure A.1.1 - Beliefs in IND and TEAM for each cooperation type.



Appendix A.1.5: Post-experimental questionnaire in condition TEAM

Set 1: Contributions by individuals and teams

Q1. Imagine Part II was not played by teams but by individuals. Do you think that individuals or teams contribute more to the public account?

Individuals contribute more	Teams contribute more
56.1%	43.9%

Q2. Why do you think so? (Open question)

Statement	Argument		Number of subjects who mention argument
Individuals contribute more	Cautious shift of teams because…	…teams think task better through, which leads to lower contributions	6
		…of the responsibility for other team members	4
	Majority rule in teams forces high contributors to adjust to other team members		8
	Stingy team members influence whole team		3
	There is social support in teams to contribute less		2
Teams contribute more	There is a peer pressure within teams to make higher contributions		8
	Majority rule in teams forces low contributors to adjust to other team members		7
	Team members encourage each other to take more risky decisions (= higher contributions)		4
	Teams think tasks better through, which leads to higher contributions		2

Set 2: Team atmosphere

Q3. How satisfied have you been with the decisions of your team on the contributions in part II?

Very unsatisfied	Unsatisfied	Neutral	Satisfied	Very satisfied
25.0%	7.2%	13.3%	36.7%	17.8%

Q4. Why?

Satisfied/ Unsatisfied	Argument	Number of subjects who mention argument
Satisfied	Team members had a similar strategy	18
	Team earned a high income	13
	Team quickly came to an agreement	12
	The decision-making process was democratic	7
	Teams contributions were close to own proposal	7
	Team behaved towards other teams in a fair way	2
	There was a team spirit	2
Unsatisfied	Team contributions were very different from own proposal	8
	Team earned a low income	3
	There was no team spirit	2
	Team acted towards other teams in an unfair way	1
	One team member did not accept majority rule	1

Q5. How was the atmosphere in your team?

Very bad	Bad	Neither good nor bad	Good	Very good
2.2%	1.1%	7.8%	41.7%	47.2%

Q6. Image you had to play part II again. Would you wish to change into another team (which is determined randomly)?

Yes	No
20.0%	80.0%

Set 3: Decision-making process within teams

Q7. To what extend did you influence the team decisions in part II?

Not at all	Hardly	To some extend	Rather heavily	Heavily
1.7%	3.3%	24.4%	56.7%	14.0%

Q8. Did your team take a vote on the contributions?

No	Partly	Yes
20.0%	43.9%	36.1%

Q9. If your team did not take a vote, how did your team come to a unanimous decision?

Decision-making process	Number of subjects who mention this kind of process
The team discussed arguments/ pros and cons	17
The team took the average	7
The proposals of team members were similar from the beginning on	7
The team chose at the beginning a long-term strategy	3
One team member made a proposal and the others agreed	3
The team took the median	1

Q10. Did all members of your team have equal rights?

No	Yes
6.7%	93.3%

Q11. Was there a person in your team who decided for the whole team?

No	Yes
92.8%	7.2%

Q12. Given your experience in today's experiment, would you prefer to take decisions rather on your own or within a team, if you could participate in a similar experiment again?

On my own	Within a team
37.8%	62.2%

Set 4: Further comments

Q13. Please write down, if you have further comments concerning the experiment, your strategy, the reasons for your decisions or the difference between team and individual decisions.

Comments on the…	Comments		Number of subjects who mention comment
Experiment	Positive	Taking decisions in teams was interesting	5
		The chat communication was interesting	1
	Negative	The examples in instructions affect decisions	1
		The earnings in part II should be the average of all period incomes	1
		The time for the chat (5 minutes) was too long	1
		The experiment was too long	1
Strategy	The strategy of free-riding/ contributing little was efficient		3
	The strategy of free-riding is not efficient, as all other teams then also free-ride		1
Effect of team decisions	Deciding within a team saves you from carelessness and extreme contributions		2
	Deciding within a team leads to higher contributions and higher outcomes		1

Chapter 2

What doesn't kill you makes you stronger.

Cooperation and Efficiency in Public Goods Games under the Threat of Peer Punishment and Group Extinction

2.1 Introduction

Social dilemmas, like the private provision of a public good, are plagued by individual incentives to free ride, despite the existence of a socially optimal outcome that involves cooperation of agents. Over the last decades, in the literature many institutions and mechanisms have been proposed that intend to sustain high levels of cooperation. One of the most successful mechanisms for achieving cooperation despite the free riding incentives, at least in laboratory studies, is peer punishment (Yamagashi, 1986; Ostrom et al., 1992; Fehr and Gächter, 2000; Andreoni et al., 2003; Masclet et al., 2003; Anderson and Putterman, 2006; Gürerk et al., 2006; Carpenter, 2007; Sefton et al., 2007; Nikiforakis et al., 2012). A costly punishment option for members of the public good provision unit does not require formal authority (hence, the alternative term *informal* punishment) and it increases contributions to the public good dramatically. However, it comes at a cost: Efficiency is often affected negatively because punishment is costly to both the punisher and the punished (Fehr and Gächter, 2000; 2002; Sefton et al., 2007; Egas and Riedl, 2008; Herrmann et al., 2008; Nikiforakis and Normann, 2008; Ertan et al., 2009)[21].

More recent laboratory experiments show that overall the availability of peer punishment can increase efficiency over a situation with purely voluntary contributions, i.e., without punishment, but it requires a long-term interaction in order to balance the cost of punishment with its benefits (Gürerk et al., 2006; Gächter et al., 2008), and it has to exclude the option to retaliate on punishment. Retaliation or a counter-punishment option leads to an extremely bad efficiency balance, even in long-term interactions (Denant-Boemont et al., 2007; Nikiforakis, 2008; Engelmann and Nikiforakis, 2013). Compared to equilibrium predictions, regardless of whether they stipulate standard assumptions or invoke social

[21] Surveys are provided by Chaudhuri (2011) and Guala (2012).

preferences, peer punishment is overused and there is anti-social (or, misdirected) punishment of cooperators (Herrmann et al., 2008). One puzzling observation, however, is that field studies show a surprisingly low level of peer punishment in the real world (Balafoutas and Nikiforakis, 2012; Balafoutas et al., 2014).

In this paper, we study one promising observation for the low level of peer punishment observed in the field. We argue that – besides the fear of retaliation that plays a role and that we can control for – efficiency concerns could be much more important than conventionally believed. We implement a particular but realistic way to make efficiency concerns salient, drawing on evolutionary ideas. More specifically, exogenously composed provision units for public goods have to surpass a specific threshold of group earnings (from the consumption of the private good and of the public good subtracted by the sanctioning cost) every period in order not to get extinct. Such a setup can be motivated by evolutionary selection arguments. More often than not, groups dissolve or are dissolved in the real world (depending on performance), whereas in the laboratory studies it is assumed so far that the group always exists until the end of the game, regardless of how much money they actually burn. Results from our laboratory experiment based on this group extinction mechanism confirm our main hypothesis. More generally, the combination between the availability of peer punishment and the fear of extinction mainly leads to higher levels of cooperation combined with lower levels of punishment. In essence, the combination makes sure that the threat of punishment is sufficient to enforce cooperation within a group, and the actual act of punishment is not necessary anymore. The psychological mechanism behind the effect could be salience.

Social dilemmas are an almost ubiquitous phenomenon in human interaction. There are several examples for social dilemmas in which peer punishment and fear of group extinction play a particularly important role: Work groups in companies or organizations often fight for survival as a group (i.e., they can easily be dissolved if unsuccessful). Thereby peer punishment is available and the effort provision for a joint project is costly and a social dilemma.[22] Thus, the situation satisfies the conditions that are present in our study. Most importantly, the extinction threat can be adapted and possibly fine-tuned by the organization itself. More generally, tribal groups or societies that have to reach a certain level of benefits from the interaction in order to survive face similar incentives as the ones that are present in our experiment. Naturally, when talking about real-world implications one has to take the limitations of experimental studies in terms of external validity into account, but they allow

[22] It almost goes without saying that peer punishment does not necessarily have to take the form of a monetary fine. Social disapproval, social exclusion, or similar forms of punishment are, however, almost always available (see also Rege and Telle, 2004; Cinyabuguma et al., 2005).

gathering controlled evidence for specific incentive effects that are hard to study when using real-world data.

More specifically, our setup implements a repeated linear public goods game (the voluntary contribution mechanism) with punishment opportunities. It features four experimental treatments in a 2x2 factorial between-subjects design. Within the first dimension, we vary whether peer punishment is available with or without a retaliation option (standard punishment, or often referred to in the following simply as punishment, versus counter-punishment). The second dimension that is varied is the existence of a survival threshold. If the survival threshold is present, groups have to reach a certain level of group income in order not to get extinct as a group. Laboratory sessions include four parts with five periods each. If a group is extinct in one part, group members are excluded from the public goods game – and therefore get an income of zero – in the following periods of that part. However, they restart the public goods game in the following part. Note that there are theoretical and behavioral links between the idea of a survival threshold in each provision unit, a contest between provision units, and, more generally, competition among provision units (e.g., Tan and Bolle, 2007; Strasser, 2012). Our focus in the current paper is on the survival threshold, which is a new element in the design of a public goods experiment. Future studies will have to look at the links between the different mechanisms in greater detail.

Our experiment allows addressing several questions regarding punishment and the negative effects of counter-punishment. Firstly, if efficiency concerns become more salient (in our case, through the survival threshold), the availability of counter-punishment is much less of a problem than when there is no survival threshold. Secondly, in our setup the total punishment cost and/or the total punishment impact are uncertain. If a punisher is pivotal in dragging the group below the survival threshold the punishment impact is very high. Hence, it is probably not (only) the fear of retaliation per se that leads to very low levels of peer punishment in the field. Given our results, the uncertainty involved in punishment plays an important role. The effects of uncertain punishment, uncertain outcomes of alternative cooperation decisions and imperfect monitoring of contributions have been analyzed in public goods games (Grechenig et al., 2010; Sousa, 2010; Ambrus and Greiner, 2012; Xiao and Kunreuther, 2012), but deserves more attention.

In the remainder of this paper, we present an overview of the relevant literature (section 2.2) and our experimental design as well as the theoretical predictions (section 2.3). Section 2.4 provides the empirical results, and section 2.5 concludes the paper.

2.2 Related Literature

The literature on punishment in the public goods game is vast. An authoritative overview is provided by Chaudhuri (2011). In the following overview of the relevant literature, we focus on efficiency aspects in public goods experiments.

Whereas the cooperation-enhancing effect of altruistic, i.e. costly, punishment has been widely confirmed (e.g. Yamagashi, 1986; Ostrom et al., 1992; Fehr and Gächter, 2000; Masclet et al., 2003; Page et al., 2005; Bochet et al., 2006; Sefton et al., 2007; Egas and Riedl, 2008; Nikiforakis and Normann, 2008; Ertan et al., 2009; Fudenberg and Pathak, 2010; Sutter et al., 2010; Nikiforakis et al., 2012), the empirical evidence on efficiency, i.e. earnings, is rather mixed. Most studies, however, find either no efficiency-enhancing effect of the punishment mechanism (e.g. Page et al., 2005; Bochet et al., 2006; Sutter et al., 2010) or an average efficiency loss (e.g. Fehr and Gächter, 2000, 2002; Sefton et al., 2007; Egas and Riedl, 2008; Herrmann et al., 2008; Ertan et al., 2009).[23] Nikiforakis and Normann (2008) find that at least a punishment effectiveness of 1:3 is necessary to achieve a welfare improvement compared to the VCM without punishment (see also Egas and Riedl, 2008). In the standard experiments on the punishment mechanism players typically interact for ten periods. The relative efficiency mostly increases over time in the treatment with a punishment option compared to a standard public goods game without punishment. Gächter et al. (2008) find for short interaction terms (ten periods) of groups a relative efficiency loss of the punishment mechanism. Only for the long interaction terms of 50 periods, the punishment mechanism leads to a relative efficiency gain. These results indicate that introducing a punishment mechanism often causes efficiency problems, unless the interaction horizon is very long.

Nikiforakis (2008) confirms the efficiency problem in a more realistic setting, namely in a public goods game with two-sided punishment, i.e. when there is a costly retaliation or counter-punishment option. The counter-punishment opportunity reduces contributions and the willingness to punish compared to the treatment with a one-sided punishment option. The level of earnings is slightly, but not significantly, lower. Therefore we also examine the effect of the group extinction mechanism in case of a two-sided punishment option.[24] Engelmann

[23] There are some experimental studies showing that peer punishment can have an efficiency-enhancing effect, especially in combination with some form of communication (e.g. Ostrom et al., 1992; Bochet and Putterman, 2009; Masclet et al., 2013). However, in real life communication among group members might not always be feasible, especially in large groups.

[24] More studies explore the effect of two-sided punishment in a public goods game (e.g. Denant-Boemont et al., 2007). They confirm the main results of Nikiforakis (2008).

and Nikiforakis (2013) extend the Gächter et al. (2008) design for long-term interactions by a treatment, in which all forms of punishment, i.e. feuds, are possible. They do not find any relative efficiency gain in this case.

The focus on the effect of survival thresholds connects our work to the literature on the provision point mechanism or step-level public goods games (e.g. Isaac et al., 1989; Bagnoli and McKee, 1991; Suleiman and Rapoport, 1992; Marks and Croson 1998; Cadsby and Maynes, 1999; Croson and Marks 2000; Rondeau et al., 2005; Spencer et al. 2009; Abele et al., 2010; Freytag et al., 2014; Fischbacher et al., 2011; Ye et al., 2013), but there are obvious differences. The provision point mechanism typically uses a step-level function that requires the group to reach a certain contribution threshold for the public good to be provided. Our group extinction mechanism differs from the provision point mechanism in at least two respects: Firstly, the threshold in our experiment is based on earnings and not on contributions alone. More specifically, our threshold concerns both contributions and (counter-) punishment. Secondly, the group extinction mechanism implies no consequences in the current period if the threshold is not reached, but leads to group extinction in future periods.

Our group extinction mechanism draws on evolutionary ideas. However, there is a clear difference to the term *group selection* used in evolutionary models to explain the evolution of reciprocity or human morality (Bowles and Gintis, 2004; Gintis et al., 2008; van den Bergh and Gowdy, 2009): The term *group extinction* means that a certain performance threshold needs to be reached by the group to ensure its further existence. We do not model the inter-group competition, i.e., earnings do not depend on the behavior of other groups. This is also why there is no direct link between our study and several empirical studies on the effect of inter-group competition on cooperation (e.g. Tan and Bolle, 2007; Puurtinen and Mappes, 2009; Burton-Chellew et al., 2010; Strasser, 2012).[25]

2.3. Experimental Design, Procedures and Predictions

2.3.1 Experimental Design

Our experiment consists of four treatments, formed by a 2x2 factorial between-subjects design:

(1) A public goods game with a one-sided punishment option (*P*):

[25] Cinyabuguma et al. (2005) allow group members in the public goods game to expel individual group members. Excluded group members potentially earn less in future periods. In contrast, in our experiment the whole group gets extinct and the extinction is determined by the group performance.

A standard linear public goods game (PGG), also known as the voluntary contribution mechanism (VCM), followed by a single punishment stage;

(2) A PGG with a one-sided punishment option and the possibility of group extinction (*P_E*):

A PGG followed by a single punishment stage with the threat of group extinction;

(3) A PGG with a two-sided punishment option (*PCP*):

A PGG followed by a punishment stage and a counter-punishment stage; and

(4) A PGG with a two-sided punishment option and the possibility of group extinction (*PCP_E*):

A PGG followed by a punishment stage and a counter-punishment stage with the threat of group extinction.

Each treatment – and therefore each session – consists of four parts which are identical within each treatment. Each part consists of five identical periods; hence, altogether 20 identical periods are implemented within each treatment. At the beginning of the first part we assign subjects randomly to 4-member-groups. We use a partner design within each part. If one part is finished, however, subjects are randomly reassigned to new groups at the beginning of the next part, and this is common knowledge. Hence we re-match groups between the parts. If applicable, group extinction takes place from the period on after the threshold is not met and it is valid throughout the remainder of the part. In the subsequent part, all participants are assigned to new groups and can participate in the game again. We implement several parts to allow for learning.

Each period in each treatment consists of several stages. The first stage is always the contribution stage, in which the standard linear PGG is played (see Marwell and Ames, 1981; Isaac et al., 1985; Isaac and Walker, 1988). Each group member receives an endowment of 20 tokens in each period. Group members, then, decide simultaneously and without communication how much of their endowment to contribute to the public good, c_i, where $0 \leq c_i \leq 20$, and how much to keep for themselves (the rest). In each period the payoff function for this stage is given by

$$\pi_i = 20 - c_i + 0.3 * \sum_{h=1}^{4} c_h \quad (2.1),$$

where the public good is equal to the sum of contributions of all group members. Payoffs during the experiment are calculated in Guilders. If a participant contributes one token to the public good, the private marginal return (MPCR) is 0.3 Guilders and the social marginal benefit is 1.2 Guilders. So the inequality $MPCR < 1 < MPCR*4$ holds and implies a social dilemma: If we assume pure money-maximizing preferences, there is a dominant strategy for each group member to free ride, that is $c_h = 0$ for all h. However, this leads to a socially inefficient outcome. Note that we deliberately chose the MPCR to be rather low compared to other public goods experiments in order to increase the relevance of the group extinction mechanism.

At the end of the contribution stage subjects receive feedback on their own contribution, the group's total contribution, their earnings from the public good and from their private account, and their entire earnings in the contribution stage. In the following we explain each of the treatments in detail. The treatments *P* and *PCP* are based on Fehr and Gächter (2000, 2002) and Nikiforakis (2008) respectively.

PGG with a one-sided punishment option (P)

In treatment *P* a punishment stage is added after the contribution stage. At the beginning of the punishment stage subjects get information on the individual contributions of their group members and have the opportunity to punish any other member of their group simultaneously and without communication. If group member i wants to punish group member j, she has to assign punishment points to group member j, p_{ij}, $i \neq j$, where $0 \leq p_{ij} \leq 10$. Subjects cannot punish members of other groups. Punishment reduces the first stage income of both players i and j in different ways: For each punishment point that j receives, her first stage income, $\pi_{1,j}$, is reduced by 3 Guilders, with a lower boundary at zero if player j's cost of receiving punishment points is higher than her first stage income. The latter rule avoids that subjects receive a negative payoff without being able to control it. Subject i faces a cost of 1 Guilder for each allocated punishment point.[26] The total cost of distributing punishment points are shown at the bottom of the screen. Thus, the income of subject i at the end of the second stage, $\pi_{2,i}$, is calculated according to equation (2.2):

$$\pi_{2,i} = \max\left\{0, (\pi_{1,i} - 3 * \sum_{j \neq i} p_{ji})\right\} - \sum_{j \neq i} p_{ij} \tag{2.2}$$

Eq. (2.2) implies that the period payoff of a participant i can be negative, if subject i is punished heavily and punishes herself. Participants are explicitly warned of this possibility.[27] At the end of the second stage subjects receive feedback on their income from the first stage,

[26] We implement the linear cost function of Fehr and Gächter (2002) and later studies instead of the nonlinear cost function used in Fehr and Gächter (2000) and Nikiforakis (2008). The linear cost function is easier to understand for participants.

[27] Our data shows that this happened only in 0.6% of all possible cases. Losses in a single period can be balanced by a show-up fee of 4€ and gains from other periods.

the total punishment points received, the total cost of punishment points assigned, their period earnings as given by eq. (2.2), as well as the sum of incomes they earned so far within the current part.

PGG with a one-sided punishment option and the possibility of group extinction (P_E)

In treatment *P_E* we expand treatment *P* by a group extinction mechanism. In order to survive, i.e., in order to reach the next period, the sum of the period incomes within a group (after the punishment stage) has to reach or exceed 80 Guilders. Obviously, this is exactly equal to the sum of the endowments of the four group members. We set the threshold at 80 Guilders in order to equalize the game-theoretic contribution predictions based on pure money-maximizing preferences among all four treatments (see section 2.3.3). At the end of the second stage subjects are additionally informed about the group's period income. In case the threshold of 80 Guilders is reached or surpassed, the group participates in the game in the next period. If not, the group does not participate in the PGG in the following periods within the current part, and all group members simply have to wait until the next part starts. In the periods in which the group does not participate the group members earn no further income. Group members participate in the next part again within a newly assembled group.

PGG with a two-sided punishment option (PCP)

In the treatment *PCP* a third stage, namely the counter-punishment stage, is added after the contribution and the punishment stage of treatment *P*. At the beginning of the counter-punishment stage subjects are informed about the number of punishment points each of the other group members assigned to them in the second stage. Then each subject i has the opportunity to counter-punish those group members $j \neq i$ who punished them in the second stage by assigning counter-points. Analogous to the punishment stage, the cost of each counter-point received is 3 Guilders. The second stage income is reduced to zero and not further, if the cost of receiving counter-points is higher than the second stage income, $\pi_{2,i}$. The cost for each counter-point that i assigns to j, z_{ij}, is 1 Guilder. The period income of subject i in the treatment *PCP* is therefore given by equation (2.3):

$$\pi_{3,i} = \max\left\{0, (\pi_{2,i} - 3 * \sum_{j \neq i} z_{ji})\right\} - \sum_{j \neq i} z_{ij} \qquad (2.3)$$

In order to avoid very negative payoffs participants with a non-positive second stage income are not allowed in the third stage to counter-punish or to be counter-punished. Subject i can only counter-punish those group members j that punished her in the second stage, in order to avoid that punishment in the second stage is delayed to counter-punishment in the third stage. At the end of each period subjects are informed about their income from the second stage, the total number of received counter-points, the total cost of distributed counter-points, the period income as given by Eq. (2.3), as well as the sum of period incomes, they earned so far within the current part.

PGG with a two-sided punishment option and the possibility of group extinction (PCP_E)
Treatment *PCP_E* is the expansion of treatment *PCP* by the group extinction mechanism. After the counter-punishment stage, group members additionally receive feedback on the group period income (after the punishment and counter-punishment stage). Group members whose group is extinct (according to the same general rule as in treatment *P_E*), cannot participate in the following periods of the current part and earn no additional money in these periods.

In all four treatments subjects know the endowment, the return from the public good, the number of group members, the payoff functions, the cost for assigning and receiving points, the number of parts and periods, and (if applicable) the possibility of group extinction. To ensure thoughtful decisions, we do not impose a time limit.

Subjects further know that they are given an identification letter (A to D) at the beginning of each period in order to distinguish their actions from those of the other players within a period. However, as subjects should not be able to identify actions of another player over periods, the identification letters are reassigned in each period. In the two treatments *P_E* and *PCP_E* a post-experimental questionnaire follows after the four identical parts to investigate the effects of group extinction.[28] An incentivized Ring-test or social value orientation questionnaire (Liebrand, 1984; Liebrand and McClintock, 1988) is added. Results of this test are not reported here. All treatments finish with a questionnaire on socio-economic variables.

The total income of a subject is equal to the sum of all period incomes and earnings from the Ring-test. Subjects are informed that they are given a show-up fee of 4€ to pay for any negative payoffs generated during the experiment.

[28] The post-experimental questionnaire on group extinction is given in appendix A.2.3.

2.3.2 Procedures

We run 12 sessions at the experimental laboratory MELESSA of the University of Munich between December 2010 and March 2011. Each session lasted on average 1 hour 45 minutes. Without telling the participants we divided the 24 subjects of each session into two matching groups of 12 participants, who are re-matched in groups of 4 at the beginning of a new part. So we have 6 independent observations per treatment (Table 2.1). The experiment was conducted anonymously.

Table 2.1 - Treatments and number of sessions.

Treatments	Punishment option	Group extinction	Sessions and subjects
P	One-sided	No	3 sessions à 24 subjects
P_E	One-sided	Yes	3 sessions à 24 subjects
PCP	Two-sided	No	3 sessions à 24 subjects
PCP_E	Two-sided	Yes	3 sessions à 24 subjects

At the beginning of each session the experimenter read the neutrally-written instructions for the four identical parts aloud.[29] Instructions for the Ring-test and the following questionnaires only followed after the first four parts were accomplished. Participants knew that there would be a test and questionnaire(s) after the four identical parts and that those would be unrelated to the first four identical parts. Instructions are written in neutral language. In order to test the understanding of the mechanics and the incentive structure participants were asked to answer control questions.[30] We did not proceed until all participants had answered all questions correctly. We can thus safely assume that the participants understood the game.

A total of 288 participants were recruited via ORSEE (Greiner, 2004) from a list of voluntary participants. Subjects were German-speaking, mostly students, of different academic backgrounds. The experiment was programmed and conducted with the experiment software z-Tree (Fischbacher, 2007). At the end of the experiment we exchanged the payoffs at an exchange rate of 1Guilder = 0.04€, which yielded an average payoff of 21€.

2.3.3 Predictions

In this section we derive hypotheses for the differences between the treatments with and without group extinction. We start with the predictions under pure profit-maximizing preferences. Under this assumption neither punishment nor counter-punishment will occur in

[29] The instruction for the treatment *PCP_E* is given in appendix A.2.1.
[30] Control questions for treatment *PCP_E* are given in appendix A.2.2.

the subgame perfect Nash equilibrium as they are costly for the punisher. By backwards induction in the sequential game individuals' dominant strategy in the contribution stage of the treatments *P* and *PCP* is therefore a contribution of zero to the public good. In the treatments with the group extinction mechanism the PGG resembles a step-level PGG (see section 2.2). The difference is that if a group does not reach the threshold the group members do not earn an income from the next period on. In the step-level game all patterns of decisions that minimally satisfy the provision point are Nash equilibria (for a game-theoretic analysis see, e.g., Abele et al., 2010). Given that punishment and counter-punishment do not take place contributions of zero provided by all group members minimally satisfy the threshold of 80 Guilders. To summarize: Under profit-maximizing preferences contributions, punishment, and counter-punishment will be equal to zero and no differences between treatments with and without group extinction are predicted.

As we have mentioned in section 2.2 the empirical studies on the VCM with punishment show that punishment takes place and contributions are high, especially in comparison with the VCM without punishment. This empirical regularity can be better explained if we assume other-regarding preferences (for a survey see Fehr and Schmidt, 2003), as e.g. inequity aversion by Fehr and Schmidt (1999, see also Bolton and Ockenfels, 2000, for a similar approach). According to their theory an inequity averse person tries to avoid inequality of individual payoffs between herself and other persons in her reference group. Hence, the threat that group members who contribute less than others in the group are punished by the more cooperative group members in order to equalize earnings is credible. Using efficiency as refinement argument Fehr and Schmidt (1999) predict very high contributions in treatment *P*, especially in comparison to the VCM without punishment. As further outlined in section 2.2, Nikiforakis (2008) shows for the PGG with two-sided punishment (treatment *PCP*) that counter-punishment takes place and, compared to treatment *P*, reduces punishment and contributions. Nikiforakis (2008) suggests that counter-punishment takes place for strategic considerations over periods and for a desire to revenge punishment. The reduced punishment and cooperation in *PCP* compared to *P* can be explained by increased cost of punishment.

Our research interest lies in the comparison between the treatments with and without group extinction. Charness and Rabin (2002) present a model in which they mainly include social welfare concerns and reciprocity. Based on reciprocity their model also allows for punishment. The social welfare concerns lead people to sacrifice own material payoff in order to increase the payoffs for all recipients, i.e. to increase efficiency. In line with the theories on social preferences one can reason as follows: With our group extinction mechanism we make the social welfare concern more salient. Counter-punishment and punishment strictly decrease efficiency. Hence, they are probably reduced in the treatments with group extinction compared to the ones without the mechanism. On the other hand, contributions strictly increase the efficiency and therefore the chance to meet the threshold. Furthermore high contributions counter-balance the risk of efficiency- reducing punishment. Hence, we predict contributions to be higher in the treatments with the group extinction mechanism than in the treatments without. In other words: Because of the threat of group extinction the cost of (counter-) punishment and low contributions increase, as they increase the risk of not reaching the threshold. The decreased demand for punishment and counter-punishment in the treatments with group extinction as well as the increased contributions imply higher payoffs than in the treatments without group extinction.[31]

We summarize our main hypotheses:

HYPOTHESIS 1: *The introduction of the group extinction mechanism will increase contributions. Hence contributions will be higher in P_E and PCP_E than in P and PCP, respectively.*

HYPOTHESIS 2: *The introduction of a group extinction possibility will decrease punishment, i.e., ceteris paribus, punishment will be lower in P_E and PCP_E compared to P and PCP, respectively.*

HYPOTHESIS 3: *The group extinction mechanism will decrease counter-punishment in PCP_E compared to PCP.*

2.4 Results

We start the results section with the analysis of the cases of group extinction in treatments *P_E* and *PCP_E* (section 2.4.1). Then the effects of the group extinction mechanism on contributions, punishment, counter-punishment and efficiency follow (section 2.4.2). In section 2.4.3 we give an overview of potential motives behind the behavior in treatments *P_E* and *PCP_E*.

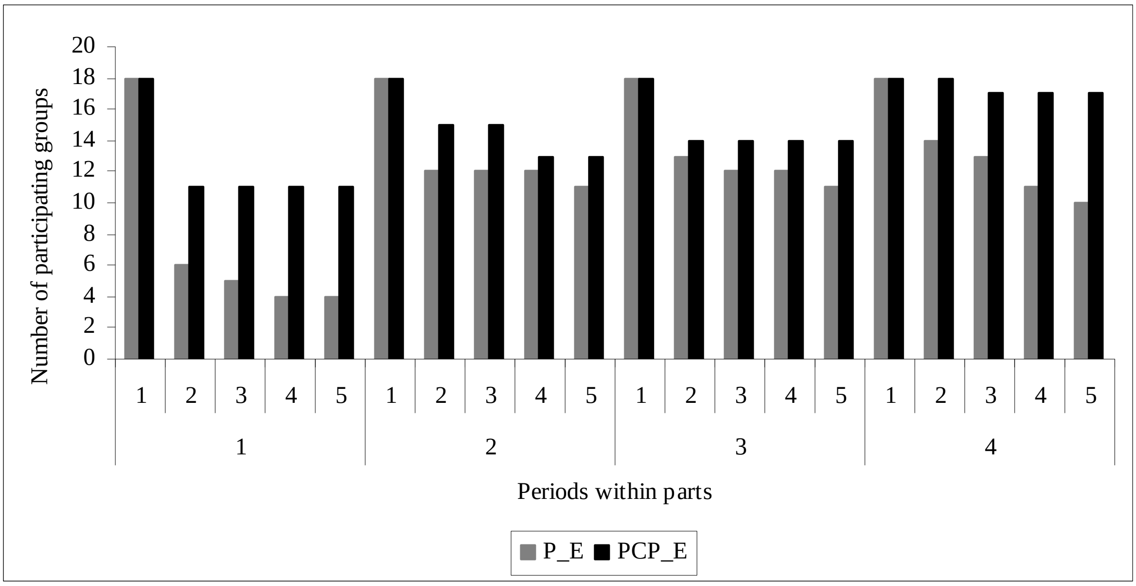
[31] Apart from these economic considerations, the threat of group extinction might also have a psychological effect: The common aim of reaching the threshold might increase the team spirit and therefore increase cooperation and reduce punishment (e.g. Aubé and Rousseau, 2005; Senécal et al., 2008).

2.4.1 Group Extinction

As we use in the following the matching group of 12 subjects as the unit of statistically independent observations and calculate probabilities of participation within these units, we apply non-parametric tests for continuous and not discrete variables.

Figure 2.1 shows the number of participating groups – groups that are not extinct – over parts and periods within parts. We find a learning effect over the parts: Whereas only 4 and 11 out of 18 groups participate in *P_E* and *PCP_E* in the last period of part 1, 10 and 17 groups still participate in the last period of part 4 (two-sided Wilcoxon signed-ranks test, $p = 0.030$ for *P_E*; $p = 0.053$ for *PCP_E*; $p = 0.004$ for both *P_E* and *PCP_E*). This shows that over time participants adapt their behavior more and more to the threshold. It is especially striking that for both treatments *P_E* and *PCP_E* in part 1 many more groups are extinct than in the following parts: The participation rate is only 26.4% for *P_E* and 61.1% for *PCP_E* in periods 2 to 5 of part 1. In contrast, in parts 2 to 4 in *P_E* on average 66.2% and in *PCP_E* 83.8% participate in periods 2 to 5 (two-sided Wilcoxon signed-ranks test, $p = 0.028$ for *P_E*; $p = 0.293$ for *PCP_E*; $p = 0.011$ for both *P_E* and *PCP_E*). This result indicates learning to coordinate across phases at the beginning of the experiment.

Figure 2.1 - Number of participating groups in the treatments with group extinction mechanism over time.



It is further noteworthy that if groups get extinct within a part, they usually get extinct already after the first period of this part: The participation rate decreases between period 1 and period 2 of each part on average by 37.5 and 19.4 percentage points in *P_E* and *PCP_E*, respectively. In contrast, between periods 2 and 5 of each part, the participation rate only decreases by 20.0 and 5.2 percentage points, respectively.

Over periods 2 to 5 of all parts, we find a participation rate of 56.3% in treatment *P_E* and 78.1% in treatment *PCP_E*. Hence, more groups participate in *PCP_E* than in *P_E* (two-sided Mann-Whitney U-test, p = 0.025), which is quite surprising. One could have expected a higher participation rate in *P_E* than in *PCP_E*, as the additional counter-punishment stage in *PCP_E* offers a further possibility to decrease the group income. Furthermore Nikiforakis (2008) shows that the income is slightly worse in his treatment *PCP* compared to treatment *P*. However, participants in *PCP_E* seem to anticipate the increased risk of going extinct and therefore seem to increase their contributions and decrease their punishment compared to *P_E* in order to ensure to reach the threshold (for details see section 2.4.2).

Table 2.2 outlines the results of a probit regressions of participation for the treatments *P_E* and *PCP_E*. As regressors we introduce the following variables: A dummy variable for the two-sided punishment option, taking value 1 in case of treatment *PCP_E* and 0 in case of *P_E*; And a control variable for the four parts and the five periods within each part, respectively; The variable *Previous Extinction* takes the value 1 if a person has been extinct in one of the previous parts and 0 otherwise. The results of the probit regressions confirm the results of the non-parametric tests: There is a learning effect over parts, as the fraction of participating groups increases with each part by about 10 to 12%. Of course, over the periods within each part the participation rate decreases. However, the insignificance of the variable *Previous Extinction* indicates that the experience of being extinct in a previous part does not have an extra effect on participation in later parts. We further find that more groups are extinct in treatment *P_E* than in *PCP_E*.

RESULT 2.1: *In treatments P_E and PCP_E there is a substantial number of groups that are extinct. The participation rate is 56.3% in treatment P_E and 78.1% in treatment PCP_E in periods 2 to 5 of each part. However, we observe a significant decrease in the number of extinct group over the course of the experiment.*

Table 2.2 - Determinants of participation in treatments *P_E* and *PCP_E*.

Dep. Var.:	Participation	Participation
Two-sided punishment	0.234*** (0.063)	0.255*** (0.079)
Part	0.123*** (0.034)	0.102** (0.040)
Period	-0.032*** (0.008)	-0.032*** (0.008)
Previous Extinction		0.084 (0.074)
N	2304	2304
Wald χ^2	88.79***	106.05***
Pseudo log likelihood	-1296.475	-1291.187

Notes: This table reports probit marginal effects of participation in treatments *P_E* and *PCP_E* in periods 2 to 5 of each part. Robust standard errors (clustered on matching groups of 12 subjects) are reported in parentheses. *, **, *** denote significance at the 10%, 5% and 1% level respectively.

2.4.2 Contributions, Punishment, Counter-punishment and Earnings

Next we analyze the effect of the group extinction mechanism on contributions, punishment, counter-punishment and earnings. Obviously, we cannot apply a conventional regression analysis, as we would not be able to differentiate between selection effects and incentive effects of the extinction mechanism. Instead we use a three-step analysis for each dependent variable: Firstly we analyze the effects of the group extinction mechanism only for the first period of each part, as in the first period no extinction has taken place, and we thus can use data from all groups. In order to analyze the effect of the group extinction mechanism over all periods we secondly calculate the Heckman two-step procedure (Heckman, 1979; see also Cameron and Trivedi, 2005), which takes the selection bias into account. The Heckman two-step procedure is a generalization of the tobit model and comprises two equations, the participation equation and the outcome equation. It assumes that the error terms of both equations are correlated. In the first step the likelihood of participation (of not being extinct) is estimated, in the second step the effects on the level of contributions, punishment, counter-punishment or earnings are estimated for those subjects who participate – corrected for the selection bias. We use an OLS regression for the selection equation and for the outcome equation.[32] We include the inverse of Mill's ratio in the outcome equation to control for the potential selection bias and bootstrap the standard errors. Thirdly, in addition to the Heckman two-step procedure we run a robustness check with the data of all periods, by dropping in

[32] Usually the first step of the Heckman selection model is calculated with a probit regression. As in the treatments without group extinction the prediction for the likelihood of participation was 100%, we calculate an OLS regression instead.

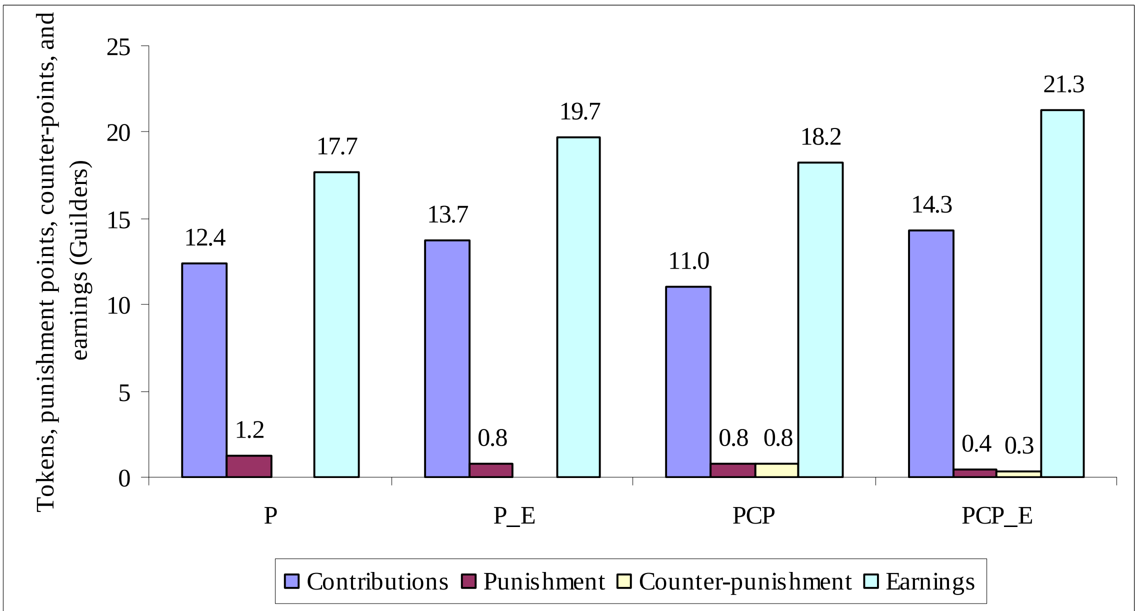
each period the same number of groups in treatments *P* and *PCP* as are actually going extinct in treatments *P_E* and *PCP_E*, respectively. In order to make the data in the treatments with and without group extinction mechanism comparable, we drop the observations with the lowest group earnings in treatments *P* and *PCP*.

For all three steps we conduct several parametric tests: We calculate OLS and tobit regressions as well as random effects panel regressions to control for panel effects. The different regression models yield similar results. In the following we report the results of OLS regressions, as they are based on minimal assumptions.

Contributions

Figure 2.2 presents mean contributions, mean number of punishment points and counter-points distributed and average earnings by treatment in the first period of each part. It shows that average contributions seem to be slightly higher in *P_E* than in *P* in the first period of each part (13.7 versus 12.4 tokens). However the difference is not significant (p = 0.262)[33]. In contrast, contributions in *PCP_E* are significantly higher than in *PCP* (14.3 versus 11.0 tokens, p = 0.056) even in period 1 of each part, which supports our hypothesis 1.

Figure 2.2 - Mean contributions, punishment, counter-punishment and earnings in the first period of each part.



Note: (Counter-) punishment refers to the average sum of punishment points (counter-points) that subjects distribute to their group members. Earnings refer to the income after (counter-) punishment.

[33] In the following we present the results of two-sided Mann-Whitney U-tests, using the unit of 12 subjects as independent observation. Reported p-values are based on this test unless otherwise noted.

If we pool over the two extinction and the two no-extinction treatments, we find a clear effect of the group extinction mechanism on contributions (11.7 versus 14.0 tokens, p = 0.030). Furthermore, in line with Nikiforakis (2008), contributions in *P* seem to be higher than in *PCP* (12.4 versus 11.0 tokens), however the difference is not significant here (p = 0.336).

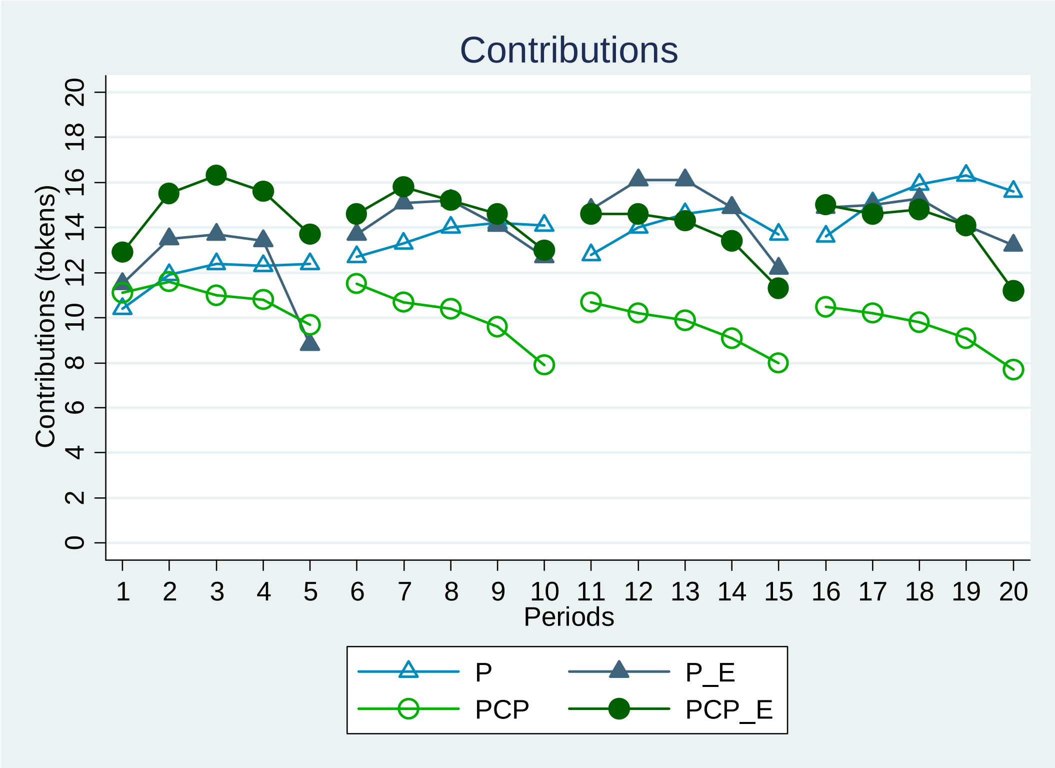
Table 2.3 - Determinants of contributions in period 1 of each part.

Dep. Var.: Contributions	(I)	(II)	(III)	(IV)
Observations from:	*P* and *P_E*	*PCP* and *PCP_E*	All treatments	All treatments
Group extinction	1.392 (1.094)	3.316* (1.696)	2.354** (1.007)	1.392 (1.070)
Two-sided punishment			-0.434 (1.007)	-1.385 (1.414)
Group extinction × Two-sided punishment				1.924 (1.974)
Part	1.060*** (0.220)	0.184 (0.338)	0.622*** (0.217)	0.622*** (0.217)
Constant	9.703*** (0.846)	10.509*** (0.773)	10.318*** (0.782)	10.799*** (0.921)
N	576	576	576	576
R^2	0.054	0.068	0.050	0.056

Notes: This table outlines the results of OLS regressions of contributions in period 1 of each part. Standard errors (clustered on matching groups of 12 subjects) are reported in parentheses. *, **, *** denote significance at the 10%, 5% and 1% level respectively.

Table 2.3 outlines the results of OLS regressions of contributions in period 1 of each part. As regressors we include the following further variables: A dummy variable for the group extinction mechanism, taking value 1 in case of treatment *P_E* or *PCP_E* and 0 otherwise; A dummy variable for two-sided punishment which takes value 0 in case of treatments *P* or *P_E* and value 1 in case of *PCP* or *PCP_E*; And a variable for the interaction effect of the group extinction possibility and the two-sided punishment option. The results of the regressions confirm the results of the non-parametric tests: The group extinction mechanism leads to by 3.3 points higher contributions in *PCP_E* compared to *PCP* in period 1 of each part. For the comparison between *P_E* and *P* the effect goes in the anticipated direction, but is not significant.

Figure 2.3 - Mean contributions over time.



Note: Observations of extinct group members are not included.

Figure 2.3 shows the dynamics of contributions over time for all treatments. If a group is extinct in a certain period, the observation is not included in the average. Contributions in the treatments with group extinction mechanism show a reversed U-form within each part: In early periods contributions increase, in the last period of each part contributions decrease sharply. Contributions might increase at the beginning of each part, as those groups with low contributions get extinct rather at the beginning of each part. The strong "end-part effect" is presumably based on the fact that there is no threat of extinction in the last period of each part and that there are no reputational concerns (as groups are re-shuffled after each part). We further find a typical restart effect at the beginning of each part in treatments *P_E* and *PCP_E*: Contributions are higher in the first period of one part than in the last period of the previous part. In line with Nikiforakis (2008) figure 2.3 further shows that contributions increase in treatment *P* over the periods of each part, whereas in *PCP* contributions decrease and show a typical restart effect at the beginning of each part.

Table 2.4 - Determinants of contributions (Heckman two-step selection model).

	(I)		(II)		(III)		(IV)	
Observations from:	*P* and *P_E*		*PCP* and *PCP_E*		All treatments		All treatments	
Dep. Var.:	Partici-pation	Contri-butions	Partici-pation	Contri-butions	Partici-pation	Contri-butions	Partici-pation	Contri-butions
Group extinction	-0.350*** (0.035)	13.414* (6.983)	-0.175*** (0.038)	1.433 (13.363)	-0.262*** (0.031)	19.635* (11.738)	-0.350*** (0.034)	8.225 (8.610)
Two-sided punishment					0.088** (0.031)	-7.767* (4.005)	0.000 (0.000)	-3.724* (1.997)
Group extinction × Two-sided punishment							0.175*** (0.050)	-0.326 (4.869)
Part	0.049** (0.020)	-0.823 (0.856)	0.042 (0.024)	0.291 (2.631)	0.045*** (0.016)	-2.691 (1.848)	0.045*** (0.016)	-0.729 (0.954)
Period	-0.054*** (0.017)	2.111** (1.021)	-0.026** (0.010)	-0.948 (1.771)	-0.040*** (0.010)	2.391 (1.707)	-0.040*** (0.010)	0.640 (0.952)
Inverse of Mill’s ratio		-89.953* (48.036)		39.998 (174.197)		-162.400 (105.449)		-54.159 (59.093)
Constant	1.040*** (0.044)	35.421*** (12.853)	0.973*** (0.040)	0.578 (51.102)	0.963*** (0.029)	62.173** (31.630)	1.007*** (0.030)	29.233* (16.547)
N	2880	2376	2880	2628	5760	5004	5760	5004
R^2	0.274	0.039	0.140	0.106	0.218	0.067	0.235	0.082
Wald χ^2		14.19***		25.32***		12.72**		23.40***

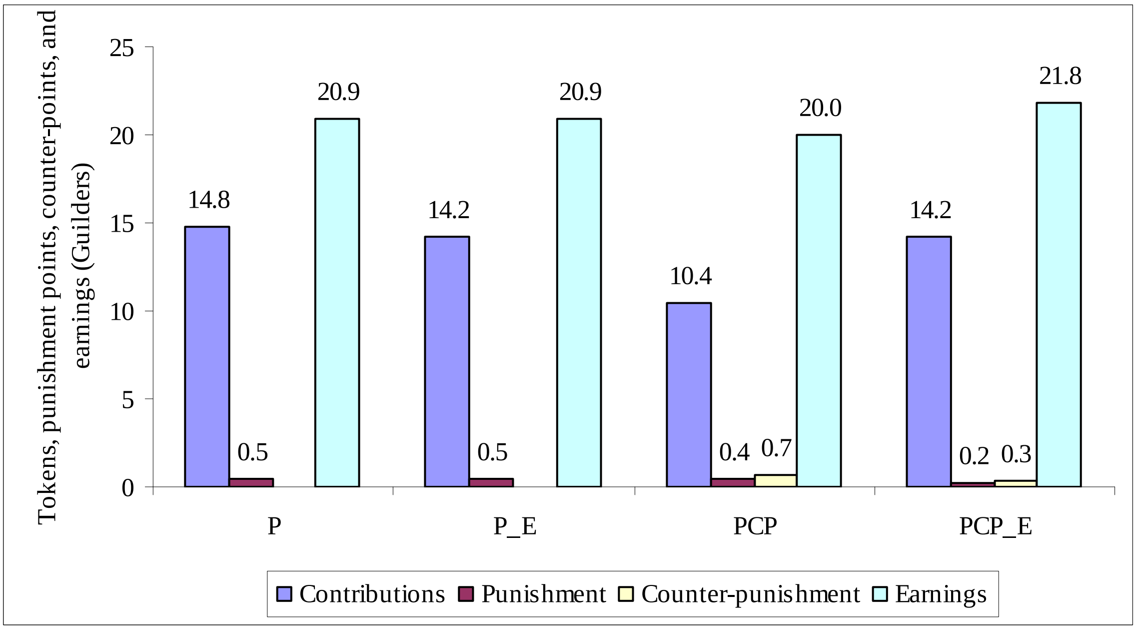
Notes: This table outlines the results of the Heckman two-step selection model of contributions. The model consists of an OLS regression of participation and of contributions. Bootstrapped standard errors (clustered on matching groups of 12 subjects) are reported in parentheses. *, **, *** denote significance at the 10%, 5% and 1% level respectively.

Table 2.4 outlines the results of the Heckman two-step estimation procedure of contributions. For participation, table 2.4 outlines in addition to table 2.2 that – not surprisingly – participation is lower in the treatments where group extinction is possible. For the level of the contributions, corrected by the selection bias, we find that the group extinction mechanism increases contributions, in particular if one-sided punishment is available. In line with Nikiforakis (2008) contributions are higher in *P* than in *PCP*.

Figure 2.4 displays the mean contributions, mean distributed punishment points, counter-points and earnings in the balanced data set (excluding observations with low group income in *P* and in *PCP*), which we create for the robustness check. We find in the balanced data set that subjects contribute 14.8 tokens in treatment *P* and 14.2 tokens in *P_E* (p = 0.631). In treatment *PCP_E* they make by 3.8 tokens higher contributions than in *PCP*, however the difference is not significant when we use non-parametric tests (p = 0.150; 12.3 versus 14.2 tokens over all treatments, p = 0.419). We analyze these data with OLS regressions (table 2.5). The effect of the group extinction mechanism is close to significant in case of a two-sided punishment option. However, we do not find a significant positive effect in case of a one-sided punishment option. This could be due to the fact that the contributions are already quite high in treatment *P*. Comparing the results on contributions in the different analyses, we can state the following result:

RESULT 2.2: *The group extinction mechanism increases contributions – at least in the first period of each part.*

Figure 2.4 - Mean contributions, punishment, counter-punishment and earnings in a balanced data set.



Notes: Observations of extinct group members are not included. (Counter-) punishment refers to the average sum of punishment points (counter-points) that subjects distribute to their group members. Earnings refer to the income after (counter-) punishment.

Table 2.5 - Determinants of contributions in a balanced data set.

Dep. Var.: Contributions	(I)	(II)	(III)	(IV)
Observations from:	*P* and *P_E*	*PCP* and *PCP_E*	All treatments	All treatments
Group extinction	-0.608 (1.657)	3.814♦ (2.136)	1.865 (1.458)	-0.607 (1.621)
Two-sided punishment			-2.133 (1.426)	-4.344** (1.898)
Group extinction × Two-sided punishment				-4.422• (2.651)
Part	0.594* (0.308)	-0.425 (0.363)	0.002 (0.260)	0.002 (0.258)
Period	0.330 (0.229)	-0.508*** (0.128)	-0.130 (0.150)	-0.130 (0.148)
Constant	12.285*** (1.351)	12.965*** (1.532)	13.863*** (1.328)	15.100*** (1.420)
N	1872	2376	4248	4248
R^2	0.023	0.094	0.049	0.078

Notes: This table outlines the results of OLS regressions of contributions in a balanced data set. Standard errors (clustered on matching groups of 12 subjects) are reported in parentheses. *, **, *** denote significance at the 10%, 5% and 1% level respectively, ♦ and • denote 10.2% and 10.9% respectively.

Punishment

We proceed with the analysis of punishment in the same way. Figure 2.2 reveals the average sum of punishment points that subjects distribute to their three group members in period 1 of each part. The number of distributed punishment points seems to be lower in treatment *P_E* compared to *P* (0.8 versus 1.2 punishment points). However the difference is not significant (p = 0.200). For the treatments with two-sided punishment the group extinction mechanism has a significant effect in the first period of each part (p = 0.025): Subjects in *PCP* distribute on average 0.8 punishment points, subjects in *PCP_E* distribute 0.4 punishment points. Including period 1 data from all four treatments, we find a clear negative effect of the group extinction mechanism on punishment (1.0 versus 0.6 punishment points, p = 0.030). These findings strongly support our hypothesis 2.

Table 2.6 outlines the results of OLS regressions of distributed punishment points. We add two additional control variables (see also Nikiforakis, 2008): $Own\_Neg\_Diff_j = \max\{0, c_i - c_j\}$ – where c_i denotes the contribution of subject *i* and c_j denotes the contribution of group member *j* in a given period – as we assume that *i* punishes *j* more severely, the higher the difference between *i*'s and *j*'s contribution (justified punishment); $Group\_Neg\_Diff_j = \max\{0, (\Sigma_{h\neq j}\, c_h / (n - 1)) - c_j\}$, as we assume that *i* punishes *j* more severely, the more *j* deviates from the group norm (pro-social punishment). Both assumptions are confirmed by the results in table 2.6. Controlling for these two variables, we find a clear negative effect of the group extinction mechanism on punishment in the first period of each part. Subjects distribute by 0.14 punishment points less in case of a group extinction option. The effect on punishment exists in case of one-sided as well as in case of two-sided punishment possibilities.

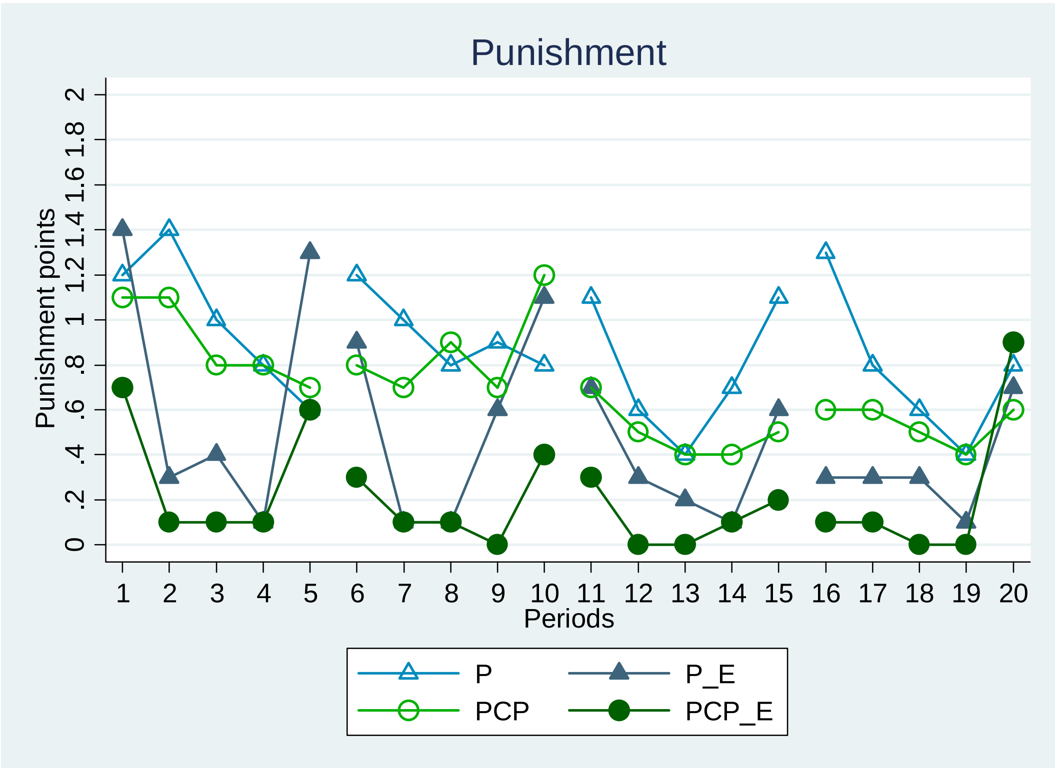
Table 2.6 - Determinants of punishment and counter-punishment in period 1 of each part.

Dep. Var.:	Punishment				Counter-punishment
Observations from:	(I) *P* and *P_E*	(II) *PCP* and *PCP_E*	(III) All treatments	(IV) All treatments	(I) *PCP* and *PCP_E*
Group extinction	-0.162* (0.089)	-0.135** (0.057)	-0.142** (0.051)	-0.155* (0.085)	-0.295 (0.207)
Two-sided punishment			-0.142** (0.051)	-0.155** (0.072)	
Group extinction × Two-sided punishment				0.026 (0.102)	
Part	-0.023 (0.029)	-0.040* (0.021)	-0.032* (0.017)	-0.032* (0.017)	0.072 (0.094)
$Own\_Neg\_Diff_j$	0.060*** (0.018)	0.026* (0.013)	0.042*** (0.011)	0.042*** (0.011)	
$Group\_Neg\_Diff_j$	0.027* (0.013)	0.021 (0.015)	0.026** (0.009)	0.026** (0.010)	
p_{ij}					0.067 0.060
$Own\_Pos\_Diff_j$					0.005 (0.028)
$Group\_Pos\_Diff_j$					-0.018 (0.044)
Constant	0.226* (0.109)	0.230** (0.090)	0.296*** (0.072)	0.302*** (0.079)	0.301 (0.264)
N	1728	1728	3456	3456	151
R^2	0.169	0.080	0.128	0.128	0.040

Notes: This table outlines the results of OLS regressions of punishment points distributed by subject *i* to *j* and of counter-points distributed by subject *j* to *i* in period 1 of each part. For the regression of counter-punishment we include only cases, in which counter-punishment is possible. Standard errors (clustered on matching groups of 12 subjects) are reported in parentheses. *, **, *** denote significance at the 10%, 5% and 1% level respectively.

Figure 2.5 shows the dynamics of punishment over time. It illustrates the average sum of punishment points that subjects distribute to their three group members over the periods of each part. If groups are extinct in a certain period the observation is again not included. In the treatments with group extinction punishment follows a U-form over periods within parts: Punishment drops at the beginning of a part – which might be at least partly due to the extinction of groups with severe punishment – and then shows a strong end-part-effect if group extinction is no threat anymore. In accordance with the contributions we find a restart effect for punishment in treatments *P_E* and *PCP_E*: Punishment is lower in the first period of each part than in the last period of the previous part.

Figure 2.5 - Average sum of distributed punishment points over time.



Notes: This figure outlines the average sum of punishment points that subjects distribute to their group members over time. Observations of extinct group members are not included.

In the next step we present the results of the Heckman two-step procedure for punishment (table 2.7). Punishment incentives seem to be not significantly affected by the extinction mechanism.

We also run a robustness check for punishment with the balanced data set. Figure 2.4 shows that in the balanced data set on average 0.5 points are distributed per subject in treatment *P* as well as *P_E* (p = 1.000). For two-sided punishment on average 0.4 punishment points in treatment *PCP* versus 0.2 punishment points in treatment *PCP_E* are distributed (p = 0.109; 0.462 versus 0.347 punishment points over all treatments, p = 0.273). Table 2.8 displays the results of OLS regressions of distributed punishment points in the balanced data set, which mainly confirm the results of the period 1 analysis: In the balanced data set group

Table 2.7 - Determinants of punishment (Heckman two-step selection model).

	(I)		(II)		(III)		(IV)	
	P and *P_E*		*PCP* and *PCP_E*		All treatments		All treatments	
Dep. Var.:	Partici-pation	Punish-ment	Partici-pation	Punish-ment	Partici-pation	Punish-ment	Partici-pation	Punish-ment
Group extinction	-0.350*** (0.035)	-0.765 (0.514)	-0.175*** (0.038)	0.064 (0.567)	-0.265*** (0.031)	-0.067 (0.381)	-0.350*** (0.034)	-0.600 (0.509)
Two-sided punishment					0.088*** (0.031)	-0.113 (0.145)	0.000 (0.000)	-0.111* (0.061)
Group extinction × Two-sided punishment							0.175*** (0.050)	0.281 (0.064)
Part	0.049** (0.020)	0.066 (0.073)	0.042 (0.024)	-0.058 (0.138)	0.045*** (0.015)	-0.027 (0.069)	0.045*** (0.015)	0.038 (0.064)
Period	-0.054*** (0.017)	-0.087 (0.074)	-0.026** (0.010)	0.045 (0.085)	-0.040*** (0.010)	0.022 (0.057)	-0.040*** (0.010)	-0.037 (0.055)
$Own\_Neg\_Diff_j$		0.069*** (0.021)		0.037*** (0.009)		0.052*** (0.011)		0.052*** (0.010)
$Group\_Neg\_Diff_j$		0.034** (0.015)		0.047*** (0.014)		0.041*** (0.010)		0.041*** (0.010)
Inverse of Mill's ratio		3.891 (3.498)		-2.738 (8.160)		-0.799 (3.605)		2.844 (3.403)
Constant	1.040*** (0.043)	-0.903 (0.978)	0.973*** (0.040)	0.853 (2.457)	0.962*** (0.029)	0.380 (1.118)	1.007*** (0.030)	-0.661 (0.986)
N	8640	7128	8640	7884	17280	15012	17280	15012
R^2	0.274	0.202	0.140	0.154	0.218	0.175	0.235	0.175
Wald χ^2		140.59***		80.60***		112.35***		138.97

Notes: This table outlines the results of the Heckman two-step selection models of punishment points distributed by subject *i* to *j*. The model consists of OLS regressions of participation and of punishment. Bootstrapped standard errors (clustered on matching groups of 12 subjects) are reported in parentheses. *, **, *** denote significance at the 10%, 5% and 1% level respectively.

Table 2.8 - Determinants of punishment and counter-punishment in a balanced data set.

Dep. Var.:	Punishment				Counter-punishment
	(I)	(II)	(III)	(IV)	(I)
Observations from:	*P* and *P_E*	*PCP* and *PCP_E*	All treatments	All treatments	*PCP* and *PCP_E*
Group extinction	-0.087* (0.046)	-0.054 (0.033)	-0.063** (0.028)	-0.077* (0.045)	-0.200 (0.118)
Two-sided punishment			-0.067** (0.028)	-0.079** (0.037)	
Group extinction × Two-sided punishment				0.024 (0.054)	
Part	-0.015 (0.013)	0.002 (0.012)	-0.006 (0.009)	-0.006 (0.009)	0.067 (0.066)
Period	-0.023 (0.013)	-0.002 (0.009)	-0.012 (0.008)	-0.012 (0.008)	-0.026 (0.040)
$Own\_Neg\_Diff_j$	0.055*** (0.015)	0.026*** (0.008)	0.039*** (0.008)	0.039*** (0.008)	
$Group\_Neg\_Diff_j$	0.024** (0.010)	0.038** (0.016)	0.031*** (0.009)	0.031*** (0.009)	
p_{ij}					0.090 (0.058)
$Own\_Pos\_Diff_j$					0.024 (0.044)
$Group\_Pos\_Diff_j$					-0.038 (0.053)
Constant	0.170** (0.064)	0.018 (0.053)	0.123**	0.129** (0.050)	0.284 (0.190)
N	5616	7128	12744	12744	352
R^2	0.183	0.129	0.152	0.152	0.036

Notes: This table outlines the results of OLS regressions of punishment points distributed by subject *i* to *j* and of counter-points distributed by subject *j* to *i* in a balanced data set. For the regression of counter-punishment we include only cases, in which counter-punishment is possible. Standard errors (clustered on matching groups of 12 subjects) are reported in parentheses. *, **, *** denote significance at the 10%, 5% and 1% level respectively.

extinction decreases punishment, in particular in case of one-sided punishment possibilities. In line with hypothesis 2 we can summarize the following result for punishment:

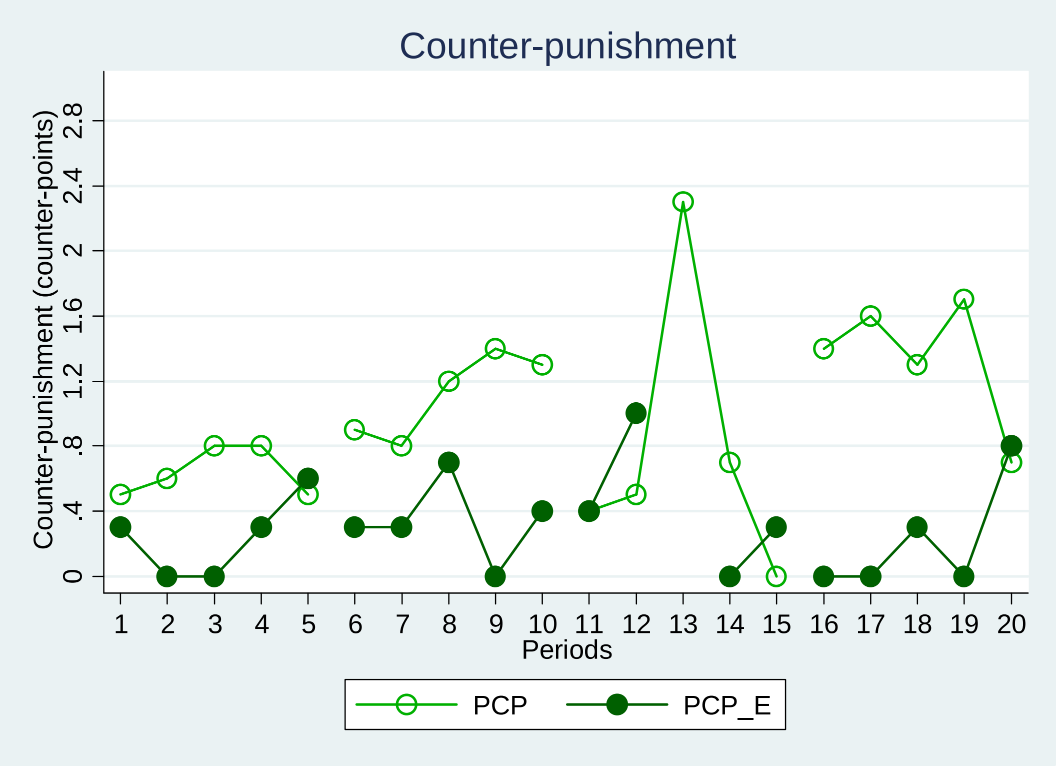
RESULT 2.3: *The group extinction mechanism leads to less severe punishment – at least in the first period of each part.*

Counter-punishment

Figure 2.2 shows the average sum of counter-points that participants distribute to their group members in the first period of each part. In the following we include only those cases, in which counter-punishment is possible, i.e. when the period income of the punishing and punished subject at the beginning of the third stage is positive and when the punished person punished the punisher in the second stage. There are significantly more cases in treatment *PCP* than in *PCP_E*, in which counter-punishment is possible in the first period of each part (101 versus 50 cases, p = 0.010), due to more frequent punishment in *PCP* than in *PCP_E*. Subjects distribute on average 0.8 counter-points in treatment *PCP* and 0.3 counter-points in *PCP_E* in the first period of each part (p = 0.121). This result goes in the direction of hypothesis 3.

Table 2.6 outlines the results of OLS regressions of distributed counter-points in period 1 of each part. We add the following control variables following Nikiforakis (2008): p_{ij}, the number of punishment points that subject i distributes in the second stage to group member j; $Own\_Pos\_Diff_j = \max\{0, c_j - c_i\}$, the positive difference in contributions between subject j and subject i; and $Group\_Pos\_Diff_j = \max\{0, c_j - (\Sigma_{h \neq j}\, c_h / (n-1))\}$, the positive

Figure 2.6 - Average sum of distributed counter-points over time.



Notes: This figure outlines the average sum of counter-points that subjects distribute to their group members whenever possible over time. Observations of extinct group members are not included. In period 3 of part 3, no possible case of counter-punishment exists.

deviation of subject *j*'s contribution from the group norm. We include the last two variables to control for unjustified and anti-social punishment. Table 2.6 indicates that the effect of the group extinction mechanism on counter-punishment goes in the direction of our hypothesis, but is not significant in the first period of each part.

Figure 2.6 illustrates the mean number of distributed counter-points per subject over time. Note that if a group is extinct in a certain period the observation is not included in the average. As we further include only those cases, in which counter-punishment is possible, the number of participants on which the period averages are based varies heavily between periods. Hence it is difficult to find particular regularities in the dynamics.

Table 2.9 shows the results of the Heckman two-step procedure for counter-punishment. The effect of the group extinction mechanism goes again in the direction of our hypothesis, but is not significant.

Table 2.9 - Determinants of counter-punishment (Heckman two-step selection model).

Dep. Var.:	Participation	Counter-punishment
Group extinction	-0.819*** (0.023)	-3.299 (2.227)
Part	0.039* (0.021)	0.219** (0.111)
Period	-0.047** (0.017)	-0.116 (0.130)
p_{ij}		0.095 (0.067)
Own_Pos_Diff$_j$		0.063 (0.055)
Group_Pos_Diff$_j$		0.005 (0.044)
Inverse of Mill's ratio		7.373 (5.721)
Constant	1.041*** (0.063)	-1.831 (1.537)
N	1238	482
R^2	0.673	0.060
Wald χ^2		16.89**

Notes: This table outlines the results of the Heckman two-step selection model of counter-points distributed by subject *j* to *i*. The model consists of an OLS regression of participation and of counter-punishment. We include only cases in which counter-punishment is possible. Bootstrapped standard errors (clustered on matching groups of 12 subjects) are reported in parentheses. *, **, *** denote significance at the 10%, 5% and 1% level respectively.

Figure 2.4 shows the average sum of counter-points distributed by subjects to their group members in the balanced data set. Subjects distribute by 0.4 counter-points less in treatment *PCP_E* compared to treatment *PCP* in the balanced data set (0.3 versus 0.7 counter-points, p = 0.065). However, if we control for several variables as outlined in table 2.8 the effect of the group extinction mechanism is not significant anymore.

RESULT 2.4: *We find a slight effect of the group extinction mechanism on the counter-punishment level; however it is not significant.*

Earnings

Figure 2.2 reveals higher earnings after (counter-) punishment in the treatments with than in the treatments without the group extinction mechanism in period 1 of each part: In treatment *P_E* subjects earn by 2.0 Guilders more than in treatment *P* (19.7 versus 17.7 tokens, p = 0.078), in *PCP_E* they earn on average 3.1 Guilders more than in *PCP* (21.3 versus 18.2, p = 0.010). If we include period 1 data from all four treatments, the group extinction mechanism has a very significant effect on earnings (20.5 versus 17.9 Guilders; p = 0.002). These findings strongly support our main hypothesis, i.e. the group extinction mechanism increases efficiency.

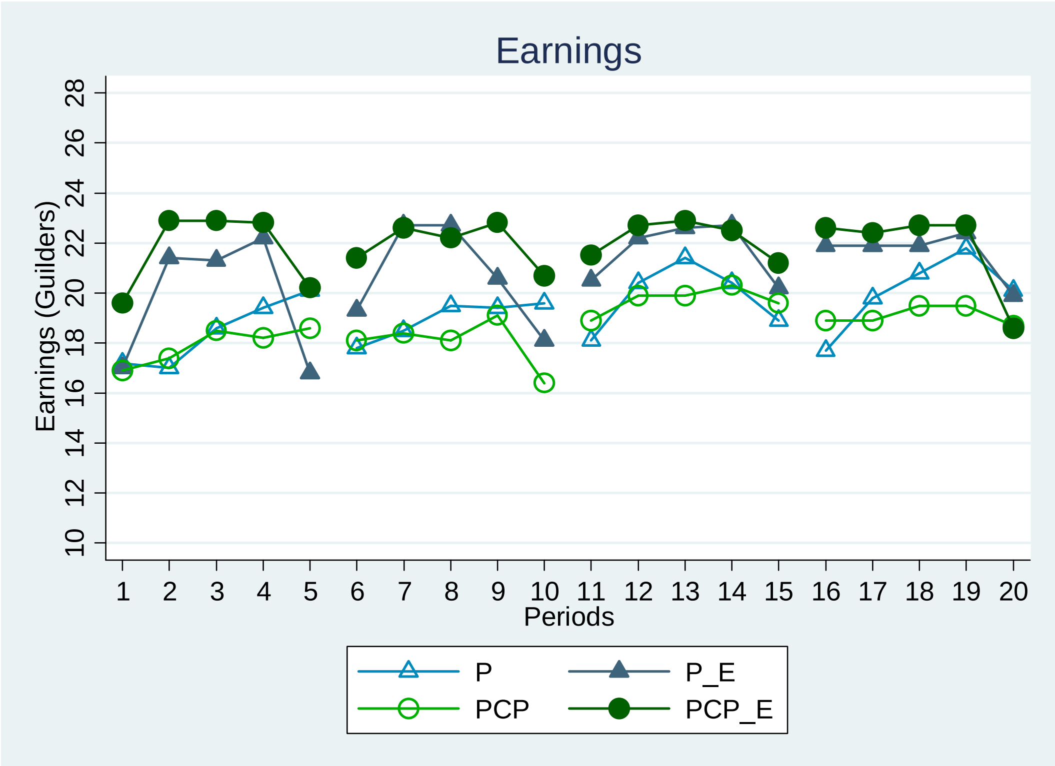
Table 2.10 - Determinants of profits in period 1 of each part.

Dep. Var.: Profits	(I)	(II)	(III)	(IV)
Observations from:	*P* and *P_E*	*PCP* and *PCP_E*	All treatments	All treatments
Group extinction	1.981* (0.942)	3.058*** (0.872)	2.518*** (0.637)	1.981** (0.921)
Two-sided punishment			1.068 (0.637)	0.529 (1.014)
Group extinction × Two-sided punishment				1.077 (1.255)
Part	0.899** (0.308)	0.787** (0.253)	0.843*** (0.195)	0.843*** (0.195)
Constant	15.426*** (1.145)	16.235*** (1.128)	15.297*** (0.834)	15.566*** (0.893)
N	576	576	576	576
R^2	0.053	0.072	0.068	0.070

Notes: This table outlines the results of OLS regressions of final profits after (counter-) punishment in period 1 of each part. Standard errors (clustered on matching groups of 12 subjects) are reported in parentheses. *, **, *** denote significance at the 10%, 5% and 1% level respectively.

The regression results in table 2.10 confirm the non-parametric results. Table 2.10 further reveals that earnings increase over parts, which is in line with the learning effect that we reported in section 2.4.1. Similar to the contributions, average earnings in the treatments with the group extinction possibility show a reversed U-form within each part (groups that are extinct are not included in the averages, see figure 2.7).

Figure 2.7 - Mean earnings over time.



Notes: This figure illustrates the earnings after (counter-) punishment over time. Observations of extinct group members are not included.

According to the results of the Heckman two-step procedure (Table 2.11) the group extinction mechanism increases profits, in particular in the case of the one-sided punishment options.

Figure 2.4 shows no differences between treatment *P* and *P_E* for the earnings in the balanced data set (20.9 Guilders in both treatments, $p = 0.873$). However, we find that subjects make higher profits in treatment *PCP_E* compared to *PCP* in the balanced data set (21.8 versus 20.0 Guilders, $p = 0.016$). If we include the data of all four treatments, we can also confirm that the group extinction mechanism has a significant effect on earnings (20.4 versus 21.4 Guilders, $p = 0.050$). Specifications (I) to (IV) in table 2.12 show that the group extinction mechanism increases earnings by about 1 Guilder in the balanced data set. The effect is especially high in case of the two-sided punishment options, where the threat of group extinction increases earnings by 1.8 Guilders.

RESULT 2.5: *The group extinction mechanism increases earnings.*

Table 2.11 - Determinants of earnings (Heckman two-step selection model).

	(I)		(II)		(III)		(IV)	
Observations from:	*P* and *P_E*		*PCP* and *PCP_E*		All treatments		All treatments	
	Partici-pation	Earnings	Partici-pation	Earnings	Partici-pation	Earnings	Partici-pation	Earnings
Group extinction	-0.350*** (0.034)	16.198** (6.915)	-0.175*** (0.038)	0.945 (6.583)	-0.263*** (0.031)	12.898** (5.719)	-0.350*** (0.034)	12.831* (7.294)
Two-sided punishment					0.088** (0.031)	-3.416 (2.149)	0.000 (0.000)	-0.627 (1.007)
Group extinction × Two-sided punishment							0.175*** (0.050)	-4.380 (3.935)
Part	0.049** (0.020)	-1.311 (0.993)	0.042 (0.024)	0.844 (1.629)	0.045*** (0.016)	-1.301 (1.000)	0.045*** (0.016)	-0.905 (0.951)
Period	-0.054*** (0.017)	2.547** (1.015)	-0.026** (0.010)	-0.358 (1.025)	-0.040*** (0.010)	1.724** (0.857)	-0.040*** (0.010)	1.369* (0.814)
Inverse of Mill's ratio		-100.489** (48.1222)		31.167 (95.816)		-99.099* (53.284)		-77.491 (51.110)
Constant	1.040*** (0.044)	43.981*** (13.417)	0.973*** (0.040)	8.671 (28.771)	0.963*** (0.029)	47.402*** (16.326)	1.007*** (0.030)	39.819*** (14.829)
N	2880	2376	2880	2628	5760	5004	5760	5004
R^2	0.274	0.053	0.140	0.074	0.218	0.058	0.235	0.061
Wald χ^2		64.78***		21.43		44.02***		45.07***

Notes: This table outlines the results of the Heckman two-step selection model of earnings after (counter-) punishment. The model consists of OLS regressions of participation and of earnings. Bootstrapped standard errors (clustered on matching groups of 12 subjects) are reported in parentheses. *, **, *** denote significance at the 10%, 5% and 1% level respectively.

Table 2.12 - Determinants of profits in a balanced data set.

Dep. Var.: Profits	(I)	(II)	(III)	(IV)
Observations from:	*P* and *P_E*	*PCP* and *PCP_E*	All treatments	All treatments
Group extinction	-0.001 (0.430)	1.781*** (0.489)	0.996** (0.373)	-0.001 (0.423)
Two-sided punishment			-0.012 (0.390)	-0.903 (0.527)
Group extinction × Two-sided punishment				1.783** (0.638)
Part	0.543** (0.184)	0.035 (0.142)	0.248* (0.123)	0.248* (0.121)
Period	0.621** (0.251)	0.205 (0.154)	0.393** (0.144)	0.393** (0.143)
Constant	17.770*** (0.840)	19.369*** (0.725)	18.673*** (0.569)	19.172*** (0.618)
N	1872	2376	4248	4248
R^2	0.052	0.035	0.027	0.035

Notes: This table outlines the results of OLS regressions of final profits after (counter-) punishment in a balanced data set. Standard errors (clustered on matching groups of 12 subjects) are reported in parentheses. *, **, *** denote significance at the 10%, 5% and 1% level respectively.

2.4.3 Results of the Questionnaire on Group Extinction

With the post-experimental questionnaire in the treatments *P_E* and *PCP_E* we examine the driving motivational forces behind the behavior in the treatments with group extinction. The detailed questions and results are presented in appendix A.2.3.

Subjects generally state that group extinction has an effect on their decisions (see statement 4). Subjects in the treatments with group extinction declare that group extinction induces them to contribute more and punish less (statements 2 and 3). This shows again that higher contributions and less punishment in the treatments with group extinction are based on a real behavioral change.

Subjects in the treatments with group extinction declare to make rather high contributions in order to reach the threshold and not to run the risk of being punished for low contributions, which would have increased the likelihood of extinction (see question 1 and statements 2 and 8). This is in line with our reasoning in section 2.3.3 and in particular hypothesis 1.

In line with our hypotheses 2 and 3 subjects agree that the threshold induces them to punish and counter-punish less (see question 1 and statement 3). The results of statements 5 to 7 show that punishment and counter-punishment take place also in the treatments with group extinction for emotional reasons like revenge for low contributions or revenge for high punishment, and for strategic reasons.

2.5 Conclusion

In our study we examine experimentally whether a survival threshold, often naturally applied in practice, has an effect on the efficiency in the context of a public goods game with (counter-) punishment. In particular we hypothesize that the group extinction mechanism increases contributions, decreases punishment and counter-punishment and therefore increases efficiency.

Table 2.13 - The effects of group extinction on contributions, (counter-) punishment, and earnings.

			Treatments *P* versus *P_E*	Treatment *PCP* versus *PCP_E*	Treatments *P* and *PCP* versus *P_E* and *PCP_E*
Contributions	First period		+	+*	+**
	All periods	Selection model	+* (+)	+ (–)	+* (+)
		Balanced data	– (–)	+♦ (+*)	+ (+)
Punishment	First period		–*	–**	–*
	All periods	Selection model	– (–)	+ (+)	– (+)
		Balanced data	–* (–**)	– (–**)	–** (–***)
Counter-punishment	First period			–	
	All periods	Selection model		– (+)	
		Balanced data		– (–*)	
Earnings	First period		+*	+***	+***
	All periods	Selection model	+** (+)	+ (–)	+** (+)
		Balanced data	– (+)	+*** (+***)	+** (+***)

Notes: This table summarizes our results on the effect of the group extinction mechanism on contributions, punishment, counter-punishment, and earnings in case of one-sided and two-sided punishment separately and all together. The results are presented for the first period of each part and over all periods – measured with the two-step selection model as well as with an OLS regression with a balanced data set. +, (–) denotes a positive (negative) effect of group extinction. *, **, *** denote significance at the 10%, 5% and 1% level respectively, ♦ denotes p = 10.2%. In parentheses we list the results for periods 1 to 4 in each part, i.e. as long as the threat of extinction really exists.

We find that group extinction takes place and that groups learn to adapt their behavior to the survival threshold over time. Table 2.13 summarizes our results on the effects of the group extinction mechanism and shows that we can largely confirm our hypotheses. We apply several different statistical methods, and they all point in the same direction (even if not every analysis is significant on conventional levels). First of all, we can clearly confirm our main hypothesis: The group extinction mechanism increases efficiency. Our main result is a consequence of both a positive effect of the possibility of group extinction on contributions and a negative effect of the group extinction mechanism on punishment. The effect on counter-punishment goes in the direction of our hypothesis. However it is not significant – which might be due to the low number of cases in which counter-punishment can take place.

We might even under-estimate the effects of extinction, as groups learn to adapt to the group extinction mechanism over time. In recent papers punishment options are extended to the possibility of feuds (Nikiforakis and Engelmann, 2011; Nikiforakis et al., 2012; Engelmann and Nikiforakis, 2013). The effect of group extinction might even be enhanced in such a setting. We leave this question for further research. As our idea of a survival threshold theoretically and behaviorally links to the idea of competition among groups (e.g., Tan and Bolle, 2007; Strasser, 2012) we further recommend examining these links in greater detail.

Our results may provide a further explanation – among others – why in field studies in contrast to laboratory studies low levels of peer punishment have been observed (e.g. Balafoutas and Nikiforakis, 2012). Often in the field – e.g. in working teams, tribal groups or societies – efficiency concerns naturally exist and therefore may lead to a decrease in punishment. Furthermore our results imply that the threat of group extinction can also be applied, adapted or made salient in the field in order to increase contributions and decrease peer punishment, e.g. in working groups. A field experiment taking up this idea would be an interesting extension of our results.

2.6 Appendix A.2

Appendix A.2.1: Instructions for treatment *PCP_E* (translated from German)

Welcome to the experiment! Thank you very much for participating!
From now on, please refrain from talking with the other participants!

General information

The purpose of this experiment is the study of decision behavior. You can earn money during this experiment which you will be paid in cash at the end of the experiment.
During the experiment you and the other participants will take decisions. Your own decisions as well as those of the other participants will affect your payment according to the rules explained in the following.
The duration of the experiment is 120 minutes. If you should have any questions, please raise your hand. One of the experimenters will come to you and answer your question in private.
During the experiment your earnings will be calculated in **Guilders.** At the end of the experiment your earnings in Guilders will be converted into Euros according to the following exchange rate:

1 Guilder = 0.04 Euro (25 Guilders = 1 Euro)

For your punctual arrival you receive 100 Guilders (4 Euros). These can be used to pay for eventual losses during the experiment. **However you can <u>always</u> prevent losses by deciding accordingly.** At the end of the experiment you will be paid your total earnings from the experiment as well as the 4 Euros in private and in cash.
While you take your decisions a clock will count down at the upper right corner of the computer screen. This countdown serves as an orientation for how much time you should need to take the decision. However there is no strict time limit. You can take more time in order to make your decision, especially in the beginning this will be the case. Only the purely informational screens without any decisions to be taken will close after the countdown.

<u>Anonymity</u>
Your input over the course of the experiment is anonymous. You will at no time receive personal information about the other participants. We will not link your name to data from the experiment. At the end of the experiment you will sign a confirmation which says that you have received the payment. This confirmation only serves accounting purposes for our sponsor. Our sponsor will not receive any data from the experiment either.

<u>Utilities</u>
For this experiment you are given a pen. We kindly ask you to leave the pen at your place after the experiment.

<u>The experiment</u>
The experiment consists of **five parts and two questionnaires**. You receive specific instructions for each segment of the experiment. We start with the identical parts I to IV. Then the first questionnaire follows. You will receive the instructions for part V and the second questionnaire after completion of the first questionnaire.
Your total earnings in this experiment consist of the sum of the earnings from parts I to V. Parts I to IV and part V are independent of one another, i.e. decisions in parts I to IV have **no** influence on the earnings in part V.

page 1 of 10

Part I to IV and first questionnaire

We will conduct **four identical parts**. The following paragraphs describe the rules of one of those four parts.

Every part consists of **five periods**. You are a member of a **4-member-group**, i.e. you are in one group together with three other members who will be chosen randomly from the other participants. **The group consists of the same four members in all five periods.**

Every period consists of **three stages**. In the first stage you decide how many points you want to contribute to a project. In the second stage you will be informed about the contributions of the other three group members. You can then decide, whether and by how much you want to reduce the earnings of the other group members by allocating points to the other group members. In the third stage you are informed about how many points you have received from the other group members. You can then decide whether and by how much you want to reduce the earnings of those group members who allocated points to you in the second stage by allocating counter-points. You should keep in mind however, that the total earnings of your group shouldn't fall below a certain threshold. Now we will describe the rules in detail. Directly afterwards we ask you to complete some sample exercises on the screen that help to make you comfortable with the decisions you have to take.

First stage

<u>Your decision</u>
You are a member of a 4-member-group. Every member of the group receives 20 points at the beginning of the stage and has to **decide how to allocate these 20 points**. The 20 points can either be **kept for oneself** or **contributed fully or in parts to a project**. Every point that is not contributed to the project is automatically kept for oneself. At the beginning of every period the following screen appears:

Remaining Time:

Period 1 – First stage

Your Endowment: 20

Your contribution to the project: [ ]

OK

HELP
Please enter an integer between 0 and 20. As soon as you have taken your decision, please click „OK“.

You decide how many points you want to contribute to the project by typing an integer between 0 and 20 (whereby 0 and 20 are also possible) into the input box. This number is your contribution to the project. After clicking the "OK" button you cannot change your decision.

<u>Your income in the first stage</u>

Once all members of your group have taken their decision, the following screen shows you your contribution to the project and, in the second line, the total amount of contributed points of all four group members, i.e. including your contribution. Furthermore the last lines show you your earnings in Guilders in the first stage.

Remaining Time:

Period 1 – First stage

Your contribution to the project:

Sum of contributions to the project in your group
(your contribution included):

Your income from the points you kept for yourself:

Your income from the project:

Your income in the first stage:

Continue

Your earnings consist of two parts:

a) **Earnings from the points you kept for yourself** (= 20 – points contributed to the project): For each point you kept for yourself you earn exactly one Guilder. No one else will receive any earnings from the points you kept for yourself. Conversely you will not receive any earnings from the points the other group members kept for themselves. Your earnings from the points you kept for yourself are shown on the screen in the third line.
b) **Earnings from the project** (= 0.3 * sum of all group members' contributions): All group members receive earnings from your contribution to the project. Conversely you also receive earnings from the other group members' contributions to the project. For every point contributed to the project you and the other group members receive 0.3 Guilders. Your earnings from the project, which are equal to those of the other group members, are shown on the screen in the forth line.

In the first stage of every period your earnings result from the sum of both components:

Your earnings (in Guilders) in the first stage =

(20 – „your contribution to the project") + 0.3*"sum of all group members' contributions"

Earnings from points you kept — Earnings from the project

The **earnings from the project** are calculated in exactly the same way for each group member, i.e. every group member receives the same earnings from the project. For example, assume that the sum of all contributions amounts to 60 points. You and the other three group members will receive 0.3 * 60 = 18 Guilders each from the project. If the total contributions amount to 9 points, every group member receives 9 * 0.3 = 2.7 Guilders from the project.

For every point you keep for yourself you receive exactly one Guilder. Assume instead of having kept the point for yourself you had contributed this point to the project. Your earnings from the project would rise by 0.3 Guilders. At the same time the earnings of every other group member would also rise by 0.3 Guilders. The total earnings of the group would increase by 1.2 Guilders. Your contribution to the project also increases the earnings of other group members. Conversely your earnings also increase if the other group members contribute more to the project. Every point that a group member contributes to the project increases your earnings by 0.3 Guilders independent of whether you contributed it to the project or another group member. For example if you contribute 5 points to the project the earnings of every group member increase by 0.3 * 5 = 1.5 Guilders. If another group member contributes, for example, 10 points your earnings as well as all other group members' earnings increase by 0.3 * 10 = 3 Guilders independent from your contribution or the contributions of the other group members.

Your earnings from the points you kept for yourself automatically result from the **difference between your initial 20 points and your contribution to the project**. For example, if you contribute 8 points to the project you receive 20 - 8 = 12 Guilders from the points you kept for yourself. If you contribute 13 points to the project your earnings from the points you kept for yourself amount to 20 - 13 = 7 Guilders.

Second stage

<u>Your decision</u>

In the second stage you are shown the contributions of every member in your group. In this stage you can **reduce the earnings of each group member or leave them unchanged by allocating points**. The other group members can reduce your earnings in the same way.

Remaining Time:

Period 1 – Second stage

	You	Group member A	Group member B	Group member C
Endowment	20	20	20	20
Contributions to the project				
Contributions in % of the endowment				
Your decision in the second stage		○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10

Your overall cost of distributing points amount to:

OK

HELP
Please indicate on the three scales, how many points you want to distribute to your three group members. As soon as you have taken your decisions, please click „OK".

page 4 of 10

On this screen you are shown the other group members' contributions to the project in the first stage (third line on the previous screen). Your contribution is shown in the left column, the contributions of the other group members are shown in the other columns. In addition the contributions are shown as a percentage of the initial endowment of 20 points.

For identification purposes each group member randomly receives a unique ID at the beginning of each period. This ID remains the same over the whole period such that you can distinguish the group members' individual actions. Hence group member B in stage two of a certain period is exactly the same person as group member B in stage three. However this ID changes between periods, i.e. group member B from period one might receive the ID C or D in period two. This change of member IDs ensures that individual decisions cannot be linked over several periods. In the first line on this screen you are shown the IDs of the three other group members for this period.

Then you decide for each group member, whether and how many points you want to allocate to this group member. In the fifth row you select on a scale the number of points you want to allocate to each individual group member. You can allocate any number of **integer** points in the range between 0 and 10 (including 0 and 10). The three scales refer to your three group members and belong to the group member, whose contribution is outlined in the third line, respectively. If you do <u>not</u> want to reduce the earnings of a group member, you have to click on 0 points.

<u>Cost of your decision</u>
Allocating points is costly, these costs depend on the amount of points you allocate. Each point you allocate costs you **one Guilder**, e.g. if you allocate two points to some group member this costs you two Guilders. Allocating another 4 points to some other group member costs you additional 4 Guilders. Allocating another point to the third group member costs you another Guilder such that the total allocation of 7 points costs you 7 Guilders. Your total costs are shown directly below the three scales in row six. Be aware that you cannot change your decision once you've clicked the OK button.

<u>Earnings of each group member</u>
If you allocate 0 points to a group member, her earnings remain unchanged. If you allocate one point (by clicking "1" on the scale) her earnings are reduced by 3 Guilders. If you allocate two points (by clicking "2" on the scale) her earnings are reduced by 6 Guilders. Every point you allocate to a group member reduces her earnings by 3 Guilders. Likewise each point that was allocated to you by another group member reduces your earnings by 3 Guilders.

Costs due to received points (in Guilders) = 3 * sum of received points

Hence the sum of received points determine by how much earnings are reduced. For example if a group member receives a total of 3 points from the other group members her earnings are reduced by 3 * 3 = 9 Guilders. If she receives a total of 4 points her income is reduced by 12 Guilders.

The only exception to this rule is that if the costs due to received points are higher than the first stage income of the group member her earnings are only reduced to 0 points and not further. The group member must however bear the costs of the points that she herself allocated to the other group members.

Earnings at the end of the second stage are calculated as follows:

page 5 of 10

Earnings (in Guilders) at the end of the second stage =
Earnings from stage 1 – 3 * (sum of points received) – sum of points allocated
if the cost of receiving points < earnings from stage 1

OR

= 0 – sum of points allocated
if the cost of receiving points ≥ earnings from stage 1

For example, imagine your earnings at the end of stage 1 amount to 20 Guilders, you further receive a total of 3 points from the other group members and allocate one point to another group member in stage 2, then your earnings at the end of stage 2 amount to 20 – 3 * 3 – 1 = 10 Guilders. If your earnings at the end of stage 1 amount to 12 Guilders, and you receive a total of 5 points and allocate two points in stage 2, then your earnings at the end of stage 2 amount to 0 – 2 = -2 Guilders.

The latter example shows that your earnings at the end of stage 2 can become negative if the cost due to allocated points are higher than the (reduced) earnings from stage 1. You can prevent losses with certainty by not allocating any points yourself. Possible losses are covered by the 100 Guilders that you received at the beginning of the experiment for arriving punctually.

After all participants have taken their decisions your earnings at the end of stage 2 are shown on the following screen:

Remaining Time:

Period 1 – Second stage

Your income from the first stage:

You received the following number of points in the second stage:

Your cost of distributing points:

Your income at the end of the second stage:

Continue

Third Stage

At the beginning of the third stage you are shown the amount of points you received from each of the other group members in stage 2:

Period 1 – Third stage

Your group members	Group member A	Group member B	Group member C
The number of points you received in the second stage			
Your decision in the third stage	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10	

Your overall cost of distributing counter-points amount to:

OK

HELP
Please indicate how many counter-points you would like to distribute to each of the other group members. You can only distribute counter-points to those group members, from whom you received points in the second stage and whose income is positive at the end of the second stage.
As soon as you have taken your decisions, please click „OK".

Then you have the option to allocate counter-points to the group members from whom you have received points in the second stage. You can only allocate counter-points to group members from whom you have received points in the second stage and whose earnings after stage 2 are positive. Further you can only allocate counter-points if your earnings after stage 2 are positive. If all three conditions are met, the scales appear on screen.

The costs due to counter-points are identical to those from allocating and receiving points in stage 2. Earnings at the end of stage 3 are calculated as follows:

Earnings (in Guilders) at the end of the third stage =
Earnings from stage 2 – 3 * (sum of counter-points received) – sum of counter-points allocated
if the costs of receiving counter-points < earnings from stage 2

OR

= 0 – sum of counter-points allocated
if the costs of receiving counter-points ≥ earnings from stage 2

Analogously to stage 2: Imagine your earnings from stage 2 amount to 20 Guilders and you receive a total of 3 counter-points from the other group members and allocate one counter-point to another group member in stage 3, then your earnings at the end of stage 3 amount to 20 – 3 * 3 – 1 = 10 Guilders. If your earnings from stage two amount to 12 Guilders, and you receive a total of 5 counter-points and allocate two counter-points in stage 3, then your earnings at the end of stage 3 amount to 0 – 2 = -2 Guilders.

page 7 of 10

After completion of stage 3 you are informed of the total amount of received counter-points and your earnings in this period:

Remaining time:
Period 1 – Third stage
Your income from the second stage:
You reveived the following number of counter-points in the third stage:
Your cost from distributing counter-points:
Hence, your period income amounts to:
Your income so far in part I (the income of this period included):
Continue

As your earnings in this period can become negative, possible losses are covered by the 100 Guilders that you received at the beginning of this experiment. Your total earnings in this part result from the sum of earnings from all five periods.

Additional rule

Please be aware of the following additional rule: **If your group's total earnings in a period are below 80 Guilders after completion of stage 3, your group will not be playing in the part's subsequent periods.** Thus if in one period the sum of individual earnings in your group is below 80 Guilders your group is excluded from the subsequent periods of the part. **You receive the earnings from that period but you will not receive any earnings in the part's subsequent periods i.e. each group member's earnings in these periods will amount to 0.** The following screen shows you whether your group's total earnings were above or below 80 Guilders in the current period. In the case below the total earnings surpassed 80 Guilders.

Period 1
The overall income of all four members in your group (you included) amounts in this period to:
Hence, the group's income amounts to at least 80 Guilders. Therefore you are allowed to participate in the following period.
Continue

The procedure of the parts II, III and IV will be exactly as in part I. Each part consists of five periods with three stages. In the first stage you decide how many of your 20 points of endowment to contribute to a project and how many you want to keep for yourself. In stage 2 you can reduce the earnings of the other group members by allocating points. In the third stage you can reduce the earnings of those group members from whom you received points in the second stage. As in part I you have to keep in mind that your group will be excluded from the part's subsequent periods if your group's earnings fall below the threshold of 80 Guilders.

The group composition <u>within</u> a part is constant, i.e. during one part you interact with the same three other group members. **<u>Between</u> parts however the group composition will be changed at random**, i.e. in part II you will interact with other group members than in part I, and in part III you will again interact with other group members. Between the different parts the experimenter announces that a new part begins.

If your group was excluded in a part because the total earnings in a period were below 80 Guilders, you will still be assigned to a new randomly chosen 4-member-group in the next part. The **additional rule** thus only applies **within a part**.

Your total earnings after part IV amount to the sum of the earnings from part I to IV. The earnings in each individual part as well as your total earnings are shown to you after completion of part IV.

Afterwards you have to fill in a **short questionnaire**. The experimenter announces the beginning of the questionnaire.

Control questions

Before starting the experiment we ask you to complete the sample exercises on the screen. If you need to do any calculations, you can click on the calculator symbol in the lower part of the screen, which will open the Windows calculator. If you have any questions please raise your hand, one of the experimenters will then come to your place and answer your questions privately.

Part V and second questionnaire

(Instructions for part V and the second questionnaire are only distributed after the accomplishment of parts I to IV and the first questionnaire)

In part V all participants will be randomly assigned to one partner. Nobody is informed about the identity of her partner. In part V you have to answer **24 questions**: you can choose one of two options A or B. Every option results in a positive or negative payoff for you and your partner. Your partner answers exactly the same questions. Your payoff in part V depends on your decision and the decision of your partner.

An example:

	Option A	Option B
Your Payoff	10.00	7.00
Partner's Payoff	-5.00	4.00

- If you choose option A you receive 10 Guilders and your partner's payoff is reduced by 5 Guilders. If your partner also chooses option A, then your payoff is reduced by 5 Guilders and your partner's payoff is increased by 10 Guilders such that you both receive 5 Guilders (10 from your own choice and -5 from your partner's choice).
- If you choose option B and your partner chooses option A, then you earn a total of 2 Guilders (7 from your own choice and -5 from your partner's choice). Your partner then earns 14 Guilders (10 Guilders + 4 Guilders).

If you take your decisions 24 values for "your payoff" from option A or option B are summed up. If your partner takes her decisions 24 values for "partner's payoff" from option A or option B are summed up. The sum of these two sums determines your earnings in Guilders, which are converted into Euro according to the exchange rate:

1 Guilder = 0.10 Euro (10 Guilders = 1 Euro)

You are not given any information about the individual decision of your partner. You are only informed about the sum of values from your partner's decisions and your resulting payoff in Guilders.

After part V we ask you to fill in another **short questionnaire**. The experimenter informs you once the questionnaire starts.

Total payoff

As soon as all participants have filled in the second questionnaire the experiment is finished and you are paid your earnings in cash individually and privately. These earnings result from:

  Your earnings from the identical parts I to IV
\+ Your earnings from part V
\+ The rest of the 4 Euros for arriving punctually

page 10 of 10

Appendix A.2.2: Control questions for treatment *PCP_E* (translated from German)

Please answer the following exercises.
These exercises serve to make you familiar with the calculation of the incomes in the experiment. They are not part of the actual experiment!

If you have questions concerning the exercises or do not come the right solution, please raise your hand. The experimenter will help you with answering the questions.

Exercise 1
You are a member of a 4-person group, in which each group member owns an endowment of 20 points.
Imagine no group member (you included) makes a contribution to the project in the first stage.
a) What will be your income in the first stage?
b) What will be the income of each of the other group members in the first stage?

Exercise 2
You are a member of a 4-person-group, in which each group member owns an endowment of 20 points.
Imagine you invest 20 points to the project. Each of the other group members also invests 20 points in the project.
a) What will be your income in the first stage?
b) What will be the income of each of the other group members in the first stage?

Exercise 3
You are a member of a 4-person-group, in which each group member owns an endowment of 20 points.
The other group members contribute together 30 points to the project. What will be your income at the end of the first stage, if
a) you contribute – additionally to the 30 points – 0 points to the project?
b) you contribute – additionally to the 30 points – 10 points to the project?
c) you contribute – additionally to the 30 points – 20 points to the project?

Exercise 4
Consider a 4-person-group, in which each group member owns an endowment of 20 points.
Imagine you invest 8 points to the project. What will be your income at the end of the first stage, if
a) the other three group members contribute together – additionally to your 8 points - further 12 points to the project?
b) the other three group members contribute together – additionally to your 8 points – further 22 points to the project?
c) the other three group members contribute together – additionally to your 8 points – further 32 points to the project?

Exercise 5
Imagine you distribute in the second stage the following points to the other group members: 9, 5, 0.
What are your total costs of distributing points?

Exercise 6
Imagine you distribute in the second stage zero points to the other three group members.
What are your total costs of distributing points?

Exercise 7
Imagine you have earned 25 Guilders in the first stage. Furthermore imagine you distribute zero points to the other group members in the second stage. What will be your income at the end of the second stage,

a) if you receive from the other three group members in the second stage 0 points?
b) if you receive from the other three group members in the second stage 4 points?
c) if you receive from the other three group members in the second stage 15 points?

Exercise 8
Imagine you distribute in the third stage the following counter-points to the other group members: 2, 0, 1.
What are your total costs of distributing counter-points in the third stage?

Exercise 9
How many Guilders must your group earn altogether, so that your group can participate in the following periods?

Appendix A.2.3: Post-experimental questionnaire in treatments *P_E* and *PCP_E* (supplements of *PCP_E* are underlined)

Numbers in square brackets indicate over both treatments P_E and PCP_E the average agreement with the statements on a 5-point Likert scale from "-2 (completely disagree)" to "2 (completely agree)" respectively.

1) Which strategic reflections determined your decisions in this experiment? Which factors do you think were important?

Strategy	Reason for strategy	Number of nominations
High contributions	…to maximize the group income	43
	…in order not to be punished in the second stage	4
	…to induce others to make also high contributions in the next periods	5
Low contributions	…because one expects not to be punished	1
	…to maximize one's own profit	18
	…because other group members also made low contributions in previous period	6
Heavy punishment	…because of small contributions of others (although group might fall below the threshold)	10
	…to induce others to make higher contributions in the next periods	5
Weak punishment	…to reach the threshold	33
	…because of costs of distributing points	9

2) The threshold of 80 Guilders group income induced me to make a **higher contribution** to the project than I would have done without the threshold. [0.4]
3) The threshold of 80 Guilders group income induced me to distribute **less points and counter-points** in the second and third stage to other group members than I would have done without the threshold. [0.5]
4) The threshold of 80 Guilders had **no effects** on my decisions. [-1.2]
5) I distributed **many points** in the second stage to other group members who contributed little to the project hoping that our group **falls below** the 80 Guilders group income and that these group members suffer a preferably high loss. [1.7]
6) I distributed **many counter-points** in the third stage to other group members hoping that our group **falls below** the 80 Guilders group income and that group members who distributed points to me in the second stage, suffer a preferably high loss. [1.9]
7) I distributed **many points** in the second stage to other group members who contributed little to the project in order to make sure that the 80 Guilders will be reached by my group also in later periods within the part. [1.2]
8) I made **high contributions** to the project so that in the second stage no points are distributed and our group does not fall below the 80 Guilders. [0.7]
9) I made **low contributions** to the project, as I assumed that the other group members would not distribute any points in the second stage (so that our group does not fall below the 80 Guilders). [-0.8]

Chapter 3

Norm Enforcement in the Lab and in the Field.
A Within-Subjects Comparison

3.1 Introduction

The enforcement of social norms is crucial for their existence in the long run. Social norms are customary behavioral rules that govern human interactions (Bicchieri, 2006; Young, 2008). They often enhance efficiency by reducing externalities or transaction costs. The enforcement of social norms in real life frequently happens by peer punishment, although it may bear some cost for the punishing party:[34] For instance, the exclusion of team members by their colleagues for little work morale, the complaints of pedestrians towards a person littering in a public place or the insults of queuers in a waiting line to a queue-jumper. Numerous laboratory experiments in behavioral economics (starting with Yamagashi, 1986; Ostrom et al., 1992; Fehr and Gächter, 2000) show that norm enforcement by peer punishment is one of the most successful mechanisms to enhance cooperation in social dilemmas.

Laboratory experiments have for a long time been the fundament of behavioral economics and they undoubtedly have several advantages in comparison to standard empirical methods: They allow a high degree of control, including a controlled variation of certain independent variables, a randomized assignment of subjects to treatments and a context-free elicitation of preferences. Hence, laboratory experiments usually possess a high internal validity by allowing causally attributing differences in behavior to differences in treatments. However, the main criticism of laboratory experiments is their external validity, i.e. it is put into question whether the results of laboratory experiments are valid in real life (List, 2006; Levitt and List, 2007a; 2007b). This criticism is especially pronounced in the domain of social preferences, as their elicitation is particularly prone to an experimenter

[34] One could also think of the enforcement of legal norms by formal punishment or of the enforcement of social norms by reward. However, we focus in this study on the enforcement of social norms by costly peer punishment and therefore use the terms norm enforcement and peer punishment, second-party punishment or informal punishment as synonyms.

demand effect. In the light of this criticism a great number of field experiments emerged that examine on aggregate level whether effects found in the lab persist in the field. In contrast, there are only a few studies that examine on individual level whether the behavior of people in the lab can be extrapolated to the field. For the latter a within-subjects comparison of behavior in the lab and in the field is necessary. Regardless of whether pure money-maximizing or social preferences (for instance Rabin, 1993; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Charness and Rabin, 2002) are assumed – the latter state that individuals do not only care about their own well-being but also about their social environment – all these models predict stable preferences within subjects between the lab and the field.

In our study we examine whether the norm enforcement behavior in the lab translates to the field. Therefore we conduct a within-subjects comparison of norm enforcement between lab and field. We carry out a natural field experiment (according to the classification of Harrison and List, 2004) near the University of Munich, in which we observe queuers when a norm violator cuts in line in front of them. In order to minimize the possibility of an experimenter demand effect we invite the subjects from the field to the lab without revealing that they participated in a field experiment. In the lab we implement three treatments dealing with second-party punishment: The standard treatment to examine norm enforcement in the lab – a three-person prisoners' dilemma game with a one-sided punishment option (see Fehr and Gächter, 2000; 2002; Falk et al., 2005) – and a treatment additionally allowing for counter-punishment, meaning that this treatment includes a two-sided punishment option (Nikiforakis, 2008). Moreover we apply a newly created game with which we try to represent the situation in the queue in an abstract way. This so-called "queuing game" includes a one-sided punishment option. In a forth treatment we elicit the cooperation types of subjects.

Our study makes several contributions: Firstly, we carry out a natural field experiment on norm enforcement, which is scarce so far[35] – although heavily requested, for instance by Guala (2012). Additionally we ex-post collect socio-demographic data of the subjects and examine their link to norm enforcement in the field. Secondly, to our knowledge we are the first ones to make a within-subjects comparison between lab and field with respect to norm enforcement. Thirdly, in the lab subjects play several economic games which allows us to run a "horse race" between different institutions. Fourthly, additionally to the behavioral data from the field and the lab experiment, we elicit of the same subjects the intended norm enforcement behavior stated in a survey. The latter presents three scenarios of real life

[35] Exceptions in economics are Noussair et al. (2011) and Balafoutas and Nikiforakis (2012).

situations, which describe different violations of concrete social norms. In addition to being able to investigate the external validity of intended norm enforcement stated in the questionnaire, we examine whether a so-called norm enforcing type exists, i.e. someone who generally behaves in a norm enforcing way across several social norms.[36]

Our main finding says that norm enforcement in the field correlates with decisions in the lab treatment including a counter-punishment option, but not with decisions in the treatments in which only a one-sided punishment option exists. Specifically we find that the likelihood of norm enforcement in the field increases with the willingness to cooperate and with the sanctioning of defectors in a prisoners' dilemma game with a two-sided punishment option. Our lab-field comparison is a first examination of the external validity of norm enforcement in the lab. Therefore generalizations regarding the interpretation and implication of our main result have to be made with care. However it suggests the following: Firstly, with respect to norm enforcement our main result confirms prior survey evidence that weighing up the danger of being counter-punished is crucial for norm enforcement decisions in the field. Secondly, concerning the methodological aspect, our main result reveals that more evidence from the field is necessary to establish the external validity of the standard treatment with a one-sided punishment option. Overall our main finding is encouraging with respect to the generalizability of lab behavior to the field. However, it also illustrates that the relevant institutional factors of the field need to be incorporated in the lab to allow for external validity.

Our further findings are the following: We find a norm enforcement rate of 32.1% in the field. The likelihood of norm enforcement in the field increases with age and with the cooperation type elicited in the lab. We find a high external validity of the norm enforcement intentions stated in the survey, as long as the same social norm is affected in the survey scenario and in the field. We do not find evidence for a norm enforcing type. Instead our questionnaire data suggest that the likelihood of norm enforcement varies with the concrete social norm affected, and specifically with the negative externality of the norm violation.

The remainder of this paper is organized as follows: In section 3.2 we present the relevant literature, before we describe in section 3.3 our experimental design. Section 3.4 outlines our results. We discuss our findings in section 3.5 and provide a conclusion in section 3.6.

[36] See also Fischbacher et al. (2001) defining several types in the domain of cooperation or De Oliviera et al. (2011) finding in the domain of charity a so-called "giving-type", i.e. donors who donate across several unrelated organizations.

3.2 Related Literature

In the following we give an overview on norm enforcement in lab experiments and in field experiments and then review the literature on within-subjects lab-field comparisons in the domain of social preferences.

Starting with Yamagashi (1986), Ostrom et al. (1992) and Fehr and Gächter (2000; 2002) several laboratory experiments show that norm enforcement by peer punishment is one of the most successful mechanisms to enhance cooperation in social dilemmas. A large number of laboratory experiments on peer punishment followed investigating for instance the influence of monetary versus non-monetary punishment (Masclet et al., 2003; Rege and Telle, 2004), the effect of punishment versus reward (Andreoni et al., 2003; Sefton et al., 2007), the influence of the punishment effectiveness (Anderson and Putterman, 2006; Carpenter 2007; Nikiforakis and Normann, 2008; Egas and Riedl, 2008), the impact of endogenous versus exogenous choice of punishment (Gürerk et al., 2006; Ertan et al., 2009; Sutter et al., 2010), the influence of counter-punishment or feuds (Denant-Boemont et al., 2007; Nikiforakis, 2008; Nikiforakis and Engelmann, 2011; Nikiforakis et al., 2012; Engelmann and Nikiforakis, 2014), the influence of uncertainty involved in punishment (Grechenig et al., 2010; Sousa, 2010; Ambrus and Greiner, 2012; Xiao and Kunreuther, 2012) or the driving forces behind punishment (Falk et al., 2005).[37]

In contrast to this huge bulk of laboratory experiments on norm enforcement, field experiments on the enforcement of social norms are scarce.[38] Exceptions in economics are Noussair et al. (2011) and Balafoutas and Nikiforakis (2012). The former carry out a framed field experiment with recreational fishermen in the Netherlands. They implement incentives similar to a linear voluntary contribution mechanism (see Chaudhuri, 2011), but closely framed to a repeated common pool resource problem. Among other treatments they implement one with a one-sided punishment option. As the game is repeated, strategic motives of punishment might play a role. The results show that non-cooperators are punished in 35.1% of all possible cases. In contrast to most laboratory findings a cooperation-enhancing effect of punishment cannot be confirmed. The authors assume that the fishermen’s norm is rather to catch as much fish as possible than to cooperate. Balafoutas and Nikiforakis (2012) conduct a natural field experiment in Athens (Greece). They examine the

[37] For an overview on norm enforcement in lab experiments see, for instance, Chaudhuri (2011).

[38] There are several studies outside the lab on the enforcement of legal norms: For instance, in the domain of TV licence fee evasion (Fellner et al., 2013) and tax evasion (Blumenthal et al., 2001; Slemrod et al., 2001; Kleven et al., 2011) the effects of audit-threats and normative appeals on compliance are investigated in field experiments. Traxler and Winter (2012) provide survey evidence on the negative relation between the willingness of people to sanction law violations and their belief about the frequency.

enforcement of two social norms in the field, the universal norm not to litter in public places and the environment-specific norm to walk on the left side of escalators. In contrast to the field experiment by Noussair et al. (2011) strategic motives of norm enforcement are minimized, as the norm violator interacts with the pedestrians only once. Balafoutas and Nikiforakis (2012) find a norm enforcement rate of 4.0% for the universal norm and of 19.3% for the environment-specific norm. Questionnaire data suggest that the rather low rates are based on the fear of counter-punishment. The authors further find that men are more likely to enforce norms than women, while the norm violator's gender and height have no effect on norm enforcement.

Psychologists conducted natural field experiments on norm enforcement in case of queue-jumping, which are closest to our field experiment. Schmitt et al. (1992) and Milgram et al. (1986) observe people in waiting lines in New York City, mostly at the ticket counter of the Grand Central Station. As the queue-jumpers and the queuers only interact once, strategic motives are minimized. Counter-punishment is theoretically possible. Schmitt et al. (1992) find that people in waiting lines are more likely to react to a queue-jumper, if there are other queuers behind her, which manifests a social obligation. In contrast, the closeness to the counter does not have any effect on norm enforcement. Both studies demonstrate that queuers directly behind the queue-jumper have a special obligation to react and are therefore most likely to respond to a queue-jumper. In both studies the authors discuss an "individual cost" versus a "moral outrage" position regarding norm enforcement in the line and confirm both: This means that norm enforcement in the waiting line occurs for individual reasons of saving time but also because of social concerns triggered by the violation of a social norm. When the experimental conditions are comparable to ours, these field experiments yield enforcement rates of 43.3% and 54.0%. In contrast to our study, none of the specified field experiments on norm enforcement conducts a lab-field comparison.[39]

Recently there is a quickly growing number of studies conducting within-subjects comparisons between lab measures and field behavior, especially in the domain of social preferences. The majority of these studies support the generalizability of lab behavior. For instance, Karlan (2005) shows that people who are more trustworthy in the trust game are more likely to repay microfinance loans. Baran et al. (2010) demonstrate that the behavior of MBA students in a trust game is positively related to their donations as alumni to their business school a few years later. In the study by Rustagi et al. (2010) conditional cooperation correlates with successful forest commons management. Fehr and Leibbrandt

[39] There is an economic field experiment on queue-jumping (Oberholzer-Gee, 2006), which examines whether a market for time exists. However, in this study the queue-jumper offers money to cut in.

(2011) demonstrate that contributions to a public goods game and patience in a simple time preference game predict limited common pool resource extraction of Brazilian fishermen. Carpenter and Seki (2011) find that several measures of social preferences of Japanese fishermen are positively linked with their productivity and the adoption of a team institution in the field.[40] In the experiment by Franzen and Pointner (2013) subjects giving in the lab dictator game send misdirected letters containing money more often back than subjects giving nothing in the lab. Englmaier and Gebhardt (2013) link a lab public goods game with three treatments in the field, whereas only one treatment shares the free-riding incentive. In this free-riding treatment they find correlations between the contributions in the lab and the effort in the field, but not in the two placebo treatments. They conclude that extrapolation from the lab to situations which share the game form but not necessarily the frame is possible.

However, there are also some within-subjects comparisons that do not – or at least do not completely – support the generalizability of lab experiments: For instance, in the study by Benz and Meier (2008) donations in the lab correlate with donations of the same people in the field, however donations in the lab are more accentuated. Carpenter and Myers (2010) find that giving in a dictator game correlates with the time spent volunteering as a fire fighter, but not with the amount of firefighting operations. Gurven and Winking (2008) find no correlation between the behavior of Bolivian indigenous inhabitants in a dictator and an ultimatum game and their observed everyday sharing of food.[41] In contrast to all these specified lab-field comparisons we conduct in our study a within-subjects comparison regarding norm enforcement.

3.3 Experimental Design and Procedures

3.3.1 Experimental Design

We first conduct the natural field experiment and then invite the participants to the lab without revealing that they have just participated in a field experiment.

[40] Strictly speaking Rustagi et al. (2010) and Carpenter and Seki (2011) compare the behavior of groups between the lab and the field. Thus, one could speak of a within-groups comparison.

[41] Recently there are many more studies that relate lab to field behavior, also in other domains than social preferences. However, these studies often make comparisons on aggregate level without conducting a within-subjects comparison (e.g. Palacios-Huerta and Volij, 2008; Palacios-Huerta and Volij, 2009; Levitt et al., 2010; Stoop et al., 2012). See Camerer (2011) for an extensive survey on the lab-field generalizability.

3.3.1.1 Natural Field Experiment

We conduct the field experiment in front of a bakery at the subway station of the University of Munich, which is frequented by roughly one thousand clients each day during the semester. Most clients frequent the bakery in the morning. Hence, the field experiment is predominantly carried out between 7.30 am and 10.30 am. The location of the field experiment has multiple advantages: Firstly, the majority of the clients are students, which is favorable, as the majority of laboratory experiments are conducted with university students and as students are most likely to come to the lab. Secondly, the lines in front of the bakery are clearly defined and mostly consist of about five to ten clients, which ensures a quick run-through. Thirdly, after having paid clients leave the bakery and therefore do not realize that the queue-jumping is repeated.

The process of cutting in line is carried out in the following way: The norm violator approaches the line and cuts in so that several persons – usually two to three queuers – stand behind her in the line. The norm violator cuts in line without trying to give the impression that the norm violation happens accidentally. In order to register the individual behavior of the queuers without any uncontrollable influence from other people, the norm violator cuts in line in front of individuals standing alone in the line. As participants interact with the norm violator only once, strategic motives of norm enforcement are reduced.[42] The role of the norm violator is played by the authoress. She jumps the queue whenever at least five persons wait in the line. We exclude queuers obviously older than 60 years, mothers with children, handicapped persons and obviously not German looking people as these characteristics presumably influence the likelihood of norm enforcement in the waiting line and of coming to the lab.

A research assistant (RA) who stands near the line, but in the back of the clients, observes the behavior of the queuers behind the norm violator. The behavior of a queuer is evaluated as norm enforcing if she directly addresses the norm violator and asks her to go to the end of the line. In this case the norm violator apologizes, goes back to the end of the line and leaves it approximately 10 seconds later. If nobody asks the norm violator to queue up at the end of the line, she also leaves the line after approximately 10 seconds. Non-verbal signals of disapproval (e.g. harrumphing loudly, looking to other clients), talking to other

[42] However, we can not completely exclude the possibility that a queuer has already seen the queue-jumper in the days before. Furthermore there is in general a strategic element in norm enforcement decisions in the field, as norm enforcers usually pursue the goal to reverse the norm violation.

clients about the queue-jumper without addressing her directly or comments on the side are not evaluated as norm enforcing.[43]

If one of the queuers enforces the norm, the RA asks this norm enforcer to participate in the laboratory experiment. In case that several queuers enforce the norm, the RA invites the one who first enforced the norm. If no norm enforcement takes place, the RA invites the client directly behind the norm violator, as she has best seen the norm violation and is therefore the main responsible person (according to Milgram et al., 1986; Schmitt et al., 1992) to enforce the norm. We call the queuer who is invited to the lab the observer.[44]

As soon as the observer has left the bakery the RA follows her for a while in order not to make the observer link the invitation to the queue-jumping. When contacting the observer, the RA explains that we search for participants fulfilling certain criteria to participate in a paid laboratory experiment on decision-making approximately one or two weeks later in the laboratory MELESSA.[45] We ensure that the participants speak German properly. At that point in time we do not reveal that the person has just participated in a field experiment. If the observer agrees to participate in the laboratory experiment we collect her name, email-address and a personal code. The latter consists of letters and numbers that only the subject knows. It allows us later in the data analysis to combine the behavioral data from the lab and the field. We tell the subjects that we need the code later to make the lab data anonymous and that the subject can already give us the code right now (what they usually do). At the end we hand out to the participant an adapted version of the MELESSA statement of participation with general information on the laboratory experiment. As soon as the observer has left, the RA records her behavior in the waiting line, the estimated age, gender and, in case of norm enforcement, whether the norm enforcer stood directly behind the norm violator in the waiting line or further behind. A few days before the laboratory session the subjects from the field receive a short reminder and further subjects from the MELESSA subject pool – i.e.

[43] One might argue that being in a hurry might affect the norm enforcement behavior in the field. Although we cannot completely exclude this point, we think that it is rather unlikely to play a major role in the field. Firstly, the field experiment is carried out at a point in time, namely at the beginning of the semester, when our main subjects, i.e. students, have a rather relaxed time. Secondly, as clients take on average only about 20 seconds to order and pay, the time gain from sending a queue-jumper to the end of the line is restricted. Thirdly, the clients mostly order rather "non-essential" food like coffee or cookies. If they were really in a hurry, they would presumably not even line up.

[44] In about 15% of all cases the norm violator does not manage to completely get into the line and the queuer next to her pushes herself in front of the norm violator without addressing her. We do not evaluate this as norm enforcing, as the queuer only solves the problem for herself (in the sense of the individual costs perspective) but not for the other queuers following her (in the sense of the social perspective according to Milgram et al., 1986, and Schmitt et al., 1992). In this case the norm violator stays in the queue directly behind this queuer and we observe the behavior of the queuers behind the norm violator.

[45] MELESSA is the Munich Experimental Laboratory for Economic and Social Sciences.

who did not participate in the field experiment – are invited to the laboratory sessions to fill them up.

3.3.1.2 Laboratory Experiment

So that the observers in the lab do not link the queue-jumping in the field to the laboratory experiment, the latter is conducted by another researcher than the authoress. Participants in the lab know that the sessions consist of four parts and several questionnaires. They are informed that the parts are independent of each other and that instructions for one part only follow if the preceding part is finished. We implement four treatment conditions, which consist of one period each. Three treatments deal with norm enforcement: In order to examine whether the standard treatment matters in the field, we implement the three-person prisoners' dilemma game with a one-sided punishment option (see Fehr and Gächter, 2000; 2002; Falk et al., 2005). To investigate the importance of potential counter-punishment in the field, we additionally implement a three-person prisoners' dilemma game with a punishment and counter-punishment option (see Nikiforakis, 2008). Furthermore, we apply a newly created game with which we try to represent the situation in the queue in an abstract way. This so-called queuing game includes a one-sided punishment option. In a forth treatment we elicit the cooperation types of subjects similar to Fischbacher et al. (2001).

At the beginning of each treatment subjects learn that they are randomly and anonymously matched to units of three members, whereby we implement a perfect stranger matching between treatments.[46] The incomes during the experiment are calculated in Guilders. At the beginning of the experiment subjects are informed that the total earnings are calculated as the sum of incomes in each of the four parts, the questionnaires and the show-up fee of 4€. The latter also serves to pay for any negative payoffs generated during the experiment. In the following we describe the treatments in detail.

Three-person prisoners' dilemma game with a one-sided punishment option (PDG+P)

This treatment is based on Falk et al. (2005). Subjects know that the PDG+P consists of two stages: The first stage includes a prisoners' dilemma game. In the second stage participants receive information on the individual first stage decisions of their unit members and can punish them, i.e. reduce their income, which is costly also for the sanctioning person.

In the fist stage subjects play a three-person prisoners' dilemma game, i.e. they simultaneously decide whether to cooperate or defect. In table 3.1 we present the payoff

[46] As in two sessions only 9 subjects participated, we could only implement a stranger matching instead of a perfect stranger matching in these sessions. No subject addressed this point.

matrix of the first stage which implies a social dilemma: If we assume pure money-maximizing preferences the dominant strategy of player *i* is to defect, independent of how the other unit members decide. However, the socially efficient outcome is achieved by cooperation of all players. We implement the prisoners' dilemma game instead of a continuous public goods game (Marwell and Ames, 1981; Isaac et al., 1985; Isaac and Walker, 1988) for two reasons: The binary version better corresponds to the situation in the field, where only two types of actions exist, namely to adhere to the social norm or to jump the queue. Secondly, the prisoners' dilemma game allows us to implement the strategy vector method and therefore to avoid learning effects in the lab.

Table 3.1 - Payoffs in the first stage of the PDG+P (Falk et al., 2005).

	Both other unit members defect	One of the other unit members defects, the other one cooperates	Both other unit members cooperate
Player *i* defects	20	32	44
Player *i* cooperates	12	24	36

In the second stage, the punishment stage, subjects are informed about each of the other unit members' first stage decision. They are told that they can simultaneously sanction each of the other unit members by assigning punishment points. Subjects cannot punish members of other units. Players can assign at most 10 punishment points to each of their unit members. If player i assigns punishment points p_{ij} to unit member j, $i \neq j$, player i's first stage income, $\pi_{1,i}$, is reduced by p_{ij}. Player j's first stage income is then reduced by 3 * p_{ij} with a lower boundary at zero if player j's cost of receiving punishment points is higher than her first stage income.[47] The latter avoids that subjects receive negative payoffs without being able to control them. Hence, player i's income at the end of the second stage, $\pi_{2,i}$, is given by

$$\pi_{2,i} = \max\left\{0, (\pi_{1,i} - 3 * \sum_{j \neq i} p_{ji})\right\} - \sum_{j \neq i} p_{ij} \tag{3.1}$$

Participants are informed about the payoff structure in Eq. (3.1). We elicit the punishment decisions with the strategy vector method (Selten, 1967). Thus we ask subjects how many punishment points they want to distribute to each other unit member for each out of four possible combinations of first stage decisions made by the other two unit members. Subjects know that they receive feedback at the end of the experiment on the actual first stage

[47] For the sake of simplicity we implement the linear cost function by Fehr and Gächter (2002) and later studies instead of the nonlinear cost function used for instance by Fehr and Gächter (2000).

decisions of their unit members, their own first stage income, the number of punishment points received by the other unit members and their own total earnings in the PDG+P.

Three-person prisoners' dilemma game with a two-sided punishment option (PDG+PCP)

The PDG+PCP builds on Nikiforakis (2008). Subjects know that this part includes the PGP+P and an additional third stage, the counter-punishment stage. In the counter-punishment stage each player i has the opportunity to counter-punish each of her unit members $j \neq i$ by assigning at most 10 counter-points to each of them. Analogous to the second stage, the cost of each counter-point that i assigns to j, z_{ij}, is 1 Guilder. The cost of each counter-point received is 3 Guilders. The second stage income is reduced to zero and not further, if the cost of receiving counter-points is higher than the second stage income, $\pi_{2,i}$. The period income $\pi_{3,i}$ of subject i is given by

$$\pi_{3,i} = \max\left\{0, (\pi_{2,i} - 3 * \sum_{j \neq i} z_{ji})\right\} - \sum_{j \neq i} z_{ij} \tag{3.2}$$

This payoff structure and the following rules are common knowledge: In order to avoid very negative payoffs participants with a non-positive second stage income are not allowed to counter-punish or to be counter-punished. In order to avoid that punishment in the second stage is delayed to the third stage, subject i can only counter-punish those unit members j who punished her in the second stage. We elicit the punishment decisions in the second stage with the strategy vector method. The counter-punishment decisions in the third stage are elicited with the direct response method, as otherwise we would have to differentiate between too many possible cases. Therefore subjects receive feedback in the counter-punishment stage on the individual first stage decisions of their unit members, the individual first stage incomes of all members of their unit and the number of punishment points received by and assigned to each of the other unit members. Subjects further know that they receive feedback on the number of counter-points received by the unit members and their own final profit in the PDG+PCP at the end of the experiment.

Queuing game with one-sided punishment option (QG+P)

The QG+P is a newly created game in which we try to reflect the field conditions in the queue in an abstract way: It consists of three stages: In the first stage in each three-person unit a ranking is established based on the number of correctly solved mathematical tasks. In the second stage the person on the last position of the ranking can decide to "jump the queue", i.e. to establish an alternative order that puts herself in the first and most prosperous position. In the third stage the person who originally reaches the first position in the first stage can sanction the “queue-jumper”, whereby sanctioning is costly also for the sanctioning person. Thus, in contrast to the PDG+P we create a ranking of subjects and allow only one player to violate the norm. Based on the results by Milgram et al. (1986) we further reverse the second-order public goods problem in the punishment stage of the PDG+P by allowing only the person “directly behind the queue-jumper” to sanction her.

In order to create a ranking in the QG+P similar to the one in the queue, subjects in the first stage have five minutes of time to calculate sums of five different two-digit numbers (see also e.g. Niederle and Vesterlund, 2007). We choose this addition task, as it is easy to explain, it does not yield any gender effects and it is influenced by effort and skill and therefore is very suitable to create a sense of entitlement among subjects. Subjects know that the number of correct answers determines their position in a ranking A, B, C, whereby the first stage income is ranked according to this order: The person on position A (whom we call unit member A) earns the highest first stage income, namely 40 Guiders. The person on position B (unit member B) receives the second highest first stage income, namely 30 Guilders, and the person on position C (unit member C) earns the lowest income of 20 Guilders. Subjects are told in the first stage that it is advantageous for them to correctly calculate as many sums as possible. They only receive the instructions for the second and third stage when the first stage is finished.

In the second stage participants are informed that unit member C decides whether to retain the first stage order or to choose an alternative order. The latter assigns unit member C position A with the highest income of 40 Guilders. The other two unit members then slip to the positions B and C and therefore receive the lower incomes of 30 and 20 Guilders respectively. If unit member C retains the first stage order, there is no third stage and the unit members earn their first stage income.

Participants know that, in case unit member C chooses the alternative order, unit member A can sanction the “queue-jumper” in the third stage by assigning punishment points to her. Unit member A can only sanction unit member C if the latter actually “cuts in line”, i.e. if unit member C chooses the alternative order in the second stage. Participants further know that unit member A can assign at most 10 punishment points to unit member C. If unit member A assigns punishment points p to unit member C, her second stage income of 30 Guilders is reduced by p, while unit member C’s second stage income of 40 Guilders is reduced by 3 * p. Subjects are informed about the payoff structure, π_i , that arises for the three unit members i in case unit member C jumps the queue:

$$\pi_A = 30 - p \tag{3.3}$$

$$\pi_B = 20 \tag{3.4}$$

$$\pi_C = 40 - 3 * p \tag{3.5}$$

We elicit the second and third stage decisions with the strategy vector method. Subjects know that they receive feedback at the end of the experiment on their own first stage position in the ranking, on unit member C's decision in the second stage, if applicable, on unit member A's decision in the third stage, and on their own total income in the QG+P.

Elicitation of cooperation types

We elicit the individual cooperation types applying a simplified version of the Fischbacher et al. (2001) method.[48] Subjects know that they play the three-person prisoners' dilemma game with the payoff structure of table 3.1. Using the strategy vector method we ask participants to decide between cooperation and defection for each possible combination of choices of the other two unit members. In order to incentivize these conditional decisions, we ask participants to make an unconditional choice before. Subjects know that at the end of this part one unit member is randomly drawn for whom the conditional choice, based on the unconditional choices of the other two unit members, is payoff-relevant. For each of the other two unit members the unconditional choice is payoff-relevant. Subjects are told that they receive feedback on the payoff-relevance of their choices, on the payoff-relevant choices of the other two unit members and on their own total income in this part only at the end of the experiment.

We start half of the laboratory sessions with treatment QG+P followed by the PDG+P, and the other half with the PDG+P followed by the QG+P. The elicitation of cooperation types always constitutes the third part. The PDG+PCP is always carried out as the last part in order to avoid learning effects from the counter-punishment stage. At the end of each laboratory session we conduct five questionnaires. For filling them in we pay the subjects 2€ altogether. The first questionnaire includes three scenarios of real life situations each describing the violation of a concrete social norm. This questionnaire conduces to examine whether subjects show a similar norm enforcement behavior across different social norms, i.e. whether a norm enforcing type exists. Appendix A.3.3 outlines the questionnaire. For each scenario subjects rate their emotions anger and fear towards the norm violator, indicate whether they would confront the norm violator and the reasons for their decision. The latter

[48] Fischbacher et al. (2001) elicit the cooperative preferences applying a continuous public goods game.

are presented in random order. The order of the scenarios is changed between sessions. In order to be able to compare the elicited emotions with a baseline, subjects rate their emotions anger and fear at the beginning of the questionnaire. One of the scenarios describes the field situation in the waiting line, when another person cuts in. This scenario conduces to measure the external validity of the scenario method. The participants fill in further questionnaires, that are not reported here, and a questionnaire on socio-demographic variables. Finally subjects indicate their personal codes.

3.3.2 Procedures

The experiment was approved by the Ethics Committee of the Faculty of Economics at the University of Munich. It was run in April and May 2014. The field experiment was carried out over 5 weeks. Appendix A.3.4 shows some photos of the field experiment. We conducted 7 laboratory sessions at the experimental laboratory MELESSA of the University of Munich. The sessions lasted up to 120 minutes and were anonymously run. We exclude the last session (with 5 participants from the field and 4 participants from the MELESSA subject pool) from the further data analysis, as the participants from the field revealed at the end that they realized a field experiment was run with them and that they discussed this in the group before the session started. This discussion might have influenced the decisions in the lab.

Table 3.2 outlines the number of subjects in the laboratory and field experiment as we also included them in the further data analysis. Thus, in the column “Participation in lab experiment” we only include the subjects from the first 6 laboratory sessions. In the field experiment 237 subjects participated. 45 of them also participated in the laboratory experiment, which yields a participation rate of 19.0%. Altogether 66 subjects participated in the lab experiment.

Table 3.2 - Number of subjects in the laboratory and field experiment.

	Participation in lab experiment	No participation in lab experiment	Total
Participation in field experiment	45 (12)	192 (64)	237 (76)
No participation in field experiment	21	---	
Total	66		

Note: The numbers in parentheses indicate the number of norm enforcers in the field.

At the beginning of each laboratory session the experimenter read the neutrally written instructions aloud.[49] In order to test the participants' understanding of the incentive structure and the decision situation, subjects were asked to answer control questions.[50] We assume full understanding, as we only started the treatments as soon as all participants answered all questions correctly.

At the end of the laboratory sessions the authoress entered the laboratory and informed the subjects that she had met each of them in the weeks before. She asked the subjects whether they recognized her and, if so, where they met her. In the first 6 sessions at most 10% of the participants from the field experiment remembered that the authoress had cut in line in front of them. Thus, we can assume that the majority did not have a clue that the queue-jumping is linked with the laboratory experiment. Finally, we informed subjects about their participation in the field experiment and asked them for a written consent to the storage and analysis of their behavioral data in the field. Subjects thereby knew that their field and laboratory behavior is anonymously linked via their personal code. All of the subjects who participated in both, the lab and the field, gave us their written consent to the storage and analysis of their behavioral data from the field.

The laboratory experiment was programmed and conducted with the experiment software z-Tree (Fischbacher, 2007). The subjects from the MELESSA subject pool were invited via ORSEE (Greiner, 2004). We did not impose any specific restrictions on them. They are similar to the subjects from the field experiment with respect to average age (subjects from the field: 24 years; subjects from the MELESSA pool: 25 years) and gender (subjects from the field: 66.7% female; subjects from the MELESSA pool: 52.4% female). Furthermore, similar to the subjects from the field experiment, the majority of them are undergraduate students (subjects from the field: 91%; subjects from the MELESSA pool: 86%). At the end of the experiment we exchanged the payoffs at an exchange rate of 1Guilder = 0.2€, which yielded an average payoff of 25.30€.

3.4 Results

The results part is divided into six sections. First we report the norm enforcement behavior in the field experiment (section 3.4.1). Section 3.4.2 gives a general overview of the lab decisions. In section 3.4.3 we check whether the selection from the field to the lab is non-

[49] Appendix A.3.1 outlines the instructions.
[50] Control questions for the QG+P are provided in Appendix A.3.2.

random and whether the participation in the field experiment affects the decisions in the lab. Section 3.4.4 provides the lab-field comparison. We examine whether further variables influence the norm enforcement behavior in the field (section 3.4.5) and present the results on the questionnaire scenarios in section 3.4.6. In the following we refer to the norm enforcers and non-norm enforcers in the field experiment, if we talk about *norm enforcers* and *non-norm enforcers*. The notation *MWU-test* refers to the two-sided Mann-Whitney U-test. We call the subjects participating both in the lab and the field the lab *and field participants*, the subjects only participating in the lab the *lab participants* and the subjects only participating in the field the *field participants*.

3.4.1 Norm Enforcement in the Field

The last column of table 3.2 indicates that in 32.1% of all cases in the field norm enforcement occurred. Typical comments directed to the norm violator were “Excuse me, I am also waiting in the line” or “Excuse me, you have to queue up at the end of the line”. Except for tapping the norm violator on the shoulder, no physical action occurred.

RESULT 3.1: *Norm enforcement in the field is observed in 32.1% of all cases.*

In 83% of all cases of norm enforcement the queuer directly behind the norm violator enforced the norm, which confirms the results by Milgram (1986) and Schmitt et al. (1992) that queuers directly behind the norm violator have a special obligation to enforce the norm compared to queuers following.

3.4.2 Overview of Laboratory Decisions

First we explain the measure we use in the following for the sanctioning decisions in the PDG+P and the PDG+PCP: The sanctioning behavior in the second stage of the PDG+P and PDG+PCP is elicited with the strategy vector method respectively. Thus, subjects decide for each of four possible combinations of first stage decisions made by the other two unit members how much punishment points to distribute. In the following we report the results on the sanctioning level in the PDG+P and the PDG+PCP each as the average sanctioning level in the two cases in which only one other group member defects and the other one cooperates, as in the field there is also only one norm violator and the other queuers adhere to the norm. However, the main results in the following also hold, if we additionally include the case in which either both other subjects defect (for the variable “sanctioning of defectors”) or both other subjects cooperate (for the variable “sanctioning of cooperators”). The same applies for the punishment likelihood. We indicate the sanctioning level in terms of sanctioning points (and not in terms of sanctioning cost).

In this section we give an overview of the laboratory decisions made by all 66 subjects participating in the lab. Column 1 of table 3.3 shows a cooperation rate of 42.4% in the PDG+P. This rate is lower than the cooperation rate of 61% reported in the corresponding treatment by Falk et al. (2005). We come to this point again in the next section. In line with the findings by Falk et al. (2005) cooperators in the PDG+P are punished by on average 0.2 points and with a punishment rate of 7.6%. For the punishment of defectors we find a slightly lower punishment level of 1.1 punishment points and a lower punishment rate of 36.4% compared to the punishment level of 1.7 points and a punishment rate of 58% found by Falk et al. (2005). We also come to this point again in the next section. In line with standard results we find that defectors are more heavily punished than cooperators (Wilcoxon signed-ranks test, $p = 0.000$) and that they are more likely to be punished (McNemar test, $p = 0.000$).

In the PDG+PCP 36.4% of the subjects cooperate. In line with the results by Nikiforakis (2008) the cooperation rate seems to be lower in the PDG+PCP compared to the PDG+P. However, the difference is not significant (McNemar test, $p = 0.317$). In line with standard results defectors are again more heavily punished than cooperators in the second stage (0.9 versus 0.2 punishment points, Wilcoxon signed-ranks test, $p = 0.000$) and they are more likely to be punished (30.3% versus 7.6%, McNemar test, $p = 0.000$). In contrast to Nikiforakis (2008) we find that the punishment severity (0.6 versus 0.5 punishment points, Wilcoxon signed-ranks test, $p = 0.398$) and the willingness to punish in the second stage (36.4% versus 30.3%, McNemar test, $p = 0.157$) are similar between the PDG+P and the PDG+PCP. Columns 2 and 3 of table 3.3 further outline the lab decisions of the lab and field subjects versus the decisions of the subjects only participating in the lab. Among the lab and field subjects we find that the punishment severity (0.8 versus 0.6 punishment points, Wilcoxon signed-ranks test, $p = 0.053$) and the willingness to punish (44.4% versus 31.1%, McNemar, $p = 0.014$) are higher in the PDG+P than in the PDG+PCP respectively, which is in line with standard results (Nikiforakis, 2008). In contrast, among the subjects only participating in the lab the punishment severity is even higher in the PDG+PCP than in the PDG+P (0.3 versus 0.2 punishment points, Wilcoxon signed-ranks test, $p = 0.046$), which is in contrast to standard results. We explain this abnormality in the next section. Punishment is avenged in 8 out of 20 (40.0%) possible cases in the third stage – which is still in line with the counter-punishment rate of 25.7% reported by Nikiforakis (2008) – with an average counter-punishment severity of 1.0 counter-points.

Table 3.3 - Descriptive statistics: Decisions made in the lab.

	All subjects in lab	Lab and field participants	Lab participants	Test statistic
PDG+P:				
Cooperation decision	0.424 (0.498)	0.489 (0.506)	0.286 (0.463)	χ2-test, p = 0.120
Sanctioning of cooperators	0.152 (0.701)	0.211 (0.843)	0.024 (0.109)	MWU-test, p = 0.519
Sanctioning of defectors	1.090 (1.780)	1.456 (2.011)	0.310 (0.680)	MWU-test, p = 0.020
PDG+PCP:				
Cooperation decision	0.364 (0.485)	0.422 (0.499)	0.238 (0.436)	χ2-test, p = 0.148
Sanctioning of cooperators	0.152 (0.701)	0.189 (0.835)	0.071 (0.239)	MWU-test, p = 0.741
Sanctioning of defectors	0.856 (1.751)	1.033 (2.021)	0.476 (0.873)	MWU-test, p = 0.582
QG+P:				
Queue-jumping	0.545 (0.502)	0.533 (0.571)	0.571 (0.507)	χ2-test, p = 0.772
Sanctioning	2.136 (3.043)	2.444 (3.223)	1.476 (2.561)	MWU-test, p = 0.251
Cooperation types:				
Free-riders	0.762	0.714	0.857	Fisher exact
Conditional cooperators	0.222	0.262	0.143	test,
Altruists	0.016	0.024	0.000	p = 0.564
Scenarios:				
Confr. in Group task scenario	5.636 (0.871)	5.578 (0.941)	5.762 (0.700)	MWU-test, p = 0.349
Confr. in Littering scenario	2.758 (1.815)	2.578 (1.802)	3.143 (1.824)	MWU-test, p = 0.258
Confr. in Queuing scenario	4.394 (1.709)	4.178 (1.862)	4.857 (1.236)	MWU-test, p = 0.231

Notes: This table outlines the laboratory decisions of all 66 subjects participating in the lab. It compares the decisions of the 45 subjects participating both in the lab and in the field with the decisions of the 21 subjects participating only in the lab experiment. The sanctioning variables in the PDG+PCP refer to the sanctioning level in the second stage. The test statistics refer to the comparison between lab and field participants versus lab participants respectively. Standard deviations are presented in parentheses.

With respect to the newly created QG+P we find that 54.5% of the subjects “cut in line”, which is comparable to the defection rate of 57.6% in the PDG+P (McNemar test, p = 0.724). 42.4% of subjects punish the queue-jumper, which is comparable to the willingness to punish defectors (36.4%) in the PDG+P (McNemar test, p = 0.206). The punishment severity is with 2.1 points higher than the punishment of defectors in the PDG+P. However this is plausible, as in the QG+P there is only one unit member that can be punished.

Depending on their decision between defection and cooperation for each combination of decisions of the other two group members we classify the lab and field subjects as free riders (who always defect), conditional cooperators (who cooperate if others cooperate) and altruists (who always cooperate). This is a simplification of the classification by Fischbacher et al. (2001), which allows for ordering the cooperation types. Table 3.3 outlines that 76.2% of the subjects in the lab are free-riders, 22.2% are conditional cooperators and 1.6% are altruists. There are three subjects left that we cannot categorize. Note that in the study by Fischbacher et al. (2001) 50% of the subjects are classified as conditional cooperators and only 30% are classified as free-riders. However, this discrepancy is plausible, as Fischbacher et al. (2001) apply a continuous public goods game, in which the free-riding option is just one out of 21 possible choices, whereas in our binary version of the public goods game, free-riding is just one of two possible options. Hence, the likelihood of free-riding is much higher in our study.

Table A.3.1 in appendix A.3.5 displays the correlations between the different decisions in the lab. It shows that there are significant and positive inter-correlations among the cooperation types and the cooperation decisions in the PDG+P and PDG+PCP as well as among the different sanctioning decisions in the QG+P, the PDG+P and the PDG+PCP, which seems plausible. Furthermore the “queue-jumping” decision in the QG+P negatively correlates with the cooperation type.

3.4.3 Selection into the Lab and Influence of the Participation in the Field

We examine whether the selection into the lab is non-random. Table 3.4 outlines the characteristics of those subjects participating both in the lab and the field and of those subjects only participating in the field. Regarding the norm enforcement rate in the field table 3.4 indicates a slightly lower rate of 26.7% among the lab and field subjects than among the subjects only participating in the field (33.2%). However, the difference is far from significant. Hence, we do not find any evidence for a selection bias with respect to norm enforcement in the field. Table 3.4 further indicates that the same applies regarding gender. In the course of the field experiment we started to estimate the age of 123 observers of whom 23 also participated in the lab, i.e. of whom we know the actual age. The estimated age is significantly correlated with the actual age by 81.8% (Spearman rank correlation coefficient, p = 0.000), although it is systematically higher than the actual age (26 years versus 24 years, Wilcoxon signed-ranks tests, p = 0.087). Table 3.4 outlines that we neither find any significant differences in the estimated age between the lab and field subjects and the subjects only participating in the field. Hence, overall we assume a random selection from the field to the lab.

Table 3.4 - Descriptive statistics: Lab and field participants versus field participants.

	Lab and field participants	Field participants	Test statistic
Norm enforcement in the field	0.267	0.332	χ^2-test, p = 0.402 (45 vs. 187 obs.)
Female	0.667	0.674	χ^2-test, p = 0.927 (45 vs. 187 obs.)
Estimated age	26	27	MWU-test, p = 0.927 (23 vs. 100 obs.)

Note: This table outlines the norm enforcement rate, the gender and the average estimated age of subjects participating both in the lab and the field and of subjects only participating in the field.

Next we examine whether participation in the field experiment has an effect on the decisions later in the lab. Therefore the second and third columns of table 3.3 present the lab decisions of subjects participating both in the lab and the field versus the decisions of subjects only participating in the lab. Regarding most lab decisions we do not find any evidence of an effect of the participation in the field experiment. However, table 3.3 outlines that lab and field participants show higher levels of sanctioning towards defectors in the PDG+P than lab participants. Furthermore the difference in the cooperation likelihood in the PDG+P is also close to significant at the 10% level. In section 3.4.2 we state that the cooperation likelihood and the sanctioning of defectors in the PDG+P over all subjects in the lab is rather low compared to the results by Falk et al. (2005). Table 3.3 indicates that this divergence is rather driven by the subjects from the MELESSA pool than by the lab and field subjects. This suggests that the difference between lab and field subjects versus lab subjects in the PDG+P might not be attributed to an influence of the participation in the field experiment. Instead there might be an influence that reduced the cooperation rate and the sanctioning of defectors among the lab subjects. When we invite the subjects from the MELESSA pool to the lab experiment we add one short paragraph at the beginning of the standard invitation email stating that another group of subjects is already invited – namely the lab and field subjects – and that this group is allowed to participate in the lab experiment at any rate, also if a person of this group appears later to the lab experiment than the subject from the MELESSA pool. Although we tried to phrase the paragraph as respectful as possible, this paragraph and the corresponding handling of subjects directly before the session might have created a feeling of being a second-order subject among the participants from the MELESSA pool. There are good reasons to assume that this might have played out at the beginning of the session, thus in the PDG+P, in the form of less cooperation and therefore also less sanctioning of defectors. This argumentation also explains the abnormal result of a lower sanctioning level in the PDG+P than in the second stage of the PDG+PCP among the MELESSA subjects. At the same time this might not have played out in the "queue-jumping" and the sanctioning decision in the QG+P because there is the mathematical task in between. Hence, we assume that the participation in the field experiment has not affected the decisions in the lab.

3.4.4 Lab-Field Comparison

In the following we compare the behavioral measures in the lab with the norm enforcement behavior in the field of the same 45 lab and field subjects.

PDG+P

Table 3.5 reveals the lab decisions of non-norm enforcers and norm enforcers in the field. We find no significant difference in the cooperation rate: 48.5% of non-norm enforcers and 50.0% of the norm enforcers cooperate in the PDG+P. Table 3.5 outlines that we neither find any difference between non-norm enforcers and norm enforcers with respect to the sanctioning of cooperators nor the sanctioning of defectors: Cooperators are punished by non-norm enforcers by 0.3 points and by norm enforcers by 0.0 points. Defectors are sanctioned by non-norm enforcers by 1.3 points and by norm enforcers by 1.8 points. The same applies for the willingness to sanction defectors (42.4% versus 50.0%, χ^2-test, p = 0.651). Moreover, the difference between the sanctioning of defectors and of cooperators does not differ between non-norm enforcers and norm enforcers. Table 3.5 accordingly outlines that we do not find any significant correlations between the norm enforcement behavior in the field and the lab decisions in the PDG+P. With respect to both decisions, the cooperation decision and the sanctioning level in the PDG+P, we do not find any effect of the order of treatments (Cooperation: χ^2-test, p = 1.000; Sanctioning level: MWU-test, p = 0.434).

RESULT 3.2: *We do not find any significant correlation between the norm enforcement behavior in the field and the lab decisions in the PDG+P.*

Table 3.5 - Descriptive statistics: Non-norm enforcers versus norm enforcers.

	Non-norm enforcers	Norm enforcers	Test statistic
PDG+P:			
Cooperation decision	0.485 (0.507)	0.500 (0.522)	χ^2-test, p = 0.928 r_s = 0.013, p = 0.930
Sanctioning of cooperators	0.288 (0.977)	0.000 (0.000)	MWU-test, p = 0.212 r_s = -0.188, p = 0.216
Sanctioning of defectors	1.348 (1.873)	1.750 (2.417)	MWU-test, p = 0.714 r_s = 0.055, p = 0.718
Diff. in sanctioning of defectors and cooperators	1.061 (1.676)	1.750 (2.417)	MWU-test, p = 0.392 r_s = 0.129, p = 0.398
PDG+PCP:			
Cooperation decision	0.333 (0.479)	0.667 (0.492)	χ^2-test, p = 0.045 r_s = 0.298, p = 0.046
Sanctioning of cooperators	0.227 (0.961)	0.083 (0.289)	MWU-test, p = 0.835 r_s = 0.031, p = 0.838
Sanctioning of defectors	0.545 (1.141)	2.375 (3.142)	MWU-test, p = 0.045 r_s = 0.302, p = 0.044
Diff. in sanctioning of defectors and cooperators	0.318 (1.144)	2.292 (3.108)	MWU-test, p = 0.015 r_s = 0.368, p = 0.013
QG+P:			
Queue-jumping	54.5% (0.506)	50.0% (0.522)	χ^2-test, p = 0.787 r_s = -0.040, p = 0.793
Sanctioning	2.576 (3.373)	2.083 (2.875)	MWU-test, p = 0.867 r_s = -0.025, p = 0.869
Cooperation types:			
Free-riders	0.800	0.500	Fisher exact test, p = 0.074 r_s = 0.315, p = 0.042
Conditional cooperators	0.200	0.417	
Altruists	0.000	0.083	
Scenarios:			
Confront. in Group task scenario	5.484 (1.064)	5.833 (0.389)	MWU-test, p = 0.393 r_s = 0.129, p = 0.400
Confront. in Littering scenario	2.455 (1.787)	2.917 (1.891)	MWU-test, p = 0.441 r_s = 0.116, p = 0.447
Confront. in Queuing scenario	3.667 (1.882)	5.583 (0.793)	MWU-test, p = 0.001 r_s = 0.494, p = 0.001

Notes: One independent observation in the table is one out of the 45 subjects participating both in the lab and in the field. The characteristics of those 12 subjects who enforce the norm in the field and of those 33 subjects who do not enforce the norm are compared. Standard deviations are presented in parentheses. The sanctioning variables in the PDG+PCP refer to the sanctioning level in the second stage. r_s denotes the Spearman rank-order correlation coefficient. Kendall rank-order correlation coefficients yield very similar results.

PDG+PCP

Table 3.5 reveals that, in contrast to the PDG+P, we find significant differences between non-norm enforcers and norm enforcers with respect to the behavior of subjects in the PDG+PCP, i.e. if two-sided punishment is possible. Table 3.5 outlines that non-norm enforcers cooperate significantly less than norm enforcers in the PDG+PCP: Only 33.3% of non-norm enforcers cooperate, whereas among the norm enforcers the cooperation rate is 66.7%. Hence, cooperation among norm enforcers is twice as frequent as among non-norm enforcers in the PDG+PCP. Accordingly the cooperation rate in the PDG+PCP significantly correlates by 29.8% with the norm enforcement decision in the field.

With respect to the sanctioning level towards cooperators in the second stage we do not find any significant difference between non-norm enforcers and norm enforcers. In contrast, defectors are punished by 0.5 points by non-norm enforcers, but by 2.4 points by norm enforcers in the PDG+PCP. The difference is significant. Accordingly the norm enforcement behavior in the field significantly correlates by 30.2% with the sanctioning level towards defectors: Thus, the more a subject sanctions defectors in the PDG+PCP the higher is the likelihood that she enforces the social norm in the field. The result is comparable for the willingness to sanction defectors (24.2% among non-norm enforcers versus 50.0% among norm enforcers, χ^2-test, $p = 0.099$). These results suggest that the sanctioning of defectors in the PDG+PCP has a high external validity with respect to the norm enforcement behavior in our field experiment. Hence, in contrast to the PDG+P, the prisonsers' dilemma game with a two-sided punishment option seems to be a valid workhorse with respect to the norm enforcement behavior in the field. Additionally table 3.5 demonstrates that norm enforcers differentiate in their sanctioning behavior more heavily between cooperators and defectors than non-norm enforcers do.

RESULT 3.3: *The decisions in the PDG+PCP correlate with the norm enforcement behavior in the field: Norm enforcers are more likely to cooperate in the PDG+PCP than non-norm enforcers. At the same time they sanction defectors more heavily than non-norm enforcers do. Furthermore, in contrast to non-norm enforcers, norm enforcers differentiate in their sanctioning behavior more heavily between cooperators and defectors.*

QG+P

Table 3.5 indicates that the "queue-jumping" decision in the lab does not correlate with norm enforcement in the field: Among non-norm enforcers 54.5% "cut in line" in the lab, among norm enforcers 50.0% do so. The sanctioning behavior in the QG+P is neither related to the norm enforcement in the field: Non-norm enforcers sanction queue-jumpers by 2.6 points, whereas norm enforcers punish by 2.1 points. The same holds for the willingness to sanction queue-jumpers (45.5% versus 50.0%, χ^2-test, p = 0.787). Thereby the order, in which we present the treatments does not affect the queue-jumping nor the sanctioning of the queue-jumper (Queue-jumping: χ^2-test, p = 0.138; Sanctioning level: MWU-test, p = 0.260)

RESULT 3.4: *We do not find any significant correlation between the norm enforcement behavior in the field and the decisions in the queuing game.*

Cooperation types

Table 3.5 reveals that the distribution of cooperation types significantly differs between non-norm enforcers and norm enforcers: Among the non-norm enforcers 80.0% of subjects are free-riders. In contrast, among the norm enforcers there are only 50.0% free-riders. Cooperation types significantly correlate with the norm enforcement behavior in the field by 31.5%. This result is in line with the enhanced cooperation likelihood in the PDG+PCP among norm enforcers compared to non-norm enforcers. It further corresponds to standard lab results showing that cooperators punish defectors more heavily than defectors do (e.g. Falk et al., 2005). Our result extends this standard result across the border of the lab.

RESULT 3.5: *The norm enforcement behavior in the field correlates with the cooperation type: Persons who are classified as more cooperative types in the lab are more likely to enforce the social norm in the field.*

In order to investigate the robustness of our findings so far to various specifications, we run several regressions of norm enforcement in the field, which are outlined in table 3.6. Specifications (I) and (VI) demonstrate that the decisions both in the PDG+P and in the QG+P do not correlate with the norm enforcement behavior in the field and therefore confirm results 3.2 and 3.4. In contrast, specifications (II) to (V) reveal that most of the decisions in the PDG+PCP are related to the norm enforcement behavior in the field. According to specification (II) subjects who cooperate in the PDG+PCP are by 26.7% more likely to enforce the norm in the field. Specification (III) shows that with each punishment point assigned to a defector in the PDG+PCP the likelihood of norm enforcement in the field increases by 7.9%, even if we control for the cooperation decision in the first stage. As outlined in specification (IV) the level of sanctioning cooperators in the PDG+PCP is not linked to norm enforcement in the field. Specification (V) confirms that norm enforcers distinguish to a greater extent than non-norm enforcers between defectors and cooperators in the PDG+PCP. Hence, specifications (II) to (V) confirm result 3.3. Specification (VII) confirms result 3.5, which states that more cooperative types are more likely to enforce the social norm in the field. With each category the likelihood of norm enforcement in the field increases by 28.4%. Table A.3.2 in appendix A.3.5 outlines the probit regressions of norm enforcement with the willingness to sanction defectors in the PDG+P, the PDG+PCP and the QG+P as regressors. The results are comparable.

In summary, we can conclude: The likelihood of norm enforcement in the field increases with the cooperation decision and the sanctioning of defectors in the PDG+PCP as well as with the cooperation types in the lab. If we compare the results of the QG+P, the PDG+P and the PDG+PCP in the sense of a “horse race”, we can state that for the norm enforcement decision in the field the possibility of counter-punishment seems to be crucial, i.e. subjects in the field weigh up the possibility of being counter-punished. In contrast, we do not find a link between measures in lab treatments that only allow one-sided punishment, may it be the PDG+P or the QG+P, and norm enforcement behavior in the field.

3.4.5 Control Variables and Norm Enforcement in the Field

We elicit several control variables in the field and several characteristics of the lab and field subjects during the lab experiment. In this section we examine their influence on the norm enforcement behavior in the field. These variables are the observer’s gender and age, her field of study (an eightfold scaled categorical variable), the weather on the day of the field experiment (a fivefold scaled ordered variable taking the value 0 in case of cloudy, rainy weather and the value 4 in case of warm, sunny weather) and the wear of the norm violator (a threefold scaled ordered variable with value 0 indicating sporty, casual wear and value 2 indicating rather chic, elegant wear). We include the latter variable, as the wear of the norm violator presumably expresses her social status.

In contrast to Balafoutas and Nikiforakis (2012) we do not find a link between the observer’s gender and the norm enforcement behavior in the field experiment: Out of the 160 women 52 enforce the norm (32.5%). Among the 77 men 24 enforce the norm (31.2%, χ^2-test, $p = 0.837$). With respect to the field of study most of our subjects participating both in the lab and the field are arts students. We find no effect of the field of study on norm enforcement (Fisher exact test, $p = 0.157$). Furthermore the weather – we find a norm

Table 3.6 - Probit regressions of norm enforcement in the field.

Dep. var.: Norm enforcement in the field	(I)	(II)	(III)	(IV)	(V)
PDG+P:					
Cooperation decision	0.005 (0.411)				
Sanctioning of defectors	0.019 (0.100)				
PDG+PCP:					
Cooperation decision		0.267** (0.421)	0.212 (0.451)	0.301** (0.138)	0.250* (0.450)
Sanctioning of cooperators				-0.112 (0.072)	
Sanctioning of defectors			0.079** (0.110)		
Diff. in sanctioning of defectors and cooperators					0.100** (0.050)
QG+P:					
Queue-jumping					
Sanctioning					
Cooperation type					
Scenarios					
N. enfor. in Group task scen.					
N. enfor. in Littering scen.					
N enfor. in Queuing scen.					
Constant					
N	45	45	45	45	45
Wald χ^2	0.34	3.81*	10.93***	5.88*	7.99**
Log pseudo-likelihood	-25.924	-24.094	-21.559	-23.475	-20.516

(*Continued*)

enforcement rate of 36.0% in case of cloudy, rainy weather versus 33.3% in case of warm, sunny weather (χ^2-test, p = 0.692) – and the wear of the norm violator – we find a norm enforcement rate of 34.8% in case of sporty, casual wear versus 26.1% in case of chic, elegant wear (χ^2-test, p = 0.601) – do not seem to be related to norm enforcement in the field. In contrast, we find a significant effect of the observers' age. Among the lab and field subjects age ranges from 18 to 44 years. We find that the norm enforcers in the field are on

Table 3.6 – Probit regressions of norm enforcement in the field (*Continued*).

Dep. var.: Norm enforcement in the field	(VI)	(VII)	(VIII)
PDG+P:			
Cooperation decision			
Sanctioning of defectors			
PDG+PCP:			
Cooperation decision			
Sanctioning of cooperators			
Sanctioning of defectors			
Diff. in sanctioning of defectors and cooperators			
QG+P:			
Queue-jumping	-0.049 (0.420)		
Sanctioning	-0.012 (0.065)		
Cooperation type		0.284** (0.389)	
Scenarios			
N. enfor. in Group task scen.			0.043 (0.060)
N. enfor. in Littering scen.			-0.001 (0.032)
N enfor. in Queuing scen.			0.138*** (0.039)
Constant			
N	45	42	45
Wald χ^2	0.33	4.78**	9.72**
Log pseudo-likelihood	-25.920	-22.855	-19.438

Notes: This table outlines the results of probit regressions of norm enforcement in the field. We report the marginal effects at the mean. An observation in the regression is one subject participating both in the lab and the field. For the PDG+P we do not include the sanctioning of cooperators as regressor, as all subjects who are willing to sanction cooperators are non-norm enforcers and therefore determine norm enforcement in the field perfectly. OLS regressions yield similar results. Standard errors are presented in parentheses. *, **, *** denote significance at the 10%, 5% and 1% level respectively.

average by 4 years older than the non-norm enforcers (27 years versus 23 years, MWU-test, p = 0.074).

Table 3.7 - Probit regressions of norm enforcement in the field (control variables included).

Dep. var.: Norm enforcement in the field	(I)	(II)	(III)	(IV)	(V)
PDG+P:					
Cooperation decision		-0.004 (0.138)			
Sanctioning of defectors		0.003 (0.033)			
PDG+PCP:					
Cooperation decision			0.197 (0.142)		
Sanctioning of defectors			0.074* (0.041)		
QG+P:					
Queue-jumping				-0.044 (0.130)	
Sanctioning				-0.023 (0.020)	
Cooperation type					0.215* (0.125)
Control variables:					
Female	-0.032 (0.405)	-0.060 (0.129)	-0.112 (0.133)	-0.021 (0.136)	-0.013 (0.136)
Age	0.033*** (0.042)	0.033*** (0.013)	0.033** (0.015)	0.036*** (0.012)	0.033** (0.015)
Field of study	0.044 (0.147)	0.042 (0.042)	0.003 (0.043)	0.041 (0.042)	0.042 (0.048)
Weather	-0.034 (0.144)	-0.027 (0.046)	-0.021 (0.051)	-0.028 (0.045)	-0.029 (0.050)
Wear	0.069 (0.309)	-0.007 (0.075)	-0.021 (0.078)	-0.016 (0.072)	-0.023 (0.073)
N	45	45	45	45	42
Wald χ^2	8.64	8.43	12.98*	13.13*	9.05
Log pseudo-likelihood	-21.274	-21.510	-18.947	-21.053	-19.097

Notes: This table outlines the results of probit regressions of norm enforcement in the field. We report the marginal effects at the mean. An observation in the regression is one subject participating both in the lab and the field. OLS regressions yield similar results. Standard errors are presented in parentheses. *, **, *** denote significance at the 10%, 5% and 1% level respectively.

Table 3.7 outlines the results of probit regressions of norm enforcement in the field. In specification (I) we examine the relation between norm enforcement in the field and the control variables. We further outline several specifications of table 3.6 whereby we additionally include the control variables. All specifications confirm the non-parametric results in this section: Among the control variables only age is correlated with norm enforcement in the field. The willingness to enforce the norm increases by about 3% with each additional year of an observer's life. The norm enforcement rate might increase with age, because the adherence of social norms might become more important for people in the course of life. Another explanation might be that people rather dare to enforce social norms with increasing age. Specifications (II) to (V) indicate that the results of section 3.4.4 hold even if we include the control variables.

3.4.6 Results on the Questionnaire Scenarios

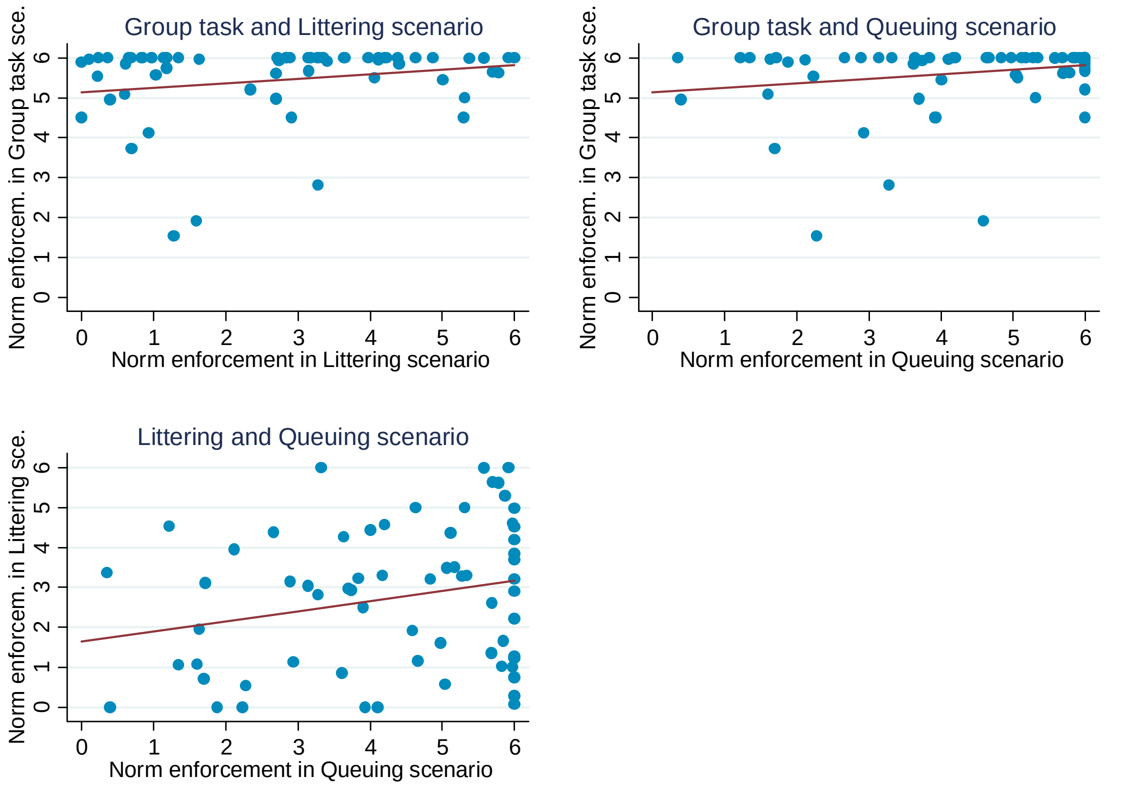
In this section we present of the 66 subjects participating in the lab the intended norm enforcement behavior stated in the three questionnaire scenarios, each describing the violation of a concrete social norm. We ask subjects to rate on a 7-point scale from 0 to 6 whether they would confront the norm violator in the scenario. The detailed description of the scenarios as well as the related questions with the corresponding answers are outlined in appendix A.3.3.

Table 3.5 outlines that non-norm enforcers and norm enforcers in the field only differ regarding the stated sanctioning behavior in the Queuing scenario (3.7 versus 5.6), but not in the other scenarios. This result is confirmed by specification (VIII) of table 3.6. Hence, stated intentions on norm enforcement behavior seem to be correlated with the actual norm enforcement behavior in the field, as long as the same social norm is affected. This finding confirms former research in social psychology: Within the framework of the theory of reasoned action several studies show that stated intentions have a highly predictive power, as long as the intention corresponds to or is compatible with the predicted behavior (Ajzen, 1991; Ajzen and Fishbein, 2011).

In the Group task scenario subjects on average completely agree to confront the norm violator (5.6), in the Littering scenario they are neutral (2.8) and in Queuing scenario they would rather confront the queue-jumper (4.4). Traxler and Winter (2012) find in the domain of legal norms survey evidence that law violations with the largest negative externalities are punished most frequently. In our study the Group task scenario describes the norm violation with the potentially largest negative externality. Accordingly we find that the intended sanctioning in this scenario is highest (pair wise Wilcoxon signed-ranks tests, $p = 0.000$, $p = 0.000$), which suggests that the result by Traxler and Winter (2012) on the punishment of law violations may be extended to the enforcement of social norms.

This leads us to the question whether a norm enforcing type exists, i.e. somebody who generally behaves in a norm enforcing way across several social norms, or whether norm enforcement depends on the concrete social norm. The scatterplots in figure 3.1 show that there is hardly any systematic relation between the scenarios with respect to intended norm enforcement behavior. This finding is supported by table A.3.1 which outlines the correlations between stated norm enforcement in the three scenarios. It shows that these correlations do not exceed 23.0% and are at best significant at the 10%-level.

Figure 3.1 - Scatterplots of norm enforcement stated in the questionnaire scenarios.



Notes: These scatterplots outline the pair wise relations between the three questionnaire scenarios with respect to intended norm enforcement. One independent observation in the figures is one out of the 66 subjects participating in the lab, respectively. The first scatterplot refers to the relation between the Group task and the Littering scenario, the second scatterplot shows the relation between the Group task and the Queuing scenario, and the third scatterplot refers to the relation between the Littering and the Queuing scenario.

We additionally calculate Cronbach's α, the widely accepted index for the internal consistency, i.e. the interrelatedness of items (Cortina, 1993). If the number of items is small, as is the case in our questionnaire, a value of at least 0.70 can be considered as satisfactory and as a necessary but not sufficient condition for unidimensionality (Green et al., 1977; Cortina, 1993). Cronbach's α for the stated norm enforcement in our three scenarios reaches a mean value of 0.423, with a one-sided 95% confidence interval at a value of 0.186 – hence clearly below the 0.70 boundary.

RESULT 3.6: *We do not find evidence for a norm enforcing type. Our questionnaire data rather suggest that norm enforcement depends on the concrete social norm violated, and specifically on the externality of the norm violation.*

Contingent on the stated norm enforcement intentions in the scenarios, we ask the subjects for their reasons for confronting or not confronting the norm violator. The results are presented in appendix A.3.3. We are especially interested in the reasons why people do not confront the norm violator. Unfortunately in case of the Group task scenario and the Queuing scenario no subject states that she would not confront the norm violator. Therefore we do not have any information on the reasons for not confronting the violator in these scenarios. In case of the Littering scenario subjects most frequently state the fear of counter-punishment for not confronting the norm violator (pair wise Wilcoxon signed-ranks tests, $p = 0.000$, $p = 0.001$, 26 obs.). This result further supports our interpretation of result 3.3: Weighing up the danger of being counter-punished is crucial for norm enforcement decisions in the field.

3.5 Discussion

In our field experiment we observe a norm enforcement rate of 32.1%. Interestingly this rate corresponds to the average expected norm enforcement rate of 37.6%, which subjects state in our lab questionnaire in case of queue-jumping (see appendix A.3.3). Our norm enforcement rate lies between the one of 43.3% and 54.0% by Milgram et al. (1986) and Schmitt et al. (1992) and the norm enforcement rate of 4.0% for the universal norm and of 19.3% for the environment-specific norm by Balafoutas and Nikiforakis (2012). There are several reasons that might explain why our norm enforcement is lower than the one by Milgram et al. (1986) and Schmitt et al. (1992): A methodological reason might be that in our study we do not evaluate as norm enforcing if the queuer next to the norm violator pushes herself in front of the norm violator. In this case we observe the reaction of the following queuers. Secondly, most of the queuers in our study are students, i.e. rather young people. We have shown that the likelihood of norm enforcement increases with age. Thirdly, the other two studies on queue-jumping are predominantly carried out at the ticket counter of the Grand Central Station in New York City, where people are in hurry, as they need to reach their trains. This might increase the likelihood of norm enforcement. Moreover we cannot exclude inter-temporal or cultural differences in norm enforcement. The latter is supported by laboratory findings (Henrich et al., 2006; Herrmann et al., 2008; Marlowe et al., 2008).

On the other hand our norm enforcement rate is higher than the ones by Balafoutas and Nikiforakis (2012), which might be due to the different social norms. Interestingly the different norm enforcement rates correspond to our results of the questionnaire scenarios that also indicate that queue-jumping is more heavily confronted than littering in public places, which is one of the norm violations in the field experiment by Balafoutas and Nikiforakis (2012). Our questionnaire data further suggest that the finding by Traxler and Winter (2012), that norm enforcement increases with the negative externality of law violations, can be extended to the domain of social norms. There are good reasons to assume that queue-jumping causes larger negative externalities than littering in public places or stopping at the left side of the escalator. Secondly, Balafoutas and Nikiforakis (2012) conduct their field experiment in the highly frequented main subway station in Athens. Hence strategic motives of norm enforcement are minimized. Our field study is carried out at the subway station near the University of Munich which is presumably frequented by less people and it might be frequented by several students everyday. Thus, we can not completely exclude the possibility that some subjects in our field experiment have seen the norm violator before. Thus, the strategic motives in our field experiment might be enhanced compared to the ones in the study by Balafoutas and Nikiforakis (2012). Thirdly, these authors use a more conservative measure of norm enforcement than in our field experiment: Only if the observer reacts twice this is evaluated as norm enforcing. Fourthly, we cannot exclude cultural differences to account for the different norm enforcement rates.

The main finding of our experiment is that norm enforcement in the field correlates with decisions in the lab PDG+PCP – specifically with the cooperation decision and the sanctioning of defectors – but it does not correlate with decisions in the PDG+P or the QG+P, i.e. the treatments which allow only for a one-sided punishment option. We discuss whether it is really the possibility of counter-punishment that is responsible for these different correlations or whether there are any design specifics that drive them. Therefore we compare the PDG+P and the PDG+PCP in detail. Firstly, one could object that the PDG+P is either played as the first or second part, whereas the PDG+PCP is always played as the last – i.e. the forth – part. Specifically one could object that learning or order effects drive the different correlations. However, as no feedback is given before the last stage of the PDG+PCP, learning effects are unlikely to drive the different correlations. Secondly, one could complain that we pay out all parts of the lab experiment. Expectations about earnings in former parts might influence decisions in later parts. In the light of these first two objections, however, it is even more surprising that the norm enforcement behavior in the field does not correlate with decisions made in the first or second part, but with decisions made in the forth part, i.e. in the PDG+PCP.

Thirdly, in the PDG+P the feedback on the first and second stage is given at the end of the laboratory session, whereas in the PDG+PCP the feedback on both stages is immediately given at the beginning of the counter-punishment stage. Thus, the time span between the decisions and the corresponding feedback is unequally large. As we are not aware of any laboratory study that finds an effect of the time span between decisions and the feedback on the decisions, we rather assume that this point is an unlikely candidate to be responsible for the main result. However, this might be an interesting methodological research question for future research.

Fourthly, one could object the slightly different feedback format regarding the first stage of the PDG+P and the PDG+PCP: In the PDG+P subjects receive feedback – at the end of the session – on the individual first stage decisions of each unit member and on their own first stage profit, whereas in the PDG+PCP subjects additionally receive feedback on the individual first stage profits of their unit members. Nikiforakis (2010) shows that the feedback format in a repeated public goods game with a one-sided punishment option affects contribution and punishment decisions over time. However, the feedback format does not affect decisions in the first period. As the PDG+P and the PDG+PCP are both only played for one period, the feedback format is an unlikely to drive our main result. Moreover in the prisoners' dilemma game in comparison to a public goods game it is much easier to derive earnings from individual cooperation decisions. In the light of this argumentation we are confident to assume that the different correlations between the norm enforcement behavior in the field and the decisions in the PDG+PCP in contrast to the decisions in the PDG+P are primarily driven by the possibility of counter-punishment.

Next we discuss the impact of the strategy vector method on the generalizability of our main result to lab decisions elicited wit the direct response method. In the lab we implement the strategy vector method (Selten, 1967) in each treatment in order to avoid learning effects across treatments and to receive information on the sanctioning behavior contingent on all possible cooperation decisions of the other unit members. Falk et al. (2005) elicit decisions in a three-person prisoners' dilemma game both with the strategy vector method and the direct response method. They find a qualitatively similar pattern of punishment decisions, similar punishment rates, but a higher punishment severity with the direct response method. Regarding cooperation the vast majority of studies (e.g. Brandts and Charness 2000; for an overview see Brandts and Charness, 2011) do not find a significant difference between the direct response and the strategy vector method. Moreover Brandts and Charness (2011) conclude that they find no case in which a treatment effect found with the strategy vector method does not hold with the direct response method. Hence, we are confident to assume that all significant results that we find for the PDG+PCP also hold with the direct response method. As cooperation and punishment rates are not found to differ between the elicitation methods, we also assume that our results in the PDG+P and the QG+P regarding cooperation and punishment rates also generalize to the elicitation with the direct response method. The only case for which we cannot exclude that the elicitation method might matter is the non-significant result on the sanctioning level in the PDG+P and QG+P. Except for the latter case we are confident to assume that our main finding also extends to laboratory decisions elicited with the direct response method.

Our main result states that norm enforcement in our field experiment correlates by 29.8% with the cooperation decision and by 30.2% with the level of sanctioning defectors in the PDG+PCP. How does this magnitude of correlations relate to findings in social psychology? In the context of the person-situation debate many psychological studies report correlations of behavior in different situations not exceeding 30% (Mischel, 1968; Ross and Nisbett, 1991). In order to obtain a more reliable measure of individual acts, psychologists suggest to aggregate behavior over several situations (Epstein and O’Brian, 1985). In the context of these social psychological results the reported correlations are quite remarkable, especially as we elicit the behavioral measures in the lab and in the field only once.

Our field experiment contains some restrictions. Specifically, the norm violation is only committed by one single norm violator, namely the authoress, as we could not exclude the possibility that the queue-jumping provokes aggressive reactions or yields a loss of reputation for the queue-jumper. Balafoutas and Nikiforakis (2012) do not find an effect of the norm violator’s gender and height on norm enforcement, but it is not clear whether this also holds for our norm violation. In the field experiment by Milgram et al. (1986) the gender of the queue-jumper is varied. However, the authors do not report whether it has an effect on norm enforcement or not. A further restriction of our study is that we consider one specific norm violation in the field. Thus, generalizations of our main result have to be made with care. Our lab-field comparison can be seen as first evidence on the external validity of lab experiments regarding norm enforcement. It opens up an interesting avenue for future research on lab-field comparisons including the violation of further social norms and varying for instance the gender, age and height of the norm violator in the field.

3.6 Conclusion

We examine whether norm enforcement in the lab translates to the field. Therefore we conduct a within-subjects comparison. In a natural field experiment we elicit norm enforcement when a norm violator cuts in line. We invite the subjects from the field to the lab without revealing that they have participated in a field experiment. In the lab three treatments dealing with norm enforcement are implemented, which allows us to run a “horse race” between institutions: The standard treatment to examine norm enforcement in the lab – a three-person prisoners’ dilemma game with a one-sided punishment option (see Fehr and Gächter, 2000; 2002; Falk et al., 2005) – and a treatment additionally allowing for counter-punishment (Nikiforakis, 2008). Thirdly, we apply a newly created game with which we try to represent the situation in the queue in an abstract way. This so-called “queuing game” includes a one-sided punishment option. In addition to the treatments dealing with norm enforcement we elicit the cooperation types of subjects.

Our main finding says that norm enforcement in the field correlates with decisions in the lab treatment allowing for counter-punishment, but not with decisions in treatments including only a one-sided punishment option. Despite the differences between lab and field with respect to anonymity, stakes, scrutiny, strategic motives etc. (see also Levitt and List, 2007b) we specifically find that the likelihood of norm enforcement in the field increases with the willingness to cooperate and with the sanctioning of defectors in a prisoners’ dilemma game with a two-sided punishment option. Our lab-field comparison is a first examination of the external validity of norm enforcement in the lab. Therefore generalizations regarding the interpretation and implication of our main result have to be made with care. However it suggests the following: Firstly, with respect to norm enforcement our main result confirms prior survey evidence that weighing up the danger of being counter-punished is crucial for norm enforcement decisions in the field. The fact that counter-punishment opportunities mostly exist in decentralized interactions in the field, whenever punishment opportunities exist, even increases the implication of our main result. Secondly, concerning the methodological aspect, our main result suggests that more evidence from the field is necessary to establish the external validity of the standard treatment with a one-sided punishment option. Overall our main finding is encouraging with respect to the generalizability of lab behavior to the field. However, it also illustrates that the relevant institutional factors of the field need to be incorporated in the lab to allow for external validity.

Our further results are the following: We find a norm enforcement rate of 32.1% in the field. The likelihood of norm enforcement in the field increases with age and with the cooperation type elicited in the lab. Presenting several questionnaire scenarios with violations of concrete social norms to the subjects in the lab, we examine whether a norm enforcing type exists, i.e. someone who generally behaves in a norm enforcing way across several social norms. We find a high external validity of the norm enforcement intentions stated in the survey, as long as the same social norm is affected in the survey scenario and in the field. However, we do not find evidence for a norm enforcing type. Instead our questionnaire data suggest that the likelihood of norm enforcement varies with the concrete social norm affected, and specifically with the negative externality of the norm violation. This result can be considered as a first indication. We recommend further research including a greater number of social norm violations presented to a larger sample of subjects. As our main result stresses that the danger of being counter-punished is crucial for norm enforcement decisions in the field we suggest more lab experiments including the possibility of avenging punishment. Moreover we recommend more “horse races” of economic institutions to get a deeper understanding of certain field behavior. Finally our lab-field comparison can be seen as first evidence on the external validity of lab experiments regarding norm enforcement. It opens up an interesting avenue for future research on lab-field comparisons including the violation of further social norms in the field.

3.7 Appendix A.3

A.3.1 Instructions (translated from German)

Welcome to the experiment and thank you for your participation!

From now on please do not speak with other participants of the experiment and please turn off your mobiles.

General remarks about the procedure

This experiment examines decision-making behavior. You can earn money, which will be paid cash after the experiment. During the experiment you and the other participants will be asked to take decisions. Your decisions as well as the other participants' decisions will determine your payment according to the following rules.
The experiment will last at most **120 minutes**. If you have questions or if something is not clear, you can just raise your hands. I will come to you and answer your question privately. For the sake of simplicity we only use male notations in this experiment.
During the experiment we will not speak of Euros but of **Guilders**. Your profit during the experiment will therefore be calculated in Guilders. At the end of the experiment all your earned Guilders will be converted into Euros according to the following exchange rate:

5 Guilders = 1 Euro
1 Guilder = 0.20 Euros (or 20 Eurocent)

For your punctual appearance you will additionally receive 20 Guilders. At the end of the experiment you will receive the sum of all your profits made during the experiment in private and in cash.
While you are coming to your decisions, a clock will run down in the upper right corner of the screen. This will give you an orientation in which time you should take your decisions. Normally you can exceed this given time frame. If exceeding the time frame is not possible, we will inform you in advance.

Anonymity
All your inputs over the whole experiment will be anonymous. You will at no time receive personal information about the other participants of this experiment. The elicited data will be encoded and treated as confidential. At the end of the experiment you will have to sign a receipt, which will only serve our accounting. Neither the sponsor of the experiment will receive any individual data of the experiment. Furthermore, at the end of the experiment we will need your written consent for saving your elicited data anonymously as well as for analysing it in aggregated form.

Tools
At your place you find a pen as well as a piece of paper. We ask you to leave both at your table after having finished the experiment.

The experiment
The experiment consists of **four parts and five questionnaires**. For each part of the experiment you will receive an instruction. The single parts are further subdivided into several steps and are independent of each other, i.e. decisions in one part do not have any influence on later parts of the experiment. Your overall earnings arise from the sum of all the earnings you gained within the parts I to IV, from the questionnaires and from the 20 Guilders for your punctual appearance. At the end of the experiment you will be informed about your overall earnings and about the purpose of this experiment. Each of you will then separately receive the payment in cash.

Part I - First Stage

You are a member of a randomly created three-person group. Your group thus consists of you and two other group members. You will at no time be told the identity of your two group members. The group formation with the other two participants is only valid within part I. It is impossible to form another group with the same two participants in subsequent parts.
Part I consists of **three stages**: Initially, we will explain to you the first step.

First stage
In the first step you have to **calculate** the sum of five randomly chosen two-digit numbers which appear on the screen. You will be given 5 minutes to calculate the correct sum of a series of these problems. For the calculation of the sums you are not allowed to use a calculator. You can, however, use the piece of paper at your place. Below you can see an example of how the screen will look like in the first stage.

Part I: First Stage

Problem 3

Please calculate the sum of the five numbers

67	65	36	77	33	

OK

In the last column, the input field, you have to type in the sum of the five given numbers. To confirm the sum please press the "OK" button in the lower right corner. After pressing "OK" you will no longer be able to change your decision. Afterwards you will be immediately informed whether your answer has been correct or not.

At the end of the 5 minutes, you and your group members will be put in a **ranking according to the quantity of correctly solved problems.** The group member that has correctly calculated the largest number of problems within the given time will obtain position A; the group member that has correctly calculated the second largest number of sums will obtain position B; the group member that has correctly calculated the lowest number of sums will obtain position C. If two group members solve the same amount of problems correctly, the PC will randomly decide between the two positions.

Your income in the first stage depends on your position within the ranking:
The group member on **position A** (we will call him "**group member A**") earns **40 Guilders**,
the group member on **position B** (we will call him "**group member B**") earns **30 Guilders**,
the group member on **position C** (we will call him "**group member C**") earns **20 Guilders**
in the first stage.

It is advantageous for you to correctly solve as many problems as possible in the first stage.

Only at the end of the experiment you will be informed about the amount of correctly solved problems by your group members and yourself and thus about your position in the ranking.

We will now start with the first stage. Afterwards you will be explained the second and the third stage of part I. Do you have any questions? If so, please raise your hand. I will come to your cubicle and answer your questions privately.

Part I – Second and Third Stage

(The instruction for the second and third stage was only presented after the accomplishment of the first stage)

The general decision situation

Based on the amount of correctly solved problems a ranking out of you and your two group members was set up in the first stage. You will get to know your position at the end of the experiment.

In the following we will first explain to you the general decision situation in the second and the third stage. Afterwards you will find on the screen some sample exercises to familiarise yourself with the decision situation. After that you will learn the precise procedure of the second and the third stage.

Second stage

In the second stage the group member on position C (group member C) decides whether

- the **original ranking of the group members' income** from the first step should be maintained or
- whether an **alternative ranking** should apply.

If group member C chooses the alternative ranking, the group members will earn the following income:
Group member C earns 40 Guilders.
Group member A earns 30 Guilders.
Group member B earns 20 Guilders.

If group member C chooses the original ranking from the first stage, the group members will earn the same income as determined in the first stage.

Third stage

If group member C maintains in the second step the original ranking (from the first stage), your income of the first stage will count as your final income from part I and a third stage will not take place anymore.

If group member C however chooses the alternative ranking of incomes in the second stage, **group member A can reduce group member C's income from the second stage or leave it unchanged**. Group member A can reduce group member C's income by **allocating deduction points to C**.

The allocation of deduction points has an impact on the incomes of group members A and C. First we describe the impact on group member C's income. If group member A allocates deduction points to C, **the 40 Guilders of group member C** from the second stage **will be reduced by three times the allocated deduction points.** If, for example, group member C receives one deduction point from group member A, then in part I group member C will earn 40 – 1 * 3 = 37 Guilders. If group member C, however, receives 3 deduction points, then group member C will earn 40 – 3 * 3 = 31 Guilders etc. Thus, every deduction point that group member C receives will reduce his income by 3 Guilders. Please note that group member A can **allocate at most 10 deduction points to group member C**.

If group member A allocates deduction points to group member C, then group member A has to bear cost as well. Every deduction point that group member A allocates to group member C costs group member A **one Guilder each**. If, for example, group member A allocates one deduction point to
group member C, then group member A earns in part I 30 – 1 * 1 = 29 Guilders. If group member A, for example, allocates 8 deduction points to group member C, then A earns 30 – 8 * 1 = 22 Guilders.

Thus, the following **income (in Guilders) for group member C** results in part I:
= 40 – 3 * (the number of received deduction points)

Thus, the following **income (in Guilders) for group member A** results in part I:
= 30 – 1 * (the number of allocated deduction points)

The **income (in Guilders) of group member B** is in any case:
= 20

Please note that these rules only apply if group member C chooses the alternative ranking in the second stage.

<u>Control questions</u>
Before continuing with the instruction, we ask you to complete the sample exercises on your screen. In case of questions please raise your hand. I will then come to your cubicle and answer your questions privately.

Part I – Second and Third Stage

Procedure

The second and the third stage contain the decision situation as described above. You will take your decisions in both stages of part I **only once**. In doing so, you take your decision in the second stage assuming to be group member C and in the third stage assuming to be group member A.

<u>Second stage</u>
In the second stage, **for the case of being group member C,** you have to indicate whether the original distribution of incomes from the first stage should be maintained or whether it should be changed as described above. If you take this decision in the second stage you do not know whether you are group member C or not. Therefore you should consider your decision carefully, as it could be relevant for you.
Only at the end of the experiment you will get to know whether you are group member C or not, i.e. whether your second stage decision is payoff-relevant for all members of your group or whether the second stage decision of another group member will be payoff-relevant.

<u>Third stage</u>
In the third stage **we assume that group member C chose the alternative income ranking in the second stage and that you are group member A**. If you make the decision in the third stage you do not know which group member you are and whether group member C chose the alternative ranking in the second stage. However, your decision in the third stage could be payoff-relevant for you. Please think about your decision carefully.

In the third stage you indicate whether you want to distribute deduction points to group member C and, if yes, how many – always assuming you are group member A and that group member C chose the alternative ranking in the second stage. The screen in this stage looks as follows:

Part I – Third stage

Assume that you are group member A and that group member C in the second stage chose the alternative ranking.
Please decide as group member A, whether you want to distribute deduction points to group member C and, if yes, how many:

OK

You have to type in a **number between 0 and 10 in the input field**. You definitely have to make an input. If you do not want to change the income of group member C you have to type in “0”. If you want to distribute, for example, 3 deduction points please type in “3” etc. You can distribute at most **10 deduction points** to group member C. If you have made your decision, please press the “OK” button. Then you will not be able to revise your decision.

You take your decision only **once**. You will be informed about the result of part I at the end of the experiment, after finishing the questionnaires. Then you will get to know how many problems your two group members and yourself correctly answered in the first stage and as which group member you arise from the first stage. Furthermore all participants of your group will learn to know the decision of group member C in the second stage and the number of deduction points group member C received. Every participant will be informed about his own income of part I.

Do you have any questions? If so, please raise your hand. I will come to your cubicle and answer your questions privately.

Part II

(Parts and instruction were presented sequentially)

The general decision situation

First we will explain to you the general decision situation. Afterwards you will find some sample exercises on the screen, which will help you to familiarise yourself with the decision situation. After that you will learn the precise procedure of part II.

You are again a member of a randomly matched **three-person group**. Thus, your group consists of you and two other group members. You will at no time be told the identity of your two group members. The group formation with the other two participants is only valid in part II. It is impossible to form a group with the same two participants as in part I or to form a group with the same persons in subsequent parts.

Part II consists of **two stages.** In the first stage **you take a decision between two options: Option A and option B**. In the second stage **you then decide whether you want to reduce the first stage income of the other two group members or not** and, if so, by how much. This can be done by allocating deduction points. In the following we will explain part II in detail.

<u>First Stage</u>

In the first stage of part II **you decide between two options: option A and option B**. The other two members of your group also decide between option A and option B. The income of each group member in the first stage of part II will be determined in the same way according to the following payoff matrix:

	Your two group members both choose Option A	**One group member chooses Option A, the other chooses Option B**	**Your two group members both choose Option B**
You choose Option A	20	32	44
You choose Option B	12	24	36

This means:
If all three group members choose option B, then all three group members receive in the first stage of part II 36 Guilders each, i.e. you as well as your two group members will earn 36 Guilders each in the first stage of part II.
If you and a second group member both choose option B, whereas the third group member chooses option A, then you and the second group member will receive 24 Guilders each. The third group member, who chose option A, will then receive 44 Guilders.
If you choose option B and your two group members choose option A, then you will receive 12 Guilders in the first stage of part II, whereas your two group members earn 32 Guilders each.

If all three group members choose option A, then all three ones will receive 20 Guilders each.
If you and a second group member choose option A while the third group member chooses option B, then you and the second group member will receive 32 Guilders each. The third group member, who chose option B, will then receive 12 Guilders.
If you choose option A and your two group members choose option B, then you will receive 44 Guilders in the first stage of part II. The two other group members will then receive 24 Guilders each.

Thus, your earnings in the first stage depend on your own decision as well as on the decision of the other two group members. As soon as all members of your group took their decision in the first stage, the second stage follows.

<u>Second Stage</u>

In the second stage **you may reduce the income of each other group member by allocating deduction points.** You may **also leave the incomes unchanged**. The other two group members can reduce your income in the same way, if they want to.

In the following we will describe how the distribution of deduction points affects the incomes: If you allocate deduction points to one of the group members, **you will reduce his first stage income by three times your distributed deduction points**. If, for example, you allocate 1 deduction point to another group member, then you will reduce his first stage income by 3 Guilders. If you allocate to another group member 2 deduction points, you will reduce his income by 6 Guilders, etc. Thus, each deduction point, which you allocate to another group member, will reduce his income by 3 Guilders. Equally, each deduction point allocated to you reduces your first stage income by 3 Guilders. Please note that you can **allocate at most 10 deduction points per group member**.

Cost of received deduction points (in Guilders) = 3 * (sum of received deduction points)

If you allocate deduction points, you have to bear cost in Guilders as well. Each distributed deduction point costs you **one Guilder**. For example, if you allocate altogether 4 deduction points, this will cost you 4 Guilders and so on. If you do not allocate deduction points, then there obviously will not arise any cost to you.

Your income at the end of the second stage, i.e. your income in part II, will be therefore calculated as follows:

Your income (in Guilders) at the end of the second stage =
= First stage income – 3 * (sum of received deduction points) – sum of distributed deduction points

For example, if your first stage income amounts to 36 Guilders and if you received from the other two group members altogether 3 deduction points in the second stage while distributing one deduction point to another group member, then your income will be calculated as 36 – 3 * 3 – 1 = 26 Guilders. In contrast, if your first stage income amounts to 20 Guilders and if you receive 2 deduction points and you do not allocate any deduction points in the second stage, then your income of part II will be calculated as 20 – 2 * 3 – 0 = 14 Guilders.

Please note that **the income of one group member cannot be reduced further than to 0 Guilders by the other two group members**. This means, if the cost of the received deduction points is larger than the group member's first stage income, his income will be reduced to 0 Guilders and not further. Nevertheless, even in this case the group member has to bear the cost of his distributed deduction points.

This means that your income (in Guilders) may be negative at the end of the second stage in part II, if your cost of distributing deduction points in the second stage exceed your (possibly reduced) first stage income. **However, you can avoid such losses in part II <u>with certainty</u> by virtue of your own decision.** Possible losses will be covered by the 20 Guilders, which you received automatically in the beginning of the experiment and which will be charged against your overall income of the experiment at the end.

<u>Control questions</u>
Before continuing with the instruction, we ask you to complete the sample exercises on your screen. In case of questions please raise your hand. I will then come to your cubicle and answer your questions privately.

Part II

Procedure

Part II includes the decision situation as described above. You take your decision in part II **only once.**

<u>First Stage</u>
You take your decision in the first stage by selecting either the field "I choose option A" or the field "I choose option B". If you have taken your decision please click on the "OK" button.

<u>Second Stage</u>
Only at the end of the experiment after finishing the questionnaires you will learn how the other two participants decided in the first stage. Therefore you have to indicate in the second stage for each possible case how many deduction points you want to distribute. There are four possible cases:

Case 1: The other two group members chose option A.
Case 2: The second group member chose option A and the third member chose option B.
Case 3: The second group member chose option B and the third member chose option A.
Case 4: The other two group members chose option B.

For all four possible cases you have to decide whether you want to allocate deduction points to the other group members and, if so, how many. So that you can take your decision, four input screens representing the four cases will arise. We show you the third screen (case 3) as an example:

Part II – Second stage

Please decide whether you want to distribute deduction points and, if so, how many:

Case 3

	You	Second group member	Third group member
First stage decision	Option B	Option B	Option A
First stage income	24	24	44
Your decision in the second stage	---		

OK

The screen shows case 3 in which the second group member chose option B and the third group member chose option A. We suppose that you chose option B in the first stage. Now you have to decide for this case, whether you want to distribute deduction points to the other two group members and, if so, how many.
The screen is built as follows:

- The **second row** indicates the decision between option A and option B in the first stage.
- The **third row** indicates the income of each group member that results from the first stage. Under the heading "You" you find your income and in the third and fourth column you see the incomes of the other two group members.
- In the **fourth row ("Your decision in the second stage")** you have to decide, how many deduction points you want to distribute to each of the other two group members by indicating a number between 0 and 10 in each input field. You definitely have to make an input. If you do not want to change the income of one of the group members, you indicate "0". If you want to distribute for example 7 deduction points you have to indicate "7" etc. **You can distribute at most 10 deduction points to each member of your group**.

First you take your decision for case 1. As soon as you made your input, please click on the "OK" button. Then the second screen (case 2) will appear, followed by the third (case 3) and the fourth (case 4) screen. Thus, with regard to the distribution of the deduction points you take 8 decisions in total on four screens.

Obviously, the other two group members only took one combination of decisions in the first stage. This means, that only one of the four cases actually applies. For your payoff it will therefore be relevant how you decided in this specific case.

For example, imagine that case 1 applies (the second as well as the third group member chose option A): This means, that your decision on the first screen will be payoff-relevant. The deduction points you indicated in case 1 determine your payoff and the payoff of the other two group members. In contrast, if case 3 applies, the decisions you took on the third screen will be payoff-relevant.

You take your decision only **once**. At the end of the experiment, after finishing the questionnaires, you will get to know how the other two group members decided in the first stage, the number of deduction points you received in the second stage as well as your earnings from part II.

Do you have any questions? If so, I will come to your cubicle and answer them privately.

Part III

The general decision situation

Again we will first outline the general decision situation. Afterwards you will learn the procedure of part III in detail.

You are again a member of a randomly matched **three-person group**. You will at no time be told the identity of your two group members. It is impossible to form a group with the same two participants from former parts or to form a group with the same persons in part IV.

In part III you have to **decide again between option A and option B**. The other two members of your group have to decide as well between option A and option B. The income of each group member will be determined in the same way according to the following payoff matrix:

	Your two group members both choose Option A	**One group member chooses Option A, the other chooses Option B**	**Your two group members both choose Option B**
You choose Option A	20	32	44
You choose Option B	12	24	36

This means:
If all three group members choose option B, then all three group members receive 36 Guilders each in part III and so on.

Procedure

Part III includes the decision situation as described above. You take your decision in part III **only once.**

<u>Your input</u>
As described above you have the choice between option A and option B. In part III, each group member has to take **two types** of decisions, which we will call in the following the **unconditional decision** and the **decision table**.

- When taking your unconditional decision you simply choose between option A and option B. As soon as you have taken your unconditional decision, please press the “OK” button.
- You will then be asked to fill in a **decision table**. In the decision table you have to indicate **for each possible combination of unconditional decisions of the <u>other</u> two group members, how <u>you</u> want to decide yourself between option A and option B**. Thus, you have to take your decision between option A and option B depending on how the other two group members made their unconditional decisions. The screen will be built up as follows:

Part III

Your conditional decision (Decision table):

Combinations of unconditional decisions of the other group members	Your two group members both choose Option A	One of your group members chooses Option A, the other group member chooses Option B	Your two group members both choose Option B
Your decision	○ Option A ○ Option B	○ Option A ○ Option B	○ Option A ○ Option B

OK

In the first row of the table above you see the possible combinations of unconditional decisions of the other two group members. For each of those three possible combinations you now select in the second row option A or option B – given the decisions of the other two group members. You have to choose an option for each possible combination of decisions of the other two group members. Thus, you have to indicate how you decide between option A and option B, if both of the other group members choose option A, if only one chooses option A and the other one chooses option B and if both choose option B. After having taken all three decisions on the screen, please press the "OK" button.

<u>**Payoff relevance of the decision table**</u>

As soon as all participants have taken their decisions, the computer randomly chooses one member in each group. For this **randomly chosen member** only his filled in **decision table** will be payoff-relevant. For the **other two group members,** who are not chosen, the **unconditional decision** is payoff-relevant. When you fill in your decision table, you obviously do not know whether you will be the group member who is randomly chosen. Thus, you have to make both types of decisions very carefully, as both could become relevant for you. We present two examples:

<u>Example 1:</u> Suppose **you are the randomly chosen group member, i.e. your decision table is payoff-relevant**. Thus, for the other two group members the unconditional decision is payoff-relevant. Suppose that one of the other two group members chose option A as unconditional decision, while the other one chose option B as unconditional decision.

If you indicated in your decision table to choose option A if one of the other group members chooses option A and the other one chooses option B, then you earn 32 Guilders in part III. The other group member, who chose option A, earns 32 Guilders as well. The group member who chose option B earns 12 Guilders in part III.

In contrast, if you selected in your decision table option B if one of the other group members chooses option A and the other one chooses option B, then you earn 24 Guilders. The group member who chose option A earns 44 Guilders. The group member, who also chose option B, earns 24 Guilders as well.

<u>Example 2:</u> Suppose **you have not been chosen**, so that for you and another group member the **unconditional decision is payoff-relevant**. We assume that your unconditional decision is option B and the one of the other group member is also option B. If the randomly chosen group member indicated in his decision table to choose option A, if the other two members chose option B, then you and the other group member, who chose option B, earn 24 Guilders each. The group member, who selected in the decision table option A, if the other two ones choose option B, earns 44 Guilders.

In contrast, if the randomly chosen group member selected in the decision table option B as well, if the other two members choose option B, then all three group members earn 36 points each.

page 10 of 14

You take your decisions only **once**. You will be informed about the result of part III at the end of the experiment, after finishing the questionnaires. Then you will get to know which decision is payoff-relevant for you (your unconditional decision or the decision table), the corresponding decisions of your group members and your earnings in part III.

Part IV

The general decision situation

Firs we will explain to you the general decision situation. Then you will find some sample exercises on the screen, which will help you to familiarize yourself with the decision situation. Afterwards you will learn the procedure of part IV in detail.

You are again a member of a randomly matched **three-person group**. Thus, your group consists of you and two other group members. You will at no time be told the identity of your two group members. The group formation with the other two participants is only valid in part IV. It is impossible to form a group with the same participants from former parts.

Part IV consists of **three stages.** In the first stage **you decide again between option A and option B**. In the second stage **you decide whether you want to reduce the individual first stage income of the other two group members or not** and, if so, by how much. This can be done by allocating deduction points. In the third stage **you can then decide, whether or by how much you want to reduce the income of those group members from whom you received deduction points in the second stage**. This can be done by allocating counter-points. In the following we describe part IV in detail.

<u>First Stage</u>
In the first stage of part IV you decide again between option A and option B. The income of each group member will be determined in the same way according to the following payoff matrix:

	Your two group members both choose Option A	**One group member chooses Option A, the other chooses Option B**	**Your two group members both choose Option B**
You choose Option A	20	32	44
You choose Option B	12	24	36

This means:
If all three group members choose option B, then all three group members receive 36 Guilders in the first stage of part IV and so on.

<u>Second Stage</u>
In the second stage you can **reduce the income of each other group member by allocating deduction points**. You can leave the **income also unchanged**. The other group members can reduce your income in the same way, if they want to.

Again the following rule applies: If you allocate deduction points to one of the other group members, **you will reduce his income by three times your distributed deduction points.** Likewise each distributed deduction point to you reduces your income by 3 Guilders. Please note that you can allocate **at most 10 deduction points per group member.**

Cost of received deduction points (in Guilders) = 3 * (sum of received deduction points)

If you allocate deduction points, you have to bear cost in Guilders as well. Each distributed deduction point costs you **one Guilder**.

Thus, your income at the end of the second stage will be calculated as follows:

Your income (in Guilders) at the end of the second stage =
= First stage income – 3 * (sum of received deduction points) – sum of distributed deduction points

Please note: **The income of one group member cannot be reduced further than to 0 Guilders by the other two group members**. This means that if the cost of received deduction points are larger than the group member's first stage income, his income will be reduced to 0 Guilders and not further. Nevertheless, even in this case the group member has to bear the cost of his distributed deduction points.

Since your income at the end of the second stage can be negative, possible losses will be covered by the lump sum of 20 Guilders.

Third Stage

Then you have the **possibility to redistribute points to those group members from whom you received deduction points in the second stage**. We will call these points **counter-points**. Please note that you can **allocate counter-points only to those group members from whom you received deduction points in the second stage**. Group members **with a negative income or an income of zero** are **neither able to distribute counter-points nor to receive any**. Again you can only distribute **at most 10 counter-points per group member**.

The income reduction based on received counter-points as well as the cost for distributing counter-points arise analogously to the second stage. This means: If you allocate counter-points to one of the group members, **you will reduce his income by three times your distributed counter-points.** Likewise, each counter-point distributed to you reduces your income by 3 Guilders. **If you distribute counter-points there will arise costs in Guilders for you**. Each distributed counter-point will cost you **one Guilder**. Thus, your income at the end of the third stage results as follows:

Your income (in Guilders) at the end of the third stage =
= Second stage income – 3 * (sum of received counter-points) – sum of distributed counter-points

For example, if your income at the end of the second stage amounts to 20 Guilders and you receive in the third stage altogether 3 counter-points of the other two group members and you distribute 1 counter-point to one group member, then your income in part IV will be calculated as 20 – 3 * 3 – 1 = 10 Guilders.

Please note: Analogously to the second stage **the income of one group member cannot be reduced further than to 0 Guilders by the other two group members**. This means, if the cost of the received counter-points are larger than the group member's second stage income, his income will be reduced to 0 Guilders and not further. Nevertheless, even in this case the group member has to bear the cost of his distributed counter-points.

Since your income at the end of the second stage can be negative, possible losses will be covered by the lump sum of 20 Guilders.

Control questions

Before continuing with the instruction, we ask you to complete the sample exercises on your screen. In case of questions please raise your hand. I will then come to your cubicle and answer your questions privately.

Part IV

Procedure

Part IV includes the decision situation as described above. You take your decision in part IV only **once**.

First stage
In the first stage all participants take their decision between option A and option B by selecting either "I choose option A" or "I choose option B" on the screen.

Second stage
You will be informed about the decision of the other two group members only at the end of the experiment after having finished the questionnaires. Therefore you have to determine in the second stage how many deduction points you want to allocate in each possible case. There are again four possible cases:
Case 1: The other two group members chose option A.
Case 2: The second group member chose option A and the third member chose option B.
Case 3: The second group member chose option B and the third member chose option A.
Case 4: The other two group members chose option B.

For all four possible cases you have to decide whether you want to allocate deduction points to the other group members and, if so, how many. So that you can take your decision, four input screens will arise again. You have to indicate a number between 0 and 10 in each input field.

Obviously, only one of the four possible cases actually applies. How you decided in this specific case will determine your payoff.

Third stage
In the third stage **you decide, whether or how many counter-points you want to allocate**. In the following we exemplarily show you how the screen will be built up:

Part IV – Third stage

Please decide whether you want to distribute counter-points and, if so, how many:

	You	Second group member	Third group member
First stage decision	Option A	Option A	Option B
First stage income	32	32	12
Your distributed deduction points in the second stage	---	3	1
Your received deduction points in the second stage	---	1	2
Your decision in the third stage	---		

OK

- The **second row** indicates the decisions between option A and option B in the first stage. We exemplarily assume that you chose option A, the second group member chose option A and the third group member chose option B.
- The **third row** indicates the income of each group member that results from the first stage decisions. Under the heading “You” you find your income and in the third and fourth column you see the incomes of the other two group members.
- In the **fourth row** (“Your distributed deduction points in the second stage”) you again see the deduction points that you allocated to the second and the third group member in the second stage.
- The **fifth row** (“Your received deduction points in the second stage”) shows the amount of deduction points that you received from the second and the third group member in the second stage.
- In the **sixth row** (“Your decision in the third stage”) you now indicate, how many counter-points you want to distribute to each of the other two group members by entering in the sixth row a **number between 0 and 10** in each input field. Please note: You can only allocate counter-points to those group members from whom you received deduction points in the second stage. Furthermore you can only distribute counter-points to another group member if your income and the income of the other member are positive. These three conditions are fulfilled if an input field appears under the header of the other two group members in the fifth row. Otherwise you cannot allocate counter-points to the particular group member and no input field will appear on the screen.

At the end of the experiment you will be informed about the amount of received counter-points as well as about your income in part IV.

Do you have questions? If so, I will come to your cubicle and answer them privately.

Five Questionnaires

Finally you will have to answer five short questionnaires. These questionnaires are an important part of our research. Please answer honestly and spontaneously. Of course, your answers are anonymous.
Partly you can give free answers. In this case you will see a blue chat box on the screen. If you write your answer in the blue box and press afterwards “Enter” your response will be saved.
On the last page we will ask you to indicate your personal code, which will be explained to you on the PC and which serves to make your data anonymous.
For filling in the questionnaires you receive in total 10 Guilders, or 2 Euros.

Feedback, Clarification and Payoff

You now receive information about the decisions of your respective group members in the parts I to IV and about your income.
Finally we need your consent that we are allowed to anonymously save and analyse your data. Therefore we inform you about the purpose of this experiment.

Payoff
At the end you individually receive your income in cash. This is made up of:
Your income from part I
+ your income from part II
+ your income from part III
+ your income from part IV
+ 2 euros for the completion of the questionnaires
+ the rest of the 4 euros for your punctual appearance.

page 14 of 14

A.3.2 Control questions for the QG+P

Control question 1:
You are a member of a three-person group. Imagine group member C chooses in the second stage the alternative ranking. What will be your income (in Guilders) at the end of the second stage, if …

a) … you come out as group member A of the first stage?
b) … you come out as group member B of the first stage?
c) What will be the income of group member C?

Control question 2:
You are a member of a three-person group. Imagine you are group member A and group member C chooses in the second stage the alternative ranking. What will be your income (in Guilders) at the end of the third stage, if …

a) … you distribute 0 points to group member C in the third stage?
b) … you distribute 3 points to group member C in the third stage?
c) … you distribute 8 points to group member C in the third stage?
d) In all these cases what will be the income of group member B at the end of the third stage, if group member C has chosen the alternative ranking?

Control question 3:
You are a member of a three-person group. Imagine you are group member C and you choose the alternative ranking in the second stage. What will be your income (in Guilders) at the end of the third stage, if …

a) … group member A distributes 0 points to you in the third stage?
b) … group member A distributes 3 points to you in the third stage?
c) … group member A distributes 8 points to you in the third stage?

A.3.3 Post-experimental questionnaire on scenarios

Numbers in square brackets indicate the average agreement with the statements on a 7-point Likert scale from "0 - completely disagree" to "6 - completely agree" respectively. Numbers in round brackets refer to the standard deviations each. If necessary, the number of observations is indicated respectively. The numbers regarding the emotions anger and fear indicate the difference to the corresponding baseline emotions respectively.

Group task scenario

Imagine you are a student and attend a university course. In this course students are divided into groups of three to accomplish in their group a student project until the end of the semester. Your group splits the project into three parts, whereby all group members agree to accomplish one part each. You arrange to meet shortly before the deadline to merge the three parts. On the day of the planned meeting, one of your group members announces per email without further explanation that he did not accomplish his part. If his part is missing, the whole group will not pass the course. You now accidentally meet this group member on the corridor of the university building.

- Please rate your emotions towards this person.
  Anger: [3.742 (2.143)]
  Fear: [0.015 (1.130)]
- Would you confront the group member? [5.636 (0.871)]
  - *In case of no norm enforcement*: Why would you not confront the group member?
    - In order to avoid tension. [--]
    - Because I do not find it especially badly if somebody does not accomplish his part of the project. [--]
    - Because the group member probably has a good reason for not having accomplished her part. [--]
    - Because I prefer the third person in the group to confront the group member. [--]
    - Other: free text.
  - *In case of norm enforcement*: Why would you confront the group member?
    - In order to make the group member accomplish his part in the future. [3.905 (2.006), 63 obs.]
    - Because it is wrong not to accomplish one's part of the project. [5.476 (1.030), 63 obs.]
    - Other: free text.

Littering scenario

Imagine you are standing at the subway station with some other pedestrians and you are waiting for the subway. Another pedestrian throws his litter on the ground when he passes you.

- Please rate your emotions towards this person.
  Anger: [2.591 (2.267)]
  Fear: [-0.106 (1.191)]
- Would you call on the pedestrian to bin his litter? [2.758 (1.815)]
  - *In case of no norm enforcement*: Why would you not call on the pedestrian?
    - In order to avoid tension. [4.423 (1.474), 26 obs.]
    - Because I don't find it especially badly, if somebody litters. [1.308 (1.644), 26 obs.]
    - Because I prefer other persons to confront the pedestrian. [2.500 (2.140), 26 obs.]
    - Other: free text.
  - *In case of norm enforcement*: Why would you call on the pedestrian?
    - In order to make the pedestrian binning his litter in the future. [4.227 (2.045), 22 obs.]
    - Because it is wrong to litter. [5.773 (0.429), 22 obs.]
    - Other: free text.

Queuing scenario

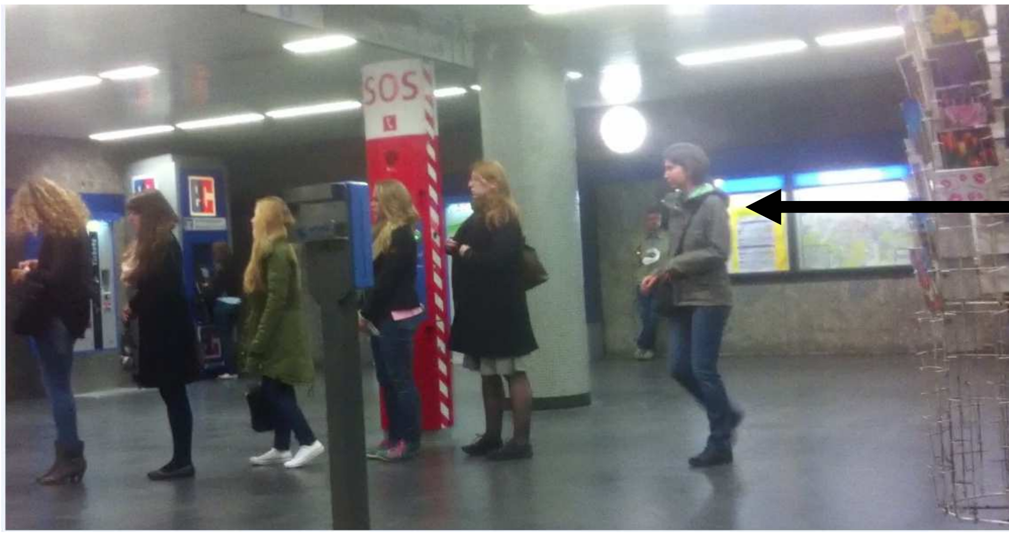
Imagine you are standing in a waiting line. Someone cuts in line directly in front of you.

- Please rate your emotions towards this person.
  Anger: [2.985 (2.072)]
  Fear: [-0.333 (1.072)]
- Would you call on this person to go to the end of the line? [4.394 (1.709)]
  - *In case of no norm enforcement*: Why would you not call on the person?
    - In order to avoid tension. [--]
    - Because I do not find it especially badly if somebody cuts in line. [--]
    - Because the person presumably accidentally cut in line. [--]
    - Because I prefer other persons waiting in the line to confront the person. [--]
    - Other: free text.
  - *In case of norm enforcement*: Why would you call on the person?
    - In order to make the other person going to the end of the line in the future. [3.688 (1.980), 48 obs.]
    - Because it is wrong to cut in line. [5.438 (1.319), 48 obs.]
    - Other: free text.

What do you think:

- Out of 100 persons how many know the social norm to line up at the end of the line? [97.0 (4.861)]
- Out of 100 persons how many confront another person who cuts in line directly in front of them? [37.6 (26.561)]

A.3.4 Photos of the field experiment



Norm violator

Norm violator

Norm violator

Table A.3.1 - Correlation matrix: Decisions made in the lab.

	PDG+P		PDG+PCP		QG+P		Coop. type	Scenarios		
	Coop. decision	Sanction. of defectors	Coop. decision	Sanction. of defectors	Queue-jumping	Sanction.		N. enfor. Group task sce.	N. enfor. Littering scenario	N. enfor. Queuing scenario
PDG+P										
Coop. decision		0.011	0.498***	-0.197	-0.017	-0.144	0.368**	0.135	0.033	-0.085
Sanctioning of defectors	0.011		0.089	0.737***	0.055	0.645***	0.036	-0.006	0.003	0.055
PDG+PCP										
Coop. decision	0.498***	0.089		0.154	-0.132	-0.009	0.245*	0.235*	-0.090	0.021
Sanctioning of defectors	-0.197	0.737***	0.154		0.091	0.541***	0.097	0.041	-0.013	0.193
QG+P										
Queue-jump.	-0.017	0.055	-0.132	0.091		-0.104	-0.310**	0.073	-0.007	-0.011
Sanctioning	-0.144	0.645***	-0.009	0.541***	-0.104		0.069	0.034	0.090	0.153
Coop. type	0.368**	0.036	0.245*	0.097	-0.310**	0.069		0.020	-0.037	-0.102
Scenarios										
N. enf.: Group task scenario	0.135	-0.006	0.235*	0.041	0.073	0.034	0.020		0.192	0.230*
N. enf.:Littering scenario	0.033	0.003	-0.090	-0.013	-0.007	0.090	-0.037	0.192		0.212*
N enf.: Queuing scenario	-0.085	0.055	0.021	0.193	-0.011	0.153	-0.102	0.230*	0.212*	

Notes: This table outlines the Spearman rank-order correlation coefficients between laboratory decisions. An independent observation in the table is one out of the 66 subjects participating in the lab . Kendall correlation coefficients yield similar results.
*, **, *** denote significance at the 10%, 5% and 1% level respectively.

Table A.3.2 - Probit regressions of norm enforcement (willingness to sanction included).

Dep. var.: Norm enforcement in the field	(I)	(II)	(III)
PDG+P:			
Cooperation decision	0.018 (0.132)		
Willingness to sanction defectors	0.062 (0.134)		
PDG+PCP:			
Cooperation decision		0.261* (0.138)	
Willingness to sanction defectors		$0.224^{§}$ (0.159)	
QG+P:			
Queue-jumping			-0.035 (0.134)
Willingness to sanction			0.035 (0.134)
N	45	45	45
Wald χ^2	0.24	6.81**	0.14
Log pseudo-likelihood	-25.985	-22.882	-26.024

Notes: This table outlines the results of probit regressions of norm enforcement in the field. We report the marginal effects at the mean. An observation in the regression is one of the 45 subjects participating both in the lab and the field. OLS regressions yield similar results. Standard errors are presented in parentheses. *, **, *** denote significance at the 10%, 5% and 1% level respectively. § denotes p = 0.135.

REFERENCES

ABBINK, K., J. BRANDTS, B. HERRMANN, AND H. ORZEN (2010): Intergroup conflict and intra-group punishment in an experimental contest game. *American Economic Review* 100, 420-447.

ABELE, S., G. STRASSER, AND C. CHARTIER (2010): Conflict and coordination in the provision of public goods: A conceptual analysis of continuous and step-level games. *Personality and Social Psychology Review* 14, 385-401.

AI, C. AND E. C. NORTON (2003): Interaction terms in logit and probit models. *Economics Letters* 80, 123-129.

AJZEN, I. (1991): The theory of planned behaviour. *Organizational Behavior and Human Decision Processes* 50, 179-211.

AJZEN, I. AND M. FISHBEIN (2011): Predicting and changing behavior: The reasoned action approach. Taylor & Francis Group.

AMBRUS, A. AND B. GREINER (2012): Imperfect public monitoring with costly punishment – An experimental study. *American Economic Review* 102, 3317-3332.

ANDERSON, C. M. AND L. PUTTERMAN (2006): Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism. *Games and Economic Behavior* 54, 1-24.

ANDREONI, J., W. T. HARBAUGH, AND L. VESTERLUND (2003): The carrot or the stick: Rewards, punishments and cooperation. *American Economic Review* 93, 893-902.

AUBE, C. AND V. ROUSSEAU (2005): Team goal commitment and team effectiveness: The role of task interdependence and supportive behaviours. *Group Dynamics: Theory, Research, and Practice* 9, 189-204.

BAGNOLI, M. AND M. MCKEE (1991): Voluntary contributions games: Efficient private provision of public goods. *Economic Inquiry* 29, 351-366.

BALAFOUTAS, L., R. KERSCHBAMER, M. KOCHER, AND M. SUTTER (2014): Revealed distributional differences: Individuals versus teams. *Journal of Economic Behavior and Organization*, in press.

BALAFOUTAS, L. AND N. NIKIFORAKIS (2012): Norm enforcement in the city: A natural field experiment. *European Economic Review* 56, 1773-1785.

BALAFOUTAS, L., N. NIKIFORIKS, AND B. ROCKENBACH (2014): Direct vs. indirect punishment among strangers in the field. mimeo.

BARAN, N., P. SAPIENZA AND L. ZINGALES (2010): Can we infer social preferences from the lab? Evidence from the trust game. NBER Working Paper No. 15654.

BENZ, M. AND S. MEIER (2008): Do people behave in experiments as in the field? – Evidence from donations. *Experimental Economics* 11, 268-281.

BICCHIERI, C. (2006): The grammar of society: The nature and dynamics of social norms. Cambridge University Press.

BLINDER, A. S. AND J. MORGAN (2005): Are two heads better than one? Monetary policy by committee. *Journal of Money, Credit and Banking* 37, 789-812.

BLUMENTHAL, M., C. CHRISTIAN, AND J. SLEMROD (2001): Do normative appeals affect tax compliance? Evidence from a controlled experiment in Minnesota. *National Tax Journal* 54,125-138.

BOCHET, O., T. PAGE, AND L. PUTTERMAN (2006): Communication and punishment in voluntary contribution experiments. *Journal of Economic Behavior & Organization* 60, 11-26.

BOCHET, O. AND L. PUTTERMAN (2009): Not just babble: Opening the black box of communication in a voluntary contribution experiment. *European Economic Review* 53, 309-326.

BOLTEN, G. E. AND A. OCKENFELS (2000): ERC: A theory of equity, reciprocity, and competition. *American Economic Review* 90, 166-193.

BORKENAU, P. AND F. OSTENDORF (2008): NEO-FFI: NEO-Fünf-Faktoren-Inventar nach Costa und McCrae, Manual. Hogrefe.

BORNSTEIN, G., T. KUGLER, AND A. ZIEGELMEYER (2004): Individual and group behavior in the centipede game: Are groups more ‘rational’ players? *Journal of Experimental Social Psychology* 40, 599-605.

BORNSTEIN, G. AND I. YANIV (1998): Individual and group behavior in the ultimatum game: Are groups more ’rational’ players? *Experimental Economics* 1, 101-108.

BOSMAN, R., H. HENNIG-SCHMIDT, AND F. VAN WINDEN (2006): Exploring group decision making in a power-to-take experiment. *Experimental Economics* 9, 35-51.

BOTELHO, A., G. W. HARRISON, L. M. C. PINTO, AND E. E. RUTSTRÖM (2009): Testing static game theory with dynamic experiments: A case study of public goods. *Games and Economic Behavior* 67, 253-265.

BOWLES, S. AND H. GINTIS (2004): The evolution of strong reciprocity: Cooperation in heterogeneous groups. *Theoretical Population Biology* 65, 17-28.

BRANDTS, J. AND G. CHARNESS (2000): Hot vs. cold: Sequential responses and preference stability in experimental games. *Experimental Economics* 2, 227-238.

BRANDTS, J. AND G. CHARNESS (2011): The strategy versus the direct-response method: A first survey of experimental comparison. *Experimental Economics* 14, 375-398.

BURTON-CHELLEW, M. N., A. ROSS-GILLESPIE, AND S. A. WEST (2010): Cooperation in humans: Competition between groups and proximate emotions. *Evolution and Human Behavior* 31, 104-108.

CADSBY, C. AND E. MAYNES (1999): Voluntary provision of threshold public goods with continuous contributions: Experimental evidence. *Journal of Public Economics* 71, 53-73.

CAMERER, C. F. (2003): Behavioral game theory: Experiments in strategic interaction. Princeton University Press.

CAMERER, C. F. (2011): The promise and success of lab-field generalizability in experimental economics: A critical reply to Levitt and List. Working Paper. Retrieved from Social Science Research Network.

CAMERON, A. C. AND P. K. TRIVEDI (2005): Microeconometrics: Methods and applications. Cambridge University Press.

CARPENTER, J. (2007): The demand for punishment. *Journal of Economic Behavior and Organization* 62, 522-542.

CARPENTER, J. AND C. MYERS (2010): Why volunteer? Evidence on the role of altruism, image, and incentives. *Journal of Public Economics* 94, 911-920.

CARPENTER, J. AND E. SEKI (2011): Do social preferences increase productivity? Field experimental evidence from fishermen in Toyama Bay. *Economic Inquiry* 49, 612-630.

CASON, T. N. AND V.-L. MUI (1997): A laboratory study of group polarisation in the team dictator game. *Economic Journal* 107, 1465-1483.

CHARNESS, G. AND M. RABIN (2002): Understanding social preferences with simple tests. *Quarterly Journal of Economics* 117, 817-869.

CHARNESS, G., L. RIGOTTI, AND A. RUSTICHINI (2007): Individual behavior and group membership. *American Economic Review* 97, 1340-1352.

CHARNESS, G. AND M. SUTTER (2012): Teams make better self-interested decisions. *Journal of Economic Perspectives* 26, 157-176.

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

CINYABUGUMA, M., T. PAGE, AND L. PUTTERMAN (2005): Cooperation under the threat of expulsion in a public goods experiment. *Journal of Public Economics* 89, 1421-1435.

COOPER, D. J. AND J. H. KAGEL (2005): Are two heads better than one? Team versus individual play in signaling games. *American Economic Review* 95, 477-509.

CORTINA, J. (1993): What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology* 78, 98-104.

COX, J. C. (2002): Trust, reciprocity, and other-regarding preferences: Groups vs. individuals and males vs. females. *In* R. Zwick and A. Rapoport (eds.), Experimental Business Research, 331-350. Kluwer Academic Publishers.

COX, J. C. AND S. C. HAYNE (2006): Barking up the right tree: Are small groups rational agents? *Experimental Economics* 9, 209-222.

CRAGG, J. G. (1971): Some statistical models for limited dependent variables with application to the demand for durables goods. *Econometrica* 39, 829-844.

CROSON, R. AND M. MARKS (2000): Step returns in threshold public goods: A meta- and experimental analysis. *Experimental Economics* 2, 239-259.

DARIEL, A. AND A. RIEDL (2010): Different social preferences, similar selfish actions: The case of public goods game and gift exchange game. mimeo.

DENANT-BOEMONT, L., D. MASCLET, AND C. NOUSSAIR (2007): Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. *Economic Theory* 33, 145-167.

DE OLIVEIRA, A., R. CROSON, AND C. ECKEL (2011): The giving type: Identifying donors. *Journal of Public Economics* 95, 428-435.

EGAS, M. AND A. RIEDL (2008): The economics of altruistic punishment and the maintenance of cooperation. *Proceedings of the Royal Society – Biological Sciences* 275, 871-878.

ENGLMAIER, F. AND G. GEBHARDT (2013): Free-riding in the lab and in the field. Working Paper.

ENGELMANN, D. AND N. NIKIFORAKIS (2014): In the long run we are all dead: On the benefits of peer punishment in rich environments. *Social Choice and Welfare*, in press.

EPSTEIN, S. AND E. J. O'BRIAN (1985): The person-situation debate in historical and current perspective. *Psychological Bulletin* 98, 513-537.

ERTAN, A., T. PAGE, AND L. PUTTERMAN (2009): Who to punish? Individual decisions and majority rule in mitigating the free rider problem. *European Economic Review* 53, 495-511.

FALK, A., E. FEHR, AND U. FISCHBACHER (2005): The driving forces behind informal sanctions. *Econometrica* 73, 2017-2030.

FEHR, E. AND S. GÄCHTER (2000): Cooperation and punishment in public goods experiments. *American Economic Review* 90, 980-994.

FEHR, E. AND S. GÄCHTER (2002): Altruistic punishment in humans. *Nature* 415, 137-140.

FEHR, E. AND A. LEIBBRANDT (2011): A field study on cooperativeness and impatience in the tragedy of the commons. *Journal of Public Economics* 95, 1144-1155.

FEHR, E. AND K. SCHMIDT (1999): A theory of fairness, competition and cooperation. *Quarterly Journal of Economics* 114, 817-868.

FEHR, E. AND K. SCHMIDT (2003): Theories of fairness and reciprocity: Evidence and economic applications. *In* M. Dewatripont et al. (eds.), Advances in Economics and Econometrics. Eighth World Congress of the Econometric Society, Vol 1. Cambridge University Press.

FELLNER, G., R. SAUSGRUBER, AND C. TRAXLER (2013): Testing enforcement strategies in the field: Threat, moral appeal and social information. *Journal of the European Economic Association* 11, 634-660.

FERI, F., B. IRLENBUSCH, AND M. SUTTER (2010): Efficiency gains from team-based coordination – Large scale experimental evidence. *American Economic Review* 100, 1892-1912.

FISCHBACHER, U. (2007): Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10, 171-178.

FISCHBACHER, U. AND S. GÄCHTER (2010): Social preferences, beliefs, and the dynamics of free riding in the public good experiments. *American Economic Review* 100, 541-556.

FISCHBACHER, U., S. GÄCHTER, AND E. FEHR (2001): Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters* 71, 397-404.

FISCHBACHER, U., W. GÜTH, AND M. V. LEVATI (2011): Crossing the point of no return: A public goods experiment. Jena Economic Research Papers No. 2011-059.

FRANZEN, A. AND S. POINTNER (2013): The external validity of giving in the dictator game. A field experiment using the misdirected letter technique. *Experimental Economics* 16, 155-169.

FREYTAG, A., W. GÜTH, H. KOPPEL, AND L. WANGLER (2014): Is regulation by milestones efficiency enhancing? An experimental study of environmental protection. *European Journal of Political Economy* 33, 71-84.

FUDENBERG, D. AND P. A. PATHAK (2010): Unobserved punishment supports cooperation. *Journal of Public Economics* 94, 78-86.

GÄCHTER, S., E. RENNER, AND M. SEFTON (2008): The long-run benefits of punishment. *Science* 322, 1510.

GALINKSKY, A. D. AND T. MUSSWEILER (2001): First offers as anchors: The role of perspective-taking and negotiator focus. *Journal of Personality and Social Psychology* 81, 657-669.

GILLET, J., A. SCHRAM, AND J. SONNEMANS (2009): The tragedy of the commons revisited: The importance of group decision-making. *Journal of Public Economics* 93, 785-797.

GINTIS, H., J. HENRICH, S. BOWLES, R. BOYD, AND E. FEHR (2008): Strong reciprocity and the roots of human morality. *Social Justice Research* 21, 241-253.

GRECHENIG, K., A. NICKLISCH, AND C. THÖNI (2010): Punishment despite reasonable doubt – A public goods experiment with uncertainty over contributions. *Journal of Empirical Legal Studies* 7, 847-867.

GREINER, B. (2004): An online recruitment system for economic experiments. *In* K. Kremer and V. Macho (eds.), Forschung und wissenschaftliches Rechnen 2003. GWDG Bericht 63, 79-93. Gesellschaft für Wissenschaftliche Datenverarbeitung.

GREEN, S. B., LISSITZ, R. W., AND S. A. MULAIK (1977): Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement* 37, 827-838.

GUALA, F. (2012): Reciprocity: Weak or strong? What punishment experiments do (and do not) demonstrate. *Behavioral and Brain Sciences* 35, 1-59.

GÜRERK, Ö., B. ILRENBUSCH, AND B. ROCKENBACH (2006): The competitive advantage of sanctioning institutions. *Science* 312, 108-111.

GURVEN, M. AND J. WINKING (2008): Collective action in action: Prosocial behavior in and out of the laboratory. *American Anthropologist* 110, 179-190.

HARRISON, G. W. AND J. A. LIST (2004): Field experiments. *Journal of Economic Literature* 42, 1009-1055.

HECKMAN, J. J. (1979): Sample selection as a specification error. *Econometrica* 47, 153-161.

HENRICH, J., R. MCELREATH, A. BARR, J. ENSMINGER, C. BARRETT, A. BOLYANATZ, J. C. CARDENAS, M. GURVEN, E. GWAKO, N. HENRICH, C. LESOROGOL, F. MARLOWE, D. TRACER, AND J. ZIKER (2006): Costly punishment across human societies. *Science* 312, 1767-1770.

HERRMANN, B., C. THÖNI, AND S. GÄCHTER (2008): Antisocial punishment across societies. *Science* 319, 1362-1367.

HOYLE, R. H., R. L. PINKLEY, AND C. A. INSKO (1989): Perceptions of behavior: Evidence of differing expectations for interpersonal and intergroup interactions. *Personality and Social Psychology Bulletin* 15, 365-376.

INSKO, C., R. HOYLE, R. PINKLEY, G.-Y. HONG, R. SLIM, B. DALTON, Y.-H. LIN, P. RUFFIN, G. DARDIS, P. BERNTHAL, AND J. SCHOPLER (1988): Individual-group discontinuity: The role of a consensus rule. *Journal of Experimental Social Psychology* 24, 505-519.

INSKO, C., R. PINKLEY, R. HOYLE, B. DALTON, G. HONG, R. SLIM, P. LANDRY, B. HOLTON, P. RUFFIN, AND J. THIBAUT (1987a): Individual versus group discontinuity: The role of intergroup contact. *Journal of Experimental Social Psychology* 23, 250-267.

INSKO, C.A., R. L. PINKLEY, K. HARRING, B. HOLTON, G. HONG, D. S. KRAMS, R. H. HOYLE, AND J. THIBAUT (1987b): Minimal conditions for real groups: Mere categorization or competitive between category behaviour. *Representative Research in Social Psychology* 17, 5-36.

INSKO, C. AND J. SCHOPLER (1987): Categorization, competition and collectivity. *In* C. Hendrick (ed.), Review of Personality and Social Psychology: Group Processes, 8, 213-251. Sage.

INSKO, C., J. SCHOPLER, R. H. HOYLE, G. J. DARDIS, AND K. A. GRAETZ (1990): Individual-group discontinuity as a function of fear and greed. *Journal of Personality and Social Psychology* 58, 68-79.

ISAAC, R., K. MCCUE, AND C. PLOTT (1985): Public goods provision in an experimental environment. *Journal of Public Economics* 26, 51-74.

ISSAC, R., D. SCHMIDTZ, AND J. WALKER (1989): The assurance problem in a laboratory market. *Public Choice* 62, 217-236.

ISAAC, R. AND J. WALKER (1988): Group size effects in public goods provision: The voluntary contributions mechanism. *Quarterly Journal of Economics* 103, 179-199.

KARLAN, D. S. (2005): Experimental economics to measure social capital and predict financial decisions. *American Economic Review* 95, 1688-1699.

KLEVEN, H., M. KNUDSEN, C. KREINER, S. PEDERSEN, AND E. SAEZ (2011): Unwilling or unable to cheat? Evidence from a randomized tax audit experiment in Denmark. *Econometrica* 79, 651-692.

KOCHER, M. G., STRAUSS, S., AND M. SUTTER (2006): Individual or team decision-making – causes and consequences of self-selection. *Games and Economic Behaviour* 56, 259-270.

KOCHER, M. G. AND M. SUTTER (2005): The decision-maker matters: Individual versus group behaviour in experimental beauty-contest games. *Economic Journal* 115, 200-223.

KOCHER, M. G. AND M. SUTTER (2007): Individual versus group behavior and the role of the decision making process in gift-exchange experiments. *Empirica* 34, 63-88.

KUGLER, T. AND G. BORNSTEIN (2013): Social dilemmas between individuals and groups. *Organizational Behavior and Human Decision Processes* 120, 191-205.

KUGLER, T., E. E. KAUSEL, AND M. G. KOCHER (2012): Are groups more rational than individuals? A review of interactive decision making in groups. *Wiley Interdisciplinary Reviews: Cognitive Science* 3, 471–482.

KUGLER, T., M. G. KOCHER, M. SUTTER, AND G. BORNSTEIN (2007): Trust between individuals and groups: Groups are less trusting than individuals but just as trustworthy. *Journal of Economic Psychology* 28, 646-657.

LANDIS, J. R. AND G. G. KOCH (1977): The measurement of observer agreement for categorical data. *Biometrics* 33, 159-174.

LEVITT, S. D. AND J. A. LIST (2007a): What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives* 21, 153-174.

LEVITT, S. D. AND J. A. LIST (2007b): Viewpoint: On the generalizability of lab behaviour to the field. *Canadian Journal of Economics* 40, 347-370.

LEVITT, S. D., J. A. LIST, AND D. REILEY (2010): What happens in the field stays in the field: Professionals do not play minimax in laboratory experiments. *Econometrica* 78, 1413-1434.

LIEBRAND, W. B. G. (1984): The effects of social motives, communication and group size on behaviour in an n-person multi-stage mixed-motive game. *European Journal of Social Psychology* 14, 239-264.

LIEBRAND, W. B. G. AND C. 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.

LIST, J. (2006): The bahvioralist meets the market: Measuring social preferences and reputation effects in actual transactions. *Journal of Political Economy* 114, 1-37.

LUHAN, W., M. G. KOCHER, AND M. SUTTER (2009): Group polarization in the team dictator game reconsidered. *Experimental Economics* 12, 26-41.

MARLOWE, F. W., J. C. BERBESQUE, A. BARR, C. BARRETT, A. BOLYANATZ, J. C. CARDENAS, J. ENSMINGER, M. GURVEN, E. GWAKO, J. HENRICH, N. HENRICH, C. LESOROGOL, R. MCELREATH, AND D. TRACER (2008): More 'altruistic' punishment in larger societies. *Proceedings of the Royal Society – Biological Sciences* 275, 587-590.

MARKS, M. AND R. CROSON (1998): Alternative rebate rules in the provision of a threshold public good: An experimental investigation. *Journal of Public Economics* 67, 195-220.

MARWELL, G. AND R. E. AMES (1981): Economists free ride, does anyone else?: Experiments on the provision of public goods. *Journal of Public Economics* 15, 295-310.

MASCLET, D., C. NOUSSAIR, S. TUCKER, AND M. VILLEVAL (2003): Monetary and nonmonetary punishment in the voluntary contributions mechanism. *American Economic Review* 93, 366-380.

MASCLET, D., C. N. NOUSSAIR, AND M.-C. VILLEVAL (2013): Threat and punishment in public good experiments. *Economic Inquiry* 51, 1421-1441.

MCDOWELL, A. (2003): From the help desk: hurdle models. *The Stata Journal* 3, 178-184.

MILGRAM, S., H. J. LIBERTY, R. TOLEDO, AND J. WACKENHUT (1986): Response to intrusion into waiting lines. *Journal of Personality and Social Psychology* 51, 683-689.

MISCHEL, W (1968): Personality and assessment. Wiley.

MÜLLER, W. AND F. TAN (2013): Who acts more like a game theorist? Group and individual play in a sequential market game and the effect of the time horizon. *Games and Economic Behavior* 82, 658-674.

NIEDERLE, M. AND L. VESTERLUND (2007): Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics* 122, 1067-1101.

NIKIFORAKIS, N. (2008): Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics* 92, 91-112.

NIKIFORAKIS, N. (2010): Feedback, punishment and cooperation in public good experiments. *Games and Economic Behavior* 68, 689-702.

NIKIFORAKIS, N. AND D. ENGELMANN (2011): Altruistic punishment and the threat of feuds. *Journal of Economic Behavior and Organization* 78, 319-332.

NIKIFORAKIS, N. AND H. NORMANN (2008): A comparative statics analysis of punishment in public good experiments. *Experimental Economics* 11, 358-369.

NIKIFORAKIS, N., C. NOUSSAIR, AND T. WILKENING (2012): Normative conflict and feuds: The limits of self-enforcement. *Journal of Public Economics* 96, 797-807.

NOUSSAIR, C., D. VAN SOEST AND J. STOOP (2011): Punishment, reward, and cooperation in a framed field experiment. MPRA Paper No. 34067, University Library of Munich, Germany.

OBERHOLZER-GEE, F. (2006): A market for time. Fairness and efficiency in waiting lines. *Kyklos* 59, 427-440.

OSTROM, E., J. WALKER, AND R. GARDNER (1992): Covenants with and without a sword: self-governance is possible. *American Political Science Review* 86, 404-417.

PAGE, T., L. PUTTERMAN, AND B. UNEL (2005): Voluntary association in public goods experiments: Reciprocity, mimicry, and efficiency. *Economic Journal* 115, 1032-1053.

PALACIOS-HUERTA, I. AND O. VOLIJ (2008): Experientia docet: Professionals play minimax in laboratory experiments. *Econometrica* 76, 71-115.

PALACIOS-HUERTA, I. AND O. VOLIJ (2009): Field centipedes. *American Economic Review* 99, 1619-1635.

PUURTINEN, M. AND T. MAPPES (2009): Between-group competition and human cooperation. *Proceedings of the Royal Society - Biological Sciences* 276, 355-360.

RABIN, M. (1993): Incorporating fairness into game theory and economics. *American Economic Review* 83, 1281-1302.

REGE, M. AND K. TELLE (2004): The impact of social approval and framing on cooperation in public good situations. *Journal of Public Economics* 88, 1625-1644.

RONDEAU, D., G. POE, AND W. SCHULZE (2005): VCM or PPM? A comparison of the performance of two voluntary public goods mechanisms. *Journal of Public Economics* 89, 1581-1592.

ROSS, L. AND R. E. NISBETT (1991): The person and the situation: Perspectives of social psychology. McGraw-Hill.

RUSTAGI, D., S. ENGEL, AND M. KOSFELD (2010): Conditional cooperation and costly monitoring explain success in forest commons management. *Science* 320, 961-965.

SCHMITT, B. H., L. DUBE, AND F. LECLERC (1992): Intrusions into waiting lines: Does the queue constitute a social system? *Journal of Personality and Social Psychology* 63, 806-815.

SCHOPLER, J. AND C. A. INSKO (1992): The discontinuity effect in interpersonal and intergroup relations: Generality and mediation. *In* W. Strobe and M. Hewstone (eds.), European Review of Social Psychology, 3, 121–151. Wiley.

SCHOPLER, J., C. INSKO, J. WIESELQUIST, M. PEMBERTON, B. WITCHER, R. KOZAR, C. RODDENBERRY, AND T. WILDSCHUT (2001): When groups are more competitive than individuals: The domain of the discontinuity effect. *Journal of Personality and Social Psychology* 80, 632-644.

SCHOPLER, J., C. A. INSKO, K. A. GRAETZ, S. M. DRIGOTAS, AND V. A. SMITH (1991): The generality of the individual-group discontinuity effect: Variations in positivity-negativity of outcomes, players' relative power, and magnitude of outcomes. *Personality and Social Psychology Bulletin* 17, 612-624.

SEFTON, M., R. SHUPP, AND J. WALKER (2007): The effect of rewards and sanctions in provision of public goods. *Economic Inquiry* 45, 671-690.

SELTEN, R. (1967): Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments. *In* H. Sauermann (ed.), Beiträge zur experimentellen Wirtschaftsforschung, 136-168. J.C.B. Mohr (Paul Siebeck).

SENECAL, J., T. M. LOUGHEAD, AND G. A. BLOOM (2008): A season-long team-building intervention: Examining the effect of team goal setting on cohesion. *Journal of Sport and Exercise Psychology* 30, 186-199.

SLEMROD, J., M. BLUMENTHAL, AND C. CHRISTIAN (2001): Taxpayer response to an increased probability of audit: Evidence from a controlled experiment in Minnesota. *Journal of Public Economics* 79, 483-483.

SOUSA, S. (2010): Cooperation and punishment under uncertain enforcement. CeDEx Discussion Paper Series No. 2010-06.

SPENCER, M., S. SWALLOW, J. SHOGREN, AND J. LIST (2009): Rebate rules in threshold public good provision. *Journal of Public Economics* 93, 798-806.

STOOP, J., C. N. NOUSSAIR, AND D. VAN SOEST (2012): From the lab to the field: Cooperation among fishermen. *Journal of Political Economy* 120, 1027-1056.

STRASSER, S. (2012): Cooperation, competition and preference heterogeneity. mimeo.

SULEIMAN, R. AND A. RAPOPORT (1992): Provision of step-level public goods with continuous contribution. *Journal of Behavioral Decision Making* 5, 133-153.

SUTTER, M., S. HAIGNER, AND M. KOCHER (2010): Choosing the carrot or the stick? Endogenous institutional choice in social dilemma situations. *Review of Economic Studies* 77, 1540-1566.

SUTTER, M., M. KOCHER, AND S. STRAUSS (2009): Individuals and teams in auctions. *Oxford Economic Papers* 61, 380-394.

TAN, J. H. W. AND F. BOLLE (2007): Team competition and the public goods game. *Economics Letters* 96, 133-139.

TRAXLER, C. AND J. WINTER (2012): Survey evidence on conditional norm enforcement. *European Journal of Political Economy*, 28, 390-398.

TVERSKY, A. AND D. KAHNEMAN (1974): Judgment under uncertainty: Heuristics and biases. *Science* 185, 1124-1130.

VAN DEN BERGH, J. C. AND J. M. GOWDY (2009): A group selection perspective on economic behavior, institutions and organizations. *Journal of Economic Behavior and Organization* 72, 1-20.

WILDSCHUT, T. AND C. INSKO (2007): Explanations of interindividual-intergroup discontinuity: A review of the evidence. *European Review of Social Psychology* 18, 175-211.

WILDSCHUT, T., C. A. INSKO, AND B. PINTER (2007): Interindividual-intergroup discontinuity as a joint function of acting as a group and interacting with a group. *European Journal of Social Psychology* 37, 390–399.

WILDSCHUT, T., B. PINTER, J. VEVEA, C. INSKO, AND J. SCHOPLER (2003): Beyond the group mind: A quantitative review of the interindividual-intergroup discontinuity effect. *Psychological Bulletin* 129, 698-722.

XIAO, E. AND H. KUNREUTHER (2012): Punishment and cooperation in stochastic social dilemmas. NBER Working Paper No. 18458.

YAMAGASHI, T. (1986): The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology* 51, 110–116.

YE, M., S. ASHER, L. CASABURI, AND P. NIKOLOV (2013): One step at a time: Does gradualism build coordination? mimeo.

YOUNG, H. P. (2008): Social norms. *In* S. N. Durlauf and L. E. Blume (eds.), The new Palgrave dictionary of economics, second edition. Palgrave MacMillan.

ZHANG, J. AND M. CASARI (2012): How groups reach agreement in risky choices: An experiment. *Economic Inquiry* 50, 502-515.

Eidesstattliche Versicherung

Ich versichere hiermit eidesstattlich, dass ich die vorliegende Arbeit selbständig und ohne fremde Hilfe verfasst habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie mir gegebene Anregungen sind als solche kenntlich gemacht. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht. Sofern ein Teil der Arbeit aus bereits veröffentlichten Papers besteht, habe ich dies ausdrücklich angegeben.

Datum: 27. Juni 2014

Bianca Kern